Quantum Reinforcement Learning for Dynamic Spectrum Access in Cognitive Radio Networks

Sinan Nuuman

Doctor of Philosophy

University of York

Electronics

April 2016

Abstract

This thesis proposes Quantum Reinforcement Learning (QRL) as an improvement to conventional reinforcement learning-based dynamic spectrum access used within cognitive radio networks. The aim is to overcome the slow convergence problem associated with exploration within reinforcement learning schemes.

A literature review for the background of the carried out research work is illustrated. Review of research works on learning-based assignment techniques as well as quantum search techniques is provided. Modelling of three traditional dynamic channel assignment techniques is illustrated and the advantage characteristic of each technique is discussed. These techniques have been simulated to provide a comparison with learning based techniques, including QRL. Reinforcement learning techniques are used as a direct comparison with the Quantum Reinforcement Learning approaches. The elements of Quantum computation are then presented as an introduction to quantum search techniques. The Grover search algorithm is introduced. The algorithm is discussed from a theoretical perspective. The Grover algorithm is then used for the first time as a spectrum allocation scheme and compared to conventional schemes. Quantum Reinforcement Learning (QRL) is introduced as a natural evolution of the quantum search. The Grover search algorithm is combined as a decision making mechanism with conventional Reinforcement Learning (RL) algorithms resulting in a more efficient learning engine. Simulation results are provided and discussed. The convergence speed has been significantly increased. The beneficial effects of Quantum Reinforcement Learning (QRL) become more pronounced as the traffic load increases. The thesis shows that both system performance and capacity can be improved. Depending on the traffic load, the system capacity has improved by 9-84% from a number of users supported perspective. It also demonstrated file delay reduction for up to an average of 26% and 2.8% throughput improvement.

Contents

List of Tables

List of Figures

Acknowledgments

I would like to express my thanks and great gratitude to the Department of Electronics and the University of York for giving me the chance and full support towards accomplishing my study. Great thanks and appreciations go also towards my supervisor, Professor David Grace for his unlimited and unforgettable support on both the scientific and personal side of the research journey. Many thanks and gratitude goes to my second supervisor Mr. Tim Clarke for his great support and inspiration.

I would like to thank the whole of my family as for my wife Dena who has been supporting me here in the UK and my beloved mother and sister, who have been supporting me with every possible way from far away back home as they always did.

My appreciation and gratitude are also for my friends in the research group for their support to me since arriving to the UK up until the end.

Many thanks go to the members of the York LJC and especially to Ben and Rachael Rich for their support to me and my family during difficult times which made carrying out the research possible.

Declaration

This work has not been presented for an award at this, or any other university. All contributions presented in this thesis as original are as such to be the best knowledge of the author. References and knowledge to other researchers have been given as appropriate.

Some of the research presented in this thesis has been published. The publications are listed in the next section.

Publications

1- "***A Quantum Inspired Reinforcement Learning Technique for Beyond Next Generation Wireless Networks***", Sinan Nuuman, David Grace, Tim Clarke, Presented at IEEE WCNC 2015, New Orleans, USA (http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=7122566&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D7122566)

In Preparation:

1- "***Accelerated Dynamic Spectrum Access for LTE Networks using Quantum Reinforcement learning***", IEEE Transactions on Mobile Computing.

Chapter 1 : Introduction

1.1 Overview

With every single year that passes, the number of the wireless enabled devices increases. These devices have created an exponential growth in the consumed data at a typical rate of 53% [1]. As a result, the infrastructure, the technology, and the intelligent schemes that tackle the capacity and performance problems within wireless networks must evolve at the same level if not better to support the growing demand.

It seems that the accomplished increase in the spectral efficiency of the 4G-based network is not enough to deliver the data rate necessary for the highest three level use cases for 5G defined by 3GPP[1]. These use cases are illustrated in figure 1.1 [1]. The ultimate goal for 5G is to provide an ubiquitous and instantaneous mobile broadband data. As a result, it is obvious that a more efficient and adaptable utilization of the radio spectrum is needed at least as an essential part of the overall solution which might also include exploring wider range of the spectrum.

The conventional licensed fixed frequency allocations have been used since the beginning of wireless mobile networks. These have become a bottleneck for more efficient spectrum utilization, under which a great portion of the licensed spectrum is severely underutilized [2]. Although the fixed spectrum assignment policy generally served well for many years, this efficiency proved to be due to limited demand in the early years. The dramatic increase in the access to the limited spectrum for mobile services later on has shown how limited these

techniques are. This increase is straining the effectiveness of the traditional spectrum policies [3-8].

The inefficient spectrum usage necessitates a new communication solution approach to exploit the existing wireless spectrum dilemma. This new networking approach is referred to as Cognitive Radio (CR) networks. It has been recognized, that, the spectrum usage is concentrated on certain portions of the spectrum while a significant amount of the spectrum remains unutilized [2, 3].

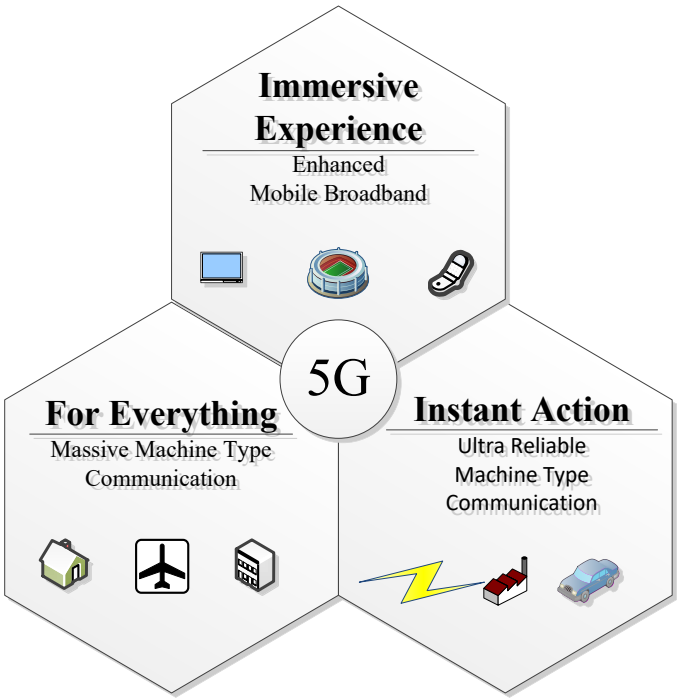


Figure 1.1. Three High Level 5G Use Cases as Defined by 3GPP and IMT 2020[1]

Dynamic Channel Assignment (DCA) was proposed to solve the problem of fluctuating traffic. In a DCA scheme, instead of implementing a fixed frequency plan, all frequencies are placed into a frequency pool and made available to all users [8, 9]. The channels are then assigned on a call-by-call basis. DCA schemes are categorized as either Centralized Dynamic Channel Assignment (CDCA) or Distributed Dynamic Channel Assignment (DDCA). A centralized controller assigns channels from the available channel pool to the calls in the case

of CDCA schemes. The central controller needs to exchange a large amount of information with the base station [10-12].

In DDCA schemes, the users do not need to communicate with other base stations or any kind of central controllers. They attempt to exploit the information usually available for users to select a suitable channel [13, 14]. This thesis concentrates on the study of a fully distributed learning-based dynamic spectrum access.

Cognitive radio networks will possibly solve this problem by providing higher bandwidth to mobile users of dynamic spectrum access networks. This is done through improving opportunistic access capability to the licensed bands without interfering with the already transmitting users [9]. Learning and adaptation are two essential features of a cognitive radio transceiver. Intelligent algorithms are used to learn the parameters of the surrounding environment. The knowledge obtained is exploited by the transceiver for the purpose of choosing the appropriate frequency band (i.e., channel) for transmission as well as its own transmission parameters to gain the best possible performance. Cognitive radio improves the capability of a wireless transceiver by using software that makes the radio transceiver capable of operating in multiple frequency bands. The cognitive radio is a special type of software defined radio which is capable of adapting itself according to the all-time-changing environment. Two main elements that affect the implementation of cognitive radio in achieving the desired system performance which are: efficient *learning* and intelligent *decision making* algorithms [2, 15]. This thesis is concentrating on improving both the decision process and learning efficiency (i.e. speed) of existing learning-based schemes.

Cognitive radio networks on the other hand, provided many research challenges[16]. These challenges are imposed by the existence of a broad range of available spectrum as well as the wide range of desired Quality-of-Service (QoS) requirements for different applications. These differences must be diagnosed and managed as mobile terminals move between wireless

networks within the available spectrum pool. The cognitive radio technology will enable the users to perform [3]:

- **Spectrum Sensing**: this determines the available part of the spectrum by sensing whether a licensed user exists in case of operating within a licensed band.
- **Selection:** selecting the best among available channels.
- **Spectrum Sharing**: this coordinates access to the channel with other users.
- **Spectrum Mobility**: this vacates the licensed channel to operate in another channel when a licensed user is detected.

Dynamic Spectrum Access was born as a technical concept with the launch of 3G cellular as a result of tensions that have appeared over the possibility of disruption and high cost of relocating existing users to make room for the 3G technology [17, 18]. At that point, it became obvious that the need for more spectrum will continue to rise as new technologies will keep arising.

One of the most effective techniques that has been used for solving network problems by learning through only trial and error is reinforcement learning (RL). Cognitive Radio (CR) networks have attracted research on such techniques to learn better spectrum utilization for which it has successfully been applied [19-23] in association with DSA schemes.

1.2 Hypothesis

The hypothesis of this thesis is that Quantum-inspired Reinforcement Learning (QRL) offers two improvements compared to more conventional reinforcement learning (RL) techniques. It essentially presents a significant improvement to the convergence speed. Thus, it causes an obvious improvement in system QoS, throughput performance, and system capacity. This causes a reduction in cooperation overhead and as a result the energy consumption needed in case of information exchange within other learning techniques.

Quantum reinforcement learning can offer up to one order of magnitude enhancement to convergence speed over conventional reinforcement learning techniques.

Learning techniques in general, including reinforcement learning, consist of two main parts. These are the decision making part and the learning and reasoning part. Serious efforts have been made to implement reinforcement learning techniques on a distributed level with dynamic channel assignment [9, 20-22, 24, 25]. Previous attempts to modify the reinforcement learning speed focused on the learning and reasoning part. Specifically, they dealt with the basis upon which weights or Q-values are updated and how are they calculated. This thesis focuses on the way decisions are made in the first place as well as modifying the way it is judged and analysed for a faster learning process.

The growing number of wireless devices makes reinforcement learning and learning in general less effective for the reason that learning needs time, particularly the time needed to find the proper resource allocation policy. As a result, more capacity can be gained for a communication system as high adaptability potentially supports system capacity. A one order of magnitude speed up in convergence speed offers a much shorter collision period among fully distributed learning agents within a Multi-Agent Reinforcement Learning process (MARL).

1.3 Thesis Outline.

The rest of this thesis is outlined as follows:

Chapter 2 provides background information as well as a literature review for the thesis work. An explanation of the fundamentals of cognitive radio networks and the types of radio resource management for dynamic spectrum access is presented. The literature review summarises the work that has been done on channel assignment techniques within wireless networks including the pre-cognitive radio and post-cognitive radio eras. A definition and an

introduction to the idea of reinforcement learning are also included. The advantages resulting from the application of Reinforcement Learning (RL) in wireless communication networks are illustrated. In addition, the disadvantages of Reinforcement Learning in case of large action space are referred to as well. A short introduction to the idea behind quantum computation which will be used as a modification to reinforcement learning in this thesis is included. The reasons and ideas behind proposing Quantum Reinforcement Learning instead of Reinforcement Learning (RL) are clarified. The few research works that included comparisons of Quantum Reinforcement Learning (QRL) with Reinforcement Learning (RL) are reviewed.

Chapter 3 illustrates the system modelling techniques and performance evaluation parameters used in the thesis work. The system simulation procedure is explained as well as the traffic model used within it. In addition, the parameters used for system performance measurements are all briefly illustrated. Finally, the methods used to verify the performance results are given. It gives the basis that has been used for comparison among different schemes that has been experimented in this thesis.

Chapter 4 includes an illustration of three traditional dynamic channel assignment techniques. Their merits are discussed as an introduction to understand the reason behind the viability of our newly proposed scheme later on. The architecture used for all the simulations in the thesis is explained. Details of the system architecture like the base stations, antennas and radio propagation model are specified. Finally, simulation results for the explained dynamic channel assignment techniques are given to compare them and identify their respective merits.

Chapter 5, introduces quantum computation (QC) and quantum search (QS) techniques. An emphasis is placed on the used Grover search algorithm and the theoretical idea behind it. The search procedure for the Grover algorithm and its advantage over the conventional search

methods are illustrated. An explanation for how to determine the number of Grover algorithm iterations is presented including in cases when multiple targets (solutions) exist. A theoretical explanation of the difference between classical and quantum searches is included to put the basis for the reason behind the proposal of quantum search application. Simulations for Grover algorithm based channel assignment schemes are illustrated and compared to conventional schemes as well as reinforcement learning scheme. Chapter 5, represents an essential basis for the development of the full quantum reinforcement learning scheme in chapter 6. It introduces the separation of the search process from the learning engine and develop it independently. It is the point when the advantages of two traditional search techniques can be gathered in one search algorithm. The Grover search algorithm is also applied as a spectrum assignment mechanism and compared to traditional assignment techniques from chapter 4.

Chapter 6, presents the modelling and performance simulation of the full quantum reinforcement learning scheme. An introduction to the theory and idea of reinforcement learning is given. An explanation for the version of the base reinforcement learning strategy used for comparison as well as for quantum search modification is illustrated. The results for quantum reinforcement learning technique as well as the results of other comparison techniques are presented with discussion. The ideas behind the performed modifications of the new QRL scheme over the RL scheme are explained. Differences between the QRL and RL schemes are also illustrated.

Chapter 7 presents some recommendations for future work based on the accomplishments of the thesis work. Suggestions for applications that can exploit the new ideas deduced from this thesis are presented. Chapter 8 illustrates key conclusions for the thesis chapters, and then sums up the novel contributions within the thesis work. It bullets the main ideas used in this thesis to gain the produced novel contributions.

Chapter 2 : Literature Review

2.1 Introduction

This chapter provides the background behind some of the accomplished works on dynamic spectrum access schemes in cognitive radio networks. It pictures the growing problem of spectrum allocation with the growing number of wireless devices.

For the reasons mentioned above, this chapter explains the concept of intelligent (cognitive) radio networks, and why they were put into action. It explains the principles of Radio Resource Management (RRM) techniques. It then illustrates previous works that investigated traditional spectrum assignment procedures. Later on, a review of some of the investigated reinforcement learning-based schemes is presented. In addition, some of the works dealing with quantum reinforcement learning within fields other than wireless networks that have inspired the proposal of this work are also illustrated.

2.2 Cognitive Radio

A reliable and efficient wireless communications system has always meant three important things: an efficient allocation of users to the available spectrum, while avoiding collisions as much as possible and providing minimum transmission delay.

Since the beginning of the era of wireless networks and wireless-enabled devices, spectrum allocation strategies became the core of interest among wireless network providers. This is mainly due to the fact that the number of users as well as the number of the network applications proved to be in a continuous growth. Different approaches and policies have been

proposed to tackle the problem of the apparently limited spectrum. For many years, wireless networks have been dominated by one essential strategy. This strategy is fixed frequency planning or fixed spectrum division among coverage areas. This strategy has proved to be quite restricting and limited. This is due to the limited viable solutions that are applicable to such strategies and their inadaptability with communication traffic fluctuations over vast service areas. Moreover, studies showed clearly that the available spectrum is extremely underutilized. Figure 2.1 is an example of the spectrum usage in a few places in the UK (directly reproduced from [26]). The temporal and geographical variations of the spectrum usage can be observed.

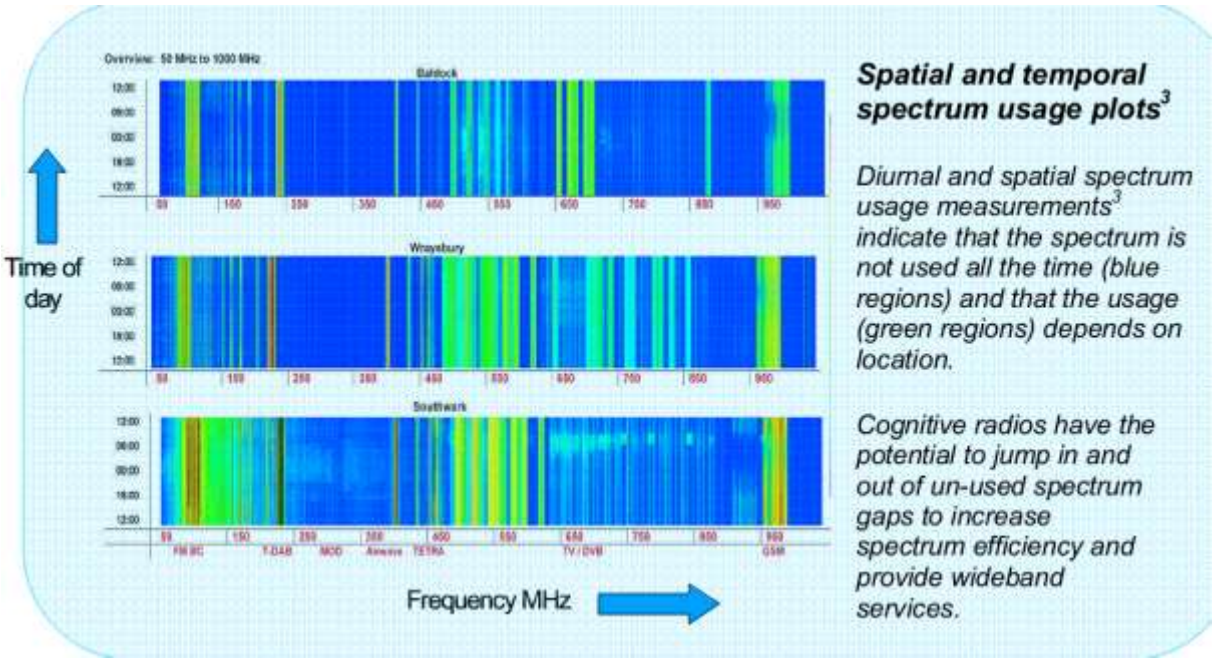


Figure 2.1. Spectrum Occupancy Measurements in a Rural Area (top), near Heathrow Airport (middle) and in Central London (bottom) (directly reproduced From [26]).

The channel usage levels are represented using colours within figure (2.1). The unused frequency is represented in blue while the red is the highly used frequency. The rest of the colours represent different degrees of usage between them. The figure shows that a large proportion of the spectrum is underutilized regardless of time and location.

The next stage approach proposed later on for better allocation of the available spectrum was an open pool strategy. The whole spectrum as an open pool is offered for the whole service area. This idea has been introduced for the first time in [7, 27]. This approach was described as dynamic spectrum allocation and known also later as cognitive radio. The concept upon which cognitive radio is based is the utilization of the available spectrum by opportunistic access to the licensed bands without interfering with other existing users. The suggested definition of cognitive radio in [26] : '*a radio system employing a technology, which makes it possible to obtain knowledge of its operational environment, policies and internal state, to dynamically adjust its parameters and protocols according to the knowledge obtained and to learn from the results obtained*'. As a result, the main motivation for the idea of cognitive radio is maximizing the efficiency of the available spectrum utilization through an all-time reliable solution. The desired solution should also be able to monitor (sense) the spectrum for available windows as well as monitoring traffic loads and spectrum usage temporally and spatially. It means that the cognitive user can use more than one frequency band for transmission depending on the length of availability time of these bands. A *spectrum hole* or *white space* is the name used for the available spectrum in this case. In other words, a spectrum hole or white space is ' *an assigned frequency band to a licensed (primary) user although at a specific time and geographical location, the same band seems to be unused by that user*' [3]. Thus, we can state that full use of the available spectrum holes (white spaces) means the utilization of the spectrum is beyond the capabilities of the traditional fixed frequency scheme plans. If a spectrum hole was requested by a primary (licensed) user while being used by a cognitive (unlicensed) user, the cognitive user shall have two options. It might either change the transmission parameters to be able to stay in the same hole without interfering with the licensed user or simply move to another spectrum hole. This process is described by figure 2.2 (directly reproduced from [3]):

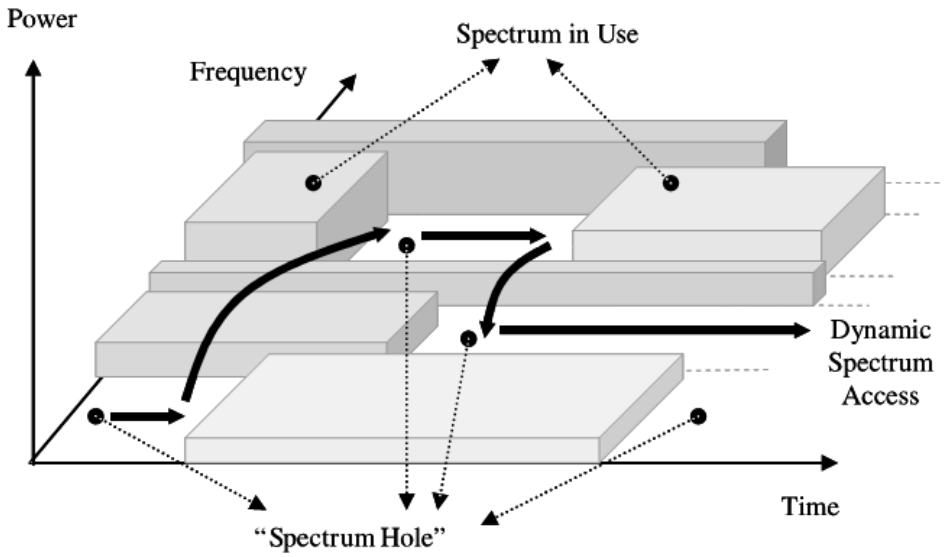


Figure 2.2 . Illustration of Spectrum Hole Utilization (directly reproduced from [3])

We can deduce that a cognitive radio has to make full use of the available spectrum and should be capable of accessing the spectrum in a fully dynamic way. The two key features that make such capability a possible task are cognition and adaptation. Further details of spectrum management techniques will be discussed in section 2.3.

The most important property that distinguishes a cognitive radio system from others that implement traditional channel assignment techniques is the embedded capability of cognition. The required capabilities from a cognitive radio system makes it essential for it to sense any variations within the radio environment over a period of time or space [5]. This property is what makes the spectrum holes discoverable and utilizable. The procedure that represents a cognitive operation which is also called a cognitive cycle is shown in figure 2.3 [2]:

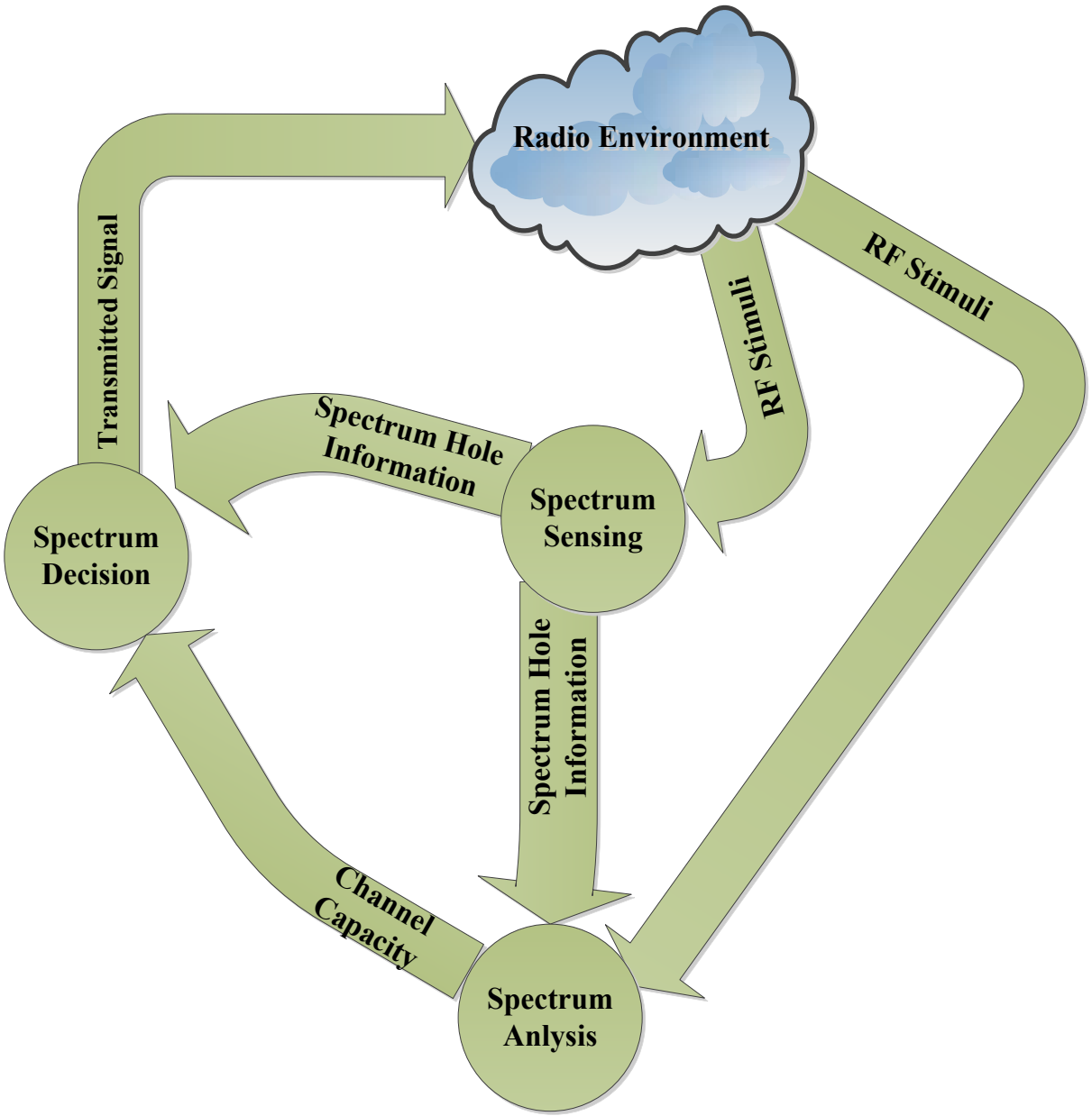


Figure 2.3. Cognition Cycle [2].

The three main functions that can be observed within the cognitive cycle in the figure are [3, 28-30]:

1. ***Spectrum Sensing*:** The process of determining the spectrum status as well as the activity of users by sensing the target frequency band. Further details are mentioned about spectrum sensing in the following section.

2. ***Spectrum Analysis***: The processes of utilizing the information gained from spectrum sensing. It includes scheduling the spectrum access and planning it. Knowledge about the spectrum holes (e.g. interference estimation) is gained by carrying out information analysis.
3. ***Spectrum Decision***: It is the decision of the cognitive user of whether access the available spectrum or not. It is based on the information arising from spectrum sensing and spectrum analysis. As a result, the cognitive radio system should be able to determine available channels as well as appropriate transmission parameters [5]. All decisions should be based on the main objectives like enhancing the system performance (e.g. maximizing the throughput of the unlicensed users) and fulfilling the constraints (e.g. maintaining the interference caused to licensed users below the target threshold).

Most learning and reasoning strategies are developed and applied for the purpose of enhancing the cognition capability of communication systems. Previous studies focused on exploiting the information gained from the spectrum to produce more successful cycles (i.e. enhancing the spectrum decision part). This thesis proposes a novel learning-based approach which focuses on both enhancing the spectrum decision made by the cognitive user and the learning part as well to reduce the number of events needed to reach optimum performance.

Adaptation is another important capability of a cognitive radio system [3, 6]. The system should be able to adapt itself to the changes within the wireless environment by adjusting specific parameters in the transmission operation. It is the speed of such adaptation that differentiates between one learning scheme and another. This thesis seeks a faster learning process through a novel change in the decision making process as well as new way of knowledge base updating as a key for a better adaptation.

2.3 Radio Resource Management

At the system level, the task of reducing channel interference as well as the control of other characteristics of radio transmission, is carried out through one of the radio resource management (RRM) methods [31]. These methods are essentially aimed to improve the utilization of the available spectrum and enhancing system energy efficiency. Several processes can be considered as a part of radio resource management. These processes include:

- Handover
- Channel allocation
- Power control

The work carried out within this thesis focuses on the aspect of channel allocation and improving it as a way of interference control to increase both system capacity and performance.

The ultimate goal for all algorithms that have been developed during the last decade investigating RRM [32] has been maximizing both system capacity and transmission rate and system energy efficiency. Radio resource management allows multiple users to use the same network through multiple techniques. All used techniques aim to allocate a shared available spectrum to the users who demand transmission using the system.

2.3.1 Multiple Access Techniques

These techniques have been developed as variable ways of allocating limited available spectrum within the wireless system to multiple users. In other words, these are ways of sharing the spectrum. They are based on a multiple access protocol and control mechanisms,

known as media access control (MAC). Different categories exist within multiple access techniques which are:

2.3.1.1 Channelization Techniques

This type of technique is based on dividing the available spectrum within the wireless system into frequency bands (channels). Such a division is usually done in different forms. It is one of the principle techniques that has always been used within wireless cellular systems. The channels can be allocated to multiple users. There might be multiple channels allocated simultaneously within a link in a data packet network. This is due to the fact that there could be data packets and relayed traffic in transmission simultaneously. As mentioned before, there are different forms of channelization. Four fundamental channelization techniques have been developed for the purpose of multiple accesses [33] which are:

- ***Frequency Division:*** In this case, the available spectrum range is divided into multiple frequency bandss (channels). These channels are allocated to several transmitting users. Interference among neighbour sub-channels might appear in frequency division techniques. Thus, a guard band (unallocated frequency range) is used for separation.
- ***Time Division:*** In this case, the frequency range is not divided as in frequency division. Instead the frequency range is divided into time periods (slots). The usage time is divided among users as time slots. In this case as well, interference might appear between neighbouring transmitters. It occurs as a result of possible delay that usually happens as a result of more than one user using the same frequency during certain portions of time divisions. Thus, guard bands are used between time slots to prevent interference.
- ***Code Division:*** This kind of division has been developed to combine the benefits of both frequency and time division. Each user can fully utilize the entire spectrum in

both frequency and time domains. Spreading codes are employed to divide signals among multiple users.

- ***Space Division:*** In this case, the wireless transmission coverage area is divided rather than dividing the frequency or time. Directional antennas are used in this case to connect users in different directions. The negative gain of antennas at the side lobes controls the interference. The negative side of this type of division is that narrow beam antennas are large in size and thus difficult to implement on mobile stations and small cell base stations. Thus, such division techniques are exclusively applied to backhaul networks among base stations.

Some communication systems might use more than one division technique. Time and frequency divisions are both used in OFDMA within 4G systems [34, 35]. The FP7 BuNGee is an example project that implements directional antennas on the backhaul network, which use both space division and OFDMA techniques [36].

2.3.1.2 Random Access Schemes

These schemes are used to provide distributed multiple access and flexibility in spectrum access. One of the basic random access schemes is ALOHA. It allows multiple users to transmit on a common channel. Collisions occur when users try to use the same time slot and also when random back off is performed for retransmission. ALOHA is a promising technique for its simplicity which makes it desirable for networks that require minimum implementation overheads for energy saving. The detection of the carrier before transmission of data is introduced within the carrier sense multiple access (CSMA) scheme [37].

In CSMA, a request to send (RTS) and clear to send (CTS) mechanism can be used. When a node intends to transmit, it will broadcast an RTS frame to the nodes in vicinity before transmission [38]. A reply with a CTS frame comes back from the destination node. In such a

case, any other nodes that received RTS or CTS frames will avoid sending data for a given time. The transmission is then started by the source node to send data packets to the destination node. To acknowledge the delivering of the data packets, the receiver replies with an ACK (Acknowledgement) frame. Any packet that is transmitted without having an ACK reply for it in a given time will be considered a lost packet. These CSMA mechanisms are performed by the IEEE 802.11 standard. To resend the lost packet, different schemes are used which include [38, 39]:

- ***1-persistent***: the transmitter continuously detects the channel and sends data once it is free.
- ***P-persistent***: the transmitter sends data on idle channels with a probability of p.
- ***Non-persistent***: the transmitter back off the lost packet and wait for a random time to resend.

The 1-persistent technique is considered effective at low traffic loads. However, it may cause high number of collisions at high traffic load. In this case, the non-persistent is applied instead [38].

2.3.2 Frequency Band Allocation

Within frequency band allocation (FA), the whole available spectrum is divided into multiple frequency bands. These bands are allocated to multiple groups of base stations. Each base station divides the allocated frequency band into multiple sub-channels. In other words, channelization is applied in this case on a portion of the whole spectrum of the whole system within each base station BS. In current cellular communication systems, this frequency band allocation (FA) mechanism is widely used for the purpose of spectrum management [40]. FA strategies can be divided into the following categories [38]:

2.3.2.1 Frequency Planning and Cell Clustering

Reducing inter-band interference is usually the main purpose for the use of frequency planning (FP) within FA strategies. The clustering algorithm used within GSM systems is a typical FP strategy [38, 40]. A group of adjacent cells that have all the available frequencies is defined as a cluster in this algorithm. To avoid inter-cell interference, a different frequency band is assigned to each cluster member (cell/base station). The network usually consists of several clusters where the same frequency pattern is applied to each cluster. The same band is shared by two cells in a neighbour cluster [33]. The number of cells (base stations) within each cluster determines the spectrum efficiency [38]. An example of clustering in GSM cellular networks that shows clusters divided into 3 cells is shown in figure 2.4.

The number of clusters within specific network coverage area and specific available frequency range defines the size of the cluster and as a result the bandwidth within each cell. As the number of clusters increases, the size of the cluster decreases and as a result more frequency reuse is applied. On the other hand, reducing cluster size reduces the distance between cells using the same bandwidth and as a result increases interference.

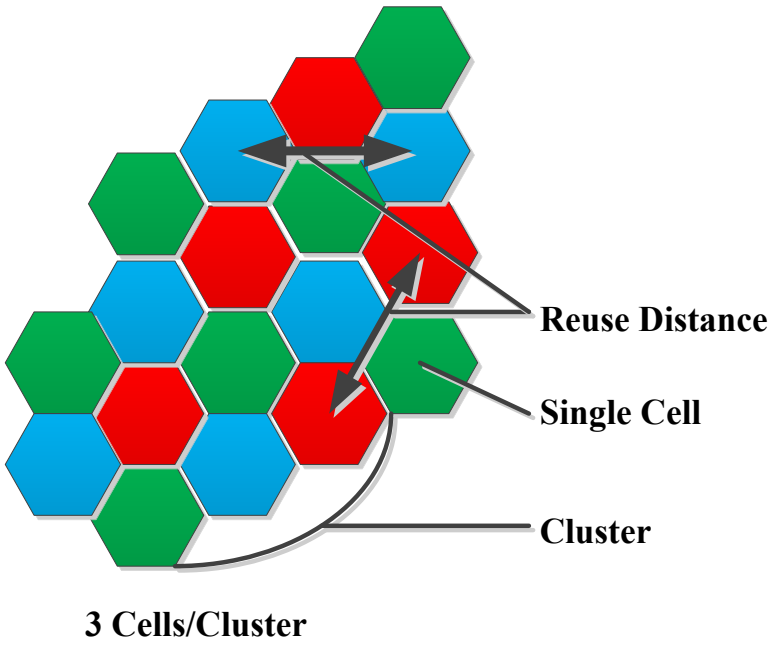
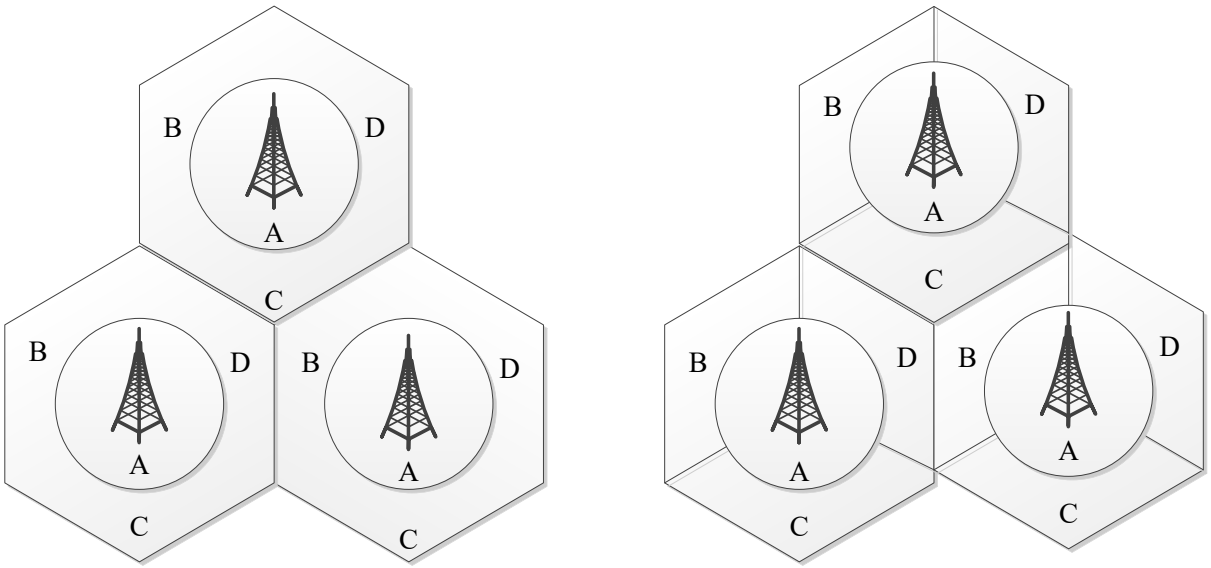


Figure 2.4. Clustering in GSM networks (3 Cells/Cluster)

2.3.2.2 Fractional Frequency Reuse

The strategy of frequency band allocation can be applied to fractional zones within the cell. A Fractional Frequency Reuse (FFR) scheme for inter-cell interference coordination has been proposed by 3GPP LTE in an OFDMA HetNet [41]. Based on the patterns of antennas of the eNBs, this scheme has been designed as omni-directional and sectored schemes as shown in figure 2.5.

With the FFR, the small cell with an omni-directional antenna is divided into inner and outer zones. Each of them has its own frequency band allocated to it. This is to make the users within the inner zones of adjacent cells capable of reusing the same frequency band. On the other hand, cluster based FA is applied on users near the cell edges within the outer zones as they receive more interference from neighbour cells. It is for this reason that, this scheme can achieve higher system capacity compared to conventional clustering based FA. This is due to the fact that the inner zone has less frequency constraints than the outer zone. The omni-directional FFR scheme has been essentially proposed for Pico or Femto cells in LTE [38].



(a) Omni-directional FFR (b) Sectorized FFR

Figure 2.5. Fractional Frequency Reuse

To include all frequencies in a single cellular area, a further sectored FFA scheme has been designed. In this case, the outer zone is further divided into three sector zones using three sectored antennas as in figure 2.5-b. Each sector of the outer zone is allocated a different frequency band. As a result, the whole available spectrum range can be used within the inner cell. On the other hand each of the outer zone parts uses a portion of the frequency band used by the inner zone. This set up is used to prevent users within outer zones of adjacent cells from using same frequency bands and as a result it reduces interference [42, 43].

2.3.2.3 Multi-beam Frequency Planning

In some network architectures like the heterogeneous mobile broadband network proposed in the FP7 BuNGee project [36], directional antennas have been used to establish wireless links.

In the BuNGee project, each ABS is supported with 4 (or 2 based on locations) directional antennas within the access network. On the other hand, HBSs are supported with multiple directional antennas to connect to several ABSs within a square area in the backhaul network.

Four frequency bands are allocated to different antenna beams on ABSs and HBSs in a special strategy that has been designed based on the antenna beams. Based on the streets to be covered and the location of the ABSs, each ABS assigns different frequency bands to four or two antenna beams. Frequency allocation for antenna beams for adjacent ABSs is coordinated to avoid interference among antenna beams covering the same street.

Four different frequency bands are allocated to a group of four adjacent antennas within backhaul network.

2.3.2.4 Protocol Architecture

The architecture where a FA strategy is used can either be centralized or coordinated or set as a distributed architecture.

A centralized FA usually used within 2G systems. In this case, the allocation of frequency bands to different cells is done by the Radio Network Controller (RNC) [32]. The Inter-Cell Interference Coordination (ICIC) strategy has been introduced within LTE systems [44-46]**.** In this case, an X2 interface is employed to exchange control information among eNBs [38, 47]. The neighbouring fractional zones can achieve band separation by exchanging information through X2 links. Dynamic FA becomes essential when coordination overhead issue occurs. Such a situation appears if the topology of the network is rapidly changed.

2.3.2.5 Spectrum Utilization and Channel Borrowing

Due to the fact that the number of channels provided by FA to a single cell, zone or an antenna is fixed and constrained to the band size, a limited number of users can be supported [32]. Within highly dense networks, traffic load can be highly dynamic in density based on time and location [48]. The uniform assignment of frequency bands is therefore unable to support the traffic dynamics. Based on queuing theory, this causes the users to be blocked

[49]. Spectrum utilization becomes difficult to accomplish in the whole network to maintain adequate network performance.

To accommodate non-uniform users' distribution and density across the network, a channel borrowing scheme has been proposed for FA. Any cell that has its allocated band fully utilized can borrow channels from neighbour or adjacent cells. Two types of channel borrowing can be recognised:

- The borrowing can include all the channels in the band for temporary use.
- A portion of the channels will be reserved for use in their allocated cell only. The remaining channels can be borrowed by adjacent cells [40].

To some extent, the borrowing scheme within FA can reduce blocking probability through dynamic scheduling of radio resources. On the other hand, this scheme contradicts with the principle of FA and might cause band overlap and thus loses the advantages of the FA scheme.

2.3.3 Dynamic Spectrum Access

RRM in Cognitive Radio networks has attracted a lot of research regarding Dynamic Spectrum Access (DSA) in the last few years [38, 50]. The rapid growth of wireless device users and the growing demand for high speed data transmission rate systems have brought up the belief that the radio spectrum available for wireless networks use has become insufficient for the growing demand in the recent years.

To overcome the problem of insufficient available spectrum for mobile networks, extra frequency bands have been used to support the necessary coverage. As an example, the 800 MHz spectrum band that has been used to provide coverage in LTE networks that have been deployed in many countries. This band has been used due to the fact that this low frequency range has favourable propagation characteristics [51]. On the other hand, this band is the UHF

band that is allocated to analogue and digital TV transmission in many countries as well. As a result and to make it possible to make this band available for LTE, Ofcom in the UK has to clear this band and reallocate another spectrum band for digital TV stations [52].

Thus, it has been shown that the capacity of frequency bands is inflexible as studies showed that the spectrum is not used all the time everywhere as usage depends on user locations [53]. Based on these facts, frequency band allocation mechanisms are limited in supporting high speed and high performance wireless networks.

As a result, the Dynamic Spectrum Access strategy has been introduced for further utilization of the spectrum that is still underutilized.

2.3.3.1 Dynamic Spectrum Access Scenarios

DSA has been first designed to facilitate Opportunistic Spectrum Access (OSA) [54]. It is meant to allow opportunistic "secondary users" to access the licensed spectrum occupied by "primary users".

Reliable QoS is assured for primary users as they are given priority to use the spectrum. On the other hand, secondary users have to find unoccupied spectrum holes to be able to transmit. In addition, the occupation of secondary users for the channel is temporary and they have to release the channel when requested by a primary user. The release of the whole licenced spectrum for all users by the operators might not be possible. This is because of the greedy usage of secondary users for the spectrum that is already causing undesired interference to primary users as well as the high cost of spectrum band purchasing. Assuring a reliable QoS for secondary users is not going to be possible in this case.

Channel assignment in the case of DSA is done either by the base stations (BS) or the mobile stations (MS). As mentioned before in section 2.2, in DSA, the available spectrum is offered

as an open pool to all BSs and MSs in the network. By this, any channel can be possibly assigned based on demand to any link and released when the demand ends [55].

As there is no FP in this case, the network might suffer from co-channel interference rather than limited frequency bandwidth.

2.3.3.2 Radio Environment Map

For the purpose of supplying reliable information about available channels within the network, the Radio Environment Map (REM) has been proposed [56, 57]. It employs a dynamic database for spectrum management purposes. The database contains information on both BS locations and spectrum usage [58].

Any BS that needs to use a channel needs to search the database for empty channels first. After finding an appropriate channel and occupies it, it updates the database. The database is dynamically updated by distributed BSs, but maintained at a central server.

Undoubtedly this scheme guarantees up-to-date spectrum occupancy information that helps to limit interference. On the other hand, information exchange among distributed BSs might be excessively high. In addition, management and storage complexities might arise as the possible growth in the number of users can make the database quite large and difficult to manage and store. In REM, spectrum awareness might be used for updating spectrum information [59]. The REM with a spectrum database is a standardized technology in IEEE 802.22 Wireless Regional Area Network (WRAN) [60] and ETSI draft [61] for TV White Space wireless access.

2.3.3.3 Spectrum Sensing

DSA has attracted a lot of research for spectrum sensing within cognitive radio networks [62]. As mentioned in section 2.2, the main reason for spectrum sensing is to support the network users with information about the unoccupied channels. Channel quality is usually

checked through interference measurements. The measurements are carried out prior to data transmission. Energy detectors are used to scan the frequency bands and gather the interference power on each channel.

Decisions about channel quality are then set by the sensing entity based on a threshold interference value that reveals whether the scanned channel has sufficient SINR for transmission [63].

2.4 Machine Learning

Machine learning is a research field which studies artificial learning systems. The main interest of this field is focused on the strategies and algorithms that enhance the performance of learning agents through experience. The learning capabilities of the agents are based upon ideas from statistics, computer science, engineering, cognitive science, optimization theory and mathematics [25, 64-66]. The cognition part within cognitive radio systems is the part within which machine learning techniques are used. This thesis focuses on a learning strategy for dynamic spectrum access in cognitive radio networks, and we here attempt to illustrate a brief definition and review of the ideas behind it.

One definition for a learning process can be as in [67], as 'the process of exploiting a set (class) of tasks T and performance measures P to gain an experience E that results in improving the performance P'. A fully complete learning process should include three main elements. These elements are: a class of tasks (target goal or output), a measuring parameter for improving performance (error rate) and a source of experience (training data or input). Based on the information used by the learning agent to learn from (training data) and their availability, we can basically categorize three main types of learning [65-69]:

- ***Supervised Learning*** [65, 68]: The learning here is accomplished through a two-stage process. First, the learning agent is given the time to observe a set of inputs accompanied with a set of desired outputs (target) for them. Then, according to that

the agent learns a function that maps from inputs to outputs. Thus, in this case, the outputs are available from the environment which acts as a teacher and the target is known in advance. Such learning is mostly used for cases where the learned objective has static properties like text recognition.

- ***Unsupervised Learning*** [65, 66]: Both stages of supervised learning are merged into one here. The learning agent in this case is given a set of inputs to learn through them without any support. No prior knowledge is available about the desired outputs, or rewards or punishments. In other words, the agent in this case learns depending on the inputs exclusively. Such type of learning is useful specifically when the learned objective is dynamic and changes all the time.
- ***Reinforcement Learning (RL)*** [64, 69]: The agent in this case interact with the environment through selecting actions. Based on these actions, the agent receives either a reward or a punishment accordingly. The goal of the agent is to learn to maximize the future rewards (or minimizes future punishments) over its lifetime. As no prior knowledge of desired results is available here, reinforcement learning can also be categorized as unsupervised learning.

2.4.1 Reinforcement Learning (RL)

Reinforcement learning (RL) is highly distinguishable from other learning techniques by the fact that the agent starts collecting information about the learned system from scratch. Such collection is made from the results of the actions made by the learning agent. These results themselves are the measure of the quality of the learning progress [64, 69]. The knowledge required for learning is gained as the agent interacts directly with the environment in this case. Rewards are given for successful actions of the agent, while failed ones result in punishments. This means that reinforcement learning (RL) does not need an environmental model as it is trial-and-error based learning. It is therefore an attractive candidate for

distributed cognitive radio scenarios [9]. This is due to the fact that no knowledge needs to be exchanged in the case of multiple agents learning the same environment. It is also because each individual agent does not need any prior knowledge about the environment which is a concerning feature in distributed cognitive radio. The main elements of a reinforcement learning (RL) system can be identified as in [64]:

1. ***Policy***: It is the element that defines the way of decision (action) making at a specific time in response to the gained environmental state.
2. ***Reward Function***: It is the function that maps each environmental state to an action (or state-action pair) to a single value which is called the reward. The desirability of the action according to a specific environmental state is indicated by the given reward. Maximizing the rewards gained by the learning agent on the long run is the optimum goal in reinforcement learning (RL).
3. ***Value Function***: It is the total amount of reward (value of the state) that the agent expects to accumulate over time starting from the specified state.

A simple model for a reinforcement learning (RL) algorithm can be mapped as in figure 2.6. It consists of the following [64]:

1. a set of possible states, represented by S;
2. a set of actions, A;
3. a set of numerical rewards R;

The learner and decision maker is called the agent. The outer part that interacts with the learning agent is called the environment.

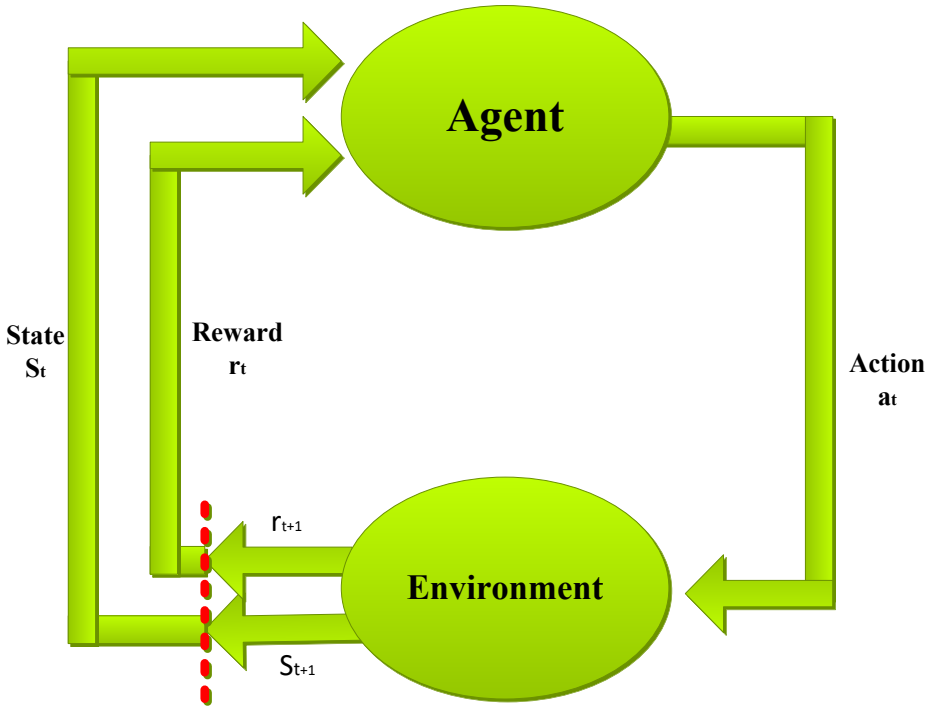


Figure 2.6: Standard Reinforcement Learning (RL) [22]

Thus, Reinforcement Learning (RL) is a very well suited technique for cognitive radio networks. In this case, the action of data transmission interacts with the radio environment and the goal is spectrum allocation.

In Reinforcement Learning, a register or table that is referred to as Q table is setup for every state with elements representing each action. In some cases, a Q table is set up for actions only. It is when one state for the system is considered. The values within Q table indicate the desirability of different actions. In other words, they represent the probability of selecting each action. Under the policy π, the action-value of a state-action pair (s, a) is defined by [64]:

$$Q_\pi(s,a) = E_\pi\{R_t | s_t = s, a_{t=a}\}$$
$$= E_\pi\left\{\sum_{k=0}^{\infty} \gamma^k \, r_{t+k+1} | s_t = s, a_{t=a}\right\} \quad (2\text{-}1)$$

where R stands for the reward which is associated with first taking action a in state s following policy π. The reward value is discounted by γ on each state $(1 > \gamma > 0)$. Values

are usually averaged over many random samples of rewards. A Monte-Carlo method is used for this purpose. The number of iterations taken in pair (s, a) is what decides the degree of accuracy of Q in a static environment. The target of solving a reinforcement learning task is to find a policy that leads to the maximum possible accumulated rewards over the long run. Based on high order Q values , the improved policy for Markov Decision Processes (MDP) can be defined:

$$Q^*(s, a) = \max_{\pi} Q^{\pi}(s, a) \quad (2\text{-}2)$$

In a dynamic channel assignment application of Reinforcement Learning (RL), a channel with the highest Q value that is not currently occupied will be selected.

One of the widely implemented algorithms of Reinforcement Learning (RL) Q learning. It is developed for the purpose of improving the action-selection policy for finite Markov Decision Processes (MDP). Initially Q returns pre-chosen arbitrary values $Q(s_0, a_0)$. Then each time an agent selects an action, it receives a reward in a new state. The Q table is updated based on rewards from the previous state and the selected action. The action-value function is defined as [63]:

$$Q_{t+1}(s_t, a_t) = \big(1 - \alpha_1(s_t, a_t)\big)Q_t(s_t, a_t) + \alpha_t(s_t, a_t)[R_{t+1} + \gamma \max_{a} Q_t(s_{t+1}, a)] \quad (2\text{-}3)$$

where $\gamma \in [0,1]$ is a discount factor that is used to decide the trades off between the importance of current and previous states. $\alpha \in [0,1]$ is the learning rate that decides the speed of convergence.

2.4.2 Quantum Computation

As a tool that was first introduced in the 1970s, Quantum computation relies on exploiting quantum physical properties of atoms or nuclei to process information using quantum bits, or

qubits. Qubits can usually perform certain calculations much faster than classical bits [70, 71]. A qubit can exist in two basic states which are $|0\rangle$ and $|1\rangle$. These states correspond to the classical logic bit states 0 and 1. The difference between the quantum and classical bit however, is that, a qubit can also lie in the superposition of both the $|0\rangle$ and $|1\rangle$ states. As a result, when expressing a qubit $|\Psi\rangle$, it is reasonable to write:

$$|\Psi\rangle = \propto |0\rangle + \beta|1\rangle \qquad (2\text{-}4)$$

where $\propto$ and β are complex coefficients that represent the probabilities of the qubit laying in the $|0\rangle$ and $|1\rangle$ states respectively. The above representation shows a dual existence phenomenon which is called the state superposition principle. It is what makes quantum computation different from classical computation [72].

As mentioned before, a qubit normally lies in a superposition of the $|0\rangle$ and $|1\rangle$ states. In case of measuring (detecting) it, the result will be either $|0\rangle$ or $|1\rangle$. However, we cannot know whether the result will be state $|0\rangle$ or $|1\rangle$. The only fact that we know that we might get the qubit in state $|0\rangle$ with probability $|\alpha|^2$, or in state $|1\rangle$ with probability $|\beta|^2$. These two parameters are the appearance probabilities of the qubit in $|0\rangle$ or $|1\rangle$ states. As a result, the sum of squared probabilities must be equal to 1 to satisfy the following equation:

$$|\alpha|^2 + |\beta|^2 = 1 \qquad (2\text{-}5)$$

In quantum computation, computational processes are carried out on the qubit using the so called unitary transformation (U). As the qubit normally exists in superposition states, the application of the unitary transformation on it means practically performing the transformation on both states simultaneously. This corresponds to the evaluation of different values of a function $f(x)$ for different values of x at the same time. Such a case is referred to as the special property of quantum computation known as quantum parallelism. This is what makes quantum computation outperform traditional computational techniques.

The idea of parallel state update is the main point of attraction in quantum computation that suggests the viability of it as an active tool for speeding up exploration phase in a reinforcement learning (RL) based algorithm. The above property suggests that we can carry out calculations with a significant speed up compared to the time it is required in case of classical computation [72]. Based on this fact, Dong and his co-workers [73-75] have presented the concept of Quantum Reinforcement Learning (QRL) inspired by the state superposition principle and quantum parallelism. Their research works included the application of QRL for robot learning.

Their proposed QRL scheme has accomplished a learning speedup due to the reduction of the number of trials needed by the agent to choose the best channel. Punishments have been restricted to only failed action selections. This has meant a smaller number of explorations by the agent. Exploration only happens in case of failed actions. Moreover, channel ranking proved to be better produced with presented QRL scheme.

The idea behind the proposed improvement in QRL which makes the difference to RL lies within the decision making process. While QRL still adopt the trial and error strategy, the decision that based on the result of this strategy is different. QRL algorithm starts by agent randomly selecting an action. Then, all later decision made by the agent are made upon channel preference. In QRL, another preference table is set up instead of Q table and works with it that is called the amplitude table. The procedure for updating the amplitude table values is explained in chapter 6. The agent in a QRL scheme keeps using a successful action and does not explore. The agent starts to explore the next preferred action within the amplitude table if the selection of the most preferred (and previously selected successfully) action has failed. This is due to the strategy that is used to update amplitude values that turns the highest amplitude value action into the lowest one in case of failure. As the exploration selections made by the agent in a QRL scheme is based on a table formed from previous

experience, success in choosing an action for exploration in this case is much more probable than in the case of RL.

2.5 Traditional Dynamic Channel Assignment Techniques

Dynamic channel assignment strategies appeared within the research literature a long time before the birth of cognitive radio. Intelligent dynamic strategies were introduced for the first time during the 1980s. The road for the birth of cognitive radio was paved through the gradual developments that improved those techniques. The following review provides an overview of the previous research work.

The first time a channel assignment process was made based on defining an interference threshold as an essential tool to improve performance was in 1989 by Akerberg [76]. The channel selection policy was based on the Least Interference Channel (LIC). That means that at the time of assignment, the channel that has the least interference power from other users using the same frequency is selected. The performance of this scheme is assisted through comparison with schemes that are based on other assignment conditions with different interference threshold settings (Non-LIC). The results showed that tighter interference thresholds enhanced the system performance. Such a conclusion was expected, as the focus was mainly on dropping probability reduction as a performance indicator. Call dropping in this paper is given 10 times the importance of call blocking. Thus, choosing better quality channels (channels with the least value of interference affecting them) definitely reduced the call dropping and as a result boosted system performance significantly.

A further study was carried out which gave both blocking probability and dropping probability equal importance for performance evaluation in [77]. In this research paper, a Local Autonomous Dynamic Channel Allocation (LADCA) with power control was proposed. The researchers showed that applying both distributed channel assignment and distributed power control combined can actually improve system performance. A conclusion

is made that almost all the restrictions that are imposed on system capacity are actually due to call dropping rather than call blocking. This was due to that the results referred to the fact that in this case almost all unsuccessful calls are dropped calls.

A proposal to replace the interference level on the channel using CIR measurements directly was made in an investigation that was published in [78]. Two assignment techniques were tested for viability based on this parameter. These techniques were the First Available (FAC) and the Best Quality (BQC) channel assignment techniques. It is clear from the titles that the FAC scheme selects the first channel in a pre-defined list that satisfies the CIR defined requirement. While the BQC scheme selects the highest CIR channel for assignment. Considering the reassignments as well when the measured CIR values fall below threshold level, the FA scheme is capable of gaining a near-optimum performance as shown in the research paper.

Further investigation has been made by Law [79] considering the previous work as a starting point. The LIC scheme mentioned earlier has been compared with the FAC scheme in this case. The principle of outage probability has been introduced, defined and adopted by the author in this paper to measure system performance. The results showed obviously that the LIC scheme proved to outperform FAC when applying different interference thresholds.

In [80] an analytical model was used to investigate the upper and lower bounds of the capacity of the distributed dynamic channel assignment schemes. Interference based distributed dynamic assignment schemes have been studied in [81]. Higher performance proved to be gained from such dynamic schemes compared to the FCA schemes in this work.

The case of rise in call dropping due to existence of mobile devices in some vulnerable regions has been investigated in [82]. Measuring the interference levels at both sides of the transmission link prove to have significant effect on reducing call dropping as shown in this paper using a scheme based on this idea.

2.6 Intelligent Dynamic Channel Assignment Schemes.

The following research reviews show the development of reinforcement learning (RL) based dynamic channel assignment schemes. It is based on improving the use of gained information through improving the policy and also changing the level of learning scheme application. The aim of this review is to give a flavour of works within reinforcement learning (RL) that preceded our presented modification to the learning process.

2.6.1 Reinforcement Learning (RL) Based Schemes

Depending on the level of scheme application, two categories of reinforcement learning (RL) based channel assignment schemes are identified. The first category is centralised schemes in which channels are assigned at a centralised server (base station). The second category is distributed schemes in which case a spectrum decision is made by an individual user only.

As centralized schemes are easier from technical point of view, they were investigated in most research works prior to cognitive radio networks. This was also encouraged by the availability of a reasonable amount of information at the network level. On the other hand, distributed learning-based schemes started to gain attention when the principle of cognitive radio networks has been introduced. This attention was supported by the capability of cognitive radio networks of working according to distributed strategies [83-85]. Local measurements stimulate decisions in the case of distributed schemes rather than centralized information.

2.6.1.1 RL-based Schemes before introducing Cognitive Radio

Encouraged by the availability of the system information at the network level, almost all of the scenarios during this period applied centralized schemes. Junhong Nie and Simon Haykin investigated the centralized dynamic channel assignment scheme based on Q-learning [86]. The target of the application of this scheme in this work was a cellular network. Through

exploiting the information gained throughout the learning process, the mentioned system channel assignment procedure has been based on a session by session basis. A successful interaction of the learning agent with the wireless environment during the learning process produces an optimal channel assignment policy. The Q-values were the action driving parameters upon which a channel assignment (action) is performed. The system states were defined according to channel availability in cells all over the service area.

A useful and reasonable comparison of the raising Q-learning approach with both a fixed channel assignment (FCA) scheme and a good Dynamic Channel Assignment (DCA) scheme MAXAVAIL [87] has been made. The comparison has been made on a 49 cell cellular communication system platform. It was obvious that the Q-learning based algorithm outperformed the FCA one resulting in a higher system capacity even with changing traffic conditions from spatially uniform to non-uniform or to time varying traffic. At the same time the Q-learning scheme seems to achieve a similar performance as MAXAVAIL.

Call admission control for cellular networks has been considered as an addition to the channel assignment part in [88]. The mentioned consideration has been made. Senouci and Pujoile attempted to further investigate the work of Nie and Haykin. The number of calls per cell, the channel availability information, and call blocking have been considered in this work. Whether in a stable system or a system with rapid and significant variations, the Q-learning scheme proved to be able to achieve an optimal policy that outperformed the traditional DCA schemes regarding system capacity due to the high adaptability of the new scheme.

2.6.1.2 RL-based Schemes after introducing Cognitive Radio

In [89] a fully distributed Q-learning scheme has been applied on a small 2 secondary user system with 2 channels. A comparison has been made against a centralised one to detect the effect of less information gain by the user on its capability of learning. For each secondary user, the other users were considered as a part of the environment. Results showed a

promising fast convergence capability of independent users when the temperature parameter was carefully chosen and tuned.

The authors in [90], introduced a scheme with two adaptive RL-based spectrum-aware routing protocols within multi-hop cognitive radio networks. Q-learning and Dual Reinforcement Learning (RL) are applied respectively for them. The cognitive nodes stored a table of Q-values that estimate the numbers of available channels on the routes and update them while routing. Based on that, they can learn the good routes which have more available channels from just local information. The proposed protocols showed according to the results, a better performance than the spectrum-aware shortest path protocol during low network loads. They also showed a learning of the optimal route 1.5 times as fast as the spectrum-aware Q-routing during low and medium network load.

In [91], a proposal for a distributed framework for spectrum assignment in the context of cellular primary networks has been made. In each autonomous cell, a reinforcement learning (RL) based dynamic spectrum assignment algorithm has been included. The presented algorithm showed a better trade-off between spectral efficiency and QoS fulfilment compared to both fixed spectrum planning and centralized strategies. It also showed a good management of the spectrum configuration of the system in case of a new infrastructure to be added.

Yang and Grace in [92], presented two distributed channel assignment schemes applied in a cognitive radio system using reinforcement learning (RL) and a weighting factor. Two schemes of channel priority and random picking were shown. These schemes were compared for different number of iterations used to derive the channel weighing factors. The reinforcement learning (RL) based distributed channel assignment schemes showed an improvement to the channel assignment speed by reducing reassignments as well as blocking and dropping rates. An improvement of the system performance by the base stations has been accomplished through the priority channel principle. In addition, after large number of weight

update iterations, the result a significant reduction of reassignments for both schemes has been noticed in the results. A performance improvement has been achieved through learning about past successful and unsuccessful assignments and increasing the acceptance threshold as the available channels increased.

A Q-learning scheme that is based on rewarding users for each data transmission is considered by the authors of [93] . The channel usage of Primary Users (PU) is assumed to be uniformly distributed on the available wireless spectrum. The success of transmission of any packet is acknowledged by a certain signal transmission response. The no response case is considered as a failed transmission. Each successful transmission is awarded with a positive value known as reward. In the case of a failed transmission, a negative value is awarded which is known as punishment. The throughput level has been enhanced by 2.84 times with the use of the Q-learning scheme in this case. Only single user is considered as a reinforcement learning (RL) secondary user (SU) in this research work. In other words, all other users are ordinary non-learning entities depending on traditional DCA scheme. Thus, the system and learning model have been significantly simplified to the minimum scale in this paper. However, the study has been transformed to a multi-agent reinforcement learning (RL) case by Yau et al. [94, 95]. A Carrier Sense Multiple Access (CSMA) based system is considered in this paper. Q-value updates are carried out after every packet transmission. System level information regarding the locations of users are used to define the states of the system. As was expected depending on the single entity case, the multi-agent Q-learning level has enhanced the system performance.

For a cognitive radio system to be able to opportunistically transmit in licenced frequencies without interfering with previously assigned users, it should predict its operational parameters such as transmit power and spectrum. This capability is called spectrum management, which is difficult to achieve when users can only make local decisions

and react to the environmental changes. In [96], the authors introduced a spectrum management approach based on multi-agent reinforcement learning (RL) for cognitive radio ad hoc networks with decentralized control. They have used value functions for the evaluation of different transmission parameters. The function was also used for enabling efficient assignment of transmission powers and spectrum through the achievement of maximizing the long-term rewards. The scheme evaluation has been made through comparison with random and greedy spectrum assignment. Results showed the outperformance of the reinforcement learning (RL) scheme over the other compared ones. In addition, a Kanareva-based function approximation has been applied to improve the scheme capability to handle large cognitive radio networks. This function approximation showed that it can reduce the used memory without loss of performance. As a result, it was concluded that interference to licenced users can be reduced with reinforcement learning (RL) based spectrum management.

An essential process for detecting the existence of primary users using licenced frequency bands is the spectrum sensing process. An option that might be applied to improve detection probability is the cooperative sensing. Such an approach is an effective way for secondary users to tackle channel impairments. Lo and Akyildiz [97], have presented a reinforcement learning-based cooperative sensing scheme. The scheme has been aimed at addressing the overhead problems like sensing delay for reporting local decisions and the increase of control traffic in the network. The scheme was designed so that the secondary user is able to learn four elements. The user learns to find the optimal set of cooperative neighbours with minimum control traffic. It also learns to minimize the overall cooperative sensing delay as well as selecting independent users for cooperation under correlated shadowing. In addition, it should learn to improve the energy efficiency for cooperative sensing. Several temporal-difference learning methods were used to show that the reinforcement learning (RL) based sensing with Q-learning gave the best trade-off between exploration and exploitation. They

also showed that the proposed scheme had the ability to converge to the optimal solution and adapt itself to the environmental changes.

A decentralized Q-learning algorithm based on multi-agent learning was introduced in [98] to tackle the issue of aggregated interference generated by multiple cognitive radio agents at passive primary receivers for wireless regional area networks (WRAN) systems. Two cases of full and partial information availability have been considered for base stations. In case of complete information, they showed that the multi-agent system is able to learn a policy to keep the interference under a desired value. In case of partial information available, the convergence to the selected policy was slower although having implementation benefits in terms of deployment and feasibility. Results have shown that constraints of primary users can be fulfilled by both schemes regardless of the geometry and scenario.

Chen and Qiu [99], have proposed a Q-learning –based bidding algorithm. In the proposed algorithm, the secondary users learn from their competitors so that they can place better bids for available frequency bands. The results showed an enhanced capability for spectrum assignment prioritizing using the proposed algorithm.

In [20], a fully distributed reinforcement learning (RL) based scheme has been proposed by Jiang, Grace, and Liu. A basic transmitter-receiver pair system model with free space propagation model has been used. The spectrum sensing which has to be done by the user has been limited to 3% of available resources at the beginning of each communication. Three different weighting schemes has been suggested which are similar according to the rewarding value but differ in their punishment values. The results showed superiority of the learning scheme over the non-learning scheme from the blocking probability point of view. The learning scheme showed 60% lower value of blocking than the non-learning scheme. However, since the dropping probability has not been taken into account in the process of the state update, it seems that learning scheme has higher dropping than the non-learning scheme.

In addition, results show that gaining better performance depends mainly on choosing an appropriate weight values.

Jiang, Grace, and Liu further proposed a 'pre-play' stage and a preferred resource set technique for the above learning scheme in [100]. In the 'pre-play' stage, a cognitive radio user explores the whole available spectrum channels with equal probability with weights of the used channels updated after each action. Preferred channels have been distinguished through defining a specific weight threshold. The exploration stage suspends when the user obtains a full set of preferred resources for the exploitation stage where user spectrum sensing for selection will be restricted to the preferred list. The user will move back to pre-play again if the weight of any of the resource of the preferred list has decreased under the pre-defined weight threshold. The results show a significant reduction in the spectrum sensing. The overall time and energy consumption of spectrum sensing in the minimum sensing scheme is about 23% of the full sensing scheme. Dropping and blocking probability showed an obvious reduction compared to the full sensing scheme as well.

Jiang, Grace, and Mitchell have further investigated the above mentioned scheme through proposing an exploration time enhancing approach [22]. The user first reserves a certain number of channels. The user then select the appropriate channel from the reserved list to communicate according to different strategies depending on which stage the user is in. Due to the fact that the whole spectrum is fully partitioned in advance, the exploration process and as a result the whole learning procedure time is reduced. However, a drawback of this technique is that some users might be constrained to a limited list of channels that might have high interference over their transmissions. This causes higher blocking and dropping than the same pre-play technique without pre-partitioning.

Jiang, Grace, and Li have then proposed two stage reinforcement learning (RL) based cognitive radio scheme with a first warm up stage in [21]. During the warm up stage, the user

supposed to explore the whole available spectrum pool with equal probability. The weights of actions are updated accordingly. A threshold weight value has been set such that if a spectrum resource selection weight exceeds the threshold, it is considered a preferred one. After some time, a whole list of several preferred resources is distinguished by the user. At this point, the user turns to the next exploiting stage. By adjusting the size of the preferred resource list and the value of the weight threshold, the exploration stage can be controlled. Results show that dropping and blocking are reduced as the preferred list gets bigger.

In [101], a Q-value based adaptive call admission control scheme (Q-CAC) for distributed reinforcement learning (RL) based dynamic spectrum access in mobile cellular networks has been proposed. The research target was to provide a good quality of service (QoS) without the need for spectrum sensing. A stateless Q-learning algorithm with Win-or-Learn-Fast (WoLF) learning rates to develop an efficient dynamic spectrum assignment scheme. The performance of the proposed algorithm has been analysed using the spatial distribution of the probabilities of call blocking and dropping across the network. The scheme was compared with a 100% accurate spectrum sensing based dynamic spectrum assignment scheme. The proposed scheme proved significant reduction in spatial fluctuations in blocking and dropping probabilities. These results provided more cells with acceptable quality of service. They also gave the advantage of each base station using only information gained from its own trials to produce comparable and competitive performance to spectrum sensing based methods.

Morozs, Clarke and Grace [102] investigated the use of case-based reinforcement learning (RL) for dynamic secondary spectrum sharing in cognitive cellular systems for temporary events. Evaluation of the performance for the proposed scheme was evaluated using system level simulations that involve a stadium small cell network. In comparison to classical reinforcement learning (RL), the case-based RL scheme showed an increased adaptability of the cognitive cellular system of the stadium to sudden changes in the environment caused by

the aerial eNB being dramatically switched on and off. The proposed scheme also showed that when applied to be able to accommodate a 51-fold increase in offered traffic without the need for additional available spectrum. It also showed no degradation in the quality of service for primary users.

A case-based RL has also been applied to dynamic topologies with dynamic spectrum assignment for cellular networks in [103]. The performance improvements expected from the scheme over classical reinforcement learning (RL) scheme have been investigated. The application of a stateless Q-learning algorithm with case-based reasoning functionality showed a significant improvement of the temporal performance of a 9 base station network with dynamic topology. The used scheme proved to reduce the performance degradation in terms of the probabilities of call blocking and dropping in case of transition among different phases of the network topology. The obtained result meant an increased usable range of traffic loads of the network.

A proposal for the concept of the Win-or-Learn-Fast (WoLF) variable learning rate for distributed Q-learning based dynamic management algorithm has been made in [104]. The authors demonstrated with the proposed scheme the importance for choosing the correct learning rate through the simulation of a large scale stadium temporary event network. The investigation results showed that using the WoLF variable learning rate has provided a clear enhancement in the quality of service in terms of blocking and interruption probabilities compared to typical values of fixed learning rates. In addition, and based on their results, the authors suggested that it is possible to provide a better quality of service using distributed Q-learning with WoLF variable learning rate that outperforms the spectrum sensing based opportunistic spectrum access scheme but without any spectrum sensing involved.

In [105], a Distributed ICIC Accelerated Q-learning (DIAQ) algorithm for dynamic spectrum access (DSA) in long term evolution cellular systems (LTE) was proposed. The

presented scheme combined distributed reinforcement learning (RL) and standardized inter-cell interference coordination (ICIC) signalling in the LTE downlink. This has been performed using the framework of Heuristically Accelerated Reinforcement Learning (HARL). In addition to the proposed scheme, a Bayesian network based approach for the theoretical analysis of reinforcement learning based dynamic spectrum assignment was also presented. A large scale stadium temporary events simulations have been performed and showed the achievement of superior quality of service over the typical heuristic ICIC scheme and a state-of-the-art distributed reinforcement learning (RL) based approach. A better quality of service (QoS) has been gained in terms of probability of transmissions and the support for higher system throughput densities of up to 59 Gbps/km^2. The probability of retransmission time response characteristics of DIAQ has been compared with distributed Q-learning which showed a significant improvement in performance at the initial stage of learning. An improvement of 44-81% was shown in the results except for the ultra-high traffic loads as a result of using of heuristics for guiding the exploration process. DIAQ also showed superior final performance and convergence speed.

Efficient spectrum management techniques as well as flexible cellular system architectures can have a major role in accommodating the exponentially increasing need for mobile data capacity in the near future. A significant increase in the efficiency of the use of radio spectrum for wireless communications can be achieved by dynamic secondary spectrum sharing. It is an intelligent approach that gives the chance for unlicensed devices to access to parts of the spectrum that are underutilised otherwise by the occupying users. In [106], a heuristically accelerated reinforcement learning (HARL)-based framework for dynamic secondary spectrum sharing in long term evolution cellular systems (LTE) was proposed. The proposal has been made to utilize the radio environment map as external information as a guidance for the learning process of the cognitive radio system. A stadium temporary event

scenario has been simulated to clarify that schemes based on the proposed HARL framework can achieve excellent controllability of the spectrum sharing autonomously. Such a result caused a dramatic reduction in in primary system quality of service degradation that is caused by the interference with the secondary cognitive system. It showed a superior performance in comparison with purely heuristic and reinforcement learning solutions. The emerged patterns of spectrum sharing when using the proposed scheme caused a significant reliability of the cognitive eNodeB on the aerial platform.

An assessment for the robustness of the distributed reinforcement learning (RL) approach for dynamic spectrum access (DSA) in cellular systems with asymmetric topologies and non-uniform offered traffic distributions has been presented in [107]. A distributed Q-learning based DSA scheme has been used when simulating a stadium small cell LTE network. Simulations have shown that such asymmetries within the network environment do not result in reduction in the level of the QoS at any location of the network. This shows that distributed Q-learning approach has good adaptability to asymmetries in the network topology and offered traffic distribution.

As a result, research gave a significant attention to the fact of RL techniques viability for DSA especially for fully distributed solutions. The possibility of learning agent independence and capability of learning with limited gained information is a desired property for fully distributed schemes. However, growing size of action space that needs to be learned by the agent imposed a challenge of reduced convergence speed due to long time needed to explore available solutions. It is based on this point that Quantum Reinforcement Learning (QRL) scheme was introduced.

2.6.2 Quantum Reinforcement Learning Based Schemes

Most research carried out using quantum computational techniques solely or in association with reinforcement learning (RL) to produce efficient learning systems were applied within

fields other than cognitive communication systems. Several quantum algorithms were presented during the 1990s to solve classical problems more efficiently. The first proposal to use a quantum algorithm as a search tool has been applied to unstructured database applications by Grover in [108]. The proposal showed that the presented algorithm which became well-known as Grover algorithm can reduce the number of test iterations required to search for an item within unstructured N elements from an average of $N/2$ times to $\sqrt{N}$ times when using Grover algorithm.

In [73], the state superposition principle was first introduced in combination with classical reinforcement learning (RL) approach to enhance a robot learning capability in finding a pre-specified root within a specially designed room. A better trade-off between exploration and exploitation has been exhibited using the newly Quantum Reinforcement Learning (QRL) scheme in comparison with classical reinforcement learning (RL) approach. The trade-off superiority resulted in significantly faster convergence.

The authors in [109], presented a multi-agent learning policy aiming to produce an efficient trade-off of exploration and exploitation in a different way than the traditional greedy and softmax action selection methods. The states and actions of multi-agent learning agents are represented as quantum superposition states. Probability amplitudes were introduced instead of traditional probability of action. A quantum search algorithm has been adopted as an action selection policy. The 2 agents introduced within experimental simulation were 2 robots trying to find their way to a specific point in a room that is divided into 9x9 steps representing probable root steps. The quantum inspired reinforcement learning algorithm adopted in the research proved a superior learning speed over traditional reinforcement learning and the Nash Q-learning schemes used for comparison within the research.

Quantum amplitude amplification is a useful technique in quantum computational techniques that can enhance the success probability for some quantum algorithms. The Q-

value reinforcement strategy, used within reinforcement learning (RL), is actually the same idea which boosts the probability of choosing the good action based on learning experience. Thus, amplitude amplification can be used in the same way with quantum search algorithms. Based on this idea, Daoyi, Chunlin, and Hanxiong [110], proposed a learning algorithm based on amplitude amplification with quantum search algorithm for a robot navigation system. Here again, the faster learning process over classical reinforcement learning approach was explained by the better trade-off between exploration and exploitation.

In [75], a fully working quantum reinforcement learning (QRL) algorithm using the Grover quantum search algorithm in combination with quantum amplitude amplification as a reinforcement technique was introduced. Again, the application platform was a robot navigation system functioning to lead the robot in a grid world with the dimensions of 20x20 blocks. An episodic learning process was simulated for the robot learning to move from a start to a goal block within the grid. An episode was defined as the robot trying to get from the start to the end block. Failing to get to the target meant eliminating the episode and starting over. Thus, when the robot (learning agent) finds an optimal policy through learning, the number of moves for one episode will be reduced. Temporal difference learning algorithms have been compared with proposed algorithm. The proposed quantum reinforcement algorithm showed superior learning speed over the traditional temporal difference algorithm.

The authors in [74, 111, 112], proposed a quantum inspired Q-learning (QIQL) algorithm for indoor robot navigation control. The simulated robot was aimed to learn the shortest route leading to a target block within a grid map. Different sized maps with different values for learning rates were used for testing the presented learning scheme. Result for all experiments proved an efficient learning speed capability of the QIQL scheme and superiority over classical Q-learning approach.

2.7 Conclusion.

In this chapter an introduction and a clarification for the aim of the chapter are provided. A brief introduction to cognitive radio is presented. Discussions about the different types of radio resource management techniques are then provided. Later an introduction to machine learning, reinforcement learning (RL) and the idea behind quantum computation is illustrated. A review of the accomplished works on using traditional channel assignment schemes is also given. A review of accomplished research on using reinforcement learning (RL) algorithms before and after the introduction of cognitive radio networks is then presented. Finally a review for the limited research done on using quantum techniques with reinforcement learning (RL) for different artificial intelligent purposes is presented.

Chapter 3 : System Modelling and Performance Evaluation

3.1 Introduction

This chapter describes the system modelling, simulation techniques, and the measurement parameters used for the work in the thesis. Simulation has become the most popular method for studying system performance. It is carried out by mimicking the behaviour of the system under investigation. The development of more sophisticated and fast computers as well as flexible programming languages has supported the popularity of simulation methods as a study tool of performance. Modelling of a system can also be done by designing of a real experimental system. Moreover, the system can be mathematically modelled as well. Producing a real system will cost a lot of time and resources. A mathematical model

consumes moderate time and can often be considered to be the cheapest. However, it might include some simplifying assumptions that reduce actual system accuracy. Computer simulation stands in the middle of other two methodologies with low cost, moderate accuracy and low time consumption [113].

The system modelling and simulation techniques are introduced in the next section. Then the key performance parameters used to evaluate the system performance are described in section 3.3. The information about simulation verification method is given in section 3.4.

3.2 System Simulation Technique

Different simulation tools are available and capable of performing the system level modelling of our wireless communication system. First of all, there are the high level programing languages like Visual Basic, Visual C#, Visual C++ or Java. In addition, specialized platforms as the OPNET are available. In this thesis, the simulation work has been carried out using the MATLAB technical programming language. MATLAB is a matrix based programming language with a high capability to perform specialized simulation tasks with relatively small sized programmes. This is due to the presence of huge library of ready programmed functions. These functions reduce the time and effort to accomplish the desired goal. Usually high level programming languages are sometimes preferred for their high execution speed. However, they are user time consuming due to the need of programming all the functions needed in the investigated system. This is because of the absence of ready specialized function libraries. This results in difficulties during debugging and editing. On the other hand, optimizing codes can improve MATLAB simulation time up to a reasonable level. Thus MATLAB has been preferred for carrying out simulation work in this thesis When simulating systems like the one investigated in this thesis, accurate measurements cannot be gained from a single measurement. Thus, a Monte Carlo simulation has been implemented. The Monte Carlo method is a technique that is used to approximate the results

of quantitative measurements through statistical sampling. It is used for dynamic, fluctuating and uncertain systems as in the case of wireless networks. As the inputs to the system (traffic load and environmental effects) are uncertain or not precise, the results of a single measurement of system performance parameters cannot be precise. Thus, in a Monte Carlo simulation, the system is simulated a large number of times. During each simulation, the uncertain performance parameters (for different system users) are measured. The final results of system performance in a single simulation, represent the overall performance of all system users for a certain period of time (or a number of events accomplished as a total). On the other hand, at the end of a certain number of simulations, the average of performance parameter values for all simulations are gained. These results in this case represent the expected (approximate) values rather than precise numbers.

An event-based simulation technique is used. This technique reduces the time needed for simulation significantly. In such techniques, measurements are taken when events occur rather than when a certain period of time passes. The timing of user arrivals is pre-generated and the time for transmissions is pre-calculated based on link quality. Thus instead of calculating time linearly second by second, events like transmission start and end are checked based on their pre-calculated time sequence and thus less time has to pass for all events to occur than in reality. The general procedure of the event based scheme in this thesis is illustrated in figure (3-1).

The simulator will first generate the locations of buildings, access base stations (ABS), hub base stations (HBS) and users. After that, the propagation environment will be generated. The departure times of files (time of events) based on a predefined parameters are also generated. Then, the simulator will go through each event according to their time sequence with the measurements taken at each event time. After the conclusion of a predefined large number of events, the results of the measurements obtained during simulations are calculated as an

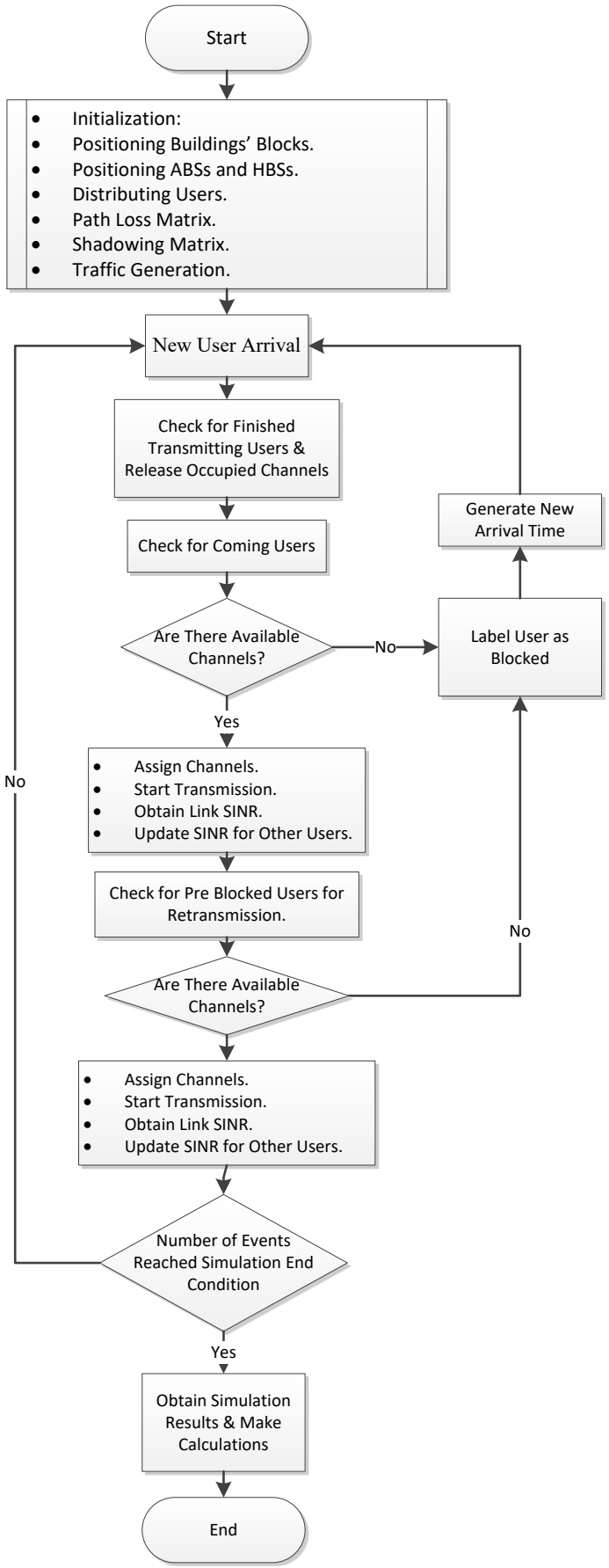


Figure 3.1. Simulation Procedure Flowchart.

illustration for the behaviour of the system. These results represent the average of performance values during the period of time required for the pre-defined number of events to occur.

3.3 Traffic Model

The Poisson traffic model is used in the simulations to generate the file traffic. In other words, the negative exponential distribution governs the inter-arrival and service time of transmissions. The generation of a file is independent of the past generations. In this thesis, only uplink (UL) traffic for the access network (MS-to-ABS) has been considered. After establishing a wireless link, it is assumed that a user will transmit data at a data rate that depends on the link SINR. As the number of the users entering the system gets higher, the transmission rate for each existing user is re-checked for any effect caused by new arrivals upon the already transmitting users through interference. Files transmitted by users are considered to have a fixed data size along the simulation. Users are randomly distributed over the whole coverage area outside the buildings. Each user is considered to generate one file at a time.

The traffic can be modelled through three different levels [114] which are:

- Session level,
- Burst level,
- Packet level.

The session level is usually characterised by the user session inter-arrival time and data file size [115]. Each user session might contain one or more burst which can be modelled at the burst level. Each burst in turn might contain one or more packets that can be modelled at the packet level. The work within this thesis considers session level modelling.

During simulations, a Signal-to-Interference-plus-Noise-Ratio (SINR) based admission control scheme is considered. According to this scheme, users are admitted to the system in

the case where their uplink SINR value is equal or greater to a pre-specified value. The minimum allowed SINR value is considered as specified by the BuNGee project.

3.4 Performance Measurements

To evaluate the system capacity and performance, specific parameters have been selected for measurement during simulations. To evaluate the link quality for determining whether a user can start transmitting over it, Signal-to-Interference-plus-Noise Ratio (SINR) is used. It is also used to determine the level of link transmission rate throughout the simulation time. Blocking probability is used to monitor the system capacity. Outage probability on the other hand is used to monitor the link SINR level drop due to interference levels which can cause transmission delay by stopping it until minimum SINR level is recovered. Throughput and delay have been used as well to determine the system performance which is highly important for data-oriented wireless applications [9].

3.4.1 Signal-to-Interference-plus-Noise-Ratio (SINR)

Signal-to-Noise-plus-Interference ratio (SINR) is an essential parameter with which the link quality of service is measured [116]. It is defined by the average received signal power (S) and the average co-channel interference power (I) plus the noise power from other sources (N). The user uplink SINR is calculated taking into account mobile station (MS) transmission power, gains for both ABS and MS in addition to effects like path loss, noise floor and shadowing. The user uplink SINR is calculated as follows:

$$Signal\ Power\ (dB) = P_{MS} + G_{MS} + G_{ABS} - L_{PL} - L_{Sh} \tag{3-1}$$

$$SINR = \frac{Signal\ Power}{P_N + P_I} \tag{3-2}$$

Where P_{MS} is the MS transmitted power, G_{MS} is the MS antenna gain, G_{ABS} is the ABS antenna gain, L_{PL} is the path loss, L_{Sh} is the shadowing (all in dB), P_N is the noise power, and P_I is the total interference received from other users transmitting on the same frequency.

3.4.2 Blocking Probability and Outage Probability

The blocking probability is defined as the statistical probability that a new file transmission request will fail to find a suitable channel that satisfies the system maximum allowed interference condition [117]. The probability of a transmission being blocked is calculated as follows:

$$Blocking\ Probability = \frac{Number\ of\ Blocked\ Transmission\ Attempts}{Total\ Number\ of\ Transmission\ Attempts} \quad 3\text{-}3)$$

In a situation when there are additional arrivals during the lifetime of ongoing transmissions, it is expected that the SINR level for some links may fall below the defined minimum value for the system for some time. The probability of the SINR value to fall below the fixed predefined value is defined as the *outage probability* [118]. Outage probability can be calculated as follows:

$$Outage\ Probability = \frac{Number\ of\ Failed\ Transmissions}{Total\ Number\ of\ Admitted\ Transmissions} \quad (3\text{-}4)$$

3.4.3 Average File Delay

In this thesis, file delay is considered as the time period starting from a file transmission request by the mobile station, through transmission (and potentially retransmission after being blocked) of the file to the ABS until successfully transmitted. The sum of the delay of all transmitted files is also calculated as the *total delay* while the division of the total delay by the number of transmitted files results in the *average file delay* for the simulated system. Comparing the minimum time required for the file to be completely transmitted (taking into account the maximum transmission rate the link can support) with the average file delay gives a good QoS parameter for real time applications like video conferencing and live video streaming.

$$Average\ File\ Delay = \frac{Total\ Delay\ of\ Success.Transmitted\ Files}{Total\ No.of\ Success.Transmitted\ Files} \quad 3\text{-}5)$$

3.4.4 Throughput

Since bandwidth utilization is a major objective of access schemes, throughput provides a measure of the percentage of capacity used in accessing the channel. The total data size of files that are successfully transmitted to the access point in a certain time interval is defined as the *throughput* in this thesis [117].

3.4.5 The Truncated Shannon Bound (TSB)

For evaluating the performance of each transmission link, the truncated Shannon bound (TSB) has been considered [119]. Accordingly, the transmission rate (throughput) of a specific link at a specific time is highly dependent on the SINR level for it at that time. According to TSB, the transmission rate of a specific link can be expressed as in [119]:

$$Throughput(Thr)\left(\frac{bps}{Hz}\right) = \begin{cases} 0; & for\ SINR < SINR_{min} \\ \alpha.S(SINR) & for\ SINR_{min} < SINR < SINR_{max} \\ Thr_{max} & for\ SINR > SINR_{max} \end{cases}$$

Where: S(SNIR) is the Shannon bound, $S(\text{SINR}) = log_2(1 + SINR)\ \ bps/Hz$

And: α= Attenuation factor, representing implementation losses (path loss) = 0.65

SNIR_{min} = Minimum SNIR value accepted in the system

(for minimum system accepted transmission quality) =1.8dB.

Thr_{max} = Maximum throughput value

(Maximum throughput the link can support) = 4.5 bps/Hz.

SNIR_{max} = SNIR at which maximum throughput (Thr_{max}) is reached

(Above which throughput will not increase due to link limitation) =21dB.

System throughput can then be defined as in [119]:

$$Thr_S = \sum_{i=1}^{N_u} \sum_{k=1}^{n_i} \sum_{t=0}^{T_k} Thr_{MIMO-TSB}(t).BW_c.P_{TDD} \qquad (3\text{-}6)$$

Where $Thr_{MIMO-TSB}(t)$ is the data transmission rate of the link obtained at time (t), and it is updated constantly in the simulation using truncated Shannon bound as mentioned before. T_k

is the transmission duration of the k^{th} transmission of the user, and n_i is the total number of transmissions that have been finished by the i^{th} user. N_u is the total number of users in the system. n_i is determined by the offered traffic level and the probability of successful transmissions [119]:

$$n_i = OT.P_s^i(t) \quad 3\text{-}7)$$

P_s^i is the probability of successful transmissions for the user i at time t, and it can be defined as in [119]:

$$P_s^i(t) = \left(1 - P_B^i(t)\right).(1 - P_D^i(t)) \quad 3\text{-}8)$$

P_B^i and P_D^i are the blocking probability and dropping probability of an entity i at a time t respectively. In the simulations of this thesis, transmissions are halted temporarily in case of SINR reduction to below predefined threshold value until their SINR recover back. This, no actual dropping is considered. The previous equation becomes:

$$P_s^i(t) = \left(1 - P_B^i(t)\right) \quad (3\text{-}9)$$

OT in equation (3-7) is the system offered traffic. The offered traffic level of a user OT_u can be defined as in [119]:

$$OT_u = \frac{T_{ser}}{T_{ser} + T_{int}} \quad 3\text{-}10)$$

Where T_{ser} is the mean transmission service time of user and T_{int} is the mean transmission interarrival time of a user. OT_u shows the percentage of transmission time in the simulation. OT therefore can be defined as in [119]:

$$OT = OT_u.N_u \quad 3\text{-}11)$$

OT shows the average number of active users at any time in the simulation. BW_c is the sub channel bandwidth. P_{TDD} is the percentage of time slots that have been allocated.

3.5 Verification of Simulation Results

Queuing theory [120] is a popular tool for analysing the performance of session based (no interference or other environmental effects are considered) communication systems. Well defined analytic models based on queuing theory, like the Erlang B and Engset formulae, have been used to describe different types of queuing systems. Usually, performance measurements like blocking probability can be calculated and analysed using queuing theory. However, these formulae describe systems of single base stations with no connection establishment obstacles other than channel availability. Moreover, interference levels and their effect on channel availability and transmission rates in modern wireless systems are not taken into account at all.

Another popular and more suitable verification method is modelling the system mathematically using Markov Chain Modelling [113]. A Markov chain, named after Russian mathematician Andry Markov (1856-1922), is one of the most popular mathematical tools that is used to model a dynamical system that changes its state over time. Its popularity comes from various reasons including flexibility, simplicity and ease of computation. However, the simplicity level depends on the modelled system complexity. Using Markov chain modelling for the work of this thesis has been considered as a future work.

3.6 Conclusion

In this chapter, a description of the simulation techniques and procedure for the work carried out in this thesis has been presented. A brief explanation of performance evaluation methods and has also been included. Moreover, the parameters of the quality of service (QoS) that are used in this thesis were defined. Finally, the validation methods used for the simulation work are explained.

Chapter 4 : Traditional Spectrum Assignment Techniques

4.1 Introduction

This chapter illustrates traditional dynamic spectrum assignment schemes that have been implemented with their results. These results are used as a base comparison against the newly presented Quantum Reinforcement Learning (QRL) technique. It is essential to recognise the fact that these channel assignment techniques are those upon which conventional learning techniques are based. Thus, replacing them by a quantum search technique within RL is what makes one of the essential proposed improvements to the learning process. The advantages of each assignment technique as a search process is the basis for forming the idea behind our proposal. The decision making process is the core enhancement that is investigated in this thesis later on in chapter 5. This will explain how the principle behind each of the presented conventional techniques in this chapter has inspired the development of the search process within reinforcement learning.

4.2 System Modelling and Architecture

4.2.1 Base Stations and Mobile Stations Layout

During this research, the Manhattan grid based BuNGee architecture [121] has been considered and is illustrated in figure 4.1. This architecture is a dual hop configuration that implements a small cell strategy that is aimed to increase system capacity and enhance energy savings. Access Base Stations (ABSs) are deployed along the streets with a 90m spacing among each other. ABS installations are made upon existing street lamp columns. Buildings are squarely shaped with the dimensions of 75m × 75m. The width of all streets is 15m. Hub Base Stations (HBS) are located at the centre of each of the big 9 cells that form the entire service area.

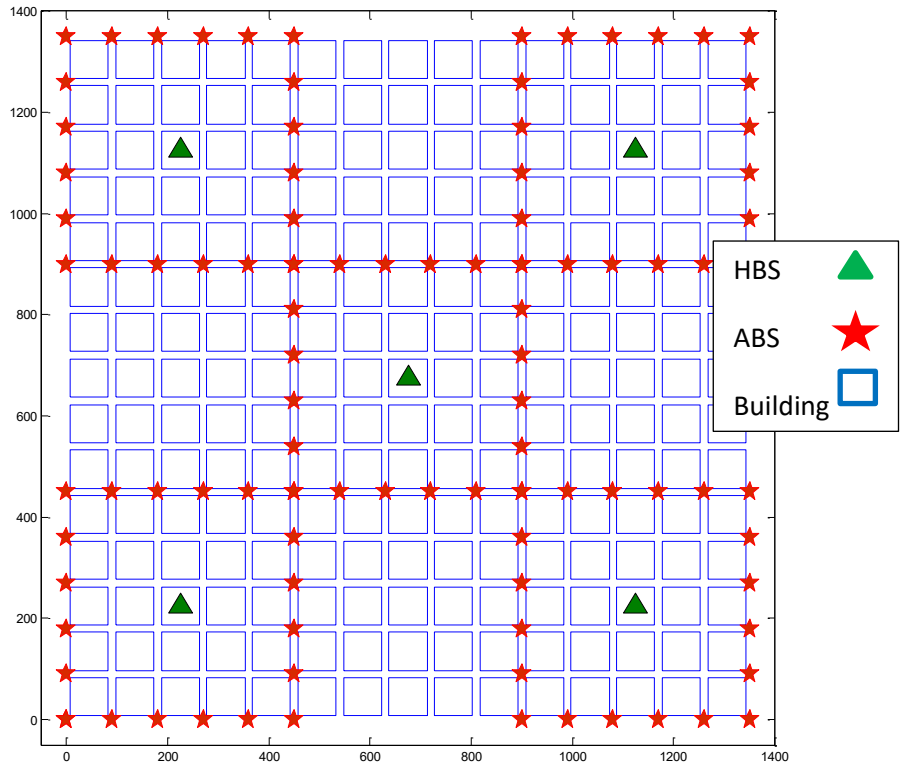


Figure 4.1. BuNGee Square Topology

4.2.2 Base Stations Antennas and Frequency Plans

Each Access Base Station (ABS) is equipped with two directional antennas pointing in two opposite directions along the street they are based on. ABSs that are deployed on the

crossings of two roads have been supplied with four directional antennas. The ABS antennas are deployed at an elevation of 5m above ground level which is lower than the roof level. All ABSs are either North-South (NS) oriented ABSs or East-West (EW) oriented ABSs according to the directions of the antennas which in turn depend on the directions of the streets. The gain of ABS antenna can be obtained from the 3D antenna pattern shown in figure 4.2 when the elevation and the azimuth angles of the MS to the ABS beams are known [122]. The mobile station (MS) antenna is assumed to be omnidirectional with a 0 dBi gain [119]. HBS antennas are deployed above roof level at an elevation of 25m.

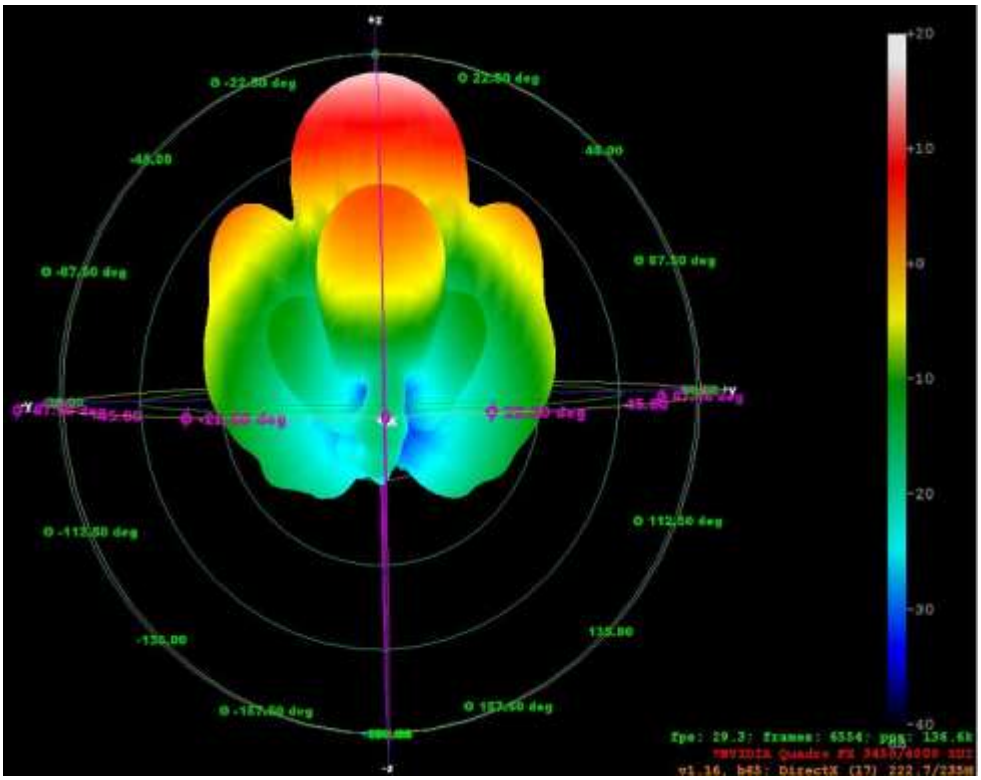


Figure 4.2. ABS 3D antenna pattern (directly reproduced from [122])

4.2.3 Dynamic Spectrum Access Schemes

Dynamic spectrum access schemes have been implemented on the BuNGee architecture. The previously mentioned ABS distribution, ABS specifications and antenna orientations which are the same of the BuNGee project have been considered. Each ABS is supposed to be capable of transmitting with all of the 20 available channels within the coverage area of the system. The transmission can be distributed over the two beams of the

two ABS antennas. The ABS is allowed to adopt any distribution of channels over the two beams including transmitting all the 20 channels over one beam. No reuse of channel (frequency) is allowed within the same ABS.

Three frequency selection techniques have been implemented for evaluation and comparison. All the techniques start with the mobile station selecting the strongest signal ABS around to send a transmission request. In practice, this choice will determine the ABS and beam since each beam within the ABS might have different signal strength on the mobile station side depending on the beam orientation. The difference between the three implemented techniques comes when the ABS choses a qualified frequency to be assigned to the requesting user. The first is the best SINR technique. In this case, the ABS scans all the available (unoccupied) channels and chooses the one with highest SINR value. The second case, is the first available channel scheme (FAC). In this case, the ABS scans all the available (unoccupied) frequencies in the same sequence every time (1 to 20) to pick and assign the first one it recognizes with SINR value that is higher than a specified threshold. The third case, is the random channel assignment (RCA) scheme. In this case, the ABS picks randomly from the 20 frequencies in the system and checks whether it is available and whether its SINR value qualifies for assignment. The ABS is allowed to make 20 random tries and assign the first appropriate frequency. If the ABS fails after 20 tries to find a qualified frequency, it announces the user request for transmission as a blocked one.

4.2.4 Radio Propagation Models

As electromagnetic signals propagate through a wireless channel, they undergo several types of effects that cause them to be weakened, changed and interfered. Effects over signals that cause them to be weakened or changed are generally referred to as noise [116, 123]. The noise sources can be categorized as multiplicative and additive effects. The additive noise arises from noise generated within receivers, such as thermal and shot noise in passive and active devices. It can also be from external sources such as atmospheric

effects, cosmic radiation and interference from other transmitters as in the case of frequency reuse.

The multiplicative noise arises from the various processes encountered by transmitted waves on their way from the transmitter antenna to the receiver antenna and illustrated as in [116]:

- The directional characteristics of both transmitter and receiver antennas.
- Reflection (like from smooth surfaces of walls).
- Absorption (by walls, trees, and atmosphere).
- Scattering (from rough surfaces).
- Diffraction (from edges such as buildings' rooftops).
- Refraction (due to atmospheric layers and layered or graded materials).

Multiplicative processes in return can be subcategorized into three types which are path loss, shadowing (or slow fading) and fast fading (or multipath fading). The path loss is an overall decrease in strength of the signal as the distance between the transmitter and receiver increases. This is regarded as the spreading of waves from the transmitting antenna and the obstructing effects of trees and buildings [117]. Shadowing is regarded as obstructions with varying nature between the transmitter and receiver such as tall buildings and dense wood. Fast fading is the result of the constructive and destructive interference between multiple waves transmitted from the transmitter to the receiver after multiple reflections from different obstacles.

In our implemented simulations path loss and shadowing have been considered. Path loss and shadowing are modelled using the WINNER II B1 urban micro-cell model [124] as all mobile stations as well as ABSs are considered outdoor in this thesis.

4.2.4.1 WINNER II B1

A Manhattan-grid based layout is considered and the antennas of both ABSs and MSs are assumed to be below the roof level of the surrounding buildings. Only outdoor ABSs

and MSs are considered. If the MS and ABS are on a same street (Line Of Sight (LOS)) like in figure (4-3), then the path loss can be calculated as in [124]:

$$PL = 10.0log_{10}(d_1) + 9.45 - 17.3log_{10}(h'_{BS}) - 17.3log_{10}(h'_{MS}) + 2.7log_{10}(\frac{f_c}{5.0}) \quad (4\text{-}1)$$

Where:

$$h'_{BS} = h_{BS} - 1 \quad (4\text{-}2)$$

And

$$h'_{MS} = h_{MS} - 1 \quad (4\text{-}3)$$

d_1 is the distance between ABS and the LOS MS, h_{BS} is the ABS antenna height and h_{MS} is the MS antenna height which is 1.5m.

On the other hand, when the MS and ABS are not on the same street (Non Line Of Sight (NLOS)) like in figure (3-4), then the path loss can be calculated as in [124]:

$$PL = \min(PL(d_1, d_2), PL(d_2, d_1)) \quad (4\text{-}4)$$

Where:

$$PL(d_k, d_l) = PL_{LOS}(d_k) + 20 - 12.5n_j + 10n_j log_{10}(d_l) + 3log_{10}(^{f_c}/_{5.0}) \quad (4\text{-}5)$$

And

$$n_j = \max(2.8 - 0.0024d_k, 1.84) \quad (4\text{-}6)$$

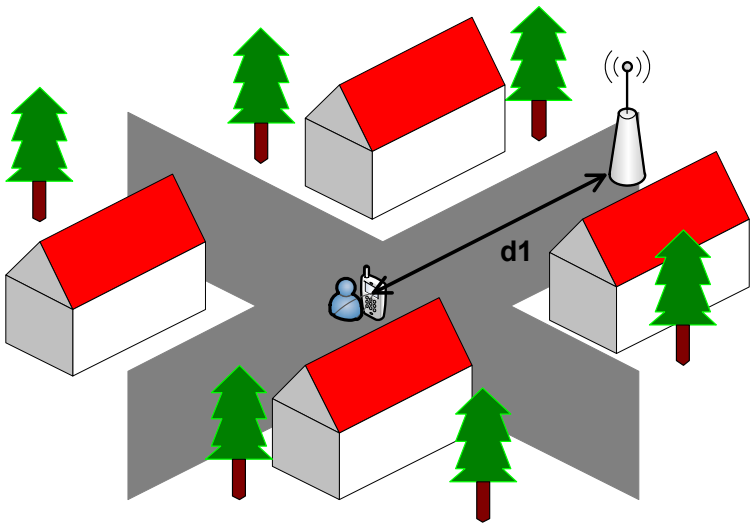


Figure 4.3. LOS path loss

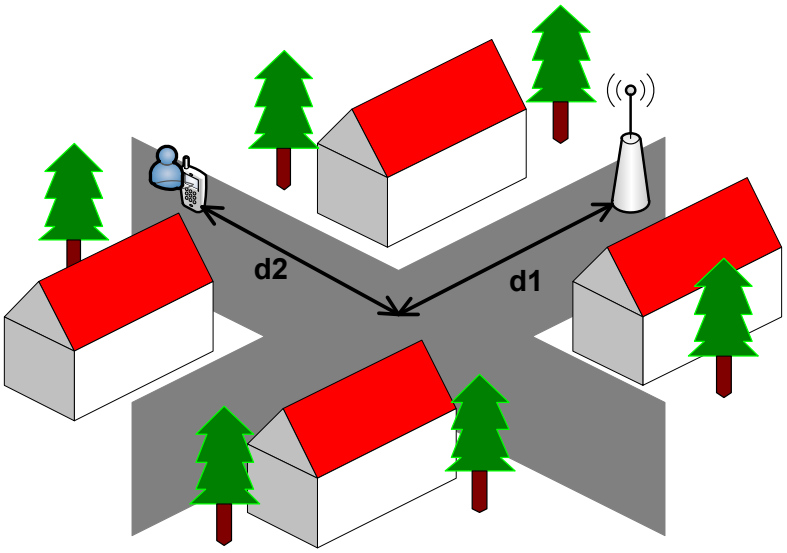


Figure 4.4. NLOS path loss

PL_{LOS} is the path loss of B1 LOS, d_1 and d_2 are the distances between the entities along the street as it is shown in figure 4.4. Experiments on the suggested assignment techniques were carried out using the Manhattan grid based BuNGee architecture as a platform for testing the system performance. An illustration of the BuNGee system topology can be seen in figure 4.1. The locations of users have been considered as fixed. The system simulation modelled the access network (user-ABS) uplink. File sizes that are transmitted by users are considered as fixed for all users. All simulation sessions start with the assumption of maximum link transmission rate (4.5 Mb/s) for the purpose of calculating

the transmission time of file. Example values of system parameters assumed for the mentioned simulations are illustrated in table (4-1). ABSs are divided into two groups depending on the direction of the streets they are deployed on.

Table 4.1. Sample System Simulation Parameters

Parameter	Value
Number of Users	3500
Number of ABSs	112
Number of Beams per ABS	2
File Size	4 MB
Minimum (Threshold) SINR	1.8 dB
Maximum SINR	21 dB
Maximum Link Transmission Rate	4.5 Mb/s
MS Antenna Gain	0 dB
MS Transmission Power	23 dBm
MS Antenna Height	1.5 m.
ABS Antenna Height	5 m.
ABS Antenna Maximum Gain	17 dBi
Street Width	15 m.
ABS Distance from Neighbour Building	7.5 m.
Block Side Length	75 m.
Channel Bandwidth	1 MHz

4.3 Single Base Station Simulation Results

A small scale simulation of a system consisting of a single base station has been performed. The reason behind it is both testing and validating of the simulation procedure. Moreover, the results can be used for a traditional comparison with Erlang B and Engset calculations for blocking probability. Since both mentioned equations do not consider several important parameters like multiple base stations, interference, noise floor,

shadowing, and signal strength, it was important to start simulating a simple case to insure starting from a solid ground. A second reason is that when simulating a more complicated system with a dynamic frequency plan that considers the mentioned parameters, it will be easier to explain the differences when comparing with the Erlang B and Engset results. The system simulation assumes the existence of one base station, 20 available channels and 50 users. The result of blocking probability as a function of traffic load is shown in figure 4.5.

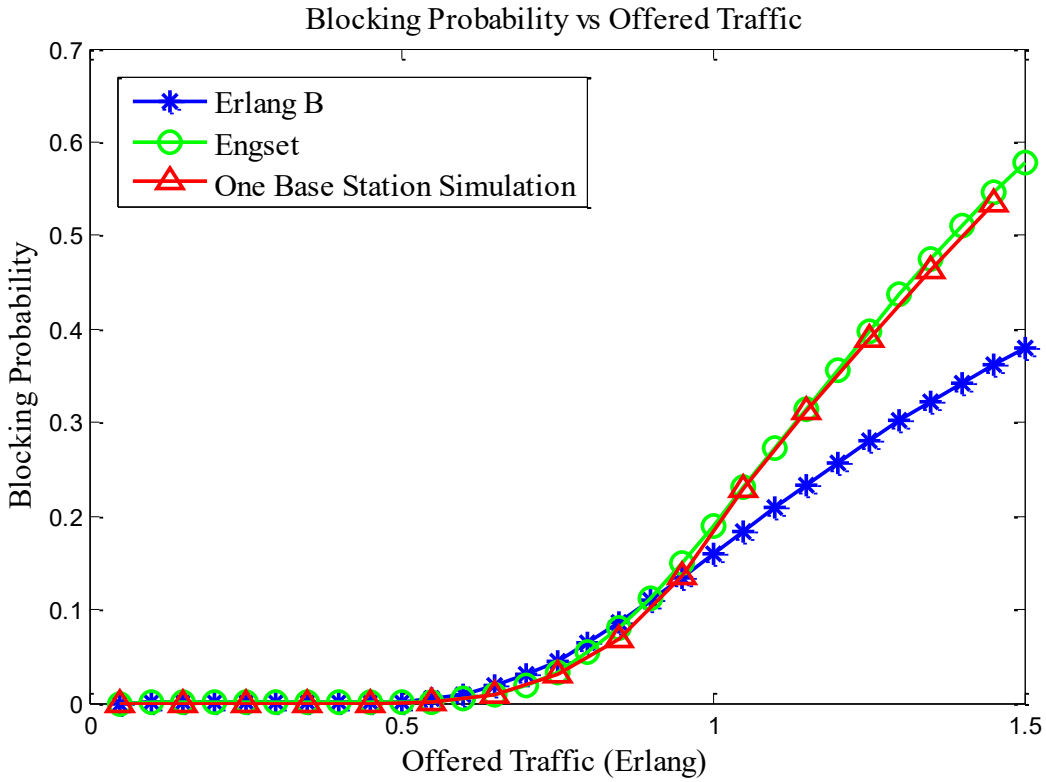


Figure 4.5. Blocking Probability as a Function of Traffic Load for a Single Base Station System.

It is quite obvious that our simulation results at this point follow the trend and shape of the curve of Engset results. The reason behind this result is that our simulated system is close in characteristics to the system assumed in the case of Engset formula. The Engset scheme assumes a limited number of users. The number of users is similar to the number of available channels. Erlang B on the other hand assumes an almost unlimited number of users or a number that is several times more than the number of available channels. The

difference between the two cases is that the blocking probability of Erlang B rises earlier than that of the Engset but after that the gradient of it become smaller and eventually the blocking probability value of Erlang B becomes lower than that of Engset.

4.4 Simulation Results

The three implemented dynamic spectrum access schemes are the best SINR scheme, first available channel assignment scheme, and random channel assignment scheme. Figures 4.6 to 4.10 represent the results of different metrics for the implemented schemes.

In figure 4.6, the blocking probability of the three traditional DCA based schemes are plotted as a function of traffic load in Gb/s. Usually blocking probability is a popular tool for system capacity testing. In most cases, a maximum value of 5% blocking probability is considered for a reliable system. Thus, the maximum traffic load value hosted by the system such that the resulting blocking probability is equal or less than 5% is considered as the system capacity limit.

It is obvious from the graph that when looking at the maximum capacity of both the first available channel assignment (FAC) and random channel assignment (RCA) schemes are almost similar. This is recognised from the fact that blocking probability curve for both schemes cross the 5% limit almost at the same level. The best SINR channel assignment scheme has a slightly better capacity although the difference is not so significant. This is due to the fact that the best quality channel is not necessary available at high traffic loads. This apparent similar performance no longer holds when observing figure 4.7. In this case, it is clear that the random assignment scheme outperforms the FAC scheme from the outage probability point of view. The reason behind that is the nature of the assignment procedure. The FAC scheme ensures that all the assigned users are packed one after the other by assigning them always to the first available channel of a fixed sequence. This causes the probability of collisions to be high as all users are assigned to the first available channel in the same way for all ABSs. As a result, such a scheme raises interference with existing users as the probability of different users requesting transmission from different

ABSs being assigned to the same channel is relatively high. As a result, a poor outage performance is expected from this scheme. In the case of the random assignment scheme, as different ABSs assign channels in a totally random sequence, the interference effect is reduced. From the blocking probability point of view, it is obvious that the best SINR based scheme outperforms the other schemes. The best SINR scheme is the best way of ensuring the assignment of the best quality channel. Choosing the best quality channel ensures minimum possible interference and as a result an obvious reduction in both blocking and outage probabilities. However, adopting a best quality channel scheme includes a considerably higher computational complexity. The system in this case is forced to check all available channels within all detectable ABSs and then compare them.

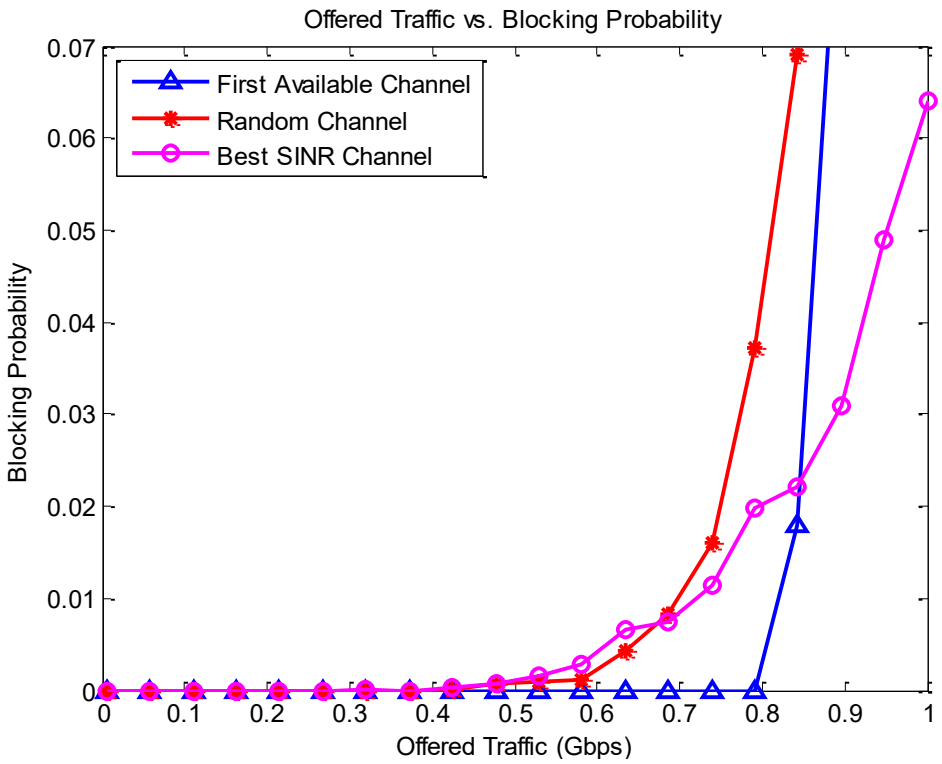


Figure 4.6. Blocking Probability as a Function of Offered Traffic.

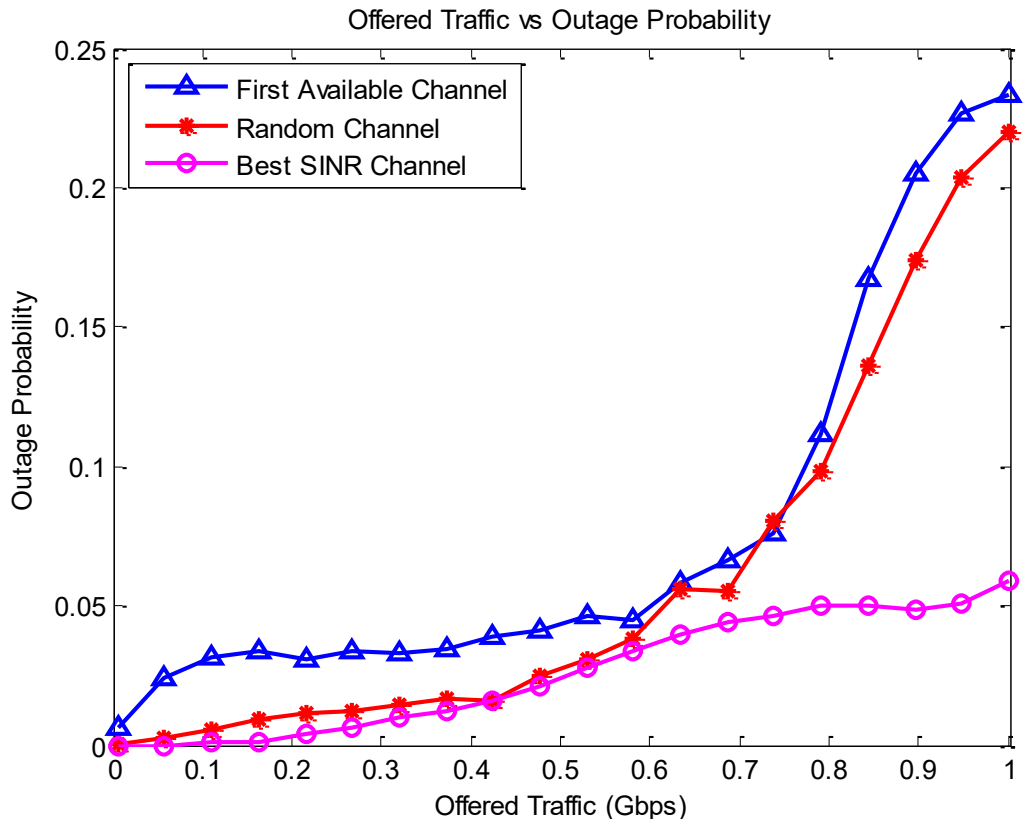


Figure 4.7. Outage Probability as a Function of Offered Traffic

Figure 4.8 shows the average additional file delay (i.e. additional transmission time caused by link transmission rate drop due to interference and also caused by blocking). The additional delay is calculated by subtracting the minimum transmission time (based on highest link transmission rate) from the total actual transmission time. From delay measures, it is obvious that best SINR based scheme outperforms the other two schemes. On the other hand, the FAC and random assignment schemes seem to perform with vary similar delay performance except for a slight outperforming of random scheme over FAC scheme due to the better outage performance of the random assignment scheme. Both schemes experience a significant increase at some point due system saturation. This case is not recognised in case of best SINR scheme. In case of the best SINR scheme, it is due to the low outage probability resulting from selecting the best quality channels which makes the system in this case more capacitive and of higher throughput.

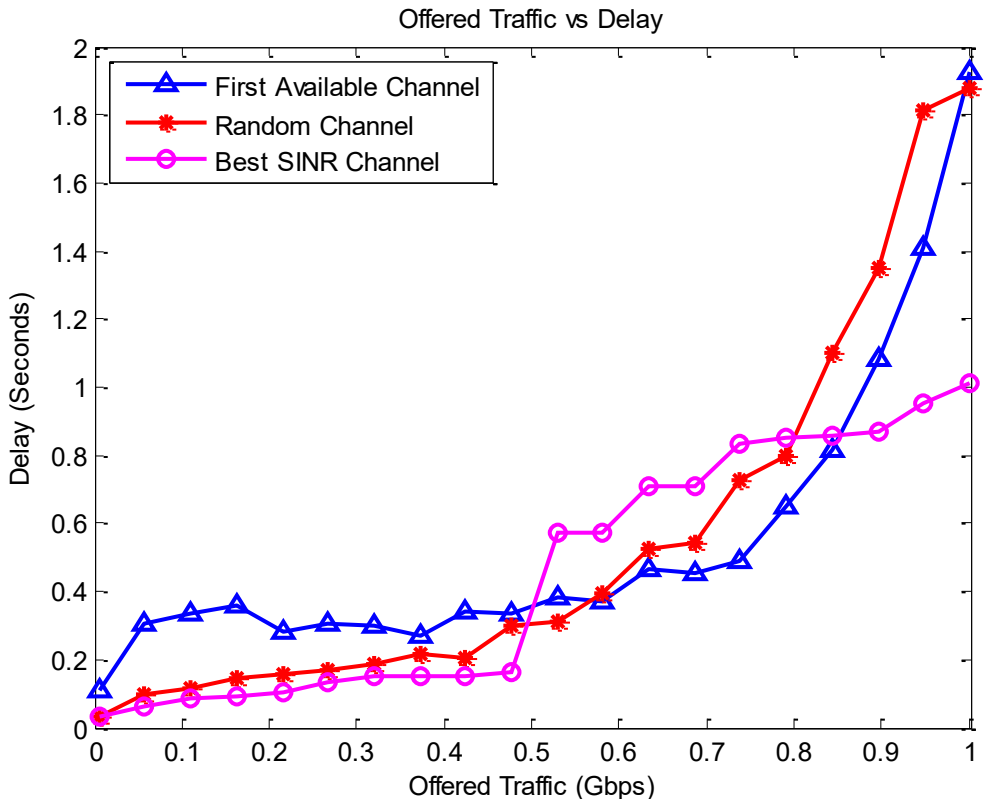


Figure 4.8. Average File Delay as a Function of Offered Traffic.

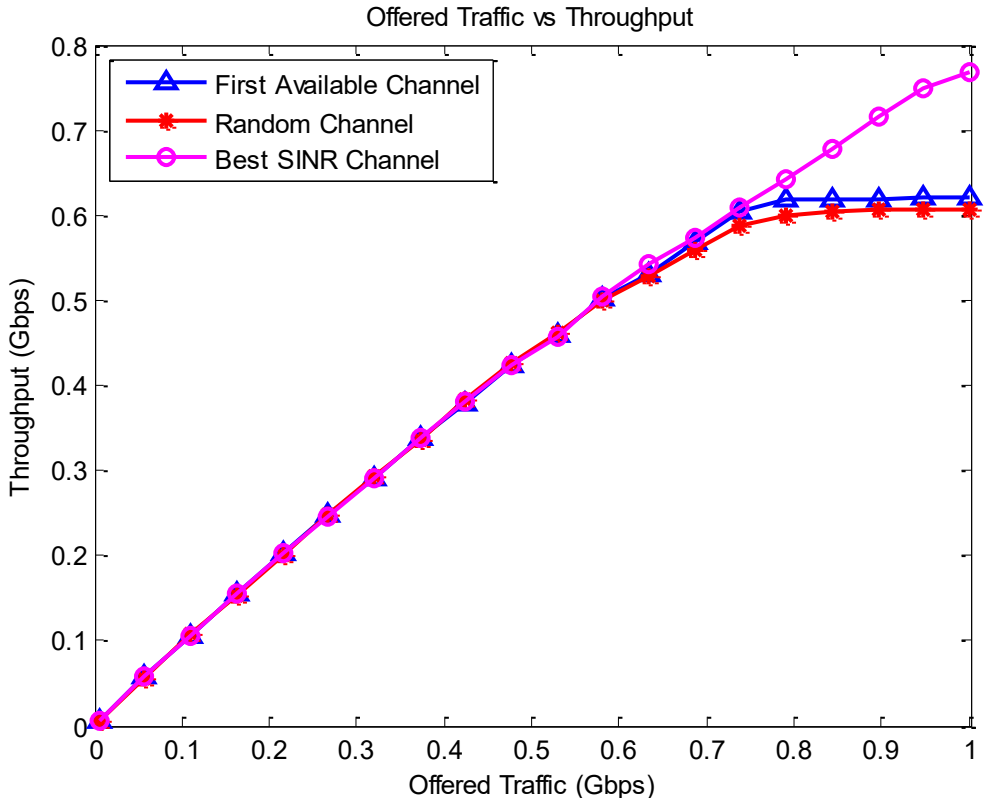


Figure 4.9. System Throughput as a Function of Offered Traffic.

The best SINR scheme also allows a gradual degradation in system performance as a result of the selection policy which is based on the best available resources. Figure 4.9,

shows that the best SINR scheme has the benefit of high quality channel choice reflected of the continuous increase of the throughput throughout the system functional range. Both the first available channel and the random channel assignment schemes have similar throughput trends with a slightly better throughput for the first available scheme due to the slightly higher capacity. This is due to the fact that blocking probability has a bigger influence on performance than outage probability. This has resulted in better performance for the lower blocking probability scheme. The same reason makes the first available scheme become better than the random assignment scheme at some point in the middle of the tested operational range from the delay point of view. The behaviour of the three schemes over the tested operational range becomes more obvious in figure 4.10 that shows the file delay value with respect to system throughput. It shows obviously the throughput limits for the three schemes as well as the increase in file delay as a price of a certain system throughput gained. Delay here is considered to be the difference between the minimum time required for file transmission (based on the maximum link transmission rate) and the actual time spent transmitting the file.

The essential reason behind using the FAC assignment technique is that it is a sequenced search procedure. In other words, it is a structured search process within which the searcher always knows what direction it should search for new channels. Usually such search procedures have certain direction and the search domain is reduced after each trial as the searcher knows what has been tested and what is left.

The RCA on the other hand is an unstructured search process. The tests for available channels completely random at every trial. As a result, the searcher has no idea about what has been tested and what is left. The search domain in this case is never reduced regardless of how many trials have been carried out.

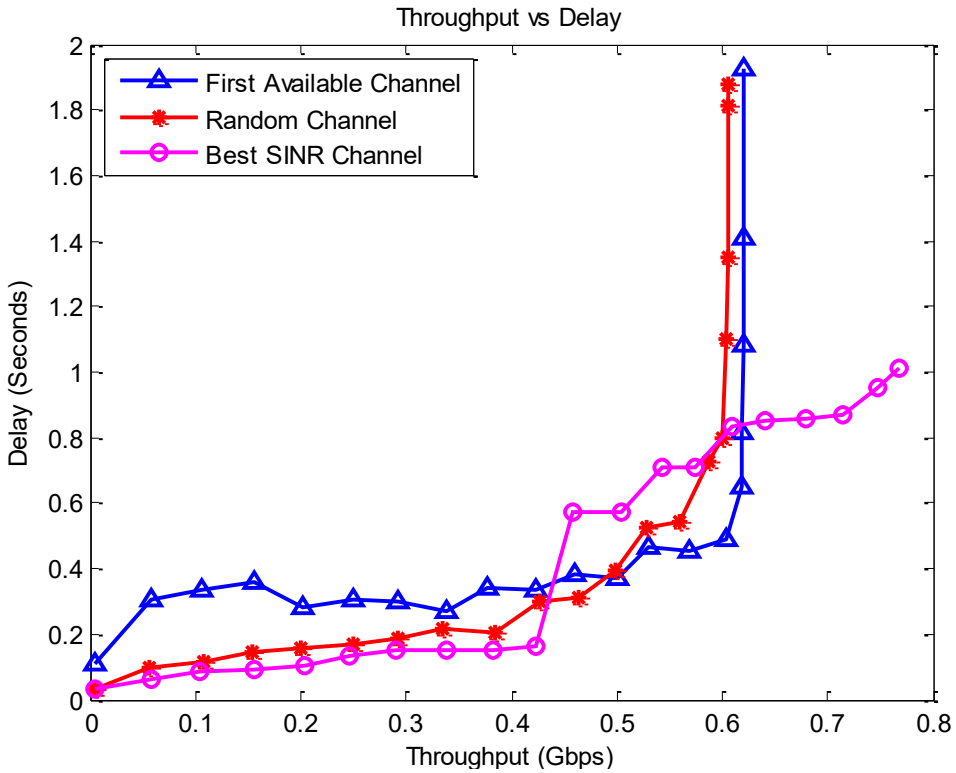


Figure 4.10. File delay as a Function of System Throughput.

Although, the sequenced search procedure in the FAC scheme provides a more consistent search that leads to a reduction in the possible solutions as the searcher tests more options, it is not practical to perform it for all ABSs as it is causes significant interference with adjacent cells. Moreover, it becomes slower as the traffic load increases and the number of possibly unoccupied channels decreases.

On the other hand, the RCA search process guarantees the variation among selected channels within different ABSs in most cases and as a result it reduces the interference effect. However, in cases of high traffic loads, the RCA becomes as slow as the FAC technique with the reduction of available channels and might become even slower occasionally as feedback from repeated channel tests do not result in reducing the search domain. This might make the search process last for much longer than expected.

The best SINR technique depends totally on the channel quality for choosing channels. However it is a relatively slow and power consuming process as it is necessary to scan and test all the available channels at every single trial.

4.6 Conclusion

This chapter provided an illustration of the simulation results of some traditional channel assignment techniques using the BuNGee architecture as a test platform. Sample parameters and the hardware used were both presented. Then, a basic simulation using Engset and Erlang B fomulas has been presented for simulation verification purposes. Three traditional dynamic channel assignment schemes were introduced. First available channel assignment, random channel assignment and best quality channel assignment schemes were simulated, compared and discussed. These simulations have been carried out for the purpose of further evaluation of the search technique modification to be presented in later chapters.

As a conclusion, the target search technique to be developed is preferred to have the advantages of the three tested assignment techniques. These are: the structure (sequence) of FAC, the variation (among searchers) as in RCA and the quality based search as in the best SINR.

Chapter 5 Quantum Computation and Quantum Search

5.1 Introduction.

In this chapter, a general theoretical background for the quantum-based development introduced into the reinforcement learning decision making process is presented. The

ideas upon which the hypothesis of the thesis is based on are provided. Improvement of the resource allocation process is accomplished through quantum search. It is used as the decision making process within the reinforcement learning engine. The new approach is aimed to re-define the search process to combine advantages of both sequential and random search techniques. This has to be accomplished through a sequential search based on channel quality sequence rather than channel number as in First Available Channel technique mentioned in chapter 4. Such approach can reduce the number of search trials even with few channels available. This makes the learning engine when combined with it a more efficient learning technique for highly dynamic problems. The present chapter has been produced to clarify the new search technique and the superiority of it over classical search procedures.

An overview of quantum computation basic elements is presented in section 5.2. An overview of the classical channel search techniques is presented in section 5.3. The definition, idea and geometric visualization of quantum search with a basic comparison with classical search are all provided in section 5.4. Simulation results for the application of the Grover search algorithm into dynamic channel access is given in section 5.5.

5.2 Quantum Computation.

5.2.1 The Bra-Ket Notation

In 1930, Dirac had developed his own approach of matrix and vector representation. He has defined two types of entities which are “Kets” and “Bras” which are simply column vectors and row vectors respectively. The elements of these vectors and matrices are generally complex numbers [125-128]:

$$\langle A| = [A_1 \quad A_2 \quad A_3]$$

$$|B\rangle = \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix}$$

The product of a bra and a ket, denoted by Dirac as $\langle A|B\rangle$ is simply the ordinary complex number given by:

$$\langle A|B\rangle = [A_1 \quad A_2 \quad A_3]\begin{bmatrix}B_1\\B_2\\B_3\end{bmatrix} = A_1B_1 + A_2B_2 + A_3B_3$$

If we have a linear operator α that is:

$$\alpha = \begin{bmatrix}\alpha_{11} & \alpha_{12} & \alpha_{13}\\ \alpha_{21} & \alpha_{22} & \alpha_{23}\\ \alpha_{31} & \alpha_{32} & \alpha_{33}\end{bmatrix}$$

We can form an ordinary complex number by taking the compound product of a bra, a linear operator, and a ket:

$$\langle A|\alpha|B\rangle = [A_1 \quad A_2 \quad A_3]\begin{bmatrix}\alpha_{11} & \alpha_{12} & \alpha_{13}\\ \alpha_{21} & \alpha_{22} & \alpha_{23}\\ \alpha_{31} & \alpha_{32} & \alpha_{33}\end{bmatrix}\begin{bmatrix}B_1\\B_2\\B_3\end{bmatrix}$$

For any given ket $|A\rangle$, there is a bra $\langle A|$, which is called the conjugate imaginary:

$$\langle A| = [A_1 \quad A_2 \quad A_3]$$

$$|A\rangle = \begin{bmatrix}\overline{A}_1\\ \overline{A}_2\\ \overline{A}_3\end{bmatrix}$$

5.2.2 The Quantum Superposition of States

The state of a physical system in quantum theory is specified by the state vector, the ket $|\psi\rangle$. Usually it is referred to $|\psi\rangle$ as the state of the system. If $|\psi_1\rangle$ and $|\psi_2\rangle$ are possible states, then the superposition of them [127, 128]:

$$|\psi\rangle = a_1|\psi_1\rangle + a_2|\psi_2\rangle \qquad (5\text{-}1)$$

is also a state of the system, where a_1 and a_2 are complex numbers. The superposition principle states that, any superposition of states is also a state. It is a fundamental concept

in quantum theory. The bra $\langle\psi|$ provides an equivalent representation of the state in equation (5-1) in the form:

$$\langle\psi| = a_1^*\langle\psi_1| + a_2^*\langle\psi_2| \tag{5-2}$$

Where a_1^* and a_2^* are the complex conjugates of a_1 and a_2 respectively. The use of probability amplitudes in quantum computation replaces the use of conventional probability values in classical computation. They can be obtained by forming an overlap between pairs of states (applying an inner product between them). The overlap between the states $|\psi\rangle$ and $|\phi\rangle$ is the complex number $\langle\psi|\phi\rangle$ or its complex conjugate $\langle\phi|\psi\rangle$, analogous to the scalar or dot product of two vectors. If this overlap is zero, then the states are said to be orthogonal, in analogy with a pair of perpendicular vectors, which have zero scalar product. The inner product of a state with itself is real and strictly positive so that:

$$\langle\psi|\psi\rangle > 0 \tag{5-3}$$

If this inner product is unity, so that $\langle\psi|\psi\rangle = 1$, then the state is said to be normalized. If the states in (5-1) are orthonormal, that is, both orthogonal ($\langle\psi_1|\psi_2\rangle = 0$) and normalized ($\langle\psi_1|\psi_1\rangle = 1 = \langle\psi_2|\psi_2\rangle$), then the amplitudes a_1 and a_2 are given by the overlaps:

$$\begin{aligned}\langle\psi_1|\psi\rangle &= a_1 = \langle\psi|\psi_1\rangle^* \\ \langle\psi_2|\psi\rangle &= a_2 = \langle\psi|\psi_2\rangle^*\end{aligned} \tag{5-4}$$

If $|\psi\rangle$ itself is normalized, then,

$$|a_1|^2 + |a_2|^2 = 1 \tag{5-5}$$

Where $|a_1|^2$ and $|a_2|^2$ are interpreted as the probabilities that a suitable measurement will find the system to be in the state $|\psi_1\rangle$ and $|\psi_2\rangle$ respectively. The generalization of (5-1) is:

$$|\psi\rangle = \sum_n a_n |\psi_n\rangle \tag{5-6}$$

Where if $|\psi\rangle$ is normalized and the states $|\psi_n\rangle$ are orthonormal, then:

$$\sum_n |a_n|^2 = 1 \tag{5-7}$$

That is $|a_n|^2$ is the probability that a suitable measurement will find a system that started in the state $|\psi\rangle$ to be in the state $|\psi_n\rangle$.

5.2.3 The Qubits

The fundamental unit used in quantum computing to represent data is the *quantum bit* (*qubit*). This replaces the classical bits used in classical computations. However, although the qubit has two states represented as $|0\rangle$ and $|1\rangle$ just like those for classical bits, the qubit can exist in either of these two states. This is in addition to the superposition state of $|0\rangle$ and $|1\rangle$ [1]. Thus, a qubit $|\psi\rangle$ is expressed as a combination of $|0\rangle$ and $|1\rangle$:

$$|\psi\rangle = \alpha\,|0\rangle + \beta|1\rangle \tag{5-8}$$

This simple equation represents the so called state superposition principle, where α and β are complex coefficients. When the mentioned qubit $|\psi\rangle$ is measured while being in a superposition state, the qubit system *collapses* into one of its basic states $|0\rangle$ or $|1\rangle$. However, no prior state determination can be made for the qubit state after collapse. The probability of the qubit to collapse to $|0\rangle$ is $|\alpha^2|$ or collapse to $|1\rangle$ with the probability of $|\beta^2|$. Both α and β are referred to as *probability amplitudes*. The magnitude and argument of the probability amplitude represent *amplitude* and *phase* respectively [1, 2]. Thus α and β should satisfy:

$$|\alpha^2| + |\beta^2| = 1 \tag{5-9}$$

A fundamental process in quantum computation is a unitary transformation U on the qubits. It can be applied to a superposition state affecting all of its basis vectors resulting in another superposition state from superposing the results of the basis vectors. This parallel effect is called *quantum parallelism*. If an input qubit $|z\rangle$ is in a superposition state:

$$|z\rangle = \alpha\,|0\rangle + \beta|1\rangle \tag{5-10}$$

The transformation U_z can be defined as:

$$U_z: |z,0\rangle \rightarrow\ |z, f(z)\rangle \tag{5-11}$$

Where $|z,0\rangle$ represents the joint input state with the first qubit in $|z\rangle$ and the second qubit in$|0\rangle$. While,$|z, f(z)\rangle$, represents the joint output state with the first qubit in $|z\rangle$ and the second qubit in $|f(z)\rangle$. From both (5-3) and (5-4), we can gain:

$$U_z|z,0\rangle = \alpha\,|0, f(0)\rangle + \beta|1, f(1)\rangle \tag{5-12}$$

The above equation represents the quantum *black box process* or *oracle*. Entering superposed quantum states into the oracle, leads to learning of what is inside it with a significant speedup compared with the case of classical inputs. In an n-qubit system represented by a tensor product of n-qubits:

$$|\phi\rangle = |\psi_1\rangle \otimes\ |\psi_2\rangle \otimes \ldots\ldots\ldots |\psi_n\rangle = \sum_{x=00\ldots0}^{11\ldots1} C_x\,|x\rangle \tag{5-13}$$

Which "⊗"means tensor product, $\sum_{x=00\ldots0}^{11\ldots1}|C_x^2| = 1$, C_x is a complex coefficient and $|C_x^2|$ represents the occurrence probability of $|x\rangle$ when the state $|\phi\rangle$ is measured. Computing the function $f(x)$ with the unitary transform U gives:

$$U \sum_{x=00\ldots0}^{11\ldots1} C_x\,|x,0\rangle = \sum_{x=00\ldots0}^{11\ldots1} C_x U\,|x,0\rangle = \sum_{x=00\ldots0}^{11\ldots1} C_x\,|x, f(x)\rangle \tag{5-14}$$

5.2.3 Quantum Gates

Quantum gates are essential arithmetic units used to accomplish quantum computational tasks. Two specific gates are used for the work in this thesis which are *Hadamard* and *phase* gates [127-131]. The Hadamard gate (or Hadamard transform) is one of the most widely used gates and can be represented as follows:

$$H \equiv \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix} \tag{5-15}$$

Using a Hadamard gate, a qubit can be transformed from state $|0\rangle$ into an equally weighted superposition state of $|0\rangle$ and $|1\rangle$

$$H|0\rangle \equiv \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}\begin{pmatrix}1\\ 0\end{pmatrix} = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle \tag{5-16}$$

The same result applies when Hadamard gate is applied to a qubit in state $|1\rangle$:

$$H|1\rangle \equiv \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}\begin{pmatrix}0\\ 1\end{pmatrix} = \frac{1}{\sqrt{2}}|0\rangle - \frac{1}{\sqrt{2}}|1\rangle \tag{5-17}$$

The other related quantum gate is the phase gate (conditional phase shift gate). This gate is quite vital within the quantum search algorithm. It is a tool to reinforce the good decision or the good search result. The transformation describing this gate is [108, 131]:

$$U_{phase} = \begin{pmatrix}1 & 0\\ 0 & e^{i\varphi}\end{pmatrix}$$

Where $i = \sqrt{-1}$, and φ is an arbitrary real number.

5.3 Classical Channel Search.

One of the universally needed tasks in engineering and computer science is to *search*. A significant number of scientific problems can be resolved by counting the number of possible solutions, and then randomly or systematically searching these possibilities to find the correct or best one. In some cases, the determination of wrong solutions can help

eliminate them earlier and thus facilitating faster and narrower search for the ultimate solution. Such type of search problems are referred to as *structured problems* [129]. A practical example of such search process is the search for available channels in the FAC assignment scheme discussed in chapter 4.

For other problems, finding some wrong solutions might not be useful for learning anything. The only learned fact in this case is that these solutions are wrong and cannot be chosen again. Such problems are said to be *unstructured problems*. Thus, the unstructured problems are the *find-the-needle-in-the-haystack* problems [129]. A good example of such type of search processes is the search for available channels in the RCA scheme discussed in chapter 4.

The concept of unstructured search can be demonstrated using our channel assignment problem performed by the ABS. A traditional way of assignment is starting the check for available (unoccupied) frequencies within an ABS starting from the first channel (frequency band). In such a case, it is obvious what the next step will be in case the tried channel is not usable. It is simply turning to the next channel (higher or lower frequency depending on the starting channel) until finding a good channel. Such a technique is well known as the *first-available channel assignment* technique. For a situation like this, it is expected to find the appropriate channel after trials of roughly $O(\log N)$ where N, is the number of channels. This problem is said to have a complexity of $O(\log N)$. This looks efficient roughly because the ABS will know which channel to try next as it tries channels sequentially regardless whether this technique yields the aimed performance in our case or not.

On the other hand, if the ABS simply starts assigning the last tried (or most successful) channel, there will be no indication which channel it will check next. This is obviously because channels have no fixed or standard quality sequence based on how successful

they are and which also might change continuously. Such a case turns the search process into a random one like in the traditional *random channel assignment scheme* and is essentially referred to as generate-and-test process. If there are N theoretically available channels within the ABS, it will take an average, $O(N/2)$ repetitions of the algorithm to find an appropriate channel. However reinforcement learning techniques have introduced the principle of a Q-table which gave a more reliable reference for qualified channels. The changes in the sequence of channel priorities within such a table proved to be much slower than changes experienced within the wireless environment [127, 128, 130].

In an unstructured problem for finding the best candidate among a set of N channels labelled with indices x in the range,

$$0 \leq x \leq N - 1,$$

And the index of the target sought after channel is,

$$x = t$$

Now, a computational function $f_t(x)$, is presented for which when an index x is presented, it can give a result showing whether it is the index of the searched after channel or not. In specific, $f_t(x)$, is defined:

$$f_t(x) = \begin{cases} 0 & if \ \ x \neq t \\ 1 & if \ \ x = t \end{cases} \tag{5-18}$$

Where 0 stands for "no" and 1 stands for "yes". If there are N indicies, they can be expressed in binary notation using $(n = log_2 N)$ qubits. To create the equally weighted superposition state, a 1-qubit Walsh-Hadamard gate H is applied to each of the n qubits prepared initially in the $|0\rangle$ state. This means performing the operation [128-130, 132],

$$|00 \ldots 0\rangle \xrightarrow{H^{\otimes n}} \frac{1}{\sqrt{2^n}} \sum_{x=0}^{2^n-1} |x\rangle \tag{5-19}$$

When this superposition is read, a single index is non-deterministically obtained. This simple process mimics the classical generate-and-test procedure.

Now, an arbitrary starting channel $|\psi\rangle$ is picked and an operator U is applied to it such that $U|\psi\rangle$ is guaranteed to have some non-zero component in $|x'\rangle$, where,

$$U = H^{\otimes n}$$

$$|\psi\rangle \in |00 \dots 0\rangle$$

This will ensure a non-zero overlap between the unknown target $|x'\rangle$ and $U|\psi\rangle$, which means,

$$\langle x'|U|\psi\rangle \neq 0 \tag{5-20}$$

Each time $U|\psi\rangle$ is measured, the probability of finding $|x'\rangle$ is given by the modulus squared of the overlap between $|x'\rangle$ and $U|\psi\rangle$ [129]. This means,

$$P_{succ}^{Classical} = |\langle x'|U|\psi\rangle|^2 \tag{5-21}$$

Based on standard statistical theory, it is inferred that we might need roughly $|\langle x'|U|\psi\rangle|^{-2}$ to find the solution with probability of $O(1)$(i.e., near certainty). Thus, this is the "classical" complexity for an unstructured search for a channel using generate-and-test.

5.4 Quantum Channel Search: Grover's Algorithm

5.4.1 The Oracle

An oracle is a unitary operator,O, that allows the estimate of the computational cost of some algorithm measured in units of "the number of calls to the oracle" [130]. It a is black box where the internal workings of it have the ability to recognize the solutions to the search problem by making use of an oracle *qubit* [128-130].

This enables the comparison of the relative costs of classical unstructured search versus quantum unstructured search in terms how many times each algorithm must call the oracle. The issue is not whether the solution to some search problem is or is not known in advance of the search, but rather how many times we must query the knowledge-holder before we learn the solution.

In the abstract unstructured search problem the knowledge holder is the oracle, or "black-box function" $f_t(x)$ [130, 133]. Suppose we want to search through a space of N elements. In such a case we concentrate on the *index* of the elements which is a number that lies in the range 0 to $N-1$. If the index is stored in $n$bits, we assume that $N = 2^n$ and the search problem shall have M solutions, with $1 \leq M \leq N$. A function of the search problem $f$which takes an integer input x in the range $0\ to\ N-1$ shall have a solution $f(x) = 1$ if x is a solution to the problem or $f(x) = 0$ if x is not a solution to the search problem [128].

Now, if we are supported with a quantum *oracle*:

$$|x\rangle\,|q\rangle \xrightarrow{O} |x\rangle\,|q \oplus f(x)\rangle \tag{5-22}$$

Where $|x\rangle$ is the index register, $\oplus$ denotes addition modulo 2, and the oracle qubit $|q\rangle$ is a single qubit which is flipped if $f(x) = 1$, and is unchanged otherwise. It can be checked whether x is a solution to the problem by preparing $|x\rangle|0\rangle$, applying the oracle, and checking to see if the oracle qubit has been flipped to$|1\rangle$.

In the oracle of a quantum search algorithm, the oracle qubit is initialized in the state $(|0\rangle - |1\rangle)/\sqrt{2}$. If x is not a solution to the problem, applying the oracle to the mentioned state does not change the state. However, if x is a solution to the problem, then $|0\rangle$ and $|1\rangle$ are interchanged by the action of the oracle, resulting in a final state $-|x\rangle(|0\rangle - |1\rangle)/\sqrt{2}$. Thus, the action of the oracle is [133]:

$$|x\rangle \left(\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right) \xrightarrow{O} (-1)^{f(x)}|x\rangle \left(\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right) \tag{5-23}$$

The state of the oracle here is not changed and remains $(|0\rangle - |1\rangle)/\sqrt{2}$ throughout the quantum search algorithm. With this fact, the action of the oracle can be written:

$$|x\rangle \xrightarrow{O} (-1)^{f(x)}|x\rangle \tag{5-24}$$

The oracle marks the solutions to the search problem, by shifting the phase of the solutions.

5.4.2 The Search Procedure

The problem of an ABS seeking to assign a proper channel among n available channels can be represented by a function $f(x)$ with $x \in \{0,1\}^n$. If the channels are in general denoted by $|x\rangle$, the sought channel is denoted by $|x'\rangle$, such that:

$$f(x) = \begin{cases} 1 & if\ x = x' \\ 0 & if\ x \neq x' \end{cases} \tag{5-25}$$

Making the task of Grover algorithm is to find $|x'\rangle$. In other words, it is to find a single (or more) bit, x that is equal to x' to get an output that is equal to 1.

The algorithm starts with the n-channels amplitude register in the state $|0\rangle^{\otimes n}$ [127-130, 133-135]. This register is to be transformed into a superposition state of equal amplitudes using the Hadamard transform:

$$|\psi\rangle = \frac{1}{\sqrt{N}} \sum_{x \in \{0,1\}^n} |x\rangle \tag{5-26}$$

Which means equal probabilities for all channels. This superposition includes the sought after channel(s) $|x'\rangle$ so that:

$$\langle x'|\psi\rangle = \frac{1}{\sqrt{N}} \sum_{x\in\{0,1\}^n} \langle x'|x\rangle = \frac{1}{\sqrt{N}} \tag{5-27}$$

By excluding $|x'\rangle$ which can be represented:

$$|x'\rangle = \frac{1}{\sqrt{M}} \sum_{x\in\{0,1\}^n, x=x'} |x\rangle$$

$$|\psi'\rangle = \frac{1}{\sqrt{N-M}} \sum_{x\in\{0,1\}^n, x\neq x'} |x\rangle \tag{5-28}$$

We define two operators. The first is the oracle which has an action that is described in (5-21) by [128-130]:

$$U_f = \sum_{x\in\{0,1\}^n} (-1)^{f(x)} |x\rangle\langle x| = \sum_{x\in\{0,1\}^n} (-1)^{\delta_{x,x'}} |x\rangle\langle x| \tag{5-29}$$

Where:

$$\delta_{x,x'} = \begin{cases} 1 & if \;\; x = x' \\ 0 & if \;\; x \neq x' \end{cases} \tag{5-30}$$

is the Kronecker delta function. Now, we define another operator:

$$U_s = 2|\psi\rangle\langle\psi| - I \tag{5-31}$$

By splitting $|\psi\rangle$ into two parts, the part containing $|x'\rangle$ and the second part is the rest $|\psi'\rangle$ as in (5-20), we get [128-130]:

$$|\psi\rangle = \sqrt{\frac{N-M}{N}} |\psi'\rangle + \sqrt{\frac{M}{N}} |x'\rangle \tag{5-32}$$

By inverting (5-25):

$$|x'\rangle = \sqrt{N}|\psi\rangle - \sqrt{N-M}|\psi'\rangle \tag{5-33}$$

Thus, the application of the reflection transform on the vector of the sought after channel(s) [129]:

$$U_s|x'\rangle = (2|\psi\rangle\langle\psi| - I)(\sqrt{N}|\psi\rangle - \sqrt{N-M}|\psi'\rangle)$$

$$= 2\sqrt{N}|\psi\rangle\langle\psi|\psi\rangle - \sqrt{N}|\psi\rangle - 2\sqrt{N-M}|\psi\rangle\langle\psi|\psi'\rangle + \sqrt{N-M}|\psi'\rangle$$

$$= \sqrt{N}|\psi\rangle - 2\sqrt{N-M}\sqrt{\frac{N-M}{N}}|\psi\rangle + \sqrt{N-M}|\psi'\rangle \quad (5\text{-}34)$$

Substituting $|\psi\rangle$ from (5-24):

$$U_s|x'\rangle = \frac{2\sqrt{N-M}}{N}|\psi'\rangle + \left(\frac{2}{N} - 1\right)|x'\rangle \quad (5\text{-}35)$$

Also the application of the reflection transform on the part that does not contain $|x'\rangle$:

$$U_s|\psi'\rangle = -\left(\frac{2}{N} - 1\right)|\psi'\rangle + \frac{2\sqrt{N-M}}{N}|x'\rangle \quad (5\text{-}36)$$

Now, we define an angle θ such that:

$$Sin\theta = \frac{2\sqrt{N-M}}{N} \quad (5\text{-}37)$$

$$Cos\theta = -\left(\frac{2}{N} - 1\right) \quad (5\text{-}38)$$

Which means that (5-27) and (5-28) are rotations that is:

$$U_s|x'\rangle = -Cos\theta|x'\rangle + Sin\theta|\psi'\rangle \quad (5\text{-}39)$$

$$U_s|\psi'\rangle = Sin\theta|x'\rangle + Cos\theta|\psi'\rangle \quad (5\text{-}40)$$

Now, if the Grover operator is defined as being:

$$G = U_s U_f$$

And apply it on the same vectors; we obtain another form of rotation [128-130]:

$$G|x'\rangle = Cos\theta|x'\rangle - Sin\theta|\psi'\rangle \quad (5\text{-}41)$$

$$G|\psi'\rangle = Sin\theta|x'\rangle + Cos\theta|\psi'\rangle \quad (5\text{-}42)$$

The result then appears to be that the Grover operator rotates the initial state into the desired result or solution $|x'\rangle$. Practically, the rotation is done for only a small angle per application. Thus we need to apply this operator for m times to reach the desired solution state:

$$G^m|x'\rangle = Cos\ m\theta|x'\rangle - Sin\ m\theta|\psi'\rangle \quad (5\text{-}43)$$

$$G^m|\psi'\rangle = Sin\ m\theta|x'\rangle + Cos\ m\theta|\psi'\rangle \quad (5\text{-}44)$$

5.4.3 The Geometrical Visualization

The Grover iteration or operator is considered to be a rotation in a two dimensional space spanned by a starting vector $|\psi\rangle$ and the state consisting of a uniform superposition of solutions to the search problem. Thus, the initial state of the quantum system is the space spanned by $|\psi'\rangle$ and $|x'\rangle$. The effect of G is understood by realizing that a reflection is made about the vector $|\psi'\rangle$ in the plane defined by $|\psi'\rangle$ and $|x'\rangle$ as in the illustrated steps in figure.5.1 [128, 130, 136]:

1- Initially, it is inclined at angle $(\theta/2)$ from $|\psi'\rangle$, a state orthogonal to $|x'\rangle$.

2- An oracle operation ($O = U_f$) reflects the state $|\psi\rangle$ about the state$|\psi'\rangle$ to be in the $U_f|\psi\rangle$ state.

3- The reflection transform U_s reflects $U_f|\psi\rangle$ around the initial state $|\psi\rangle$ to be in the $U_s|\psi\rangle$ state.

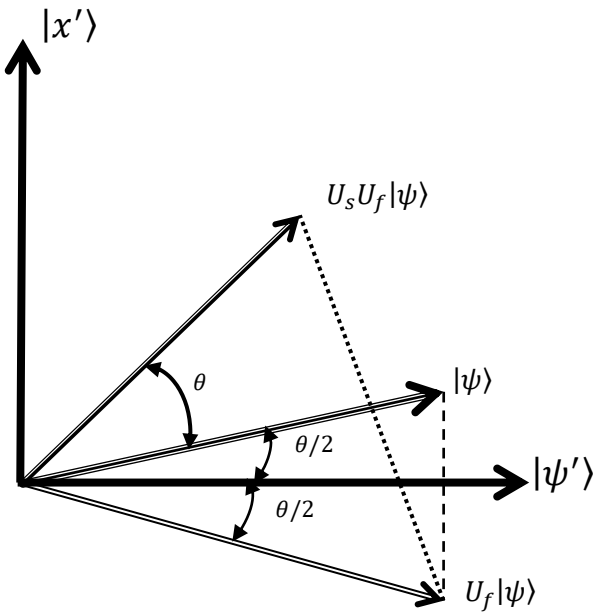


Figure 5.1. The Action of a Single Grover Iteration $G = U_sU_f$

The product of the mentioned two reflections is a rotation. The ultimate goal is to reach the solution $|x'\rangle$ state. Only a small rotation is made per iteration. Thus multiple Grover operator applications are needed. This means, that the state $G^k|\psi\rangle$ remains in the space spanned by $|\psi'\rangle$, and $|x'\rangle$ for all k. It also gives the rotation angle [128]. Let:

$$cos\,\theta/2 = \sqrt{(N-M)/N} \tag{5-45}$$

So that:

$$|\psi\rangle = cos\,\theta/2|\psi'\rangle + sin\,\theta/2\,|x'\rangle \tag{5-46}$$

Figure (5-1) shows the two reflections which comprise Gtake $|\psi\rangle$ to:

$$G|\psi\rangle = cos\frac{3\theta}{2}|\psi'\rangle + sin\frac{3\theta}{2}|x'\rangle \tag{5-47}$$

So, the rotation angle is in fact θ. It follows that continued application of G takes the state to:

$$G^k|\psi\rangle = cos\left(\frac{2k+1}{2}\theta\right)|\psi'\rangle + sin\left(\frac{2k+1}{2}\theta\right)|x'\rangle \quad (5\text{-}48)$$

Summarizing, G is a rotation in the two dimensional space spanned by $|\psi'\rangle$ and$|x'\rangle$, rotating the space by θ radians per application of G [128, 129]. Repeated application of the Grover iteration rotates the state vector closer to$|x'\rangle$. When this occurs, an observation in the computational basis produces with high probability one of the outcomes superposed in $|x'\rangle$, that is, a solution to the search problem [128-130].

Now, as explained in section 5-3, each measurement of channel test process is actually an overlap between $|\psi\rangle$ and $|x'\rangle$ such that [3]:

$$\langle x'|U|\psi\rangle \neq 0 \quad (5\text{-}49)$$

In the quantum search case, this step comes with the application of the Grover operator as shown in (5-41) which shows that:

$$G^k|\psi\rangle = \langle x'|G^kU|\psi\rangle \approx \left(\frac{2k+1}{2}\right)\langle x'|U|\psi\rangle \quad (5\text{-}50)$$

Which indicates that the overlap (i.e. chance to find a proper channel) grows roughly linearly with the number of Grover operator applications k. Thus the probability of finding a good channel grows quadratically with the number of channel checks after each application of Grover operator which means [130]:

$$P_{succ}^{Quantum} \sim k^2|\langle x'|U|\psi\rangle|^2 \quad (5\text{-}51)$$

Compared with a probability for the classical generate-and-test that is as mentioned before:

$$P_{succ}^{Classical} = k|\langle x'|U|\psi\rangle|^2 \quad (5\text{-}52)$$

Therefore, the process of amplification resulted from the application of the Grover operator has the effect of increasing the chance of finding the proper channel within a frequency pool.

The other important feature of amplitude amplification that can be recognized from (5-41) is that the overlap between the target channel and the amplitude amplified state oscillates. As a result, it is quite possible to over-amplify and reduce the probability to find the channel.

5.4.4 The Number of Needed Grover Iterations

The initial state of the system is [128-130]:

$$|\psi\rangle = \sqrt{(N-M)/N}\,|\psi'\rangle + \sqrt{M/N}\,|x'\rangle \tag{5-53}$$

So, rotating through $arccos\sqrt{M/N}$ radians takes the system to $|x'\rangle$. Let $CI(x)$ denote the integer closest to the real number x, where by convention we round halves down. Then repeating the Grover iteration [128, 130]:

$$R = CI\left(\frac{arccos\sqrt{M/N}}{\theta}\right) \tag{5-54}$$

Times rotates $|\psi\rangle$to within an angle

$$\theta/2 \leq \pi/4$$

of $|x'\rangle$. Observation of the state in the computational basis then yields a solution to the search problem with probability at least one-half. Indeed, for specific values of M and N it possible to achieve a much higher probability of success. As an example, when $M \ll N$ we have [128, 129]:

$$\theta \approx sin\theta \approx 2\sqrt{M/N},$$

giving a probability of error of at most M/N. Note that R depends on the number of solutions M, but not on the identity of those solutions, so provided we know M we can apply the quantum search algorithm as described. The equation (5-42) is useful as an exact expression for the number of the oracle calls used to perform the quantum search algorithm, but it would be useful to have a simpler expression summarizing the essential behaviour of R. Noting that from (5-42) that

$$R \leq [\pi/2\theta],$$

so a lower bound on θ will give un upper bound on R. Assuming for the moment$M \leq N/2$, we have:

$$\frac{\theta}{2} \geq sin\frac{\theta}{2} = \sqrt{\frac{M}{N}} \tag{5-55}$$

From which we obtain an elegant upper bound on the number of iterations required [128, 129, 136, 137],

$$R \leq \left[\frac{\pi}{4}\sqrt{\frac{N}{M}}\right] \tag{5-56}$$

That is, $R = O\left(\sqrt{N/M}\right)$ Grover iterations (and thus oracle calls) must be performed in order to obtain a solution to the search problem with high probability, a quadratic improvement over the $O(N/M)$ oracle calls required classically. The Grover quantum search algorithm [127-130, 132] is summarized next, for the case of $M = 1$.

Algorithm: Grover Quantum Search

Inputs

1 A black box oracle O which performs the transformation $O|x\rangle|q\rangle = |x\rangle|q \oplus f(x)\rangle$, where $f(x) = 0$ for all $0 \leq x < 2^n$ except x_0, for which $f(x_0) = 1$.

2 $n + 1$ qubits in the state $|0\rangle$.

Outputs x_0

Runtime $O(\sqrt{2^n})$ operations. Succeeds with probability $O(1)$.

Procedure

1	$\|0\rangle^{\otimes n}\|0\rangle$	Initial state
2	$\rightarrow \frac{1}{\sqrt{2^n}} \sum_{x=0}^{2^n-1} \|x\rangle \left[\frac{\|0\rangle - \|1\rangle}{\sqrt{2}}\right]$	Apply $H^{\otimes n}$to the first n qubits,
3	$\rightarrow [(2\|\psi\rangle\langle\psi\| - I)O]^R \frac{1}{\sqrt{2^n}} \sum_{x=0}^{2^n-1} \|x\rangle \left[\frac{\|0\rangle - \|1\rangle}{\sqrt{2}}\right]$ $\approx \|x_0\rangle \left[\frac{\|0\rangle - \|1\rangle}{\sqrt{2}}\right]$	Apply the Grover iteration $R \approx \left[\pi\sqrt{2^n}/4\right]$ times.
4	$\rightarrow x_0$	Measure the first n qubits

5.4.5 Cases When More Than Half The Channels are Good Channels

From the expression

$$\theta = \arcsin(2\frac{\sqrt{M(N-M)}}{N}) \quad (5\text{-}57)$$

It is obvious that the rotation angle θ gets smaller as M increases from $N/2$ to N. Thus, the number of needed Grover iterations needed by the search algorithm increases with M, for $M \geq N/2$ which the opposite to one might expect in such a case as many solutions might indicate easier search [128, 129, 132].

There are two ways to look at this problem in such a situation. If M(number of suitable channels) is known in advance to be larger than $N/2$ then it might be viable to randomly pick an item from the search space, and then check that it is a solution using the oracle. This approach has a success probability at least 50% and only requires one application of the oracle. A disadvantage of such an approach is that we may not know the number of available or good channels M in advance.

In the case when it is not clear whether $M \geq N/2$ or not, another solution can be used. The idea used in such a case is to double the number of elements in the search space by adding N extra channels to the search pool. None of the added channels are possible solutions. This doubles the number of channels to be searched to $2N$. The new search problem has only M solutions out of $2N$ entries. Thus, applying the quantum search algorithm yields the result that at most [128]:

$$R = \pi/4\sqrt{2N/M}$$

Grover iterations are required, and it follows that $O(\sqrt{N/M})$ applications of G are required to perform the search.

5.5 Simulations for Quantum Search

5.5.1 System Performance

The Grover search algorithm has been tested as a dynamic spectrum allocator. Plots for system performance metrics were done and compared to the previously tested traditional channel assignment schemes in chapter 4. A weight-driven reinforcement learning algorithm [9] that will be discussed in the following chapter has been compared to quantum search as well. The Grover algorithm has been tested as a search mechanism to explore its validity for use to develop learning techniques later on. The topology used for quantum search simulations is illustrated in figure 5.2. Some sample parameters used within simulations in the current section are given in table 5.1. The flow chart that represents the application of QS algorithm is shown in figure 5.3.

Table 5.1. Parameters for Quantum Search Simulations

Parameter	Value
No. of ABSs	20
Number of Beams/ABS	2
No. of Users	260
Number of Channels /ABS	8
Maximum ABS gain	17 dBi
SINR Threshold	1.8 dB
Maximum SINR	21 dB
Maximum Transmission Rate	4.5 MB/s
Noise Floor	-112 dBm/MHz
MS Antenna Gain	0 dB
MS Transmission Power	23 dBm
MS Antenna Height	1.5 m
ABS Antenna Height	5 m
File Size	4MB

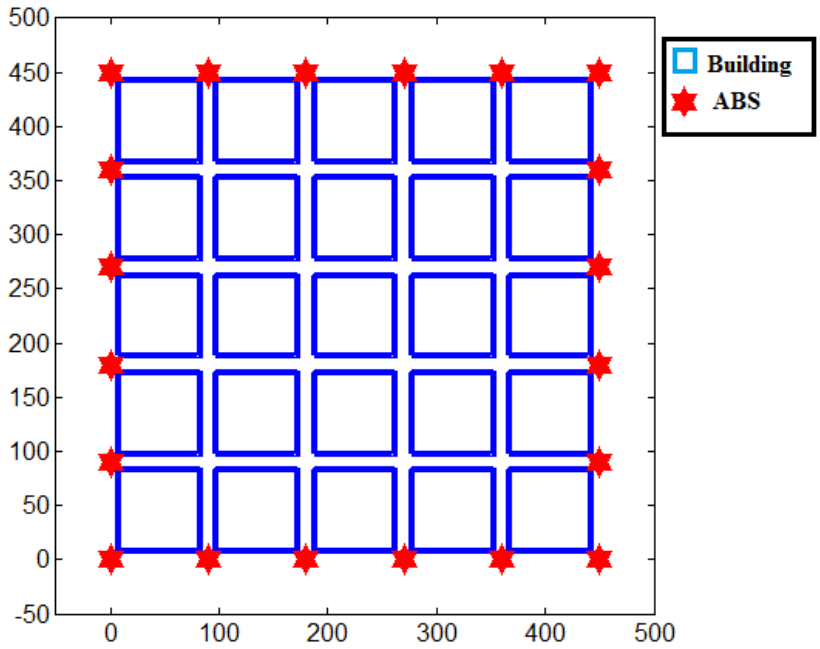


Figure 5.2. The Square Topology Used for Quantum Search Simulations

As seen from figure 5.2, a smaller scale square topology has been used for simulations. This is to insure a fare comparison for classical techniques against the QS scheme within an environment that they can highly perform within.

Figure 5.4, represents blocking probability for different traffic load values for the tested schemes.

It can be seen from the graph that although the Grover algorithm is not considered as a learning scheme, it is a competent scheme to the reinforcement learning scheme used for comparison. The small value of blocking probability at low traffic load values for the reinforcement learning schemes (RL) are due to collisions during learning period. The same behaviour is exhibited by the Grover quantum search (QS) scheme as it acts as a learner. In the case of the QS scheme, a probability table (register) is created [138]. This table is referred to as the amplitude table. It represents an alternative to the weight (or Q-table) in reinforcement learning. It includes an amplitude (action desirability) value for each action (channel). It is set initially to equal amplitudes (equal probability of all channels). The ABS starts by choosing a channel randomly only once.

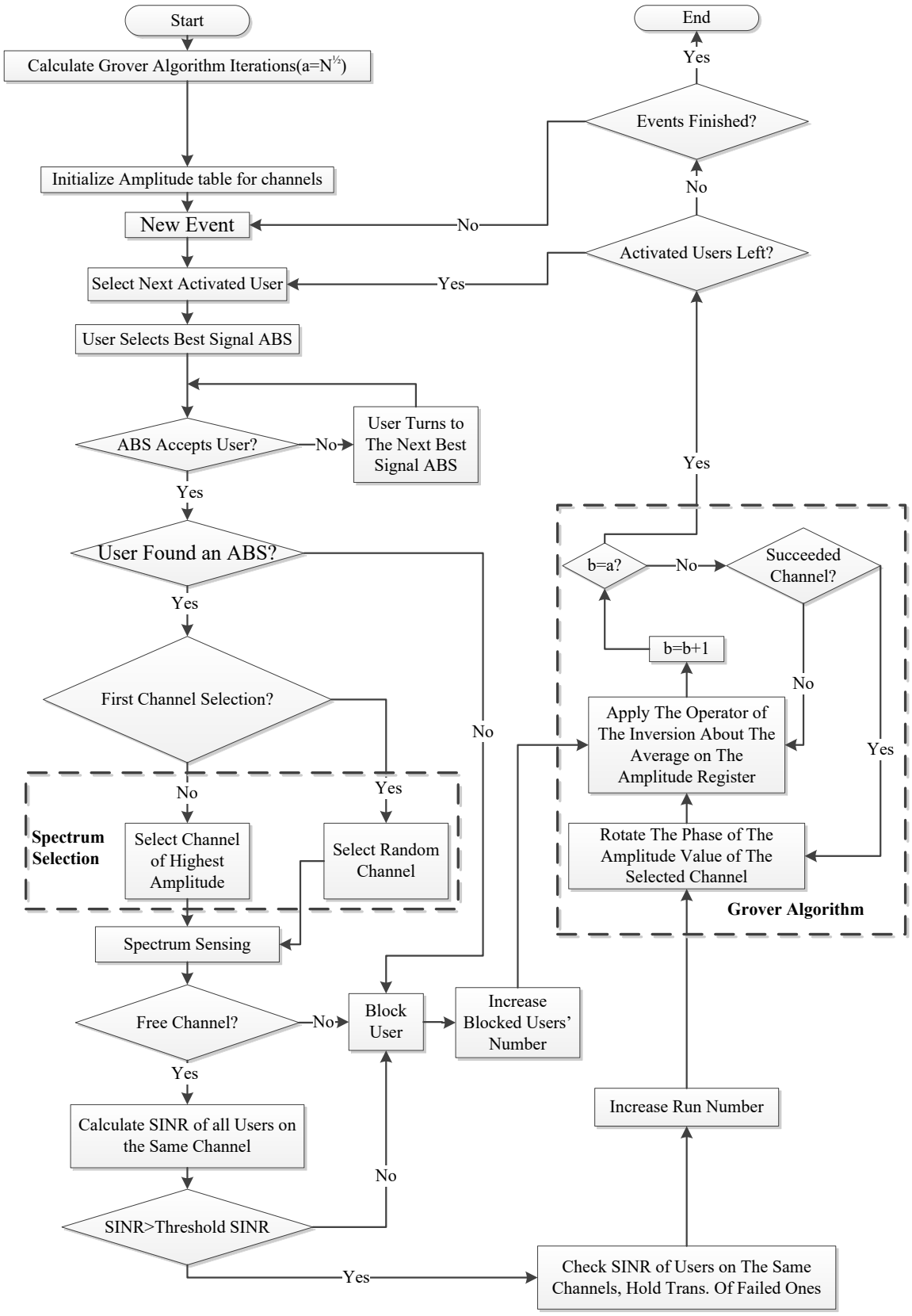


Figure 5.3. Flowchart of the Quantum Search (Grover Algorithm) Channel Allocation Scheme

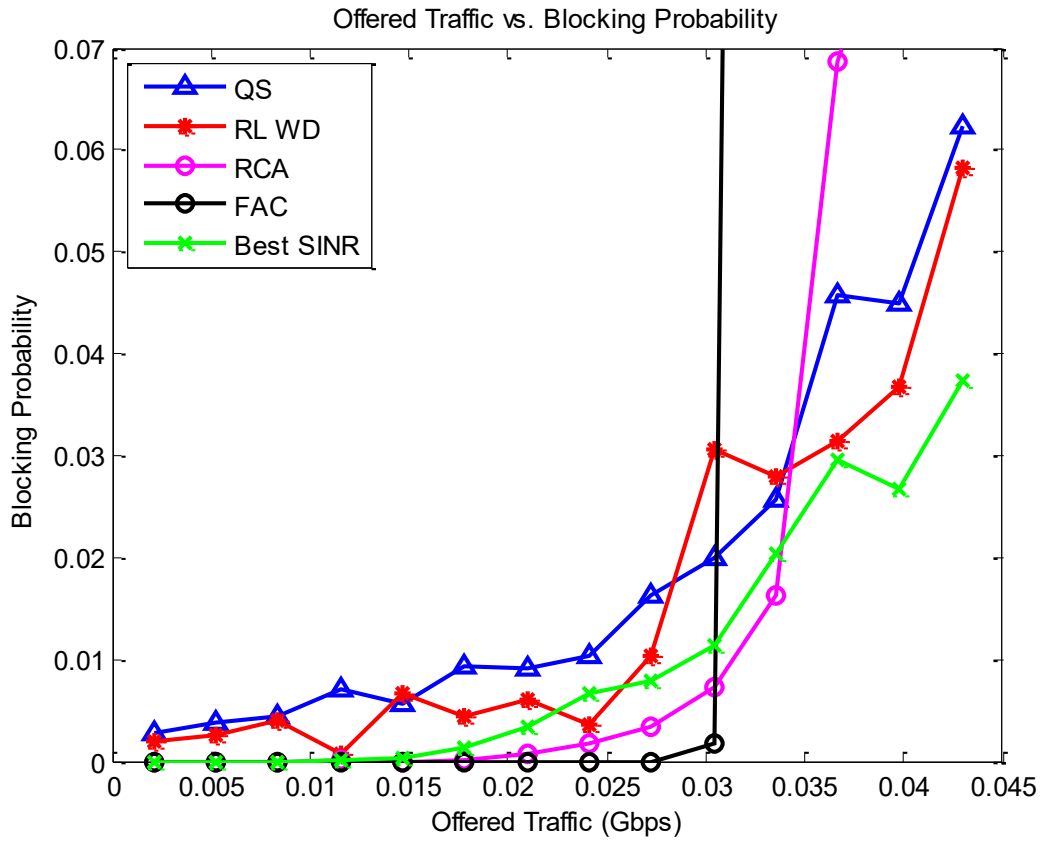


Figure 5.4. Offered Traffic vs. Blocking Probability

Later, the highest amplitude value channel is chosen all the way. When a channel is successfully chosen, the phase (amplitude sign) is inverted. This is a representation for the selection process that discriminates the successfully chosen channels. The phase of the unsuccessfully chosen channel is not inverted. An inversion operator (equation 6-8) is applied later on the amplitude table to invert values about their mean. It is a form of normalization process that gives rise to the successful channel while supresses the rest. This process of phase inversion and normalization has a mild amplification effect as the Grover algorithm has essentially been designed as a single solution finder tool. In other words, it has not been designed for a repetitive search process where the solution might change continuously. However, solutions are ranked by this process as more than one solution might exist. The same algorithm is explained later on as a part of the complete quantum inspired reinforcement learning scheme.

The adopted amplitude table within Grover search algorithm gives the algorithm a similar behaviour result to a learning-based scheme. In other words, it is obviously adaptable to the increase in traffic load and does not tend to collapse as in the case of the traditional channel assignment schemes. Traditional assignment schemes exhibits a point of collapse where a significant drop in performance occurs at some point when the system becomes unable to efficiently allocate the available spectrum.

Figure 5.5, illustrates the behaviour of outage probability with different traffic load values for the tested schemes. The efficient performance of the introduced Grover quantum search algorithm is obvious in comparison to the conventional assignment scheme. This indicated a reasonable control over interference levels which keeps an almost stable outage level over the system functional range (blocking probability<5%). This result is reflected on the average file transmission value delay in figure 5.6 and figure 5.7. The average of file transmission delay is usually a result of interference that

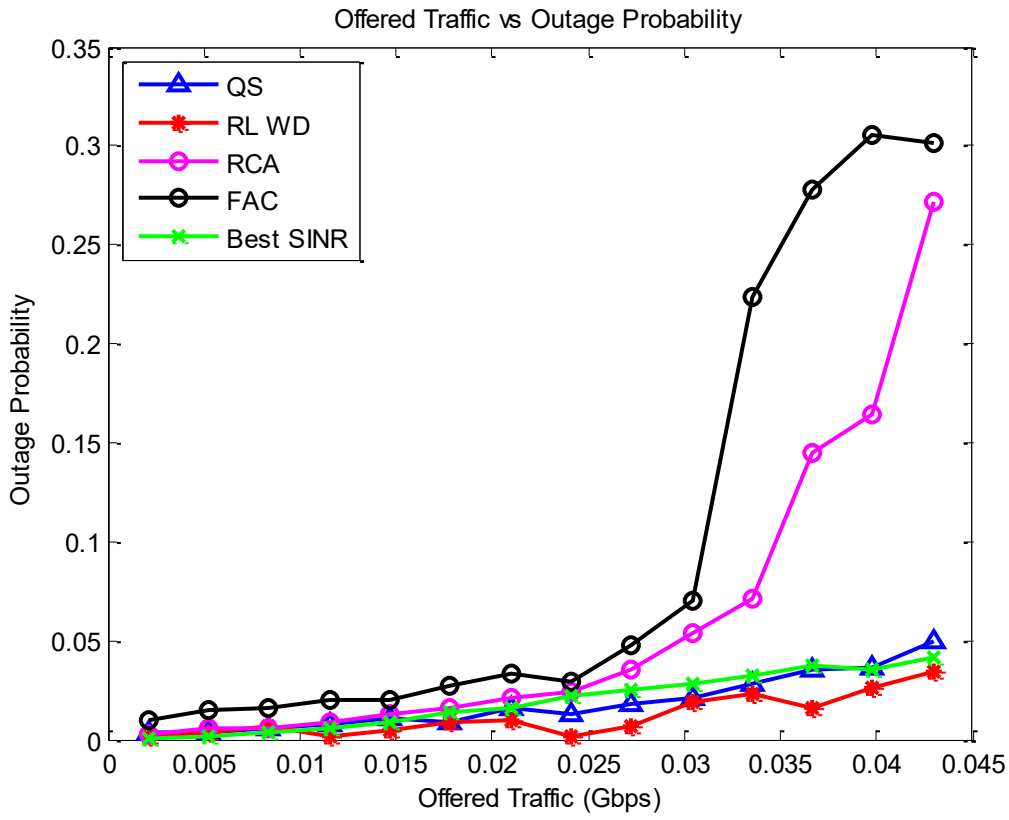


Figure 5.5. Offered Traffic vs. Outage Probability

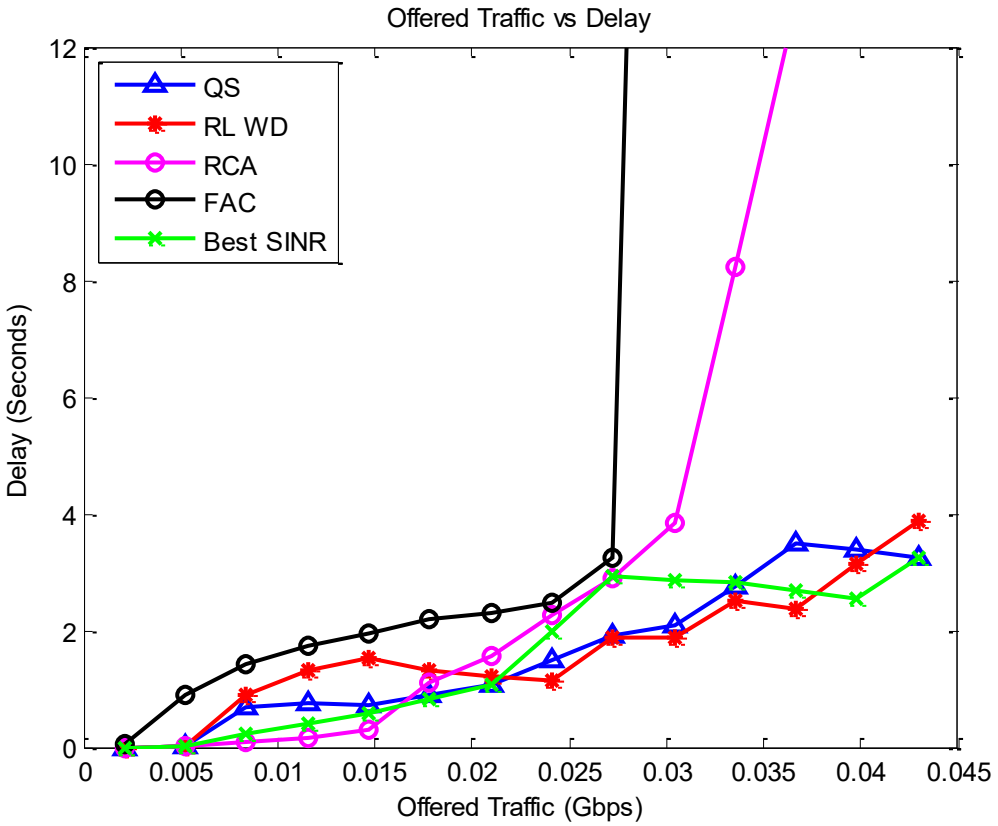


Figure 5.6. Offered Traffic vs. Delay

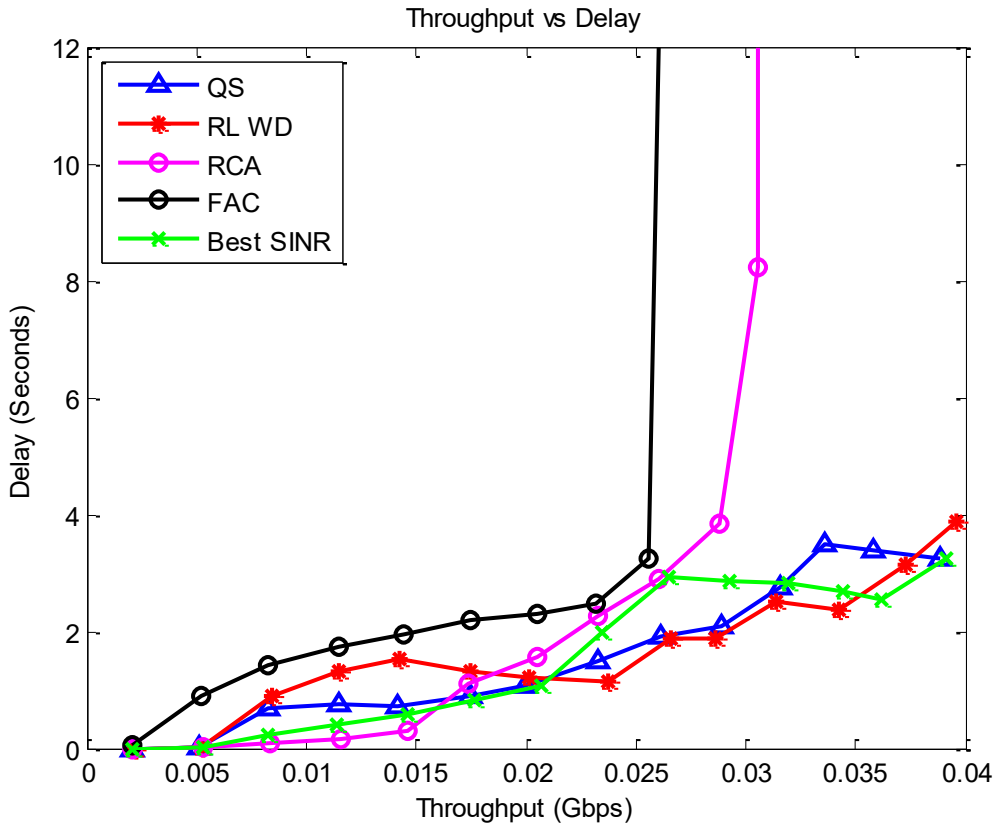


Figure 5.7. Throughput vs. Delay

reduces transmission rate (as a result of SINR value reduction when new users enter the system) and collisions resulting from failed channel selections. Figure 5.6 and figure 5.7 are both indications of throughput limits for the tested schemes within the used test platform. As a result of efficient channel allocation carried out by the QS scheme, the system throughput performs almost as good as the RL based scheme and the best SINR scheme as shown in figure 5.8. This indicates a promising scheme that if supported with a proper amplitude reinforcement (update) strategy is expected to outperform a similar conventional RL scheme. A distinctive behaviour that differentiates the QS scheme from conventional assignment schemes as well as conventional RL schemes is the consistency in choosing the channel that was successfully chosen during a previous assignment.

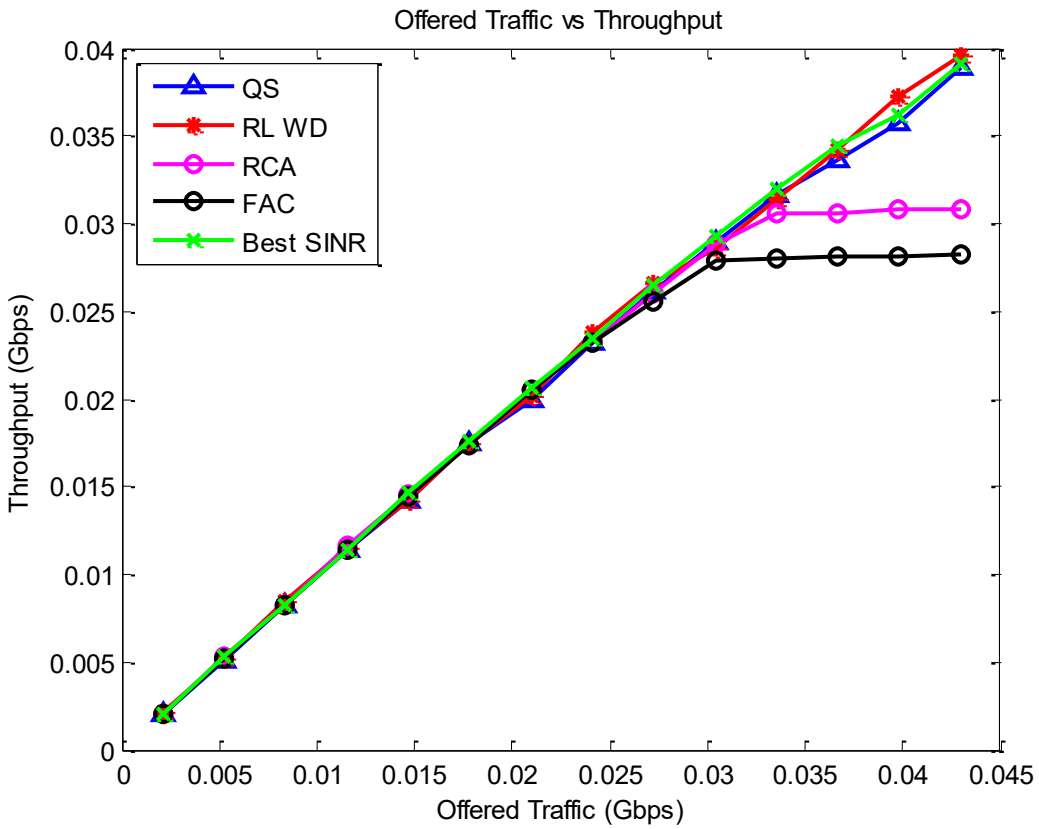


Figure 5.8. Offered Traffic vs. Throughput

The channel choice is based on the last experience where this selection is not changed unless failed. In other words, the learning agent keeps using the same channel until it fails to use it (blocked). If blocked, the channel preference changes and the agent selects

the second preferred channel. This behaviour within modified quantum inspired reinforcement learning is described in chapter 6 and is illustrated in figure 6.3.

5.5.2 Channel Partitioning

Channel partitioning is usually a desirable result within learning based schemes. It defines the efficiency of the learning engine in performing an efficient spectrum allocation. As the QS scheme has the ability of prioritizing channels based on successful selection, it is expected to be able to perform channel partitioning among users within the wireless network. A useful and efficient method of monitoring the channel partitioning process done by the QS algorithm is through monitoring user choices of available channels. A small number of random users (6 users) have been picked for channel selection (of 8 available channels) monitoring at different points during simulation time. The number of events (successful transmissions in this case) is used to define the points when channel usage is recorded. Channel usage by users has been recorded at (500, 1000, 2000, and 3000) events respectively. Channel usage by users is illustrated in figures 5.8-11.

The channel usage ratio in the plots is the ratio at which a specific channel has been selected by the specified user from the total number of channel selection by the same user. By recording the selection of the available channels by those 6 random users, an idea can be formed about the channel usage during simulations and the way the available spectrum is used.

It can be seen that there is a fare trend towards priority distribution of channels among users. Each of the users has clearly appointed a specifically preferred channel with other less favourable channels. Channels are clearly ranked in terms of usage for each user. It

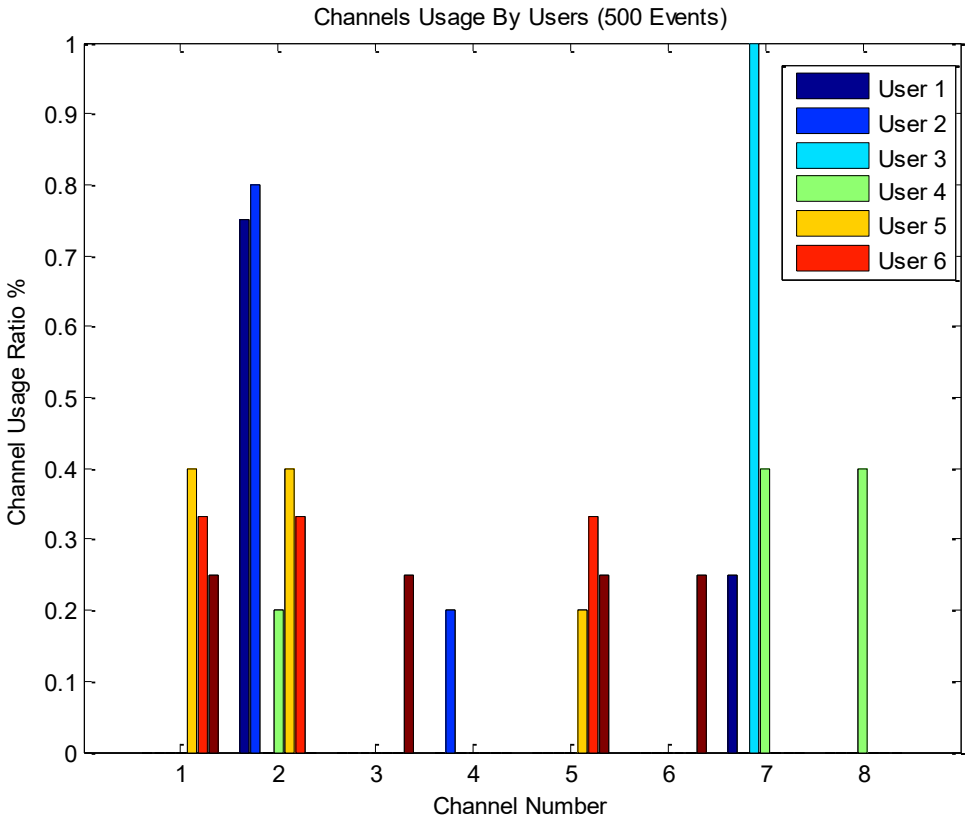


Figure 5.9. Channel Usage by Different Users (500 Events)

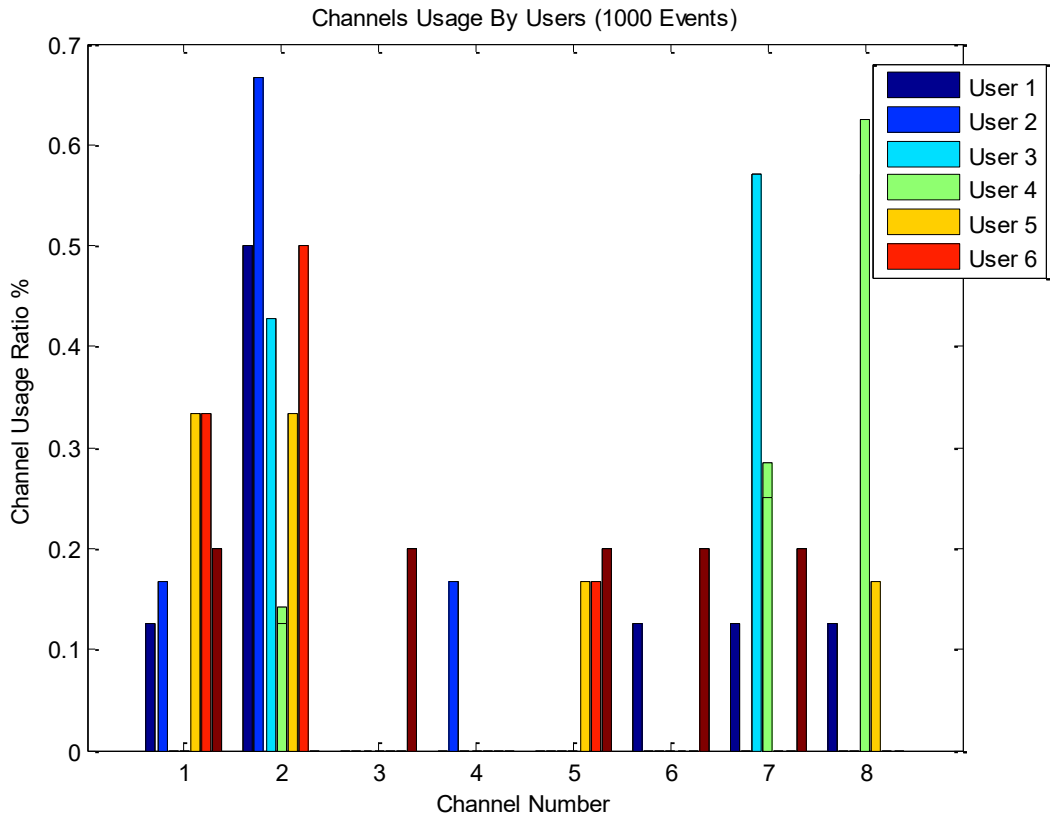


Figure 5.10. Channel Usage by Different Users (1000 Events)

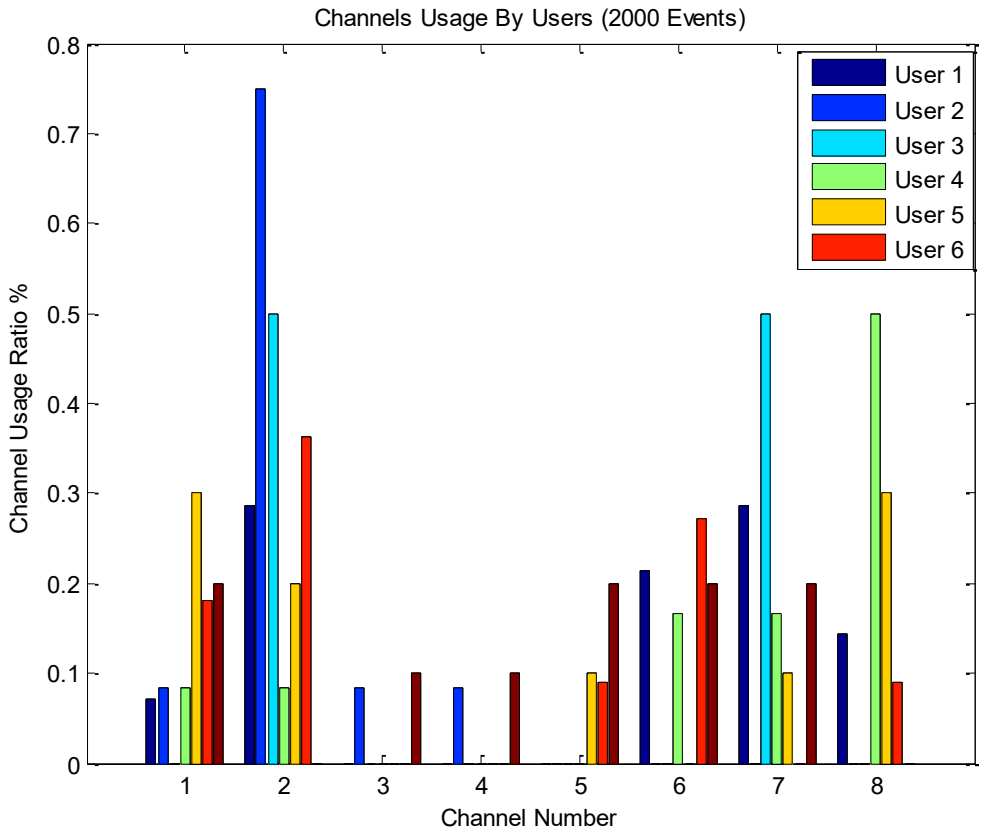


Figure 5.11. Channel Usage by Different Users (2000 Events)

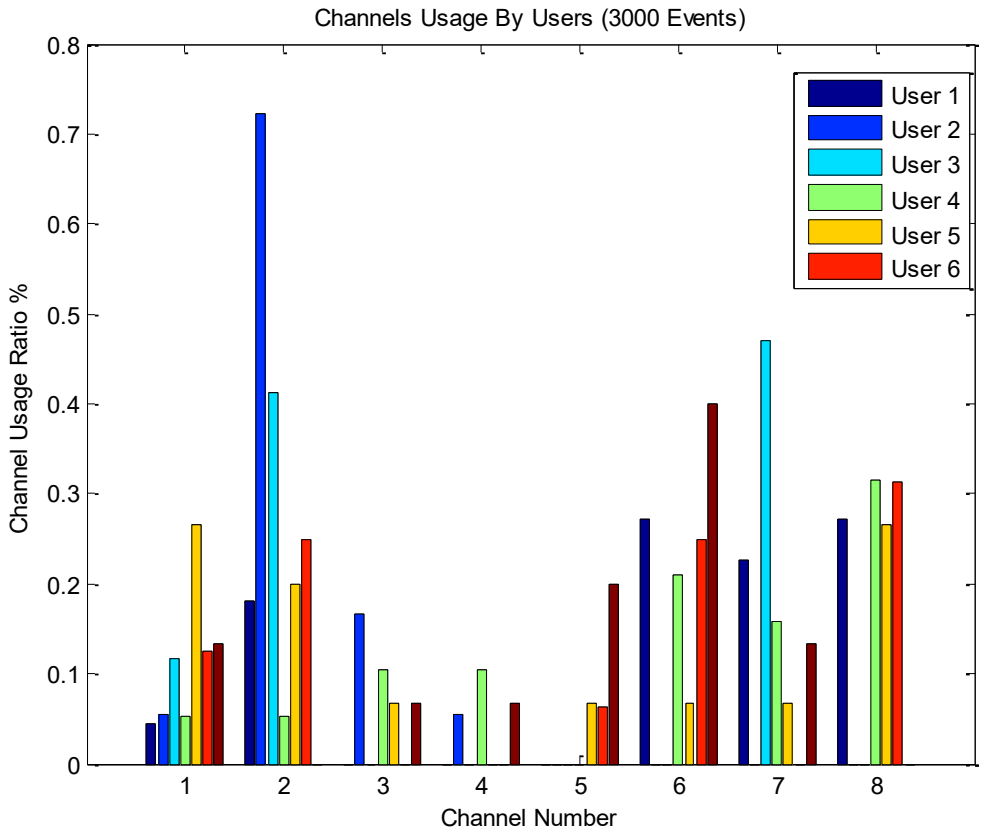


Figure 5.12. Channel Usage by Different Users (3000 Events)

is also clear that the total number of used channels differs from user to another depending on their need to change channels.

5.6 Conclusion.

This chapter has investigated the search section within the dynamic spectrum access. A comparison has been made between a classical pick and test searching technique and the quantum search technique using Grover algorithm.

The introduced search algorithm has made a novel enhancement to the cognitive wireless network channel assignment scheme by introducing an enhancement to the basic channel search and pick procedure. The introduction of the Grover search algorithm as a channel allocator is another novel contribution

The Grover algorithm turns the search procedure into a semi-learning scheme by having a discriminating effect on the correct answer of the search problem. The effect takes the shape of phase shifting rather than value changing. Simulation result for system performance working by the QS scheme proved to be competent and successful channel allocator. Channel partitioning by the QS scheme has been recorded. However, reinforcement for the search results whether they are negative or positive is needed as the Grover search is essentially designed for a single element search process.

The quantum search algorithm turns the unstructured search problem to a structured search which makes decision making process within any search process an easier and faster one.

In general, this chapter introduces the search technique that can be added to reinforcement learning technique to enhance the decision making process. As a result, it improves the speed of learning.

It will be clearly shown in the next chapter that a more effective and faster searching technique might be an even more viable way to increase the learning efficiency and speed for conventional reinforcement learning.

Chapter 6 : Quantum Reinforcement Learning

6.1 Introduction

In Multi-Agent Reinforcement Learning schemes (MARL), all exploring agents learn simultaneously and independently of each other. They try all channels in the spectrum pool with equal probability. This can give rise to increased blocking probability and delay, as a result of poor selections being tried too often. Therefore, the convergence of MARL slows and the decision making efficiency is reduced as the size of action space expands (i.e. increasing the number of learned solutions (channels)) [21, 24]. Many approaches have

been used to solve this problem [23, 106]for two main reasons. First, the RL in principle takes multiple trials to change a decision preference in most cases. Second, exploration is random and not directly based on decision ranking. Therefore, as traffic load increases, environmental non-stationarity increases. This motivates an efficient and fast exploration decision. Several procedures and modifications were proposed in [21, 23, 24, 106] to enhance both performance and time of exploration.

This chapter presents a novel quantum reinforcement learning (QRL) technique for Multi-Agent Reinforcement Learning (MARL) problems in cognitive radio systems. This has been done through the application of the quantum search theory into the reinforcement learning scheme of the wireless network. The ultimate goal for introducing a quantum searching reinforcement learning system is a modified RL algorithm that is a fully self-organized engine. The learning engine is intended to explore new decisions conditionally based up on failures rather than randomly. The exploration process should not need the tuning any predefined parameters, such as in the case of the $\epsilon - \text{greedy}$ technique. A QRL explored decision depends on a decision preference ranking which leads to reduced collisions and delays, delivering much a faster convergence.

The expected result for the proposed merge process is to reduce convergence time through a more efficient decision making process leading to a more viable multi-agent learning process.

Section 6.2 and 6.3, introduce a brief introduction to traditional and quantum reinforcement learning respectively. Section 6.4, illustrates the spectrum assignment algorithm practically used in our work. The results obtained from simulations are discussed in section 6.5.

6.2 Traditional Reinforcement Learning (RL)

The main goal of a RL algorithm is to establish an action policy based on the expected return when taking actions following that particular policy [1]. The learning process is

accomplished by direct interaction between the agent and the environment through trial and error.

The function of the reinforcement learning scenario developed for our wireless cognitive radio environment can be illustrated as in figure 6.1 [9]. The learning agent here is the access base station (ABS). The environment is represented by the wireless spectrum.

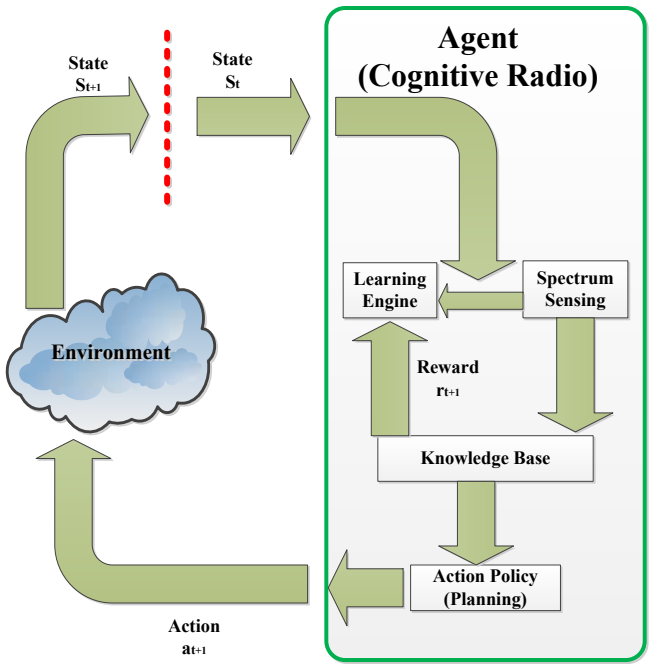


Figure 6.1. Reinforcement Learning Model in a Cognitive Radio Scenario[9].

In reinforcement learning, the value of the state s under certain policy π which is denoted by $V^{\pi}(s)$ is what the agent depends on in its action selection $A(s)$. The learning agent (ABS) aims to develop an optimal policy. Such a policy is supposed to maximize $V^{\pi}(s)$ at each learning epoch (trial). $V^{\pi}(s)$ is usually defined as in [64]:

$$V^{\pi}(s) = R\big(s,\pi(s)\big) + \gamma \sum_{s'} P(s'\,|s,\pi(s))V^{\pi}(s') \qquad (6\text{-}1)$$

Where $R\big(s,\pi(s)\big) = E\{r\big(s,\pi(s)\big)\}$ is the mean value of $r\big(s,\pi(s)\big)$. s' represents the destination states towards which the agent in state s might end up in. This is in the case of implementing the action $\pi(s)$. $P(s'|s,\pi(s))$ here is the probability of the state s making a transition to a different successor state s'.

The optimal value function $V^{\pi^*}(s)$ under the optimal policy π^*can be defined as:

$$V^{\pi^*}(s) = \max_{a \epsilon A}(R\big(s, \pi(s)\big) + \gamma \sum\nolimits_{s'} P(s' \,|s, \pi(s)) V^{\pi}(s')) \tag{6-2}$$

The optimal policy π^*can be specified as:

$$\pi^*(s) = arg \max_{a \epsilon A}(R\big(s, \pi(s)\big) + \gamma \sum\nolimits_{s'} P(s' \,|s, \pi(s)) V^{\pi}(s')) \tag{6-3}$$

Where:

$R\big(s, \pi(s)\big)$: Is the cumulative reward for the agent while being in the state s.

$\gamma \sum_{s'} P(s' \,|s, \pi(s)) V^{\pi}(s'))$: The expected feedback of its successor state s'.

The reinforcement learning strategy for the scenario we have adopted from [9] for our quantum based scheme later on maps weights to actions $\pi: W \rightarrow A$ instead of another approach that maps a state to an action $\pi: S$.[21]. The action value being updated by the ABS depending on trial and error is what decides the desirability of the action. In other words, our scenario is a weight driven or stateless scenario.

The ABSs (learning agents) are fully distributed which means that decisions are made based on local information (spectrum measurements). The reason behind choosing such a base system as a quantum developed scenario is to avoid exchanging unnecessary information on the network level even when possible. Limiting the exchange of information supports fully distributed solutions for further energy saving schemes. In addition, such scenarios limit the computational complexity for the whole system as one indication (the action value) is used for optimum policy selection.

The used base reinforcement learning model in our work consists of the following:

1- A weight table W for the performed actions by the ABS stored within the knowledge base of it.
2- A set of actions A which are in this case the set of available channels (frequency bands).
3- Numerical rewards R.

The ABS will access the communication resources (spectrum) based on the updated memory of the reinforcement learning system.

The level of success of a particular action which defines the desirability of that specific action based on its suitability for the communication request is assessed by the reinforcement learning engine. The assessment is performed by assigning a positive numerical reward to the action weight in case of success. This is done to reinforce the action weight within the ABS knowledge base. A negative numerical reward referred to as punishment is assigned to the action in case of failure to reflect the result of assessment on the action weight.

The reason behind the above mentioned scenario [9] is to develop an optimal policy that maps weight to action $\pi: W \rightarrow A$ that maximizes the value of the current memory $V^{\pi}(w)$. Based on a set of available weights for the used resources (channels) and a policy π, the action selection process can be denoted as $a = \pi(w)$. On the other hand $V^{\pi}(w)$ can be defined as:

$$V^{\pi}(w). = \sum_{w'} P\big(w'\big|w, \pi(w)\big). w' \qquad (6\text{-}4)$$

Where w is the weight of the used resource (channel) for the agent (ABS) at time t, w' is the expected weight values after taking the action $\pi(w)$ by the agent. $P\big(w'\big|w, \pi(w)\big)$, is the probability of selecting an action after performing the action $\pi(w)$. The optimal value function under the optimal, policy π^* is defined as [9]:

$$V^{\pi^*}(w). = \max_{a \epsilon A}(\sum_{w'} P(w'|w, \pi(w)). w') \quad (6\text{-}5)$$

Thus, the optimal policy can be represented as:

$$\pi^*(w). = \arg\max_{a \epsilon A}(\sum_{w'} P(w'|w, \pi(w)). w') \quad (6\text{-}6)$$

Based on its current memory, at each transmission request, the agent (ABS) chooses a resource (channel) which can result in maximizing $V^{\pi}(w)$. The result of the transmission request process will decide the type of the reward r to the knowledge base of the reinforcement learning engine. No more information is needed for the update process which proceeds within the inner loop of cognitive radio in figure 6.1 that will keep updating the knowledge base.

6.3 Value Function

One of the essential elements in reinforcement learning is the value function [139]. Reinforcement learning in principle is meant to map actions to specific situations. It is for this reason a good solution for tackling cases of trade-off between long term and short term rewards. The knowledge updating performed by the cognitive radio user is mainly based on the feedback of the value function. Based on this idea, the value function is also the weight function in our base reinforcement learning scenario used to update the spectrum sharing strategy which can be represented as follows [20, 140]:

$$W_t = f_1 W_{t-1} + f_2 \quad (6\text{-}7)$$

Where W_{t-1} is the weight of the channel at time $t-1$, and W_t is the weight at time t according to both the weight W_{t-1} and the updated feedback from the system. f_1 and f_2 are the weighting factors at time t with their values depending on the action assessment by the learning engine. In case of weight update, either a reward or punishment value is assigned to both f_1 and f_2.

6.4 Weighting Factors

The weighting factors have a major role in the learning process and as a result on the system performance. They determine the degree of response of a learning agent included in each ABS towards changes of the environment. In the case of high reward or punishment values, the changes of the wireless environment cause the learning ABS to adjust its actions swiftly in response. On the other hand, a mild reward or punishment, causes the ABS to adapt itself through gradual adjustments based on the interactions with the environment [20]. The used values for f_1 and f_2 in our work are shown in table 6.1:

Table 6.1. Weighting Factor Values

	Reward	**Punishment**
$\boldsymbol{f_1}$	1	0.5
$\boldsymbol{f_2}$	1	-1

The reason for choosing the mentioned values in table 6.1 is that they proved to result in a good system performance for our scheme. It represents an average of both “mild punishment “and “discounted punishment” schemes in [9]. In this case, consideration and memory of the past experience is reduced by 50% in case of punishment. A trade-off between fast weight update supporting values and consistent learning is another reason for choosing the above mentioned values specifically for performance comparison against QRL scheme.

6.5 Reinforcement Learning Based Resource Allocation Scheme

The base reinforcement learning algorithm [20] used in our work is illustrated in figure 6.2. The weight values for actions are initialized according to a random uniform distribution. At the beginning, any user that aims to transmit, sends a transmission request to the best signal ABS (the ABS signal sensed from the user side) to connect to. Failure to

connect to the ABS, makes the user search for the next best signal ABS and so on until it succeeds to connect to an ABS. After accomplishing the connection process between the user and the ABS, the algorithm performs following main steps:

1- ***Channel Selection***: The action selection strategy is based on $\epsilon -$ greedy technique ($\epsilon = 0.01$). The ABS in this case chooses a channel either randomly for (1%) of trials or chooses the highest weight channel for (99%) of the trials. The channel weights are randomly generated at the beginning of simulation then get updated based on the ABS learning process.

2- ***Spectrum Sensing***: The ABS senses the SINR level for the connected user. If the SINR is above the threshold level, then the user starts transmitting. If the SINR level is below the threshold level, then the user transmission is blocked and assigned a later activation time. In this case, the weight of the allocated channel is updated by a punishment factor value.

3- ***SINR Measurement:*** After the spectrum sensing step, all users using the same allocated channel in step 2, measure the SINR level at their receivers to update the link transmission rate for each of them. As a result, the remaining transmission time for each file for each transmitting user is updated at that instant. Any user that has SINR level below the threshold level stops transmission temporarily until the level of its SINR is recovered back to above threshold level.

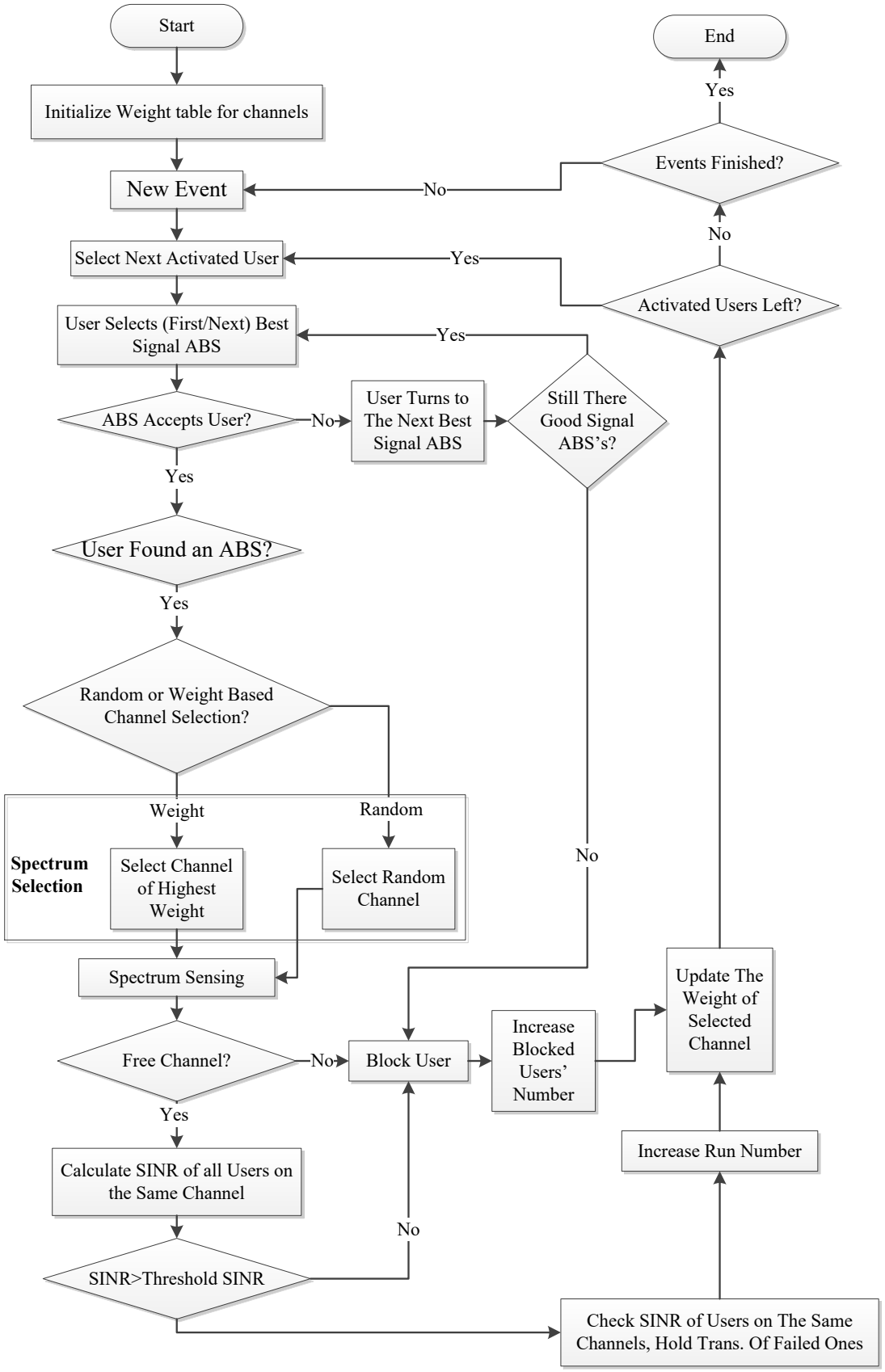


Figure 6.2. Reinforcement Learning Based Spectrum Sharing Algorithm

6.6 Quantum Reinforcement Learning (QRL)

6.6.1 Introduction

QRL aims to introduce a more efficient and self-organized decision making policy into traditional RL. The ultimate goal is a faster converging technique which makes it more adaptable for dynamic problems.

6.6.2 Grover Quantum Search Algorithm:

As previously discussed in chapter 5, the Grover algorithm has essentially been developed for fast information search purposes [108, 131]. It introduces amplitude and amplitude amplification principles upon which the decision making policy in QRL is based. The idea of introducing the Grover algorithm into reinforcement learning was first proposed in [73, 75, 141]. The Grover algorithm has been used in that case for a single agent, episodic RL problem. It significantly improved the convergence time in comparison with the pure RL algorithm. The effect of the Grover algorithm on amplitude values is illustrated in figure 6-3. In this thesis, the formation of the new QRL algorithm has been carried out by adding two important elements. These are the Grover algorithm and an amplitude amplification exponential equation to enforce the effect of the Grover search result. They include the following steps within each epoch of the RL algorithm:

1- Create a register for amplitude values for all available channels (representing the quantum form of channel desirability) and initialize it to equal values as in figure 6.3.a.

2- Discriminate the successful channel selection (if successful) by inverting the sign of the corresponding amplitude value as in figure 6.3.b.

3- Multiply a diffusion operator A by the amplitude register as in figure 6.3.c. This inverts the amplitude values about their mean which is a form of normalization. This operator is represented by the matrix:

$$A_{ij} = \begin{cases} 2/N & \text{if } i \neq j \\ -1 + 2/N & \text{otherwise} \end{cases} \quad (6\text{-}8)$$

4- Where N is the number of actions (channels) and is also equal to the length of amplitude register.

5- Update the amplitude value of the chosen decision. This is represented by an exponential equation:

$$C_a \leftarrow e^{\lambda(r+W(s'))}C_a \quad (6\text{-}9)$$

6- If the chosen channel failed, then step 2 is skipped. Step 5 is applied using a punishment value instead of reward. This will result in reducing the amplitude value of the failed channel to be the lowest one as in figure 6.3.d.

Where C_a is the amplitude value of the chosen channel, λ is a discount parameter ($0<\lambda<1$), and r is the value of reward or punishment. This gives a high increase (or reduction) in the amplitude of the action. Usually QRL uses a high reward and mild punishment values when updating the amplitude.

The idea behind applying the Grover algorithm within the reinforcement learning algorithm is to present an additional value table for the learning agent (the amplitude table) that is updated, normalized, and ranked in a way that prioritizes channels in a best-to-worst sequence which is updated along with the learning. The system then depends exclusively on this table as a decision making reference for spectrum selection. The agent starts always by selecting the highest amplitude channel for spectrum allocation. The agent selects the next best channel each time it fails to choose one. This makes exploration a pre-decided

decision by the result of spectrum assessment made by the learning engine. The exploration in this case is made through choosing the next best channel from the amplitude table rather than picking a channel randomly. Thus, there is no need for any exploration parameter tuning. As a result the following changes are made to the traditional RL procedure when adding Grover algorithm:

1- It becomes a 100% exploitation process depending solely on the best amplitude value channel.

2- In case of failure, the channel is updated with a negative reward (punishment) that turns it into the worst channel. This makes re-selecting the same channel impossible in this case. Thus, the selection preference does not depend on how many times a channel has been successfully selected before failure. Consequently, the next best channel will immediately be selected as it will become the best preferred.

3- As long as decision fails, the agent (ABS) keeps exploring through choosing the next highest amplitude channel.

4- There is no exploration without failure.

5- It turns the multi-agent reinforcement learning (MARL) engine into a self-organized one as exploration is automatically stimulated by failure rather than by tuned parameters.

6- It reduces the convergence time as only agents that need to explore do so; others do not. This reduces collisions during exploration.

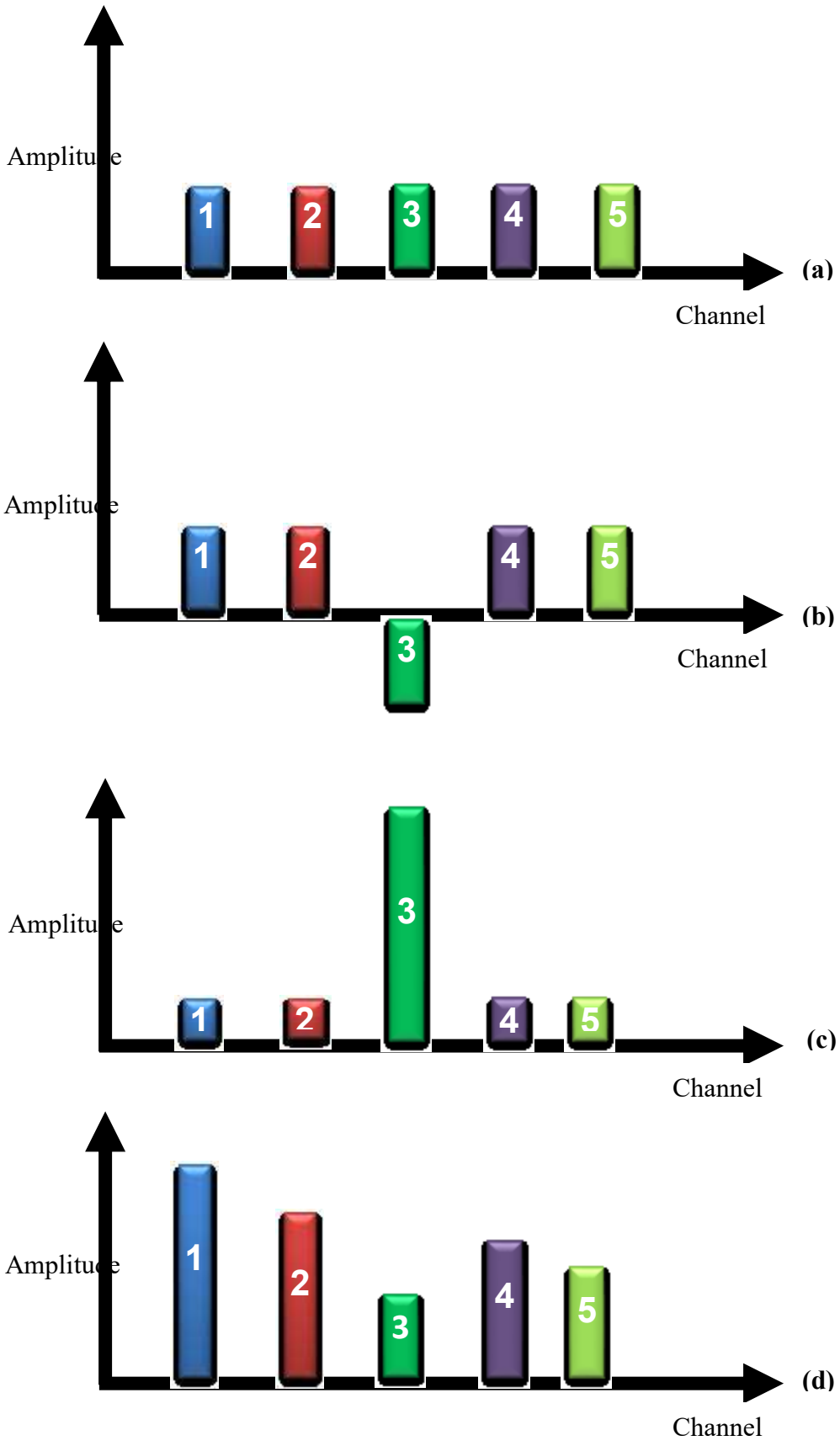


Figure 6.3. Quantum Reinforcement Learning Algorithm

6.7 Spectrum Assignment Algorithm

The spectrum assignment algorithm is illustrated in figure 6.4. The overall algorithm is a modification of the standard reinforcement learning algorithm illustrated earlier. Thus, figure 6.4 is the same as figure 6.2 with the addition made by the quantum reinforcement learning scheme. Two other learning schemes have been used for comparison. All three learning schemes tested here follow the same main algorithm in terms of weight updating, ABS selection, spectrum sensing procedure and essential system parameters. Two critical changes differentiate these schemes from each other. The first is that both the RL and random RL schemes follow an $\epsilon - \text{greedy}$ procedure for channel selection. The QRL scheme on the other hand:

1- Only initializes with a random selection, later on, the system follows the best amplitude value channel at all times. This is shown in left-hand part of figure 6.3 as the spectrum assignment part. It is this part that makes QRL an exploitation exclusive process by depending on the amplitude table for channel selection.

2- The second difference is the QRL scheme is in the right-hand part of figure 6.4 excluding the weight value updating. This part is responsible for the normalization of the amplitude table after each successful or failed action. It is also responsible for the phase rotation of the amplitude value of a successfully selected channel. Finally, it is here where the amplitude table is updated by imposing either a reward or a punishment to the channel amplitude value.

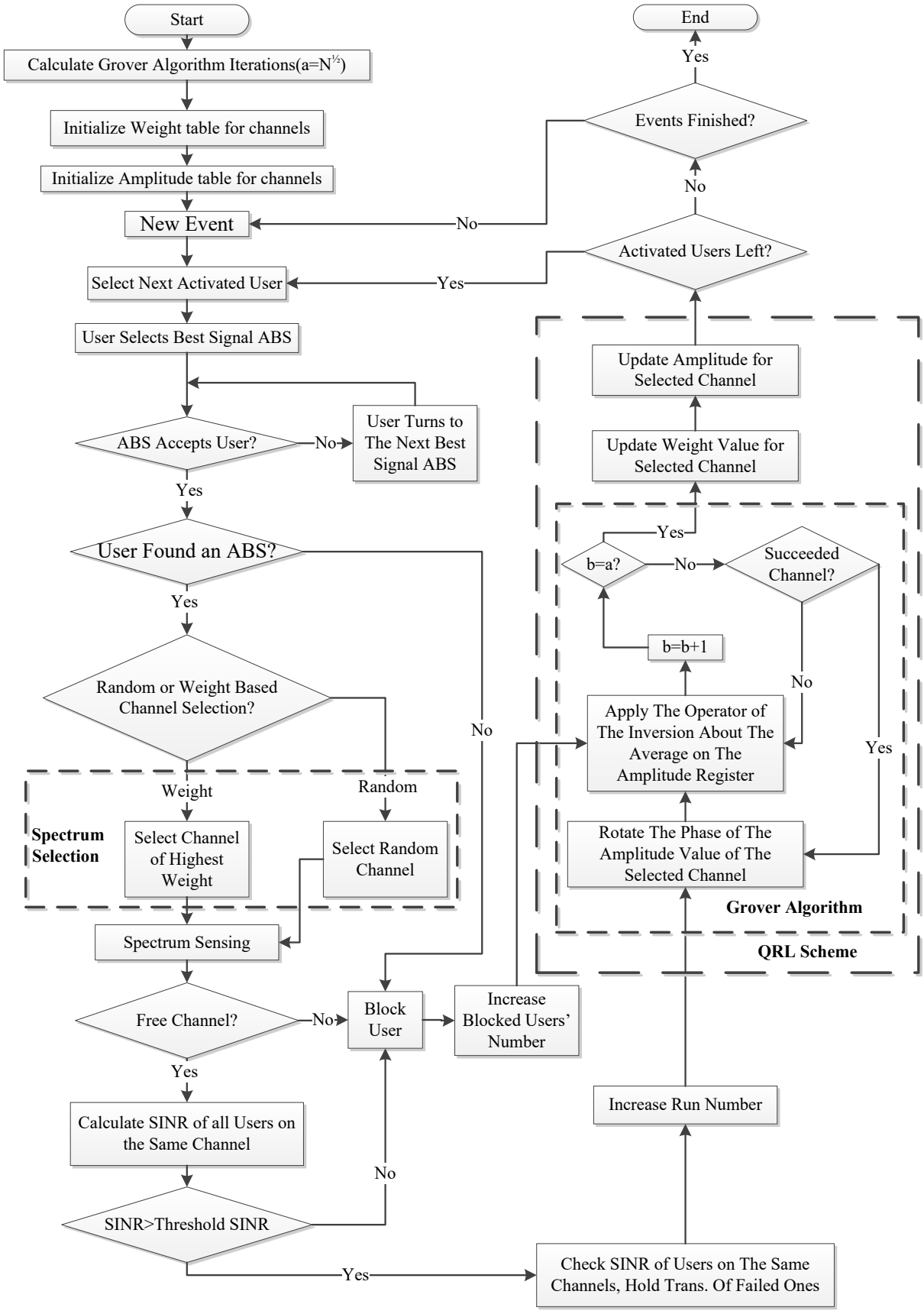


Figure 6.4. Flowchart of QRL based spectrum assignment algorithm

6.8 Results

6.8.1 System Performance

A BuNGee architecture over a 1500 m by 1500 m area is used with 3500 mobile stations (users) distributed randomly over all parts of the service area, external to buildings. The total number of channels available for the whole service area is 20 with 112 ABSs used to communicate with users. The inter-arrival time for all users within the system has been generated such that arrivals follow a Poisson distribution, and the WINNER II propagation model [3] is used. Other sample parameters of the simulated system are found in table 6.2. Reward and punishment have been set to 100 and -1 in case of QRL.

Table 6.2. System and Learning Parameters

Parameter	Value
Maximum ABS gain	17 dBi
SINR Threshold	1.8 dB
Maximum SINR	21 dB
Noise Floor	-112 dBm/MHz
File Size	4MB
QRL reward	100
QRL punishment	-1
QRL discount factor	0.9

The choice of reward and punishment values for QRL scheme in our case is based on several experiments that have shown that the most appropriate reward-to-punishment ratio is -100 (i.e. like reward is 100 and punishment is -1) . The value of discount factor is chosen based on the test results illustrated in figure 6.5 that showed a better performance than other tested values.

The proposed QRL algorithm is aimed at improving the decision making process and action evaluation and update method. It is an independent procedure of the main conventional RL algorithm that is used with it. It works through making the decision, applies the RL learning process, and then uses the parameters from RL in association to the QRL ones to form a separate preference table. As a result, it is not necessary to compare the QRL algorithm with all pure RL ones. Comparisons in our case are relative and aim to show the improvement accomplished by the added quantum technique to any conventional RL algorithm. Different values for the discount parameter λ have been tested for the purpose of performance comparison. Figure 6.5 shows the blocking probability for the system as a function for the number of events for different λ values (0.1, 0.3, 0.6, and 0.9).

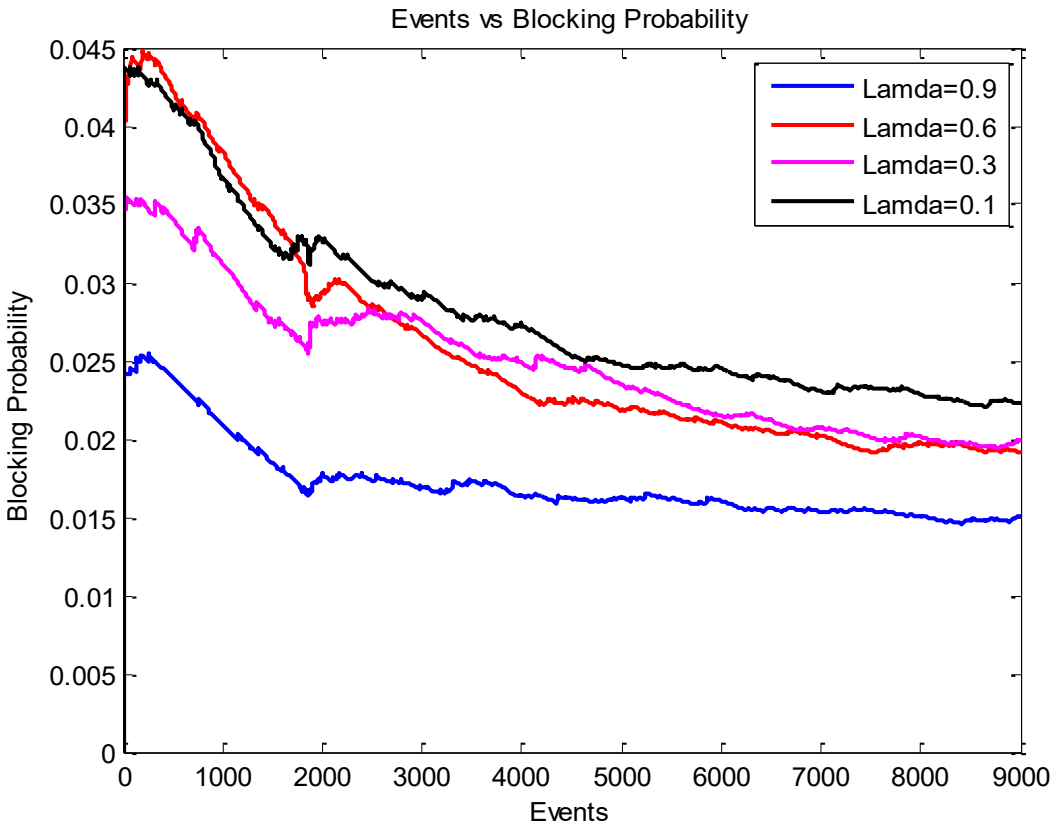


Figure 6.5. System Blocking Probability as a function of No. of Events for Different Values of Discount Factor.

It can be seen from the figure that a high value for λ gives better performance to the learning agent regarding learning speed and convergence in addition to the blocking

probability value. Thus, the value of 0.9 has been used for the remaining of the system simulations. A comparison of performance has been carried out among weight-driven reinforcement learning (WDRL), quantum reinforcement learning (QRL) and non-learning random dynamic channel assignment algorithm (RDCA). In addition a reinforcement learning technique based on random exploration upon each decision failure (Random RL) is also used to check whether such an approach can result in the same performance as QRL. All simulations are run until the same number of files that have finished transmitting. Consequently, it is an event-driven simulation rather than time-driven.

The normalized root mean square difference (NRMSD) of the channel weight (or amplitude in case of QRL) is used to measure the convergence speed. It helps to show when the learning process is starting to stabilize. It is calculated from the root mean square difference (RMSD) as follows:

$$\mathrm{RMSD} = \sqrt{\frac{\sum_{t=1}^{n}(W(s') - W(s))^2}{n}} \tag{6-10}$$

where $W(s)$ and $W(s')$ are the channel weights before and after updating respectively. The number of learning agents (ABSs) is represented by n. The normalized RMSD is calculated (to unify both RL and QRL within one plot of the same scale) using the following formula:

$$\mathrm{NRMSD} = \frac{\mathrm{RMSD}}{x_{max}} \tag{6-11}$$

where x_{max} is the maximum RMSD value. Figure 6.6 shows the test results. QRL shows a significant enhancement in convergence speed over the other schemes. It reaches convergence in 1% of the time needed for the traditional reinforcement learning system. The Random RL scheme, does not show comparable convergence performance to the QRL. It also does not have the same QoS performance. This result reflects the expectation

that the quantum algorithm can react quickly to changes in the learning agent environment. The essential quality that leads to this adaptability is a reliance on the result of the last ABS experience. The immediate turn to exploration of the next best channel in case of failure instead of the traditional random exploration is what differentiates QRL from the random RL scheme. As a consequence, it is highly expected that random RL will result in a higher system blocking probability compared to QRL.

In figure 6.7, we have used a temporal blocking probability graph to monitor the behaviour of QRL system in comparison with both RL and random RL systems. The results in this graph are consistent with those of figure 6.6. Early stabilization of blocking probability is recognized for QRL against slower stabilization for both RL and random RL. In QRL, an abnormal peak of blocking value is recognized at the very beginning of learning due to the early choices of ABSs for favourite channels without periodic exploration. Soon, due to blocking-stimulated explorations, all ABSs, reach an optimum channel choice of their own. On the other hand, in RL, the exploration process is slower as it depends on random probability. Thus, the blocking probability level in RL case stabilizes at a much slower rate. Although the random RL scheme explores depending on channel failure, it still explores randomly. Thus, no performance or convergence enhancement is recognized in it.

Figure 6.8 shows the system blocking probabilities against offered traffic for the different strategies. It is clear that QRL outperform the traditional techniques. The fast convergence makes the system able to re-adjust the Q-values of channels. Fast convergence plays a vital role in dynamic systems like wireless networks.

As users enter and leave the system randomly, fast system adjustments are important to keep up with these fast environmental changes. At high traffic loads, fluctuations in the number of users and average transferred data might make a trial and error based system incapable of retaining a workable policy. With a growing number of ABSs, a faster learning technique becomes of even greater importance. In the case of traditional RL,

changing the ranking of the channels via weight values requires several trials to reform the accumulated experience of the agent which is fundamental to any RL system. In other words, the RL system needs multiple trials over a specific decision to decide how good it is. As a result, the wrong decisions may result in several blocked users each while gaining sufficient experience. On the other hand, QRL changes the last favourite decision as soon as it fails to the second best one leaving a lower probability of repeated failure.

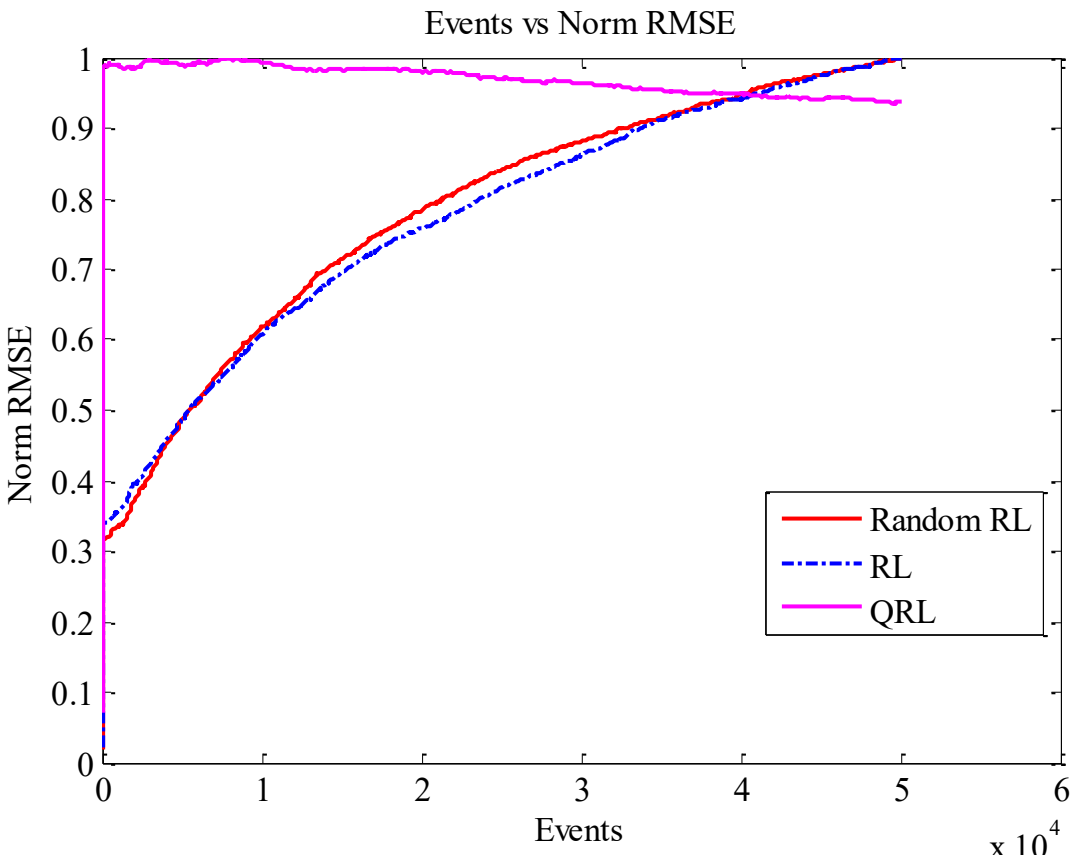


Figure 6.6. Normalized RMSD vs. No. of Events

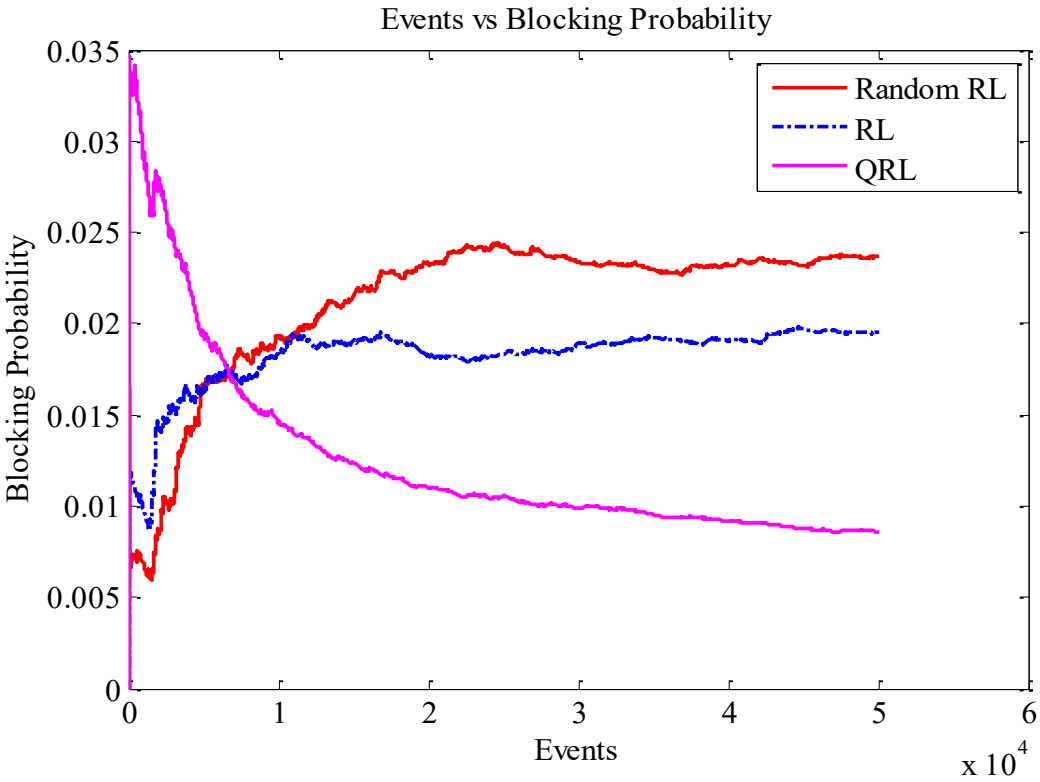


Figure 6.7. Normalized RMSD vs. No. of Events

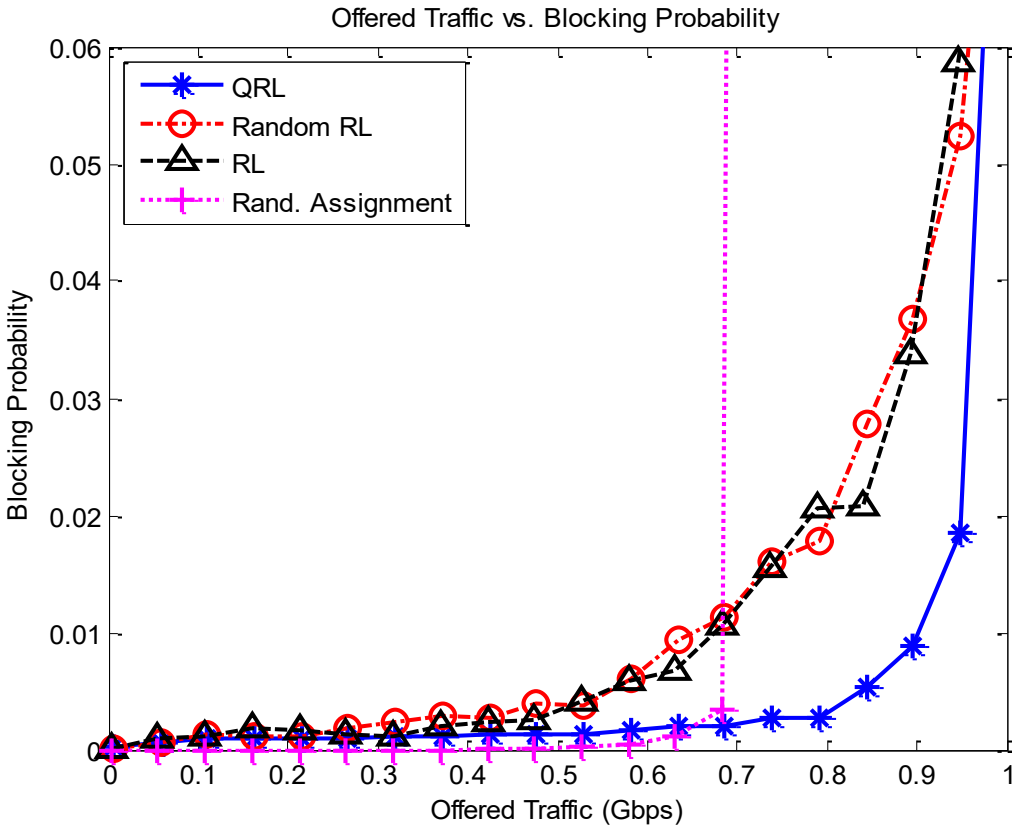


Figure 6.8. Blocking Probability vs Offered Traffic

The same effect of QRL is not gained from Random RL. The second best channel is unlikely to be selected. Thus, a worse channel outcome might be expected. This would result in overall increased blocking. As a result, an almost same performance outcome is gained from random RL and RL schemes.

Because of the ability of the new QRL scheme to accommodate more users, it is expected to see a rise in interference which in turn and in most cases would reduce the throughput per user. However, the significant reduction of blocking probability means that, the average file delay stays the same at low traffic and is even better than traditional RL at high traffic loads, and also results in increased system throughput.

In figure 6.9, the file delay plot shows non-zero delay values at the low traffic values for learning-based schemes.

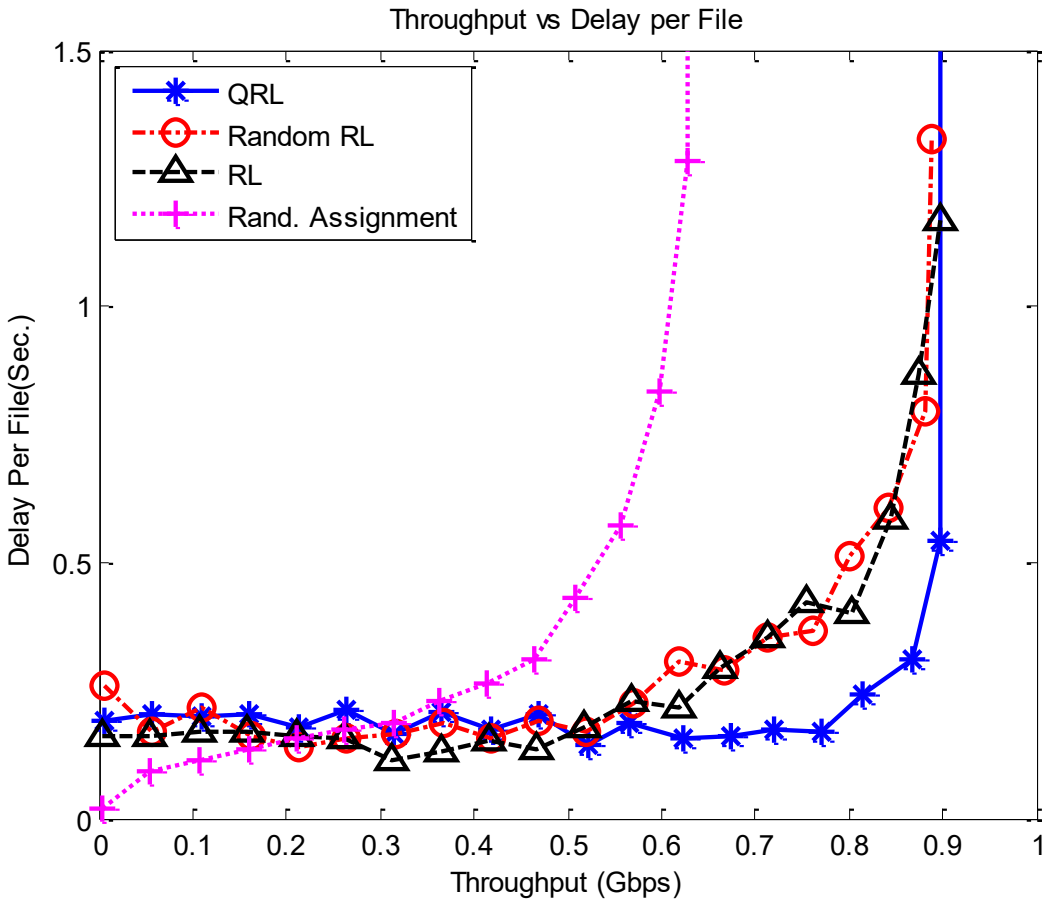


Figure 6.9. System Throughput vs Delay per File

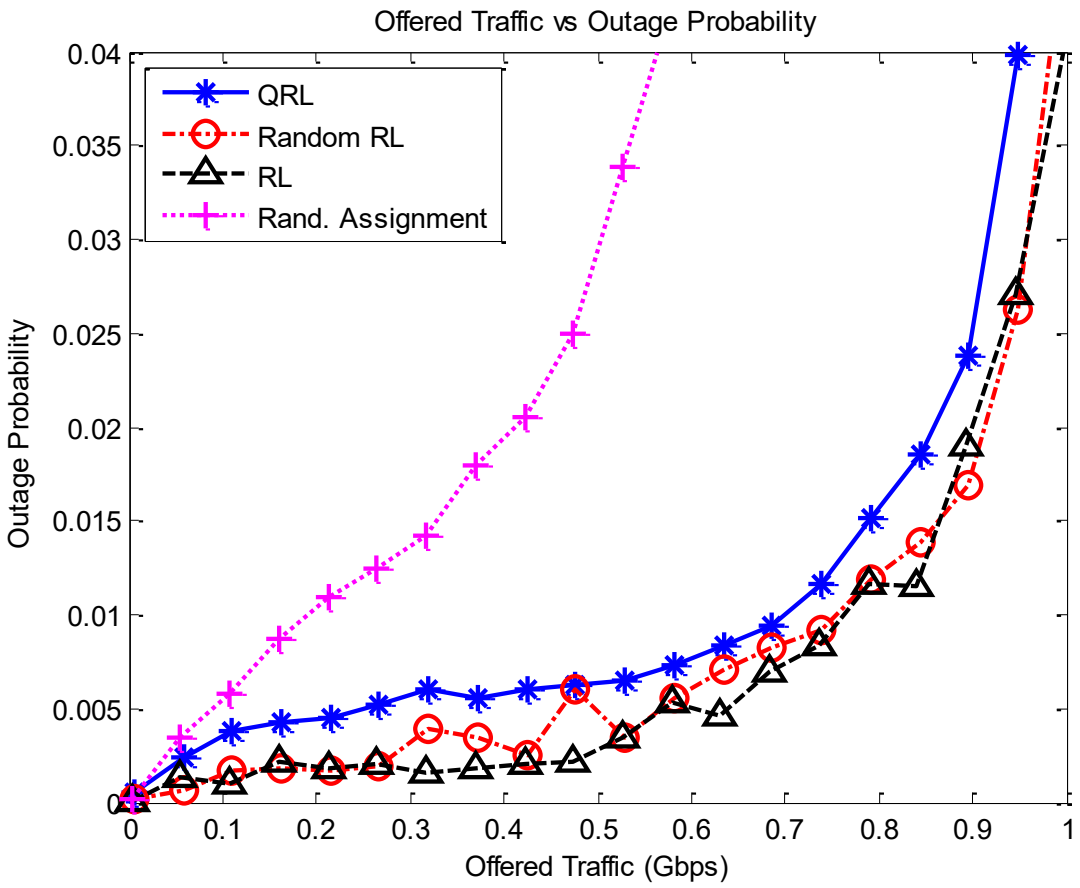


Figure 6.10. Outage Probability vs. Offered Traffic

This is a consequence of delays, caused during the learning period, also being counted. As expected file delay for QRL shows better results, with more users in the system within the functional system range (<5% blocking probability). Once again we show that the greater adaptability of QRL results in a better trade-off between capacity and delay due to interference. Random RL exhibits poorer performance than traditional RL as it fails to choose the appropriate channel. The classical RCA scheme fails at even lower traffic loads due to the inability to choose a suitable channel at high loads because of its random selection policy.

In figure 6.10, the effect of accommodating an excessive number of users compared to the other compared techniques is obvious. It is the only plot that shows a less efficient performance for a QoS parameter of QRL. Outage probability in this case shows a higher

level than RL as higher number of users is able to enter the system which increases the interference level due to higher frequency reuse level.

However, it is important to notice that such an outage probability level did not affect the more efficiently performing QRL from delay and blocking probability points of view. This is due to the success of the QRL-based scheme to transmit much higher amount of date due to the increase of system capacity.

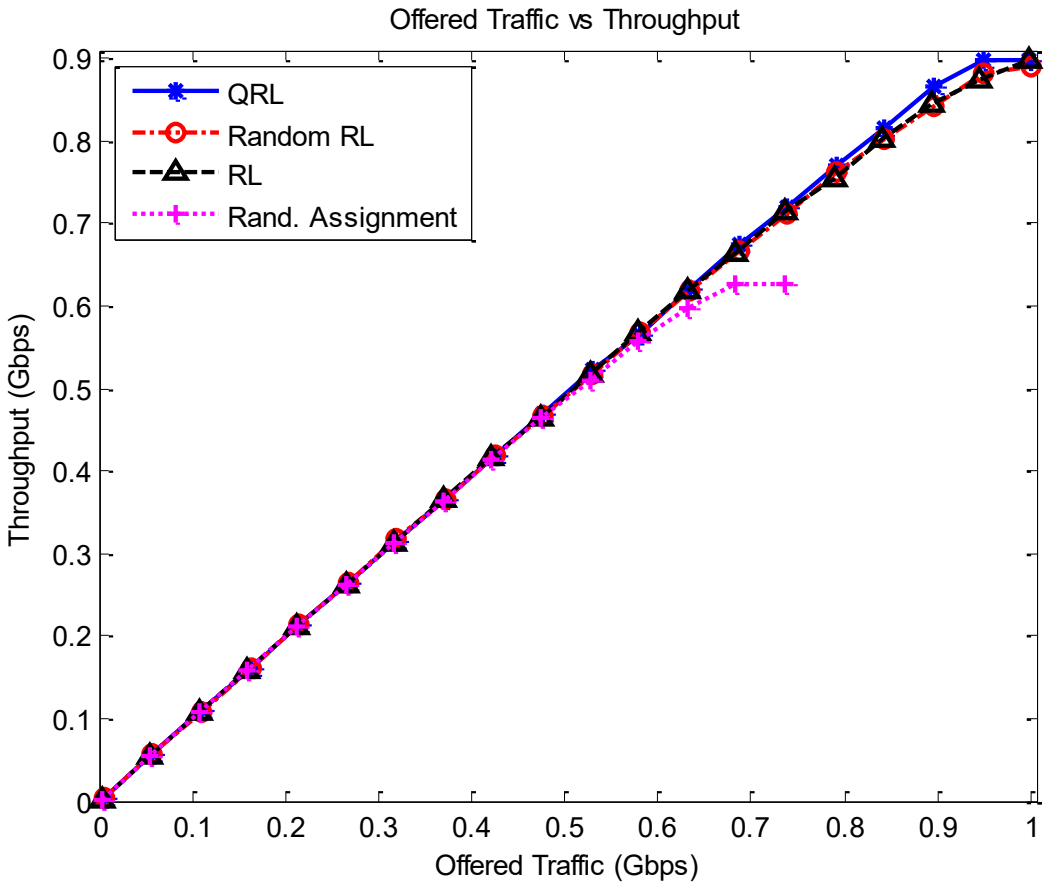


Figure 6.11: Offered Traffic vs. System Throughput

It is clear from figure 6.11, that QRL has potentially better throughput due to the enhanced blocking and delay parameters. This is consistent with the enhanced blocking and delay regions in figure 6.8 and figure 6.9 respectively.

6.8.2 Channel Partitioning

The usage of available channels by users within the system has been monitored by recording the percentage of usage of specific users for the available channels at different times during simulations. The number of events has been used as an indication of the period the recordings have been made at. Recordings have been made at (500, 1000, 2000, and 4000) events. A system of 8 available channels for every ABS has been simulated for this purpose. The results are given in figures 6.12-15. A number of random users (6 users) have been selected for channel usage monitoring. Such recordings give an idea about the distinctive channel portioning way of QRL scheme. Most users are noticed to use as few channels as possible. This is due to the QRL scheme that supports the continuous usage of a good channel as long as it does not fail to connect.

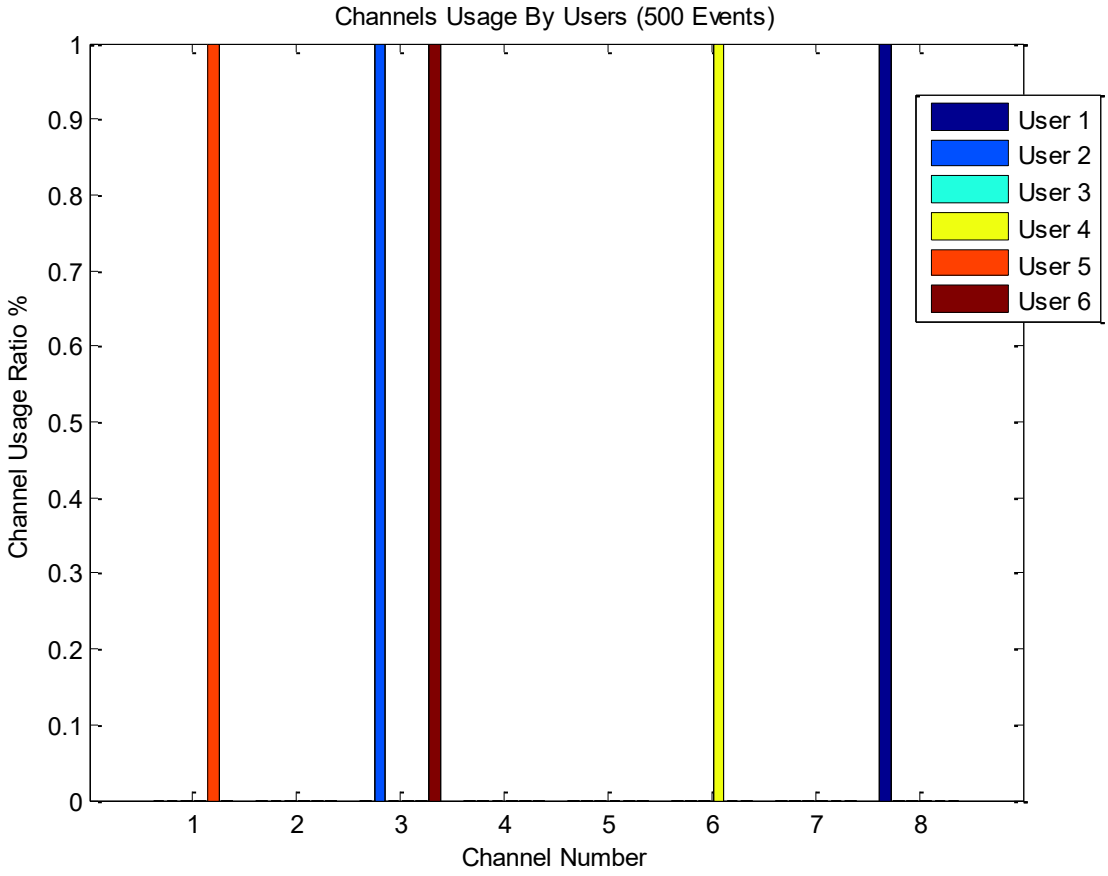


Figure 6.12. Channel Usage by Users (500 Events)

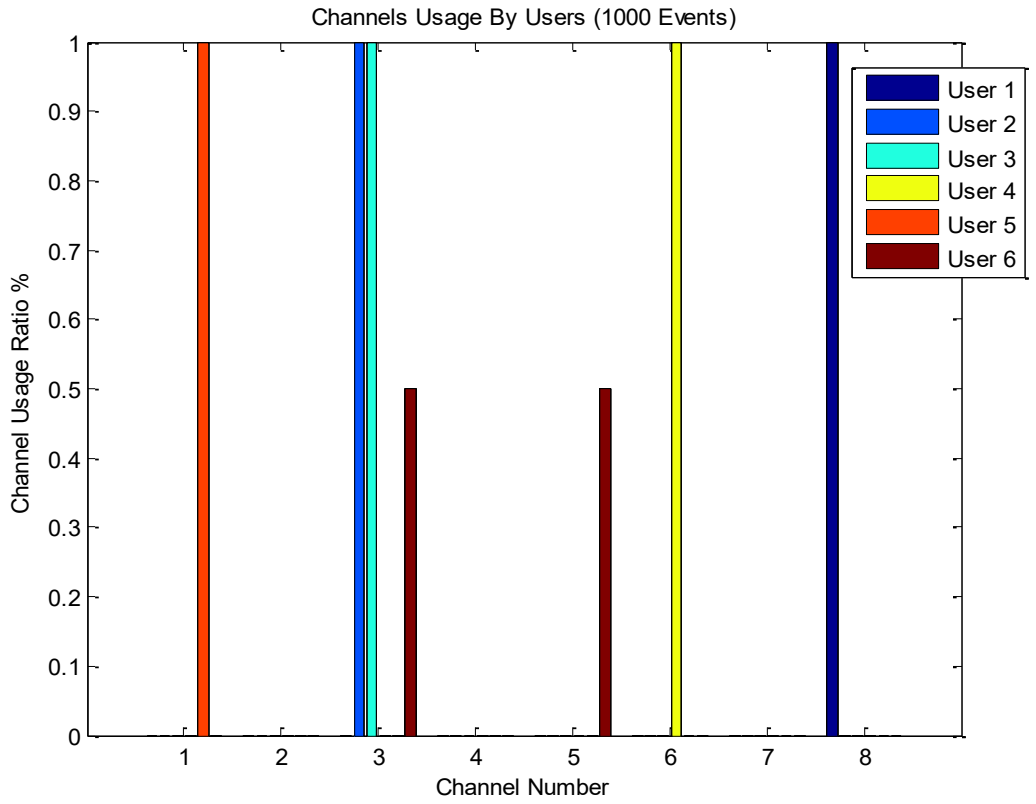


Figure 6.13. Channel Usage by Users (1000 Events)

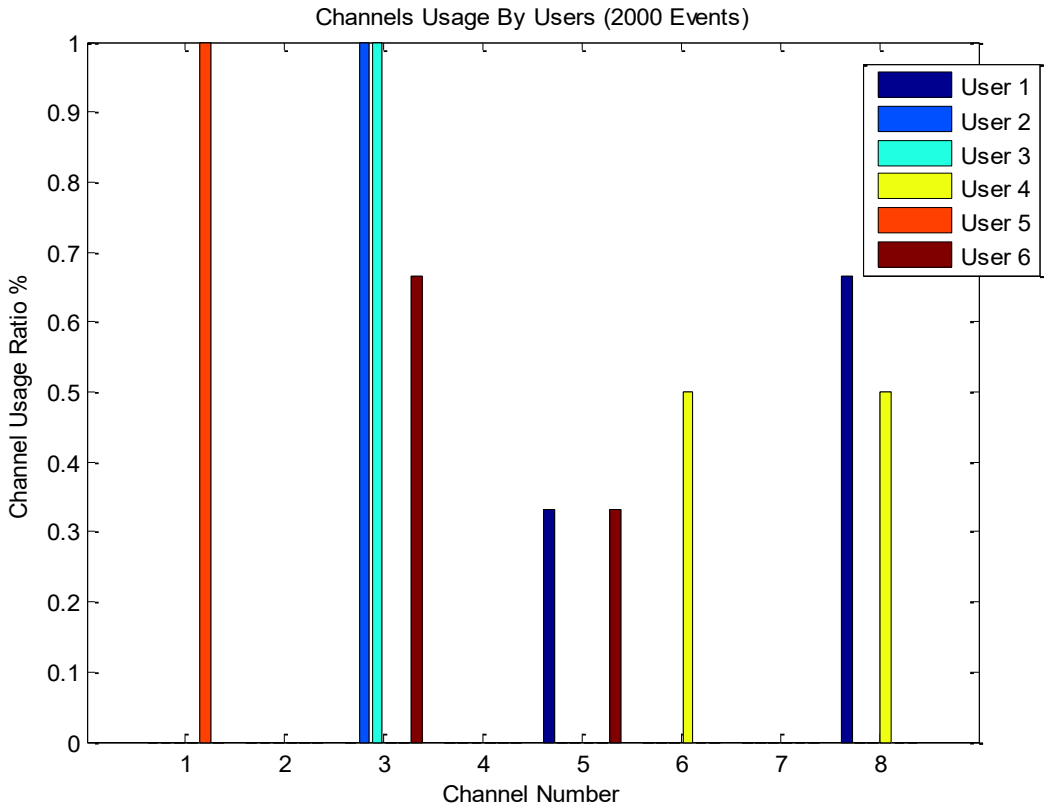


Figure 6.14. Channel Usage by Users (2000 Events)

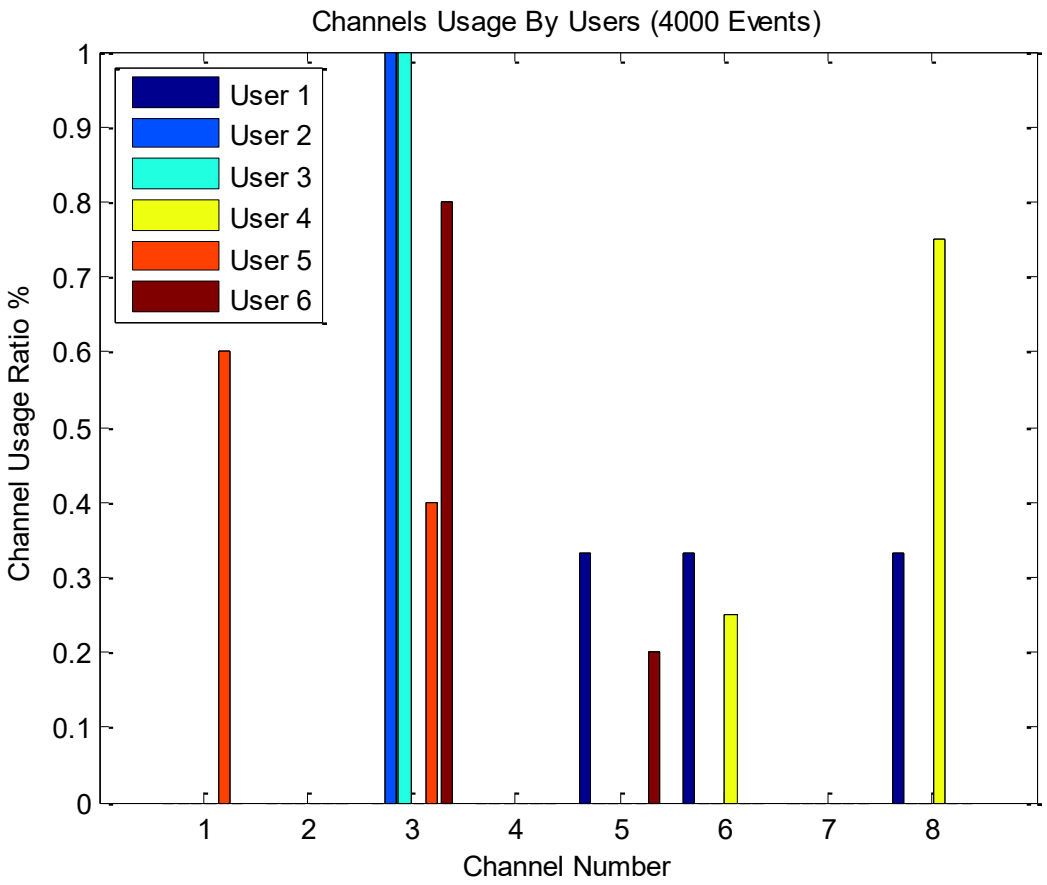


Figure 6.15. Channel Usage by Users (4000 Events)

6.9 Conclusions

This chapter presents a new quantum inspired reinforcement learning technique, QRL, for dynamic spectrum access of wireless communication networks. QRL is shown to increase the speed of learning convergence by up to one order of magnitude by introducing two essential modifications to the traditional reinforcement learning algorithm. The first, is, that the decision making process is dependent on a separate newly-introduced and well-ranked amplitude table. In this table the ranking is updated depending on the success or failure of decisions. The second is that, the exploration process is exclusively and immediately induced by the failure of the channel choice and is directed to the next best channel instead of random exploration, as used in traditional reinforcement learning

algorithms. As a consequence, and due to the improved adaptability of the new technique, the system capacity is improved in terms of blocking probability by 9% on the lowest tested traffic load value. This improvement is raised up to 84 % on the highest traffic load value where the difference in adaptability becomes clearer. This improvement is associated with a significant average file delay reduction of 26%. A system throughput improvement of up to 2.8% has also been gained.

Chapter 7: Future Work

7.1 Introduction

This chapter proposes some of the future research work possibilities based on the accomplishments of the work in this thesis. Dynamic Spectrum Access (DSA) has a central role for ultra-dense cognitive radio networks within 5G communication systems. The proposed Quantum Reinforcement Learning (QRL) technique demonstrated an ultra-high learning speed in certain circumstances. These specific criteria might solve problematic aspects within many learning systems when facing dynamic environments. It also supports fully distributed learning strategies that include a large number of agents learning together with a high possibility of reducing conflicts or collisions. The resulting fast learning showed that by relying exclusively on local information gained by the learning agent within the Access Base Station (ABS) it is possible to some extent to not lose the benefit of the high learning speed.

7.2 Intelligent LTE Systems

The most important differences between LTE and former systems like the 3G system are the base stations [142, 143]. Before LTE, there has been a need for an intelligent central node

like a RNC (Radio Network Controller) in 3G for example. The central node needed to control all the radio resources and mobility over multiple NodeB (3G base stations) underneath. NodeBs functions are based on the commands of RNC through Iub interface.

In LTE, on the other hand, Radio Resource Management is carried out in the eNBs (evolved NodeB), with signalling information exchange within the control plane over X2 interface as shown in figure 7.1. The eNBs in this case are allowed to use the entire frequency band. They manage the frequency allocations as described earlier in section 2.3.2.2 in the cell and sector to optimize all the UE's communication.

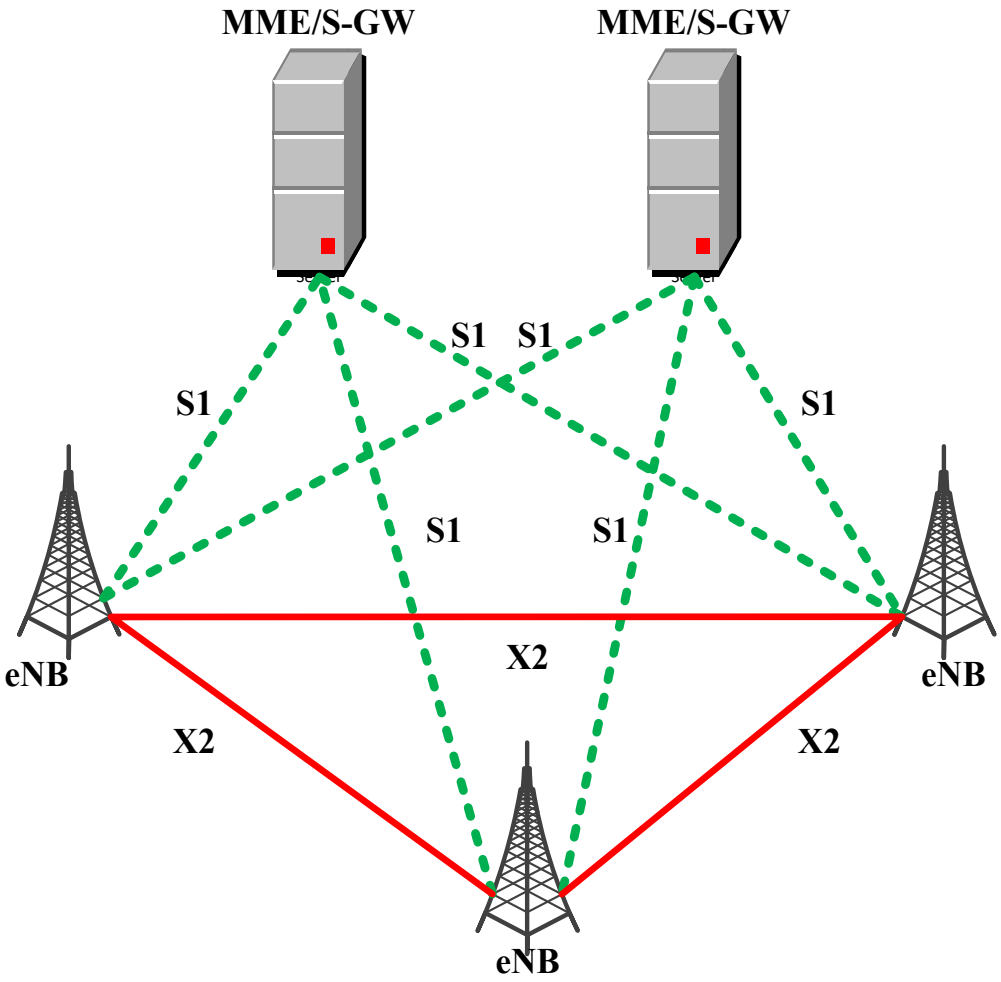


Figure 7. 1. E-UTRAN Architecture in LTE Systems

According to overview of 3GPP Release 8, the eNB functions include Radio Resource Management (RRM) which in turn includes for example:

- Radio Admission Control
- Connection Mobility Control
- Dynamic Spectrum Access (DSA) for UEs in both uplink and downlink (scheduling)

The performance of LTE eNB is highly affected by the radio resource management algorithm and its implementation. Based on the above mentioned functions of eNBs, learning techniques might be implemented within LTE systems and within eNBs as follows:

7.2.1 Intelligent Fractional Frequency Reuse (FFR)

FFR, is used within LTE as a frequency planning strategy to avoid interference among adjacent cells. It divides coverage space around eNBs into inner and outer zones to ensure full frequency reuse within inner zones (described in section 2.3.2.2). However, such strategy imposes frequency constraints that might limit the system capability to deal with dynamic environments with different spectrum demands from cell to another.

The mobility of UEs might be continuous and rapid. Their locations might change from outer to inner frequency allocation zones (or vice versa) around each eNBs. As a result, the number of users served by outer and inner zones might change which might need temporary frequency re-planning that does not cause interference with other cells which might suffer from the same fluctuation. In such a case, learning becomes a necessity to keep all UEs within each cell coverage area well served through removing the possibility of the lack of channel availability. The Learning agent within each eNB in this case has to learn how many and which channels to be allocated to each zone. Interaction between eNB and the surrounding environment through assigning channels to UEs and observing the resulting performance will build preferences that improves decision making process.

As has been discussed in chapter 6, QRL has been shown to be able to make the system more capable to be adaptable to fluctuations in traffic demand. Such capability might reduce the period of any possible drop in QoS due to rapid and dramatic change in traffic. Each fractional zone might have a learning engine within it. This will help to form a QRL-based FFR that is able to change frequency allocation policy and bandwidth according on location and amount of demand. In other words, QRL can be used to control the frequency reuse policy. The local interference environment can be learned fast and thus the frequency allocation can be changed accordingly.

7.2.2 Intelligent Connection Mobility Control:

Connection mobility control (CMC) is the function that is responsible for the management of radio resources in both connection (Handover) and idle modes of the UE. Handover decisions might be based on the following:

- UE mobility.
- eNodeB measurements.
- Neighbour cell load.
- Traffic distribution.
- Hardware resources.
- Operator defined policies.

Handover happens as a result of EU mobility between two different coverage areas of two cells that requires handing over to maintain QoS due to interference or signal strength fading. In this case, the cell where the UE starts from is referred to as the source cell while the cell that the UE ends being served by is referred to as the target cell. One of the most important points upon which the decision about the determination of the target cell is based on is the cell availability (i.e. can accommodate an additional UE). This case is very similar to the case

presented in this thesis of choosing a channel by the ABS. The other point is the QoS provided to the user by the target cell compared to that delivered by the source cell and other surrounding cells. Such information can be reported back to the source cell afterward.

A learning agent within each eNodeB in this case can be used to form an intelligent cell behaviour in choosing the best target cell in reference to UE location and used channel. A register (Q-table/Amplitude table) can be used within each cell to record and update the result of handing over each UE to a specific target cell through comparison between the performance delivered to the UEs in both cells. The accumulative experience over time can cause improvements in choosing target cells by source cells. Common knowledge among eNBs is gradually formed about performance of different channels by different cells. Such result will create a target cell preference list within each cell for each channel. An intelligent handover might present an improvement to the overall network service in case of dense and dynamic network.

7.3 Intelligent Topology Management

Topology management is usually applied as a method of controling the number of actively working ABSs within the wireless communication system. Such control aims to limit the system energy consumption without a significant loss in QoS performance. In other words, it is a trade-off between the energy saving and QoS performance.

In the literature, the blocking probability value is used mostly as a trigger for ABS activation. However, as the system capacity is mostly not constrained by the spectrum size but rather by interference, the interference level can be used as additional parameter. Moreover, the policy that is used to handover users from a deactivated ABS to another might not be necessarily the same all over the network coverage area as spectrum demand and traffic load might differ from an area to another.

A centralized learning agent might be set to learn preferences regarding ABS activations as well as target ABSs. Based on traffic load on each ABS, the desirability of deactivating this ABS changes. Other factors might be set as desirability affecters for choosing a target ABS like interference level and overall throughput.

For a system like BuNGee used in this thesis, setting two learning agents within each HBS for learning both which ABSs to turn off and which to use as target ABSs (giving 4 potential target ABS for each deactivated one) might help in trading of between energy saving and QoS level.

In a distributive scenario, an ABS among each four neighbouring ABSs might be set as a controller with two learning agents to decide which to deactivate and where to accommodate the served users. The choice for target ABS might include the controller ABS as well.
In both centralized and distributive scenarios, the results of the handover processes for MSs can be used to update preference list for target ABSs.

Quantum inspired RL can be used as a learning agent in this case which will support a fast and adaptive learning engine as found out before.

Moreover, the activation threshold for an ABS might not necessarily be a constant parameter (like blocking probability) under varying traffic loads, user applications or data demand. This requires a learning based engine for controlling the condition (threshold) of activating or deactivating the ABSs as well. This approach might be used with a different parameter like interference level or throughput level to avoid significant QoS drop in case where the threshold blocking probability level has not been reached.

7.4 Intelligent Power Control

Transmission power control has been used as a radio resource management (RRM) technique. Transmission power levels are regulated for both BSs and MSs[144]. Such

regulation gives the ability to decide the power level required for a successful transmission with reasonable QoS level as well as saving energy, that is:

1- Using the right power level saves energy and avoids unnecessary interference.
2- Increasing or lowering transmission power levels might increases or decreases the levels of frequency reuse for adjacent cells depending on traffic load and user location.

Power control is used in both uplinks and downlinks [145-147].

A learning agent can be used within each node to learn power levels required for certain coverage zones based on information exchanged among nodes. Using reinforcement learning might support learning based on repetitive channel allocation for users and feedback evaluation for these allocations using preference tables for each zone.

Communication systems including sectored frequency allocation zones like LTE might be able to benefit significantly from intelligent power control. This is because it is easier to create a preference table based on zones rather than distance ranges which might impose storage complexity.

Signal strength, interference level and throughput value for the transmitter can be used as calibrating parameters that can facilitate the learning of a proper power control policy. Based on the fact that the environmental parameters including interference levels and MS locations are all dynamic, it is expected to have a significant variation in power requirements.

Quantum inspired RL can be used as an example of a fast and efficient learning agent within eNBss to support mapping an ultimate policy for power levels for different transmission within different eNBs. An intelligent and energy-aware eNB may result in high energy savings over large covering areas.

An amplitude table might be set for each zone representing power level preferences. This is especially viable for moving users among inner and the outer 3 zones (based on interference level with the adjacent cells) as shown in figure 2.4. This is because moving from an outer

zone to the inner allows reducing transmission energy level to save energy. On the other hand moving further toward the outer zones, requires increasing the power to maintain QoS level.

Chapter 8: Conclusions

This thesis has studied the improvement of dynamic spectrum access using quantum inspired reinforcement learning techniques for highly dense capacity wireless networks. The focus has been made on speeding up existing reinforcement learning techniques to be able to learn an optimal policy faster through a better decision making process and more efficient knowledge base update.

Quantum inspired RL has improved the conventional RL algorithm in two ways. First, it improved the way the channels are searched and turned it into a structured and less complex search process. Moreover, a new way of channel ranking and preference list formation has been introduced through the amplitude table. The new table improved channel ranking and made it an independent process that does not need any parameter tuning unlike conventional RL algorithms.

The results showed a significant enhancement in both system capacity and performance. System capacity has been raised by an average of 9% at the lowest traffic load value point and reached 84% at the highest traffic value point from blocking probability perspective (i.e. the decrease in blocking probability). Average file delay has been reduced by an average of 26%. Moreover, throughput level has been increased by 2.8%.

8.1 Summary and Conclusions for Thesis Chapters

A short introduction to the subjects investigated in the rest of the thesis has been provided in chapter 1. Chapter 2, included background information as well as literature review for the

subjects of cognitive radio, dynamic spectrum access and machine learning. The introduction of the evolution and principles of cognitive radio in general have been presented. It also provided some insight into radio resource management techniques including the aspects that were in use within the rest of the thesis. An introduction to machine learning and its application to wireless communication networks has been illustrated. Definitions and a short introduction for both reinforcement learning and quantum computation concluded chapter 2.

Chapter 3 provided explanations about the modelling and programming techniques used for the models used to generate the results. A list and definitions of the performance parameters used for system evaluations have been presented.

Chapter 4 illustrated the three standard dynamic spectrum access techniques which can be used as a basis of comparison. First available channel (FAC), random channel assignment, and best SINR channel assignment techniques were discussed and applied for the system architecture used in this thesis. This chapter discussed the properties of each of the three conventional channel assignment techniques and introduced the goals of the proposed learning technique that are used in later chapters. The advantages of these three techniques were used as target inputs of our present quantum technique proposal.

Chapter 5 introduced quantum computation principles and the search techniques that have been embedded into the reinforcement learning engine. Properties of quantum search and the reason for its superiority over classical search has been discussed from a theoretical perspective. A theoretical comparison between the efficiency of classical and quantum search as a proof of its viability has been presented. This chapter has supported the necessary theoretical background and foundation for improving the decision making process within reinforcement learning. Moreover, it built the basis for the development of a fully quantum-inspired reinforcement learning scheme in the following chapter. Simulations for a spectrum allocation scheme using Grover search algorithm has been introduced for the first time. The

results are important introduction for the idea behind quantum inspired reinforcement learning.

Chapter 6 introduced the quantum inspired reinforcement learning technique. An introduction and a brief theoretical background for the fundamentals of reinforcement learning have been presented. The used weight driven reinforcement learning technique in this thesis has been outlined with its parameters explained. An introduction to quantum reinforcement learning has been presented with a flowchart that explains the modifications made to conventional RL. The QRL-based dynamic spectrum access algorithm is explained. Simulation results for the proposed QRL scheme along with the comparative results of two other different reinforcement learning techniques are presented. A weight driven reinforcement learning and a reinforcement learning algorithm that is based on random exploration after failure have been used for comparisons.

Chapter 7 presented some recommendations for further research work that can be accomplished based on the study of the current thesis and that can make use of its accomplishments. Recommendations for further research in intelligent LTE network spectrum management, intelligent topology management and intelligent power control schemes for energy saving networks were presented.

8.2 Summary of Novel Contributions

1- ***Introducing Quantum Inspired Reinforcement Learning into Wireless Communication Networks:*** Quantum inspired reinforcement learning has been introduced as a solution to improve the efficiency of the dynamic spectrum access in a wireless communication system for the first time. So far, the only practical aspect where this technique has been used is robotics.

2- ***Application of Quantum Inspired Reinforcement Learning to Multi Agent Reinforcement Learning:*** the only application involved the use of quantum

reinforcement learning has been for one or two agents and as a result much simpler applications.

3- ***Application of Grover Quantum Search Algorithm for Dynamic Spectrum Access in Wireless Communication Networks:*** This algorithm has been introduced both as a spectrum assignment scheme and as a decision making part within a much more efficient quantum reinforcement learning algorithm.

4- ***Separating The Searching Process from Learning:*** The process of channel search within reinforcement learning has always been considered as a part of the learning technique that changes only by changing the whole learning technique. A novel step has been made in the work of this thesis by dealing with the search as a separate process. As a result, it could be developed and improved without having to change anything within the learning part. Thus, the improvement that has been introduced into search can be used as a separate additional part that can be used straight forward within any type of reinforcement learning to have the same relative effect on it.

5- ***Redefining Convergence:*** Previous works carried out to speed up the reinforcement learning process, included reducing the searching domain size. The available spectrum has been divided into frequency bands for different learning engines to ensure that they need a shorter time to converge and make a preference list of channels. However, such methods imposed frequency band restrictions on the learning agents. The novel contribution of this thesis in this aspect is changing the definition of convergence into the first good quality channel to be found. No further exploration is performed until channel failure. This strategy resulted in a significant reduction in learning time needed which explains the difference in the results of the same RL technique with and without applying quantum techniques.

6- ***Adding a New Separate Preference Table:*** The amplitude table has been added as a separate preference list that benefits from the weight table (or Q-table) without actually changing anything in the way it is usually updated. The amplitude table has successfully improved the channel ranking as it ensured that channel ranking is based on quality rather than randomness.

The conclusion for the above mentioned novel proposals is a unified modification for all reinforcement learning algorithms. This is due to the fact that there are two separate parts that are added to the original algorithm. One serves as a searcher, and the other as a basis to choose a preference channel. Thus, the same algorithm can easily be added to any other RL algorithm.

Definitions

Cognitive Radio

A radio system employing a technology, which makes it possible to obtain knowledge of its operational environment, policies and internal state, to dynamically adjust its parameters and protocols according to the knowledge obtained and to learn from the results obtained.

Cognitive Agent

A wireless entity which that has the ability to observe the radio environment, making decisions regarding radio parameters, performing actions on the data transmission, learning from current and previous experiences, and training a knowledge base within it to improve future decisions. In this thesis, it represents the access base station (ABS).

Probability Amplitude

In quantum mechanics, probability amplitude is a complex number used in describing the behaviour of systems. The modulus squared of this quantity represents a probability or probability density.

Quantum Gate

In quantum computing and specifically the quantum circuit model of computation, a quantum gate (or quantum logic gate) is a basic quantum circuit operating on qubits. It represent the essential building block of quantum circuits, like classical logic gates are for conventional digital circuits.

Qubit

It is the fundamental unit for representing data in quantum computing. It has the function of a bit in classical computation with the difference that it can have the two values (1) and (0) at the same time.

Amplitude

A complex number used to describe the desirability of a certain action in quantum reinforcement learning.

Probability Amplitude

A complex number used to describe the behaviour of a system. The modulus squared of its value represents the probability or probability density,

Value Function

It is a function of states, it is used to estimate how good is it for an agent to be in a given state or how good it to perform a given action in a given state is. In this thesis, it represents the weight function.

Optimum Policy

It is the policy upon which the learning agent can achieve maximizing the gained rewards and minimizing the punishments on the long run.

Tensor Product

It is a multiplication process that is used with matrixes along with other applications. It is done by multiplying each element of the first matrix with each single element of the other.

Inner Product

It is a generalization of the dot product and a way of multiplying vectors with the result being scalar.

Glossary

ABS	**A**ccess **B**ase **S**tation
ACK	**Ack**nowledge
ALOHA	A random access protocol
BQC	**B**est **Q**uality **C**hannel
BS	**B**ase **S**tation
BuNGee	Beyond Next Generation
CAC	**C**all **A**dmission **C**ontrol
CIR	**C**ommitted **I**nformation **R**ate
CR	**C**ognitive **R**adio
CSMA	**C**arrier **S**ense **M**ultiple **A**ccess
CTS	**C**lear **T**o **S**end
DCA	**D**ynamic **C**hannel **A**ssignment
DIAC	**D**istributed **I**CIC **A**ccelerated **Q**-Learning
DSA	**D**ynamic **S**pectrum **A**ccess
eNB	**E**volution **N**ode **B**
FA	**F**requency **A**llocation
FAC	**F**irst **A**vailable **C**hannel
FCA	**F**ixed **C**hannel **A**ssignment
FFA	**F**ractional **F**requency **A**llocation
FFR	**F**ractional **F**requency **R**euse
FP	**F**requency **P**lanning
HARL	**H**euristically **A**ccelerated **R**einforcement **L**earning
HBS	**H**ub **B**ase **S**tation

HetNet	**H**eterogeneous **N**etworks
ICIC	**I**nter-**C**ell-**I**nterference **C**oordination
LADCA	**L**ocal **A**utonomous **D**ynamic **C**hannel **A**llocation
LIC	**L**east **I**nterference **C**hannel
LOS	**L**ine **O**f **S**ight
LTE	**L**ong **T**erm **E**volution
MAC	**M**ultiple **A**ccess **C**ontrol
MARL	**M**ulti-**A**gent **R**einforcement **L**earning
MME	**M**obility **M**anagement **E**ntity
MS	**M**obile **S**tation
NLOS	**N**on-**L**ine **O**f **S**ight
OFDMA	**O**rthogonal **F**requency **D**ivision **M**ultiple **A**ccess
OSA	**O**pportunistic **S**pectrum **A**ccess
PU	**P**rimary **U**ser
Q-CAC	**Q**-learning based **C**all **A**dmission **C**ontrol
QoS	**Q**uality **o**f **S**ervice
QRL	**Q**uantum **R**einforcement **L**earning
QS	**Q**uantum **S**earch
RCA	**R**andom **C**hannel **A**ssignment
REM	**R**adio **E**nvironment **M**ap
RL	**R**einforcement **L**earning
RNC	**R**adio **N**etwork **C**ontroller
RRM	**R**adio **R**esource **M**anagement
RTS	**R**equest **T**o **S**end

SINR	**S**ignal-to-**I**nterference plus **N**oise **R**atio
S-GW	**S**erving **G**ate**W**ay
SU	**S**econdary **U**ser
UHF	**U**ltra-**H**igh **F**requency
UL	**U**p **L**ink
WoLF	**W**in-**o**r-**L**earn **F**ast
WRAN	**W**ireless **R**egional **A**rea **N**etworks
WSN	**W**ireless **S**ensor **N**etworks

References

1. Yost, S. *mmWave: Battle of the Bands*. CISCO VNI 2016 2016; Available from: http://www.cisco.com/c/en/us/index.html.

2. Wang, B. and K.R. Liu, *Advances in cognitive radio networks: A survey.* Selected Topics in Signal Processing, IEEE Journal of, 2011. **5**(1): p. 5-23.

3. Akyildiz, I.F., et al., *NeXt generation/dynamic spectrum access/cognitive radio wireless networks: a survey.* Computer Networks, 2006. **50**(13): p. 2127-2159.

4. Gibson, J.D., *Mobile communications handbook*. Vol. 45. 2012: CRC press.

5. Haykin, S., *Cognitive radio: brain-empowered wireless communications.* Selected Areas in Communications, IEEE Journal on, 2005. **23**(2): p. 201-220.

6. Mitola III, J., *Cognitive radio architecture*. 2006: Springer.

7. Mitola III, J. and G.Q. Maguire Jr, *Cognitive radio: making software radios more personal.* Personal Communications, IEEE, 1999. **6**(4): p. 13-18.

8. Zhao, Q. and B.M. Sadler, *A survey of dynamic spectrum access.* Signal Processing Magazine, IEEE, 2007. **24**(3): p. 79-89.

9. Jiang, T., *Reinforcement Learning-based Spectrum Sharing for Cognitive Radio.* 2011.

10. Beck, R. and H. Panzer. *Strategies for handover and dynamic channel allocation in micro-cellular mobile radio systems*. in *Vehicular Technology Conference, 1989, IEEE 39th*. 1989. IEEE.

11. Sallberg, K., B. Stavenow, and B. Eklundh. *Hybrid channel assignment and reuse partitioning in a cellular mobile telephone system*. in *Vehicular Technology Conference, 1987. 37th IEEE*. 1987. IEEE.

12. Nettleton, R.W. and G.R. Schloemer. *A high capacity assignment method for cellular mobile telephone systems*. in *Vehicular Technology Conference, 1989, IEEE 39th*. 1989. IEEE.

13. Serizawa, M. and D.J. Goodman. *Instability and deadlock of distributed dynamic channel allocation*. in *Vehicular Technology Conference, 1993., 43rd IEEE*. 1993. IEEE.

14. Zander, J., *Radio resource management in future wireless networks: requirements and limitations.* Communications Magazine, IEEE, 1997. **35**(8): p. 30-36.

15. Niyato, D. and E. Hossain, *Cognitive radio for next-generation wireless networks: an approach to opportunistic channel selection in IEEE 802.11-based wireless mesh.* Wireless Communications, IEEE, 2009. **16**(1): p. 46-54.

16. Srinivasa, S. and S.A. Jafar, *Cognitive radios for dynamic spectrum access-the throughput potential of cognitive radio: A theoretical perspective.* Communications Magazine, IEEE, 2007. **45**(5): p. 73-79.

17. Kolodzy, P. and B. Fette, *Communications policy and spectrum management.* Cognitive radio technology, 2009. **64**.

18. Marshall, P., *Quantitative analysis of cognitive radio and network performance*. 2010: Artech House.

19. Teng, Y., et al. *Reinforcement learning based auction algorithm for dynamic spectrum access in cognitive radio networks*. in *Vehicular Technology Conference Fall (VTC 2010-Fall), 2010 IEEE 72nd*. 2010. IEEE.

20. Jiang, T., D. Grace, and Y. Liu. *Performance of cognitive radio reinforcement spectrum sharing using different weighting factors*. in *Communications and Networking in China, 2008. ChinaCom 2008. Third International Conference on*. 2008. IEEE.

21. Jiang, T., D. Grace, and Y. Liu, *Two-stage reinforcement-learning-based cognitive radio with exploration control.* IET communications, 2011. **5**(5): p. 644-651.

22. Jiang, T., D. Grace, and P.D. Mitchell. *Improvement of pre-partitioning on reinforcement learning based spectrum sharing*. in *Wireless Mobile and Computing (CCWMC 2009), IET International Communication Conference on*. 2009. IET.

23. Morozs, N., *Accelerating Reinforcement Learning for Dynamic Spectrum Access in Cognitive Wireless Networks*. 2015, University of York.

24. Jiang, T., D. Grace, and P. Mitchell, *Efficient exploration in reinforcement learning-based cognitive radio spectrum sharing.* IET communications, 2011. **5**(10): p. 1309-1317.

25. Jiang, T. and D. Grace, *Reinforcement Learning-Based Cognitive Radio for Open Spectrum Access.* Cognitive Communications: Distributed Artificial Intelligence (DAI), Regulatory Policy and Economics, Implementation, 2012: p. 195.

26. Ltd, Q., *Cognitive Radio Technology:A Study for Ofcom.* 2007. **1**(1.1).

27. Iii, J.M., *An integrated agent architecture for software defined radio.* 2000.

28. Fette, B.A., *Cognitive radio technology*. 2009: Access Online via Elsevier.

29. Hossain, E., D. Niyato, and Z. Han, *Dynamic spectrum access and management in cognitive radio networks*. 2009: Cambridge University Press.

30. Akyildiz, I.F., et al., *A survey on spectrum management in cognitive radio networks.* Communications Magazine, IEEE, 2008. **46**(4): p. 40-48.

31. Zander, J., et al., *Radio resource management for wireless networks*. 2001: Artech House, Inc.

32. Katzela, I. and M. Naghshineh, *Channel assignment schemes for cellular mobile telecommunication systems: A comprehensive survey.* Personal Communications, IEEE, 1996. **3**(3): p. 10-31.

33. Rappaport, T.S., *Wireless Communications--Principles and Practice, (The Book End).* Microwave Journal, 2002. **45**(12): p. 128-129.

34. JAVAN, A.A.K. and M.R. POURMIR, *Resource Allocation in OFDMA Wireless Communications Systems Supporting Multimedia Services.* Cumhuriyet Science Journal, 2015. **36**(3): p. 3495-3505.

35. Choi, K.W., W.S. Jeon, and D.G. Jeong, *Resource allocation in OFDMA wireless communications systems supporting multimedia services.* Networking, IEEE/ACM Transactions on, 2009. **17**(3): p. 926-935.

36. *BuNGee, Broadband Radio Access Networks (BRAN); Very high capacity density BWA networks; System architecture, economic model and derivation of technical requirements*. ETSI TR 101 534 V1.1.1:[Available from: www.ict-bungee.eu.

37. Tobagi, F.A. and L. Kleinrock, *Packet switching in radio channels: Part ii--the hidden terminal problem in carrier sense multiple-access and the busy-tone solution.* Communications, IEEE Transactions on, 1975. **23**(12): p. 1417-1433.

38. Zhao, Q., *Intelligent Radio Resource Management for Mobile Broadband Networks*. 2013, University of York.

39. Tanenbaum, A.S., *Computer networks, 4-th edition.* ed: Prentice Hall, 2003.

40. Zhang, M. and T.-S.P. Yum, *Comparisons of channel-assignment strategies in cellular mobile telephone systems.* Vehicular Technology, IEEE Transactions on, 1989. **38**(4): p. 211-215.

41. Saquib, N., E. Hossain, and D.I. Kim, *Fractional frequency reuse for interference management in LTE-advanced hetnets.* Wireless Communications, IEEE, 2013. **20**(2): p. 113-122.

42. Ghaffar, R. and R. Knopp. *Fractional frequency reuse and interference suppression for OFDMA networks*. in *Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt), 2010 Proceedings of the 8th International Symposium on*. 2010. IEEE.

43. Ali, S.H. and V. Leung, *Dynamic frequency allocation in fractional frequency reused OFDMA networks.* Wireless Communications, IEEE Transactions on, 2009. **8**(8): p. 4286-4295.

44. Holma, H. and A. Toskala, *LTE for UMTS: Evolution to LTE-advanced*. 2011: John Wiley & Sons.

45. Lindbom, L., et al., *Enhanced inter-cell interference coordination for heterogeneous networks in LTE-Advanced: A survey.* arXiv preprint arXiv:1112.1344, 2011.

46. Pedersen, K.I., et al., *Enhanced inter-cell interference coordination in co-channel multi-layer LTE-advanced networks.* Wireless Communications, IEEE, 2013. **20**(3): p. 120-127.

47. Sesia, S., I. Toufik, and M. Baker, *LTE–The UMTS Long Term Evolution*. 2001, Wiley Online Library.

48. Niu, Z., *TANGO: traffic-aware network planning and green operation.* Wireless Communications, IEEE, 2011. **18**(5): p. 25-29.

49. Kleinrock, L., *Queueing Systems - Volume 1 Theory*. 1975: John Wiley & Sons.

50. Song, M., et al., *Dynamic spectrum access: from cognitive radio to network radio.* Wireless Communications, IEEE, 2012. **19**(1): p. 23-29.

51. Karimi, H.R., et al. *European harmonized technical conditions and band plans for broadband wireless access in the 790-862 MHz digital dividend spectrum*. in *New Frontiers in Dynamic Spectrum, 2010 IEEE Symposium on*. 2010. IEEE.

52. Ofcom, *The award of 800 MHz and 2.6 GHz spectrum*. 2012.

53. QINETIQ, *Cognitive Radio Technology: A Study for Ofcom*. 2007.

54. Grace, D., H. Zhang, and M. Nekovee, *Cognitive communications [Editorial]*. Communications, IET, 2012. **6**(8): p. 783-784.

55. Subramanian, A.P., et al. *Near-optimal dynamic spectrum allocation in cellular networks*. in *New Frontiers in Dynamic Spectrum Access Networks, 2008. DySPAN 2008. 3rd IEEE Symposium on*. 2008. IEEE.

56. Yilmaz, H.B., et al., *Radio environment map as enabler for practical cognitive radio networks.* Communications Magazine, IEEE, 2013. **51**(12): p. 162-169.

57. Zhao, Y., et al. *Radio environment map enabled situation-aware cognitive radio learning algorithms*. in *Proc. SDR Forum Technical Conference*. 2006.

58. Zhao, Y., et al. *Overhead analysis for radio environment mapenabled cognitive radio networks*. in *Networking Technologies for Software Defined Radio Networks, 2006. SDR'06.1 st IEEE Workshop on*. 2006. IEEE.

59. FARAMIR, F. *Enabling Spectrum-Aware Radio Access for Cognitive Radios*. Available from: www.ict-faramir.eu.

60. Networks, I.W.G.o.W.R.A.; Available from: www.ieee802.org.

61. ETSI, W.S.D.W., *Wireless Access Systems operating in the 470 MHz to 790 MHz frequency band; Harmonized EN covering the essential requirements of article 3.2 of the R&TTE Directive*. 2013.

62. Yücek, T. and H. Arslan, *A survey of spectrum sensing algorithms for cognitive radio applications.* Communications Surveys & Tutorials, IEEE, 2009. **11**(1): p. 116-130.

63. Grace, D., A. Burr, and T. Tozer, *Distributed channel assignment strategies using coexistence etiquettes for land based radio environment.* Electronics Letters, 1996. **32**(21): p. 1956-1956.

64. Sutton, R.S. and A.G. Barto, *Reinforcement learning: An introduction*. Vol. 1. 1998: Cambridge Univ Press.

65. Russell, S.J., et al., *Artificial intelligence: a modern approach*. Vol. 74. 1995: Prentice hall Englewood Cliffs.

66. Ghahramani, Z., *Unsupervised learning*, in *Advanced Lectures on Machine Learning*. 2004, Springer. p. 72-112.

67. Mitchell, T.M., *Machine learning. WCB*. 1997, McGraw-Hill Boston, MA:.

68. Ghahramani, Z. and M.I. Jordan. *Supervised learning from incomplete data via an EM approach*. in *Advances in Neural Information Processing Systems 6*. 1994. Citeseer.

69. Kaelbling, L.P., M.L. Littman, and A.W. Moore, *Reinforcement learning: A survey.* arXiv preprint cs/9605103, 1996.

70. N. P. Barde, P.P.B., D. P. Thakur, K. M. Jadhar, *Basics of Quantum Computation.* 2013.

71. Lo, H.-K., T. Spiller, and S. Popescu, *Introduction to quantum computation and information*. 1998: World Scientific.

72. Yanofsky, N.S. and M.A. Mannucci, *Quantum computing for computer scientists*. Vol. 20. 2008: Cambridge University Press Cambridge.

73. Dong, D., C. Chen, and Z. Chen, *Quantum reinforcement learning*, in *Advances in Natural Computation*. 2005, Springer. p. 686-689.

74. Dong, D., et al., *Robust quantum-inspired reinforcement learning for robot navigation.* Mechatronics, IEEE/ASME Transactions on, 2012. **17**(1): p. 86-97.

75. Dong, D., et al., *Quantum reinforcement learning.* Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 2008. **38**(5): p. 1207-1220.

76. Akerberg, D. and F. Brouwer. *On channel definitions and rules for continuous dynamic channel selection in coexistence etiquettes for radio systems*. in *Vehicular Technology Conference, 1994 IEEE 44th*. 1994. IEEE.

77. Foschini, G.J. and Z. Miljanic, *Distributed autonomous wireless channel assignment algorithm with power control.* Vehicular Technology, IEEE Transactions on, 1995. **44**(3): p. 420-429.

78. Saunders, S.R. and F. Bonar, *Prediction of mobile radio wave propagation over buildings of irregular heights and spacings.* Antennas and Propagation, IEEE Transactions on, 1994. **42**(2): p. 137-144.

79. Law, A., L.B. Lopes, and R.S. Saunders. *Comparison of performance of FA and LIC DCA call assignment within DECT*. in *Networking Aspects of Radio Communication Systems, IEE Colloquium on*. 1996. IET.

80. Gamst, A., *Some lower bounds for a class of frequency assignment problems.* Vehicular Technology, IEEE Transactions on, 1986. **35**(1): p. 8-14.

81. Zander, J., *Performance of optimum transmitter power control in cellular radio systems.* Vehicular Technology, IEEE Transactions on, 1992. **41**(1): p. 57-62.

82. Grace, D., T.C. Tozer, and A.G. Burr, *Reducing call dropping in distributed dynamic channel assignment algorithms by incorporating power control in wireless ad hoc*

networks. Selected Areas in Communications, IEEE Journal on, 2000. **18**(11): p. 2417-2428.

83. Čabrić, D., et al. *A cognitive radio approach for usage of virtual unlicensed spectrum*. in *14th IST Mobile and Wireless Communications Summit*. 2005.

84. Mody, A.N., et al., *Recent advances in cognitive communications.* Communications Magazine, IEEE, 2007. **45**(10): p. 54-61.

85. Thomas, R.W., et al., *Cognitive networks: adaptation and learning to achieve end-to-end performance objectives.* Communications Magazine, IEEE, 2006. **44**(12): p. 51-57.

86. Nie, J. and S. Haykin, *A dynamic channel assignment policy through Q-learning.* Neural Networks, IEEE Transactions on, 1999. **10**(6): p. 1443-1455.

87. Sivarajan, K.N., R.J. McEliece, and J.W. Ketchum. *Dynamic channel assignment in cellular radio*. in *Vehicular Technology Conference, 1990 IEEE 40th*. 1990. IEEE.

88. Senouci, S.-M. and G. Pujoile. *Dynamic channel assignment in cellular networks: a reinforcement learning solution*. in *Telecommunications, 2003. ICT 2003. 10th International Conference on*. 2003. IEEE.

89. Li, H. *Multi-agent Q-learning of channel selection in multi-user cognitive radio systems: A two by two case*. in *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on*. 2009. IEEE.

90. Xia, B., et al. *Reinforcement learning based spectrum-aware routing in multi-hop cognitive radio networks*. in *Cognitive Radio Oriented Wireless Networks and Communications, 2009. CROWNCOM'09. 4th International Conference on*. 2009. IEEE.

91. Bernardo, F., et al. *Distributed spectrum management based on reinforcement learning*. in *Cognitive Radio Oriented Wireless Networks and Communications, 2009. CROWNCOM '09. 4th International Conference on*. 2009.

92. Yang, M. and D. Grace. *Cognitive radio with reinforcement learning applied to heterogeneous multicast terrestrial communication systems*. in *Cognitive Radio Oriented Wireless Networks and Communications, 2009. CROWNCOM'09. 4th International Conference on*. 2009. IEEE.

93. Yau, K.-L., P. Komisarczuk, and P.D. Teal. *A context-aware and intelligent dynamic channel selection scheme for cognitive radio networks*. in *Cognitive Radio Oriented Wireless Networks and Communications, 2009. CROWNCOM'09. 4th International Conference on*. 2009. IEEE.

94. Yau, K.-L., P. Komisarczuk, and P.D. Teal. *Enhancing network performance in distributed cognitive radio networks using single-agent and multi-agent reinforcement learning*. in *Local Computer Networks (LCN), 2010 IEEE 35th Conference on*. 2010. IEEE.

95. Yau, K.-L., P. Komisarczuk, and P.D. Teal. *Applications of reinforcement learning to cognitive radio networks*. in *Communications Workshops (ICC), 2010 IEEE International Conference on*. 2010. IEEE.

96. Wu, C., et al. *Spectrum management of cognitive radio using multi-agent reinforcement learning*. in *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: Industry track*. 2010. International Foundation for Autonomous Agents and Multiagent Systems.

97. Lo, B.F. and I.F. Akyildiz. *Reinforcement learning-based cooperative sensing in cognitive radio ad hoc networks*. in *Personal Indoor and Mobile Radio Communications (PIMRC), 2010 IEEE 21st International Symposium on*. 2010. IEEE.

98. Galindo-Serrano, A. and L. Giupponi, *Distributed Q-learning for aggregated interference control in cognitive radio networks.* Vehicular Technology, IEEE Transactions on, 2010. **59**(4): p. 1823-1834.

99. Chen, Z. and R.C. Qiu, *Q-learning based bidding algorithm for spectrum auction in cognitive radio.* Proceedings of the IEEE SoutheastCon 2011 Building Global Engineers, 2011: p. 409-412.

100. Jiang, T., D. Grace, and Y. Liu. *Cognitive radio spectrum sharing schemes with reduced spectrum sensing requirements*. in *Cognitive Radio and Software Defined Radios: Technologies and Techniques, 2008 IET Seminar on*. 2008. IET.

101. Morozs, N., T. Clarke, and D. Grace. *A novel adaptive call admission control scheme for distributed reinforcement learning based dynamic spectrum access in cellular networks*. in *Wireless Communication Systems (ISWCS 2013), Proceedings of the Tenth International Symposium on*. 2013. VDE.

102. Morozs, N., T. Clarke, and D. Grace, *Case-Based Cognitive Cellular Systems for Temporary Events.*

103. Morozs, N., D. Grace, and T. Clarke. *Case-based reinforcement learning for cognitive spectrum assignment in cellular networks with dynamic topologies*. in *Military Communications and Information Systems Conference (MCC), 2013*. 2013. IEEE.

104. Morozs, N., et al. *Distributed Q-learning based dynamic spectrum management in cognitive cellular systems: Choosing the right learning rate*. in *Computers and Communication (ISCC), 2014 IEEE Symposium on*. 2014. IEEE.

105. Morozs, N., T. Clarke, and D. Grace, *Distributed heuristically accelerated Q-learning for robust cognitive spectrum management in LTE cellular systems.* 2015.

106. Morozs, N., T. Clarke, and D. Grace, *Heuristically Accelerated Reinforcement Learning for Dynamic Secondary Spectrum Sharing.*

107. Morozs, N., T. Clarke, and D. Grace, *Short Paper: Intelligent Dynamic Spectrum Access in Cellular Systems with Asymmetric Topologies and Non-Uniform Traffic Loads.*

108. Grover, L.K. *A fast quantum mechanical algorithm for database search*. in *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*. 1996. ACM.

109. Meng, X., et al. *A novel multiagent reinforcement learning algorithm combination with quantum computation*. in *Intelligent Control and Automation, 2006. WCICA 2006. The Sixth World Congress on*. 2006. IEEE.

110. Daoyi, D., C. Chunlin, and L. Hanxiong. *Reinforcement strategy using quantum amplitude amplification for robot learning*. in *Control Conference, 2007. CCC 2007. Chinese*. 2007. IEEE.

111. Chen, C. and D. Dong. *Quantum parallelization of hierarchical Q-learning for global navigation of mobile robots*. in *Networking, Sensing and Control (ICNSC), 2012 9th IEEE International Conference on*. 2012. IEEE.

112. Chen, C., et al. *A quantum-inspired q-learning algorithm for indoor robot navigation*. in *Networking, Sensing and Control, 2008. ICNSC 2008. IEEE International Conference on*. 2008. IEEE.

113. Mo, J., *Performance Modeling of Communication Networks with Markov Chains.* Synthesis Lectures on Data Management, 2010. **3**(1): p. 1-90.

114. Aniba, G. and S. Aïssa, *A general traffic and queueing delay model for 3G wireless packet networks*, in *Telecommunications and Networking-ICT 2004*. 2004, Springer. p. 942-949.

115. Klemm, A., C. Lindemann, and M. Lohmann. *Traffic modeling and characterization for UMTS networks*. in *Global Telecommunications Conference, 2001. GLOBECOM'01. IEEE*. 2001. IEEE.

116. Saunders, S. and A. Aragón-Zavala, *Antennas and propagation for wireless communication systems*. 2007: John Wiley & Sons.

117. Prasad, R., *Simulation and Software Radio for Mobile Communications (Book) with CDROM*. 2002: Artech House.

118. Grace, D., *Distributed dynamic channel assignment for the wireless environment*. 1999, Citeseer.

119. T. Jiang, A.P., D. Grace, A. Burr, *Interim Simulation.* Available: http://www.ict-bungee.eu/, 2011.

120. Saadawi, T.N., M.H. Ammar, and A. El Hakeem, *Fundamentals of telecommunication networks*. 1994: Wiley-Interscience.

121. M. Goldhamer, e.a., *Baseline BuNGee Architecture.* 2010.

122. Tao Jiang, P.L., Chunshan Liu, Nizabat Khan, David Grace, Alister Burr, Claude Oestges, *Simulation Tool(s) and Simulation Results.* Available: http://www.ict-bungee.eu/, 2010.

123. Molisch, A.F., *Wireless communications*. 2007: John Wiley & Sons.

124. al., P.K.e., *D1.1.2 V1.1WINNER II Channel Models.* 2007.

125. N. P. Barde, P.P.B., D. P. Thakur, K. M. Jadhar, *Basics of Quantum Computation*. 2013.

126. Busemeyer, J.R. and P.D. Bruza, *Quantum models of cognition and decision*. 2012: Cambridge University Press.

127. Yanofsky, N.S., M.A. Mannucci, and M.A. Mannucci, *Quantum computing for computer scientists*. Vol. 20. 2008: Cambridge University Press Cambridge.

128. Nielsen, M.A. and I.L. Chuang, *Quantum computation and quantum information*. 2010: Cambridge university press.

129. McMahon, D., *Quantum computing explained*. 2007: John Wiley & Sons.

130. Williams, C.P., *Explorations in quantum computing*. 2010: Springer Science & Business Media.

131. Grover, L.K., *Quantum mechanics helps in searching for a needle in a haystack.* Physical review letters, 1997. **79**(2): p. 325.

132. Bruß, D. and G. Leuchs, *Lectures on quantum information.* 2007.

133. Jones, J.A. and D. Jaksch, *Quantum information, computation and communication*. 2012: Cambridge University Press.

134. Wichert, A., *Principles of quantum artificial intelligence*. 2013: World Scientific.

135. Pittenger, A.O., *An introduction to quantum computing algorithms*. Vol. 19. 2012: Springer Science & Business Media.

136. Jaeger, G., *Quantum information*. 2007: Springer.

137. Barnett, S., *Quantum information*. Vol. 16. 2009: Oxford University Press.

138. Nuuman, S., D. Grace, and T. Clarke. *A quantum inspired reinforcement learning technique for beyond next generation wireless networks*. in *Wireless Communications and Networking Conference Workshops (WCNCW), 2015 IEEE*. 2015. IEEE.

139. Kapetanakis, S. and D. Kudenko, *Reinforcement learning of coordination in cooperative multi-agent systems.* AAAI/IAAI, 2002. **2002**: p. 326-331.

140. Bublin, M., et al. *Distributed spectrum sharing by reinforcement and game theory*. in *5th Karlsruhe workshop on software radio, Karlsruhe, Germany*. 2008.

141. Chen, C.-L. and D.-Y. Dong, *Superposition-Inspired Reinforcement Learning and Quantum Reinforcement Learning.* 2008.

142. Ghosh, A., et al., *Fundamentals of LTE*. 2010: Pearson Education.

143. Ghosh, A. and R. Ratasuk, *Essentials of LTE and LTE-A*. 2011: Cambridge University Press.

144. Han, Z. and K.R. Liu, *Resource allocation for wireless networks: basics, techniques, and applications*. 2008: Cambridge University Press.

145. Biton, E., et al., *Distributed inter-cell interference mitigation via joint scheduling and power control under noise rise constraints.* Wireless Communications, IEEE Transactions on, 2014. **13**(6): p. 3464-3477.

146. Jin, Y., F. Cao, and R.A. Dziyauddin. *Inter-cell interference mitigation with coordinated resource allocation and adaptive power control*. in *Wireless Communications and Networking Conference (WCNC), 2014 IEEE*. 2014. IEEE.

147. Chen, J., C.-C. Yang, and S.-T. Sheu, *Downlink femtocell interference mitigation and achievable data rate maximization: using FBS association and transmit power-control schemes.* Vehicular Technology, IEEE Transactions on, 2014. **63**(6): p. 2807-2818.