



**AUTHORSHIP**

*During the project, I have read and consulted many different materials from textbooks, monographs to articles published at home and abroad. I hereby certify that what I have written below is completely authentic, true, the experimental measurements obtained in the thesis are not reproduced from any document in any way. Those results are what I have researched, accumulated in the process of making this thesis.*

*I take full responsibility for any sign of copying results from other documentation.*

*Hanoi, May 10th 2021*

Student

Hoang Vu Long

**ACKNOWLEDGEMENTS**

*First of all, I would like to express my deepest gratitude and gratitude to TS. Lam Sinh Cong, who enthusiastically instructed and guided me throughout the graduation thesis process.*

*I sincerely thank to all teachers who created favorable conditions for me to study and research at University of Technology - Hanoi National University, has helped me a lot in supporting the professional knowledge as well as the facilities to complete the thesis well.*

*Finally, I would like to thank my family and friends, loved ones who are always there and encourage me throughout the graduation thesis process.*

*In the process of making this thesis, due to limitations in knowledge, unavoidable shortcomings, I look forward to receiving comments and corrections from teachers and friends to complete this thesis.*

*Sincerely thanks!*

*Hanoi, May 10th 2021*

Student

Hoang Vu Long

**TABLE OF CONTENTS**

LIST OF FIGURES 5

LIST OF ACRONYMS 6

ABSTRACT 7

**CHAPTER 1.** OVERVIEW OF 5G MOBLIE NETWORK 8

1.1. Introduction 8

1.2. 5G network architecture 9

1.3. Approaching methods and technical requirements 11

1.3.1. Methods of approach 12

1.3.2. Technical requirements 13

1.4. New insights in 5G Development 14

1.4.1. Extremely short Wave (mmWave) 14

1.4.2. Massive MIMO and Beamforming 15

**CHAPTER 2.** 5G ULTRA-DENSE NETWORK 18

1. 2.1. Overview 18
2. 2.2. 5G UDN System model 19
3. 2.3. Energy efficiency and Spectral efficiency 21
4. 2.4.Optimization problem of resource allocation in 5G UDN 21

**CHAPTER 3.** SOLUTIONS FOR POWER OPTIMIZATION 23

3.1. Fundamentals of Reinforcement Learning (RL) and Q-Learning 23

3.1.1. Overview of RL 23

3.1.2. Overview of Q-Learning 25

3.1.3. Q–Learning based resource allocation in 5G UDN 26

3.1.4. Q–Learning drawbacks 27

3.2. Deep Q Network (DQN) 27

3.2.1. Overview 27

3.2.2. Model structure 28

3.3. Dueling DQN 29

3.3.1. Overview 29

3.3.2. Model structure 30

**CHAPTER 4.** SIMULATION AND RESULTS 33

4.1. Constructing the program of simulation 33

4.1.1. Python libraries for ML used in simulation 33

4.1.2. Dynamic resource allocation model of DQN and Dueling DQN 35

4.1.3. Simulation Parameters 36

4.2.Simulation Results 38

CONCLUSION AND FUTURE ORIENTATION 41

APPENDIX 42

PREFERENCES 43

**LIST OF FIGURES**

**CHAPTER 1**

Figure 1.1. 5G Architecture based on area distribution and

connection types [6] 9

Figure 1.2. 5G Overall Architecture [11] 11

Figure 1.3. Three approaches for future IMT [12] 12

Figure 1.4. Length of spectrums for 5G [13] 14

Figure 1.5. Massive MIMO Model with 3D Beamforming [14] 17

**CHAPTER 2**

Figure 2.1. 5G UDN Model [17] 19

**CHAPTER 3**

Figure 3.1. Reinforcement learning model [22] 24

Figure 3.2. Memory table of Q-Learning [9] 26

Figure 3.3. Deep Q Learning model [9] 28

Figure 3.4. Deep Q Network Structure [9] 28

Figure 3.5. Dueling DQN model diagram 30

**CHAPTER 4**

Figure 4.1. The illustrated diagram of Dueling DQN [25] 35

Figure 4.2. The illustrated diagram of DQN [25] 36

Figure 4.3. Total power consumption when UE demand rises 38

Figure 4.4. Average power consumption in time varying user demand scenario 39

**LIST OF ACRONYMS**

|  |  |
| --- | --- |
| NR | New Radio |
| UDN | Ultra-dense Network |
| EE | Energy Efficiency |
| SE | Spectral Efficiency |
| IoT | Internet of Things |
| MDP | Markov Decision Process |
| ML | Machine Learning |
| RL | Reinforcement Learning |
| DQN | Deep Q Network |
| UE | User Equipment |
| EMBB | Enhanced Mobile Broadband |
| MMC | Massive Machine Communication |
| V2V | Vehicles-to-vehicles |
| URC | Ultra–reliable Communication |
| IMT | Information Management Technology |
| QoE | Quality of Experience |
| MIMO | Multiple Input multiple output |
| AP | Access Point |
| BS | Base Station |
| QoS | Quality of Service |
| MgNB | Macro gNB |
| SgNB | Small gNB |
| RB | Resource Block |
| SUE | Small UE |
| NN | Neural Network |
| API | Application Programming Interface |
| SINR | Signal to interfere-plus-noise ratio |

**ABSTRACT**

***Abstract****:* Currently, the rapid development of Internet of Things (IoT) devices have driven a sharp increase in the density of UEs in the wireless networks, particularly 5G cellular system [1]. Thus, ultra-dense network (UDN) is being introduced as a potential network topology of the 5G network [14]. In this thesis, we study the dynamic resource allocation method in 5G UDN which considered two indicators, called energy efficiency (EE) and spectral efficiency (SE). In a network environment, EE and SE are always inversely proportional to the other, consequently. Hence, dynamic resource allocation has to deal with the problem of power optimization [15]. This optimization is also described as a Markov decision process (MDP) [8]. At the moment, some Deep Learning methods and algorithms have been introduced and used as traditional Q-Learning, its modified version such as Deep Q Network and dueling DQN have been introduced to optimize power consumption and spectral efficiency. This thesis aims to provide a brief overview about 5G UDN, EE and SE optimization problem and a comparison between traditional Q-learning and dueling DQN.

***Index: UDN, EE, SE, MDP, Deep RL, Dueling DQN.***

**Chapter 1. OVERVIEW OF 5G MOBILE NETWORK**

Chapter 1 provides an overview of 5G mobile networks. We start with a description of the architecture of 5G networks and consider the technical requirements and methods of approach. In addition, chapter 1 deals with several problems and solutions in building and maintaining network operations. Then, we go into detail with the typical new technologies used for data transmission in 5G.

* 1. **Introduction**

5G is the 5th generation mobile network. This is the new global wireless network model after 1G, 2G, 3G, and 4G networks. 5G is designed to connect everyone and everything together, including machines, objects, and equipment [1].

5G is a research project for the next communication generation based on the current communication technology 4G. 5G is expected to be practical in the 2020s. 5G has not been officially used by any official organization or document. Research is being promoted to ensure that the rollout of 5G is accelerated [1].

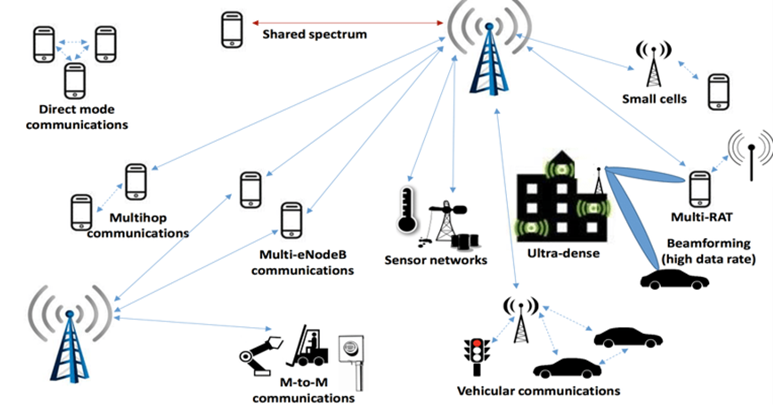
Thanks to the higher speed and data density, 5G offers many businesses and opportunities to grow. As current deployed system 4G, we have seen that companies take advantage of this by developing motorbike, taxi services, tech-taxis, food delivery, etc. 5G is 10 times faster than 4G and much more capacity - this means a lot of new businesses could emerge. The IoT (Internet of Things) is also a feature that seems particularly relevant to 5G technology [1]. Currently, IoT is being popularly used in the manufacturing sector to track factories, as well as in the transportation sector to track vehicles.

5G has the potential to be faster than any Wifi network, and it can operate as long as you get a signal there. That means the IoT equipments can be operated anywhere without being limited to local wireless networks. Today, the rise of connected objects and devices will create the way for a host of new services and related business models that enable automation in different industries such as quality, e-health, smart city, connected car, industrial manufacturing, etc. [1]

In addition to applications that focus on responding to network services, especially virtual and augmented reality, 4K video transmission, etc. 5G networks will support our communication needs to make our lives more affordable, safer, and more convenient.

**1.2. 5G Network architecture**

All changes made by the mobile generations have been based on a new radio link concept and have provided an increase in data rates. The rate of increase and decrease of latency are the main requirements of the 5G network system. In addition, integration of new services and application areas such as IoT, Vehicle-to-Vehicle communications, …. is also important. Therefore, in addition to serving human needs, 5G must cater to the requirements of different machine-style communications [1]. For example, the data rates will range from very low for sensor to very high for high-quality video.



***Figure 1.1. 5G Architecture based on area distribution and connection types* [6]**

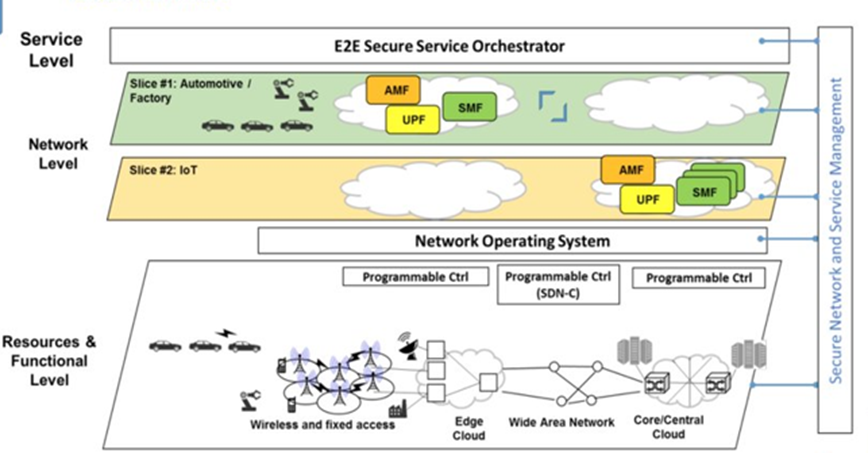
The fluctuation of system latency depends on the package size of the application. 5G is expected to be a radio multi-access technology system which integrates effectively the following basic building blocks:

* Enhanced Mobile Broadband (EMBB) will provide higher data rates and lower latency communication that improve the quality of the experience (QoE) for users [1].
* Massive Machine Communications (MMC) will provide connectivity solutions, support expansion for a huge number of network devices, which can improve connection for mobile and wireless communication systems in the future [1].
* Vehicles-to-vehicles (V2V) equipment, infrastructure, and driver assistance services require the interaction of vehicles with each other and their environment to improve traffic safety and future traffic efficiency [1]. V2V services for cellular networks demand highly reliable communication links and allow the transmission of data packets with maximum latency even when the vehicle is at high speed.
* Ultra–reliable Communication (URC) will provide scalable and cost-effective solutions for networks with high requirements of reliability and availability. When the number of users increases, by utilizing a supportive system architecture, the highly reliable service reduces speed and increases latency mechanisms to deploy media [1].
* MBB, voice service, and SMS. Because of the increase in the number of customer requests, mobile network management will significantly improve automation for management to boost the entire operation of the network cells [1].

The existence of automated communication device applications will generate traffic to the mobile network and support 5G networks performance metrics (KPIs).

Therefore, the division of network layers (network levels) will meet the needs of industries, which require separate telecommunications services, by dividing them into slices by modeled operators. The network system is divided into 3 layers: Service layer, Network layer, and Resource & Functional layer [1]. Each layer is divided into separate slices shown in Figure 1.2:

* Service Layer: Automation & V2X, IoT
* Network Layer: Network Operating System
* Resource & Functional Layer: Wide area network, wireless network, core network on Cloud



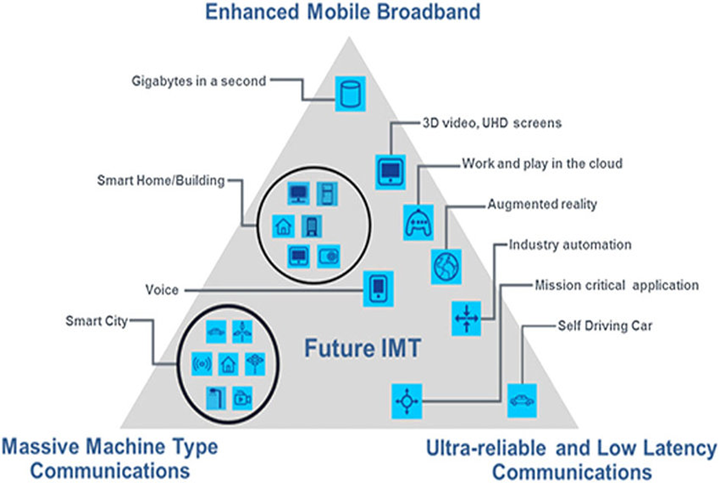
***Figure 1.2. 5G Overall Architecture*. [11]**

**1.3, New approaches and technical requirements**

The current trend is that 5G networks aim to connect almost anything, gradually satisfy human needs for new layers of service and higher levels of reliability and latency. Developing and exploiting 5G networks requires not only brighter ways of logical thinking but also a different type of network knowledge. The demand on existing and new services differ in many respects. As a result, we cannot have high reliability at a low cost, so 5G needs to be scaled to the level of performance suitable for one service while reducing costs for another [1].

**1.3.1, New approaches**

ITU-Radiocommunication, one of three units of the International Telecommunication Union (ITU) has summarized three approaches to use, showing the difference between the properties of development aspects in 5G in Figure 1.3 [2].



***Figure 1.3. Three approaches for future IMT.* [12]**

* EMBB: allows access to multimedia content, services or data. Enhanced mobile broadband demand will continue to increase, results in the growth of mobile broadband. Modified version of mobile broadband improves user performance and experience for new fields of application beyond the capabilities of existing mobile broadband applications [2].
* Ultra-reliable and Low Latency Communications: This connection has strict requests for network and device capabilities such as throughput, latency, and availability [2]. Some examples include industrial manufactory or medical surgery, intelligent grid distribution automation, traffic safety, etc.
* Massive machine-type communications: This connection is used data transmissions are not sensitive to delay. Devices are proposed to supply a low-cost and long-term service [2].

**1.3.2, Technical Requirements**

Several requirements are indicated if the current network is utilized to handle the explosive growth of the mobile Internet and IoT:

* Energy efficiency, total cost per bit, and complexity of network deployment & maintenance cannot handle traffic growth [2].
* Existing networks must monitor network resources accurately and perception of services effectively [2].
* Control the complexity when many access technology devices significantly affect the user's QoE [2].
* The widely distributed frequency spectrum will cause interference and complexity, so it must be handled with measures [2].

To do this, 5G requires the following capabilities to achieve sustainability:

*A, In terms of network construction and deployment, 5G networks need to:*

* Provide more capacity and better coverage and reduce network deployment complexity and costs [2].
* Build a flexible and scalable structure to fulfill the needs of users and services [2].
* Be flexible and efficient to various spectrum resources, including both low-frequency and high-frequency range [2].
* Connect more powerful devices to handle the access requests of a large number of IoT devices [2].

*B, In terms of operations and maintenance, 5G networks need to:*

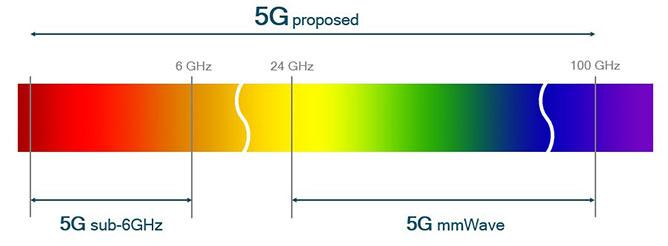
* Improve energy efficiency and cost per bit to deal with the rapid growth of data traffic and diversified needs of services [2].
* Reduce complexity due to the coexistence of multiple radio access technology, and introduce new features and functions to improve the user experience [2].
* Offer a wide range of security solutions to meet the needs of all types of devices and services of the mobile Internet and IoT [2].
* Spectral efficiency, energy consumption and cost are the three primary factors that must be addressed in sustainable mobile communications networks. To achieve sustainability, 5G needs to significantly improve the following aspects: Spectral efficiency, energy efficiency, cost performance. Compared with 4G, 5G is 3-5 times more advanced in spectrum efficiency and more than 100 times in energy efficiency and cost [2].

**1.4, New insights in 5G Development**

**1.4.1, Extremely short Wave (mmWave)**

Demand for access to more spectrum is accelerating as demand for eMBB mobile services continues to grow globally. Frequency is a decisive factor for a cellular connection as well as access to a broader spectrum. Then, the network will perform better, which means faster data processing speed and enhanced user experience. 5G networks will use new higher bands than previous mobile communications networks. The frequency band of 5G is divided into Low band, Mid band and High band (mmWave) [3] with:

* Low band: from 1 GHz to 6 GHz
* Mid band: from 6 GHz to 24 GHz
* High band: above 24 GHz



***Figure 1.4. Length of spectrums for 5G*. [13]**

Up to now, high bands above 24 GHz have been used for a long time in well-designed communications for wireless and satellite infrastructure networks [3]. However, mobile networks are only deployed in frequencies below 3 GHz because, at too high frequencies, especially mmWave bands, the mobile broadband application will be ineffective due to the narrow transmission range and easy to get clogged [3].

With mmWave, multiple antenna elements in relatively small size are used with small wavelengths at high frequencies. This feature of mmWave will be applied in a 5G system when large MIMO antenna arrays are used to create highly focused beams, capable of transmitting waves with higher energy to handle congestion and loss on both uplink and downlink down [3]. These directional beams can also be reused in space.

In mmWave implementations, it is possible to capture reflected signals or non-line-of-sight signals and use them to complement line-of-sight signals to increase the channel capacity [3]. Therefore, the reflected signal can be applied to maintain link to the mobile device even when it moves completely out of view of the base station. This is one reason mmWave is so important for 5G mobile broadband development.

Rapid adaptation to rapidly changing channel conditions is essential to 5G mmWave. At mmWave frequencies, even a small variation in the environment, such as head rotation, hand movement, or a passing car, can also change the channel and impact performance [3]. In the mmWave system in 5G, redirection and switching techniques are applied to detect and quickly transfer connections both inside and outside the access points.

**1.4.2, Massive MIMO and Beamforming**

MIMO stands for multiple-input-multiple-output. The MIMO technique meets the network demand when the number of antennas on the device is increased by 100 times. With MIMO, spectral efficiency is improved through multiple transmission paths between the base station and the established terminal. MIMO has been adopted for both Wi-Fi and LTE networks [4]. A huge step forward was to upgrade to Massive MIMO in the TDD band. Massive MIMO allows us to use a large antenna system at the same time. Massive MIMO is one of the main platforms for 5G, along with larger channel bandwidth, access to new mmWave bands and slice networks. Massive MIMO is a key component to enable wider signal coverage for mmWave bands used in 5G [4]. Massive MIMO will help improve both the coverage and capacity of 5G networks.

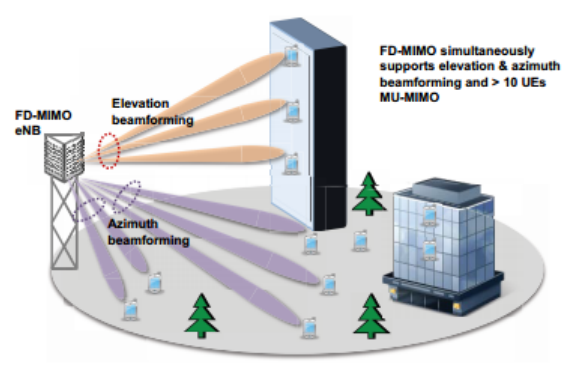
Using multiple antennas at the same time, combined with the nature of the wave, when using high frequencies such as mmWave, the emitted wave will be shaped in beamforming and will aim towards a fixed target. Beamforming uses multiple antennas to control the direction of a wavefront by balancing the magnitude and phase of the signal on individual antennas [4]. Hence, coverage is better provided for the specific areas located at the cell margins. Every single antenna in the array contributes strength to the directional signal, resulting in the array gain (beamforming gain). Beamforming can prevent interference by applying a null beam pattern in the direction of the noisy signal [4]. Adaptive beamforming is a technique that applies continuous beamforming to a moving receiver, able to keep the connection even at speeds of several hundred km/h. This is only 5G able to fulfill the requirements of a fast signal processor and powerful algorithms [4].

For beamforming-based Massive MIMO, the radio signal is demand-focused to areas with specific terminals. The spectral performance is mainly improved because beamforming in 5G is more advanced than the previous generation, resulting in greater signal accuracy and improved call quality at the cell edge [4]. Massive MIMO's difference essentially comes from the fact that the service provider can increase the network capacity without using additional transmitters or base stations. Below is a summary of the 2 biggest benefits that Massive MIMO brings:

- Improve spectral performance, increase network throughput, and upgrade network capacity. The system sends and receives multiple data signals on the same radio channel, increasing spectral efficiency per cell and the number of UEs served at the same time [4]. This increases the peak cell throughput and an improved mean efficiency

- Beamforming makes the signal stronger, reduces noise for better coverage, and improves signal quality, especially at the cell edge. Beamforming allows for extended cell range compared to traditional antennas [4]. This is especially true for higher frequencies, where the beam compensates for higher path losses.

5G is the two largest platforms that use Massive MIMO. In markets with 5G technology with huge demand for data usage, and operators needing more capacity, Massive MIMO is one of the most efficient ways to do this. The main deployment areas are dense urban traffic hotspots and by applying Full dimension MIMO (Figure 1.5) where the beam is better access to users in tall buildings [4]. In addition, upgrading networks using Massive MIMO will improve mobile operators and supply a sustainable network infrastructure.



***Figure 1.5. Massive MIMO Model with 3D Beamforming*. [14]**

**Chapter 2. 5G ULTRA-DENSE NETWORK**

In chapter 2, we will be introduced new technology in 5G Network to deal with the rapid growth of the UEs density, 5G UDN. This chapter represents the dynamic resource allocation process over 5G UDN, two categories of efficiency SE & EE to evaluate network performance. Lastly, we discuss the optimization problem described as MDP to ensure the long-term performance of the UDN system.

**2.1. Overview**

The communication requirements of smart devices and environmental mass IoT drive the density as well as the performance of 5G network infrastructure. As a result, the network needs to be very dense with many layers in the future. Ultra-dense networks (UDNs) are introduced to be very important in 5G networks development [15].

The main difference between UDN and traditional cellular networks depends on the AP density. In UDN, the AP's coverage radius is only about 10m, and there will be thousands of APs per km2. But in traditional cellular networks, the cell range is more than 500 m and there are usually less than 3-5 base stations (BS) in 1 km2 [5]. Correspondingly, terminals are connected to a UDN AP, while there are hundreds or even thousands of users residing and operating in a macro cell of the traditional network. Table below shows a comparison between UDN and traditional mobile networks [5].

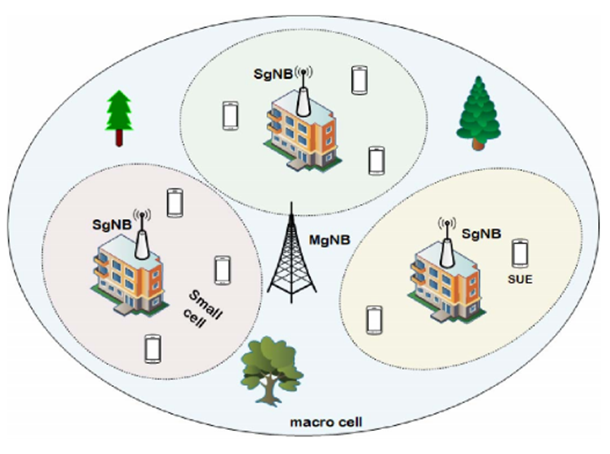


In UDN architecture, the radius of the cells reduces further and causes an increase in intercellular interference. Furthermore, due to the heterogeneity of device users over time and space, it is more difficult to manage the UDN's resources [5]. Therefore, the adaptive resource allocation in the UDN is worth further discussion. The resource allocation policy in the UDN affects the network in terms of form and user testing.

In general, 5G UDN is a new wireless networking solution that supplies higher throughput and better QoE. In UDN, the AP density can be even higher than the user density [5]. Different types of APs will work closely to achieve higher spectral efficiency, lower power consumption, and more seamless mobility.

**2.2, 5G UDN System model**

We consider downlink 5G UDN model shown in Fig. 2.1, where one macro cell contains *N* small cells.



***Figure 2.1. 5G UDN Model.* [17]**

In a macro cell, there is a deployment of a macro gNB (MgNB). A small gNB (SgNB) is deployed in each small cell. The set of SgNB as N = {1,2, …., *N*}. There are M available RBs with the set of M = {1,2, …, *M*} and the bandwidth of each RB is *Bm* [18].

The function of MgNB is to collect information and determine which RBs are available. Each SgNB selects one RB from the set of RBs and assigns a small UE [18].

The user arrival and departure process can be described as two independent processes in every network. According to the Poisson formula, UEs arrive to each small cell with parameter *λt* in each time slot [18]. Hence, we have the probability of the arrival of *x* new SUE to small cell during the period τ:

1. *P(x) =*  [18]

In similar way, SUE depart from small cell with parameter *μt*. Thus, we have the probability of the departure of *y* SUE from small cell during the period τ:

*(2)* *P(y) =*  [18]

It is assumed that the completion of SUE associated with small cells occurs just before the resource allocation [18]. Then, the set of UE linked with SgNB *n* in time slot *t* can be expressed as:

*(3)* *Un(t)* = {1, 2, 3, …*sn(t)…., Sn(t)*} [18]

In this thesis, it is allowed that the same RB can be simultaneously reused by a number of SgNBs. Thus, SgNBs of other small cells will interfere with each SUE of the small cell [18]. After the allocation of the RB *m* to the *sn(t)*, its SINR in small cell *n* can be given by:

*(4)* sn(t) = [18]

* : channel gains between SgNB *n* and SUE *sn* while RB *m* is reused.
* Pn: power to each SUE assigned by SgNB n; σ2: Noise.
* : a binary variable representing the RB allocation indicator and it receives value ‘1’ when the RB *m* is allocated to the small cell *n*, and value ‘0’ otherwise [18].

To sum up, the downlink Sum throughput in the *n*-th small cell can be given by according to Shannon’s formula [18]:

*(5)* Rn(t) = = [18]

In conclusion, we have the total throughput of the whole system:

*(6)* R(t) = = [18]

**2.3, Energy efficiency and Spectral efficiency**

In terms of dynamic resource allocation in 5G UDN, energy efficiency (EE) and spectral efficiency (SE) are considered. They are weighted due to the dynamic nature of the environment. As a result, the issue of resource allocation at different moments is described as a joint optimization [18].

Energy efficiency (EE) is the ratio between the sum throughput and total power consumption [18]. In the time plot t, we can calculate EE as:

*(7)* [18]

where PM: power of MgNB

PC: power consumption of circuit

Spectral efficiency (SE) is the ratio between the sum throughput and total bandwidth of the system [19]. In time plot t, we can calculate SE as:

*(8)* [19]

**2.4, Optimization problem of resource allocation in 5G UDN**

When resources are allocated in each period, only newcomers receive. If a resource block has been occupied in a small area at a time, its allocation will not be considered. Thus, the future state depends on the present state.

In order to guarantee the long-term performance of EE and SE in UDN system, the optimization problem is defined as a Markov decision process (MDP) [16]. When SE is maximized, the amount of available resources in the system increases. Whereas, system EE reduces and it contributes to boost the power used in the system. In contrast, maximizing EE causes a decrease in SE [19]. Because of the mentioned reasons, it is unable to meet the requirements only by optimizing SE or EE solely. The questions are raised about the tradeoff between EE and SE. Briefly, solving the tradeoff between SE and EE is the ultimate concern [16].

However, during peak and off-peak periods, SE and EE tend to deal with different demands in the system. According to access requests from UE, dynamic weights are given for EE and SE as the stability of the system is determined by the balance between them [19]. Once the number of users goes up, the rate of SE grows while the importance of EE drops. To conclude, we have the expression of the weight between SE and EE in the time plot *t*:

*(9)* [19]

The tradeoff between SE and EE can be calculated as:

*(10)*  = [19]

satisfy 4 conditions:

*(10a)* (t) ≥ R0, ∀t ∈ T, ∀n ∈ N, ∀sn ∈ *Un* [19]

*(10b)* ∈ {0,1}, ∀t ∈ T, ∀n ∈ N, ∀m ∈ M [19]

*(10c)* = Sn(t), ∀t ∈ T, ∀n ∈ N, ∀m ∈ M [19]

*(10d)* < *Smax*, ∀t ∈ T, ∀n ∈ N, ∀sn ∈ Un [19]

Condition *(10a):* ensure QoS of SUE by requesting higher than expected throughput thresholds [19].

Condition *(10b):* at time plot *t*, the small cell has two options (‘1’ means Yes và ‘0’ means No) and it can only choose one to reuse RB *m* [19].

Condition *(10c)*: it can assign each RB to maximum of one UE in each small cell. A small cell can only serve a maximum number of UE Smax [19].

Condition *(10d)*: only a finite quantity of UE can be held in each small cell [19].

**Chapter 3. SOLUTIONS FOR POWER OPTIMIZATION PROBLEM**

In this chapter, we will discuss methods to solve the resource allocation problem known as MDP. Reinforcement Learning (RL) is introduced as a basic algorithm (Q-Learning) to solve this problem, however, there are several disadvantages. Therefore, deep learning is adopted to improve the system performance and we propose new algorithms that can be defined as Deep Q Network (DQN) and its upgraded version, Dueling DQN.

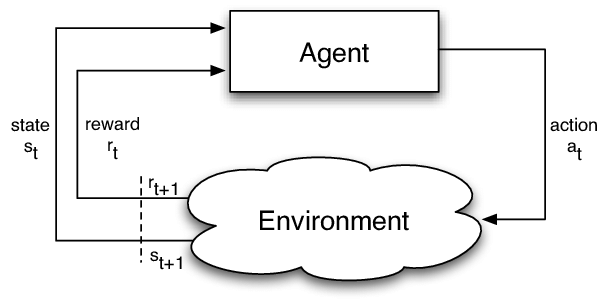
**3.1, Fundamentals of Reinforcement Learning (RL) and Q-Learning**

**3.1.1, Overview of RL**

To solve the problem of MDP in optimizing power consumption, deep Reinforcement Learning (RL) method is applied. RL is an area of Machine Learning, consisting of 7 main elements: ***Agent, Environment, State, Action, Reward, Episode, Policy*** [8]***.***

RL plays a crucial role in finding a method to select the actions of an agent in environment and then, maximize a long-term reward. RL algorithms will find a strategy to map states of environment to actions chosen by the agent [8]. RL is a trial-and-error method, this means trying again and again and learning from each other [8]. To clarify it, we have a specific process:

* Agent take actions while the environment is the world in which agent exists and operates [7].
* Agent choose the action for the environment based on current state st
* The environmental state changes after the action at has been accepted in the observations step [7].
* The agent continues to choose the next action based on the reward and the current state of the environment is st+1 at time *t+1* [7].
* At the same time, a reward feedback rt is sent to Agent
* The selection principle is to increase the reward probability to get the optimal strategy and finally, we can calculate total reward throughout the process [19]



***Figure 3.1. Reinforcement learning model.* [22]**

As it can be seen, environment, state and reward are the three key factors in RL. In this study, we define the state space, action space, and the reward function in time slot *t* based on the framework of RL [19].

* ***State space***: agent MgNB creates the decisions. The agent should learn the state of each small cell to establish the action [19]. So, we have state of the agent in time slot *t*:

*(11)* *st = {s1(t), R1(t), …, sn(t), Rn(t), x(t)}* [19]

It is concluded that the agent will obtain the information of number and throughput of all small cells in the system [19].

* ***Action space***: The decisions are made by the agent that small cells reuse RBs and we have a set of actions:

*(12)* *at = {x1(t), x2(t), …, xn(t)}*  [19]

Due to the growth of SgNB, action space increases considerably. It is clearly a problematic issue when action space explodes [19]. The number of state spaces becomes immense because of action’s effects to a state.

* ***Reward Function***: agent MgNB will receive instant reward rt by observing state st [19]. As our goal is to optimize , reward function η (t) at time t can be calculated as:

*(13)* rt = ɳ(t) = (1 - ) + [19]

Due to the huge number of RBs and the rapid increase in the number of cells, the explosion of space occurs. We have to implement a pre-screening step before learning and it results in the limitation of the size of action space [19]. For some times t, the action can only be executed if throughput Rn(t) of the corresponding action satisfy condition (10a). The outcome of this method is the reduction of possible value of *xn(t)* to control the MgNB’s action space [19].

**3.1.2, Overview of Q-Learning**

A RL algorithm can learn the value of an action in a particular state. It does not require modeling of the environment and can handle problems with random conversions and rewards without modifications [8].

Q-Learning can learn the value function of action Q (s, a), assess of the efficiency level to act in a specific state. We'll assign a scalar value based on the benefit of taking an action. Q is called value function of the action or the value function Q [9].

*(14)* *Q(st,at) = r(st,at) + maxQ(st’,at)* [9]

* *Q(st,at)*: Q target
* *r(st,at)*: Reward of action at that state
* *maxQ(st’,at)*: Discounted maximum Q value of all possible action from the next state

In the Q-learning method, we construct a memory table *Q [s, a]* to store Q values for all possible combinations between state and action. Q table shown in Figure 3.2 contains the best moves for state and action [9].



***Figure 3.2. Memory Table of Q-Learning.* [9]**

**3.1.3, Q–Learning based resource allocation in 5G UDN**

At each time t, action *at = π(st)* is determined through π policy at current state *st* by agent MgNB. Allocating the available resource blocks will reward MgNB to SgNB [19]. In RL, the expression of expected return of the state-action value function *Qπ(st,at)* can be define as:

*(15)*  *(st,at)* = [ ] [19]

where γ is the discount factor.

On purpose of making future returns less relevant to the present, sum of [] converges [19]. The target of MDP is to find an optimal strategy which receives the most rewards. For all actions selected by strategy *π∗(st) = argmax Q (st+1, at+1),* the value function *Qπ(st, at)* will be optimized by *at = π∗(st)* and it has to satisfy *Qπ(st, a­t) < Qπ∗(st, at)* [19]*.*

Q-learning is a popular method to solve MDP and according to Behrman’s equations, *Q (st, at)* can be defined as:

*(16)* *Q(st,at) = (1-α)Q(st,at) + α* [*ɳ(t)* + ] [19]

with *rt*is the reward received when state *st*transform to *st+1* and α is learning rate

(0 < α <1) [19]

* *Q(st,at)*: current value weighted by learning rate. The closer to 1 α is, the faster Q-Value will change.
* *ɳ(t)*: achieved reward if action at is taken in state *st* (weighted by learning rate)
* : maximum reward can be obtained from state *st+1*(weighted by learning rate and discount factor)

**3.1.4, Q–Learning drawbacks**

Q-learning is a simple yet powerful algorithm to create an agent memory table. This helps the agent to take the exact action. Therefore, memory requirements and computation for Q will be significant because of the combination of states and actions [32].

Everything will quickly get out of control. It is clear that we cannot infer the Q value of new states from discovered states and it leads to two problems:

* First, the amount of memory required to store and update the increase in the number of states [32]
* Second, the amount of time it takes to create the required Q-table will be expanded [32]

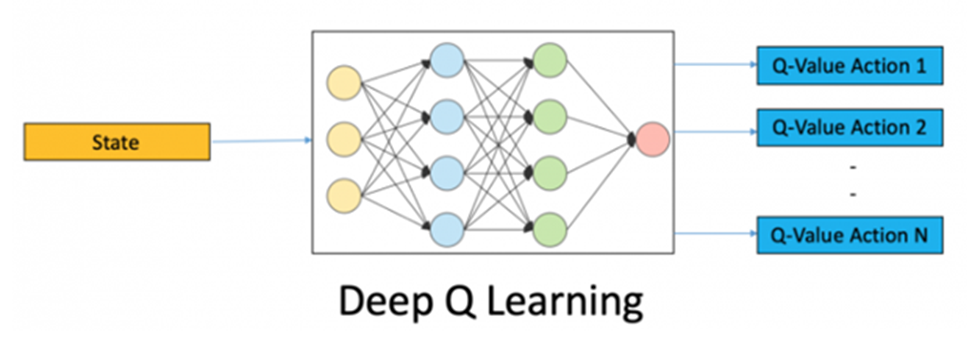
**3.2, Deep Q Network (DQN)**

**3.2.1, Overview**

To overcome those limitations, we choose the appropriate action for a certain state. In other words, we take state as input, the output is an action, and we have a new, powerful algorithm, Deep Q Network (DQN) [9]. DQN is a new deep RL strategy based on Deep Learning, combining the process of RL and a type of Neural Network (NN). A deep NN can compute the approximation of the Q value function instead of recording and storing the solutions [9]. DNN can be used to match the optimal strategy and the optimal value function:

*(17)*  *Q\*(st,at,θ) Q(st,at)* [19]

where θ is the parameter of neural network [19].



***Figure 3.3. Deep Q Learning model*. [9]**

In order to make NN learn how to estimate Q-Value for actions correctly, we have the loss function. Loss function calculates the error between the actual and predicted Q-value [9].

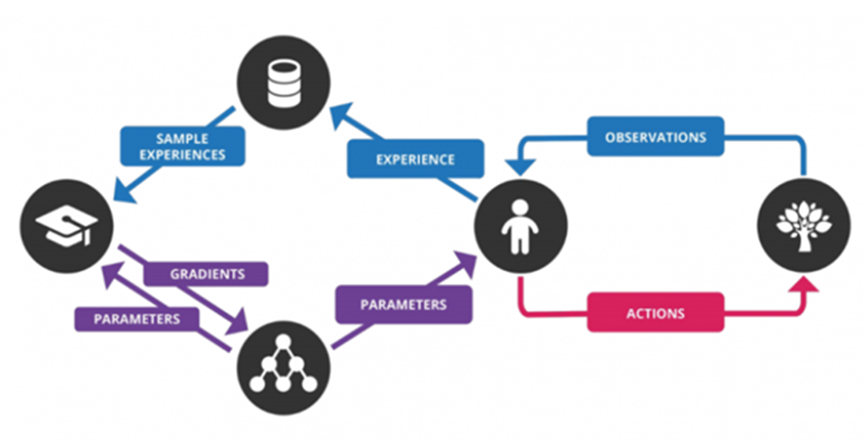
To guarantee the stability of *Q\*(st, at; θ)*, the neural network is trained to minimize the function loss L(θ), to approach the real *Q(st,at)* [19]*.* ***L(θ)*** can be expressed as:

*(18)*  *L(θ)* = E [(ɳ(t) + ] [19]

with θ-: target network parameter

θ: behavior network parameter [19]

**3.2.2, Model structure**



***Figure 3.4. Deep Q Network Structure.* [9]**

The steps related to Deep Q Network (DQN) are below:

* Preprocess and provide the state st, which will return the Q value of all possible action at in the state
* Choose an action using the ɛ-greedy policy. With ɛ probability, we choose a random action at and an action with the largest Q value, such as at = argmax

*Q (s, a, θ)* [23]

* Perform this action in one state to receive the reward. The next action will be transited in this state. These transitions are stored in the format replay buffer

<s, a, r, s’>

* Next, sample some random conversion batches from the replay buffer and calculate the loss
* Perform gradient descent step to our actual network parameters to minimize this loss
* After each iteration of C, copy the actual weight to the target network number
* Repeat these steps for M number of episodes

**3.3, Dueling DQN**

**3.3.1, Overview**

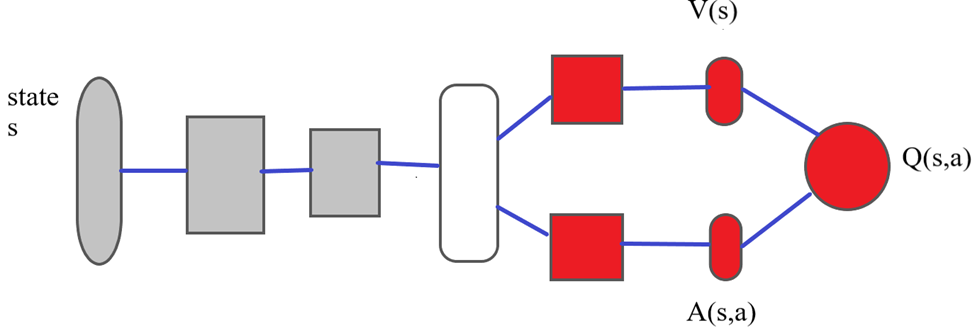
Optimizing network capacity has to face with the menace of status and action space explosion of traditional Q-learning and the instability of conventional DQN. We use Dueling DQN to solve MDP. The dueling DQN is an advanced algorithm developed on DQN structure to achieve better results [24]. Once there, the burst of state and action space is resolved and the stability of the system performance is guaranteed [23]. So, we can decompose Q (s, a) into the sum of:

*(20)* *Q(s,a) = A(s,a) + V(s)* [23]

* ***V(s):*** The value of state
* ***A(s,a):*** The advantage of action in that state (how better it does to this action than any other possible actions in that state).
* ***Q Value***: represents the value of choosing a particular action in a certain state.

**3.2.2, Model structure**

To calculate Q-Value, we use a formula with V function of state and advantage function of dependent action at [24]. Like the standard DQN architecture, we have composite classes to handle the frames. From there, we divide the network into two separate streams, one for estimating state values, and the other to estimate state-dependent action advantage. After two streams, the last module of the network combines advantage and state value outputs [23]. The process is illustrated in Fig 3.5.



***Figure 3.5. Dueling DQN model diagram***

In some states, users receive different value function sizes of RBs and identical value functions are caused by many allocation policies. Dueling DQN is upgraded to perform the function in detail. Therefore, this model has a better performance [19]. Especially, a state-action function is separated into 2 streams with a function based on states and advantages, respectively:

(21) *Q\*(st,a­t;θ,µ,ω) = V(st;θ,µ) + A(st,at;θ,ω)*  [19]

where *μ, ω* and *θ* represents the parameters of state value streams, action advantages streams and remaining parts of model, respectively.

However, two problems with this strategy are indicated:

* It is problematic to assume that and make reasonable estimates of state values and action advantages, respectively. Therefore, intentionally adding these two values can cause problems [23].
* The simple sum of both is unknown, where for a value of Q, we cannot restore ***V*** or ***A*** solely. This lack of recognition leads to poor practical performance [23].

Therefore, the last module of the neural network performing the forward mapping is shown below:

(22) *Q\*(st,a­t;θ,µ,ω) = V(st;θ,µ) + A(st,at;θ,ω) – A(st,at‘;θ,ω)* [24]

In practice, however, the advantage stream is usually equal to the value of the advantage function of the individual actions minus the mean of all the advantage functions in a state [20]. Therefore, an alternative module will replace the max operator with the mean, the real advantage function is shown as:

*(23) A(st,at;θ,ω) = A(st,at;θ,ω) -*  [24]

with  is average advantage value.

As a result, we have a function to calculate ***Q-Value***:

*(24) Q\*(st,a­t;θ,µ,ω) = V(st;θ,µ) + A(st,at;θ,ω) -* [24]

Not only can this operation ensure that the function of each action in this state remains unchanged, but also reduce the Q-value range and eliminate the degrees of redundancy to improve stability [19].

Because of sharing similar input-output interface as DQN, the training process of dueling is exactly the same. Hence, the loss function is expressed as:

*(25)* L(θ) = [23]

where *Q(st,at) = r(st,at) + maxQ(st’,at)* and performed a gradient descent step to update model parameters.

**CHAPTER 4. SIMULATION AND RESULT**

In this chapter, we will go into details with programming, simulating the technique of deep Q Network and dueling DQN by the application of deep learning. Then, we can calculate, analyze simulation results, evaluate their effect on the system and compare the efficiency of the 2 methods in optimization.

**4.1. Constructing the program of simulation**

**4.1.1. Python libraries for ML used in simulation**

Machine learning (ML) is a technology that evolved from the field of artificial intelligence. ML algorithms are computer programs that learn about how to complete tasks and how to improve performance over time. ML still requires human judgment in understanding the underlying data and choosing the right techniques to analyze the data. At the same time, before use, data must be clean, free of false positives and fake data. ML models require a sufficient amount of data to "train" and evaluate the model. In the past, ML algorithms lacked access to the large amount of essential data. Growth in big data contributes greatly to improve model and prediction accuracy [30].

Python has quickly gained popularity in the area of software development. With wealthy and powerful open-source readability and backend, developers find it easy to implement. One of the most well-known application of Python is Machine Learning. To study more about ML, it is indispensable to contribute from the generated libraries to support data analysis [10].

The wide range of use, familiarization, and understanding make it a perfect tool to handle advanced applications including complex algorithms and tools. That's why Python is an optimum method for Machine Learning and artificial intelligence solutions [10]. Thus, we will take a closer look at open-source libraries of Python for Machine Learning used in this simulation program below:

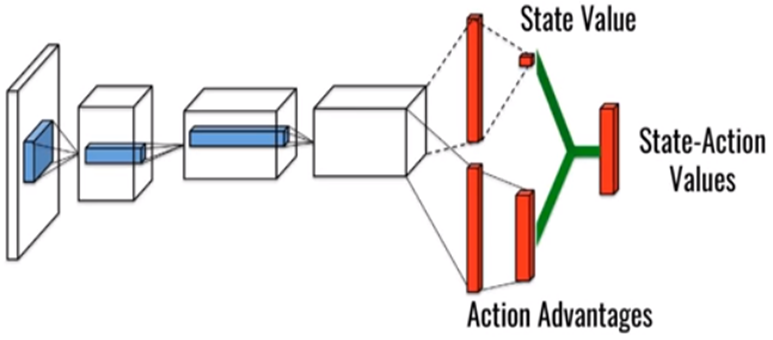
* ***Numpy***: The most basic package, when the scientific computation stack is built up is NumPy (short for Numerical Python), which provides a lot of features that are useful for operations on n-arrays & matrics in Python. This library provides the ability to vectorize math operations in the NumPy type array, improving performance and consequently execution speed [10].
* ***Cvxopt***: is a Python-based software package for convex optimization. It is built on Python’s extensive standard library and widely used to develop software for convex optimization applications straightforward [31].
* ***Matplotlib***: A core package of SciPy Stack and another Python library built specifically for the generation of simple, powerful visualizations. You can create any visualization: Line graphs; Scatter plots; Bar charts and Histograms; Pie charts, etc. However, the library is low-level, meaning you will need to write more code to reach high levels of visualization [10].
* ***Seaborn***: mostly focused on visualization of statistical models; such visualizations include heat maps that aggregate data but still describe the overall extent of the dispersion. Seaborn is based on Matplotlib [10].
* ***Tensorflow***: is an open-source library of data-flow graphs computations suitable for Machine Learning. An important feature of TensorFlow is the multi-layer node system, allowing the training of neural networks on large datasets quickly [10].
* ***Tflearn***: is a deep learning library module built on top of Tensorflow. It is designed to supply a higher-level API to TensorFlow to facilitate and speed up tests [28]. In comparison with straight Tensorflow, TfLearn method seems a little cleaner. One disadvantage of Tflearn is the lack of pre-trained models which are easy to integrate.
* ***Keras***: is an open-source library written in Python used to build Neural Networks at a high level of interface. This library is simple and highly extensible. Keras uses backends and has a simple design approach aimed at quick and easy experimentation from building compact systems [28]. The general idea of Keras is based on layers and everything else is built around these layers as well. Data is prepared in tensors, the first layer is responsible for the input of tensors, the last layer is responsible for the output, and the model is built in middle [10].

**4.1.2. Dynamic resource allocation model of DQN and Dueling DQN**

To deploy and evaluate the efficiency of power optimizing based on deep reinforcement learning, we establish simulation by 2 dynamic resource allocation algorithms based on DQN and Dueling DQN:

***Input:*** Q(st,at), set of available RBs M = {1,2,…,M}

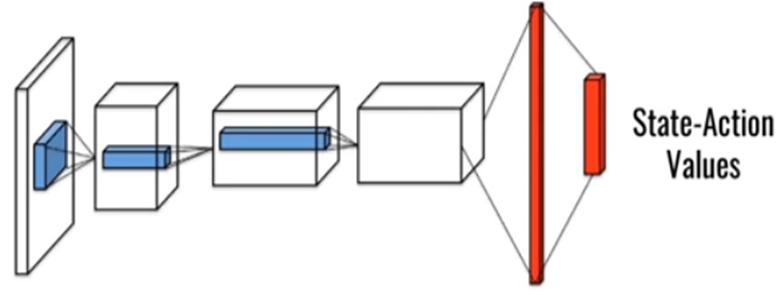
***Output:*** the tradeoff between EE and SE



***Figure 4.1. The illustrated diagram of Dueling DQN.* [25]**

In term of Dueling DQN shown in Figure 4.1, to begin with, we initialize state-action values network *Q(st,at,θ)* with target network *Q\*(st’,at’,θ-)* and the replay buffer **D** with the capacity of N. In each episode, we initialize the 5G UDN system environment, and MgNB receives the initial state s1 [20].

Started with t = 1, MgNB chooses an action at at the state st from at = max *Q\*(st,at,θ).* Then, MgNB executes action at to allocate selected RBs to SUE and calculates the immediate reward rt. The next step is at st+1, MgNB receives the system state and stores experience *{st, at, rt, st+1}* in replay buffer **D**. When the capacity of **D** reaches **N**, means the flow is divided into 2 streams in fully connected layers. In output layers, MgNB randomly selects a batch of samples *{sj, aj, rj, sj+1}* from **D** and calculates state function *V(st, θ,µ)* and action advantage *A(st, at, θ, ω).* Finally, based on (), we combine them as *Q\*(st, at, θ, µ, ω)* and complete target network parameter [20].



***Figure 4.2. The illustrated diagram of DQN.* [25]**

In comparison with Dueling DQN, as it is shown in Figure 4.2, the traditional DQN algorithm is relatively simpler. Started with initialization step, we also have the replay buffer **D** with the capacity of N and state-action values as well as the target network. In each episode, MgNB of DQN has almost the same executing sequence as its counterpart. However, the difference occurs when the capacity of replay buffer reaches N, the flow still remains 1 stream. In output layers, MgNB selects randomly a batch of samples *{sj, aj, rj, sj+1}* from **D** and calculates directly *Q\*(st,at,θ)* [20]*.*

**4.1.3. Simulation Parameters**

|  |  |
| --- | --- |
| **Parameters** | **Value** |
| System total bandwidth | 10 MHz |
| Total number of RBs | 50 RBs |
| Small cell inter site distance | 50 m |
| Number of cells | 6,8,10,12,14,16 |
| Maximum number of UE in a small cell(*Smax*) | 30 |
| Learning rate alpha | 0.1 |
| Starting speed R0 of SgNB | 10 Mbps |
| Time slot *t* | 15 ms |
| Thermal noise power | -174dBm/Hz |
| Shadow fading variance of SgNB | 10 dB |
| Power of MgNB | 46 dBm |
| Power of SgNB | 30 dBm |
| Power of circuit transmissions | 6.8 W |

In this table, a resource block (RB) is the smallest resource unit allocated to a user. One RB commonly contains 12 carrier waves, each carrier wave has a frequency band of 15 kHz, therefore, the bandwidth of 10 MHz used in this simulation is divided into 50 RB or 601 carrier waves at downlink according to 3GPP standards [29]. The possible number of small cells can be seen in the table and the distance between them is 50m. In each cell, there are no more than 30 UEs allowed [21].

In terms of system operation, the learning rate is 0.1 and the algorithm performs in 15s in 1000 time slots, therefore, we have a time slot t of 15ms. We also set up the starting point of UE demand R0 at 10 Mbps [21].

To calculate SINR, two essential parameters are thermal noise power and shadow fading. Thermal noise is effective and extensive white noise over a very wide spectrum. Thermal noise occurs due to the vibration of the carriers within the conductor and is proportional to the temperature, independent of the applied voltage [26]. The thermal noise power *P* is proportional to the bandwidth:

(26) *P = k.T.B* [26]

k: Boltzmann’s constant, equivalent to 1.38 x 10-23m2kg s-2K-1

T = temperature in degrees Kelvin

B = bandwidth in Hz

This figure is then normally expressed in terms of dBm. Thermal noise in a 50 Ω system at room temperature is -174 dBm/Hz [26].

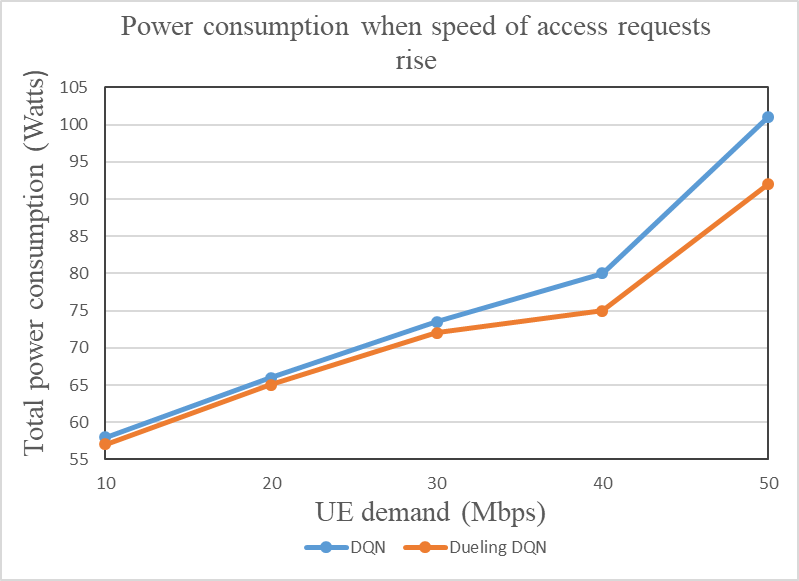
In terms of shadow fading, in wireless systems, fading can be caused by multipath propagation, known as multipath dimming, weather (especially rain) or obscuration from interfering obstacles. Shadow fading is also called slow fading because the fade time can last for seconds or minutes [27]. Moreover, the shadow fading is a large-scale discoloration and expressed with variance unit of dB. In this simulation, the shadow fading variance of SgNB is 10 dB [21].

With total bandwidth and SINR, we can have the outcome of total throughput and spectral efficiency. To evaluate energy efficiency, three different categories of power are required, namely power of MgNB PM (dBm), power of SgNB PN (dBm), and power consumption of all circuit transmissions PC(W) while PM = 46 dBm, PN = 30 dBm according to [21].

**4.2. Simulation Results**

In this simulation, we will investigate 2 scenarios of network access: Total power consumption when UE demands grow and Average power consumption in varying time slots.

After the simulation is completed, the results are illustrated in 2 different line graphs shown in Figures 4.3 and 4.4. From these outcomes, we proceed to evaluate and compare optimum efficiency of power between two Deep RL algorithms, namely DQN and dueling DQN. The outcome of Fig 4.3 will indicate which method has better available resource utility as it is more power-saving. Whereas, the result of Fig 4.4 will assess the stability of two Deep Learning algorithms and show the method with a more effective tradeoff between EE and SE.



***Figure 4.3. Total power consumption when UE demand rise***

From the line graph of Fig 4.3, it can be seen that when the speed of access (Mbps) increases, power consumption goes in a similar trend. That means the rise of user demand causes an increase in total system throughput. As a result, SE is maximized whereas EE is reduced and it is equivalent to the maximization of resources.

To begin with 10 Mbps, the number of Watts both methods have consumed is approximately 55 W. Then, when UE demand reaches 20 Mbps, there is an insignificant difference between DQN and Dueling DQN as they obtain just over 65 W. With 10 Mbps more, we can hardly see any differences as the figures of 2 algorithms stay in the range from 70 to 75 W. When access requests reach 40 Mbps, there is a noticeable change as there is a slight rise in the quantity of power consumption for Dueling DQN, meanwhile, DQN goes up rapidly to about 80 W. Nonetheless, both methods experience a dramatic increase when the requirement climb up to 50 Mbps. As it is indicated in the graph, Dueling DQN has better performance because its power only obtained just over 90 Mbps, nearly 10 Mbps lesser than its counterpart. Therefore, we can conclude that resource allocation based on Dueling DQN needs less power and maximization of available resources to operate.

***Figure 4.4. Average power consumption in time varying user demand scenario***

The line graph of Fig 4.4 shows that in each time slot, both lines of DQN and Dueling DQN have noticeable changes. In this chart, the average power consumption goes in different ways during 4-time slots and the value of each time slot *t* is 15 ms (shown in parameters table). In this situation, we consider the balance between the maximization of EE and SE to investigate the stability of the system.

Firstly, the average consumed power of both algorithms begins at nearly the same point at over 77 W and Dueling DQN is slightly higher. However, they continue in opposite trends. The line of Dueling DQN experiences a gradual decrease and at the third time slot, it hit the bottom at just under 76 W. Then, it grows a little and finishes at approximately 76.3 W. In contrast, the quantity of average power consumption of DQN fluctuates throughout the process. It increases a bit in the first time slot before drops slightly to about 77.5 W. At the third time slot, there is a dominance of DQN as its line goes up dramatically and reaches a peak at nearly 81 W. However, it follows a downward trend and the system ends with 79 W. Therefore, when time varies, the average power consumption of Dueling DQN is well lower and more stable in comparison with the other. Briefly, resource allocation-based Dueling DQN has a better tradeoff between SE and EE to improve the stability of the system.

In conclusion, Dueling DQN is the algorithm that has better trade between the EE and SE and saves more available resources. To sum up, this method can be proposed to deal with power-saving issues when user demand of access speed and time climb up.

**CONCLUSION AND FUTURE ORIENTATION**

In this thesis, we investigated the dynamic resource allocation problem over 5G UDN. Both EE and SE are considered in the network. Thus, the resource allocation issue at different moments can be described as a power optimization problem, which is known as MDP. In order to ensure the long-term performance of the UDN system, the reinforcement learning method is used with the traditional Q-Learning method. However, this algorithm still has several drawbacks. Therefore, deep learning is applied to improve the performance of the system. The Deep Q Network (DQN) and its upgraded version, Dueling DQN were proposed.

This thesis has mentioned the 5G UDN system model, and simulated and made a comparison of the efficiency between two deep learning algorithms, DQN and Dueling DQN. It is shown from the simulation results that the Dueling DQN has a better performance compared with its counterpart as it is more power-saving.

In future research, there are several recommendations:

* Research and evaluate the efficiency of another algorithm to optimize power consumption, Double DQN.
* Calculate and compare the average power consumption in time varying user demand scenario.
* Consider and research power allocation and multi-resource joint configuration in 5G UDN

**APPENDIX**



**PREFERENCES**

**Vietnamese documents**

[1] Đinh Việt Anh, *Nâng cao hiệu quả sử dụng tài nguyên mạng mật độ siêu cao trong hệ thống 5G thông qua tối ưu hóa bản tin Paging*, Luận văn Thạc sĩ, Trường Đại học Công Nghệ, Đại học Quốc gia Hà Nội, 2018, tr.3-6.

[2] Đinh Việt Anh, *Nâng cao hiệu quả sử dụng tài nguyên mạng mật độ siêu cao trong hệ thống 5G thông qua tối ưu hóa bản tin Paging*, Luận văn Thạc sĩ, Trường Đại học Công Nghệ, Đại học Quốc gia Hà Nội, 2018, tr.7-9.

[3] Đinh Việt Anh, *Nâng cao hiệu quả sử dụng tài nguyên mạng mật độ siêu cao trong hệ thống 5G thông qua tối ưu hóa bản tin Paging*, Luận văn Thạc sĩ, Trường Đại học Công Nghệ, Đại học Quốc gia Hà Nội, 2018, tr.17-18.

[4] Đinh Việt Anh, *Nâng cao hiệu quả sử dụng tài nguyên mạng mật độ siêu cao trong hệ thống 5G thông qua tối ưu hóa bản tin Paging*, Luận văn Thạc sĩ, Trường Đại học Công Nghệ, Đại học Quốc gia Hà Nội, 2018, tr.18-20.

[5] Đinh Việt Anh, *Nâng cao hiệu quả sử dụng tài nguyên mạng mật độ siêu cao trong hệ thống 5G thông qua tối ưu hóa bản tin Paging*, Luận văn Thạc sĩ, Trường Đại học Công Nghệ, Đại học Quốc gia Hà Nội, 2018, tr.23.

[6] <https://cellphones.com.vn/sforum/giai-ngo-ve-5g-dung-de-cac-thuat-ngu-danh-lua-goi-la-5g-nhung-khong-han-da-la-5g-thuc-su?utm_source=sforum&utm_medium=thi-truong>

[7] <https://viblo.asia/p/gioi-thieu-ve-reinforcement-learning-rl-djeZ1GEY5Wz>

[8] <https://forum.machinelearningcoban.com/t/tutorial-implement-cac-thuat-toan-reinforcement-learning-ddpg-bai-1/3276>

[9] <https://viblo.asia/p/gioi-thieu-ve-hoc-tang-cuong-va-ung-dung-deep-q-learning-choi-game-cartpole-Az45bYy6lxY>

[10] <https://topdev.vn/blog/top-15-thu-vien-python-tot-nhat-cho-data-science-trong-2017/>

**Foreign documents**

[11] F. Barletta, F. Massimiani, F. Pierangeli, M. Spada, A. Marotta, F. Graziosi, *5G Networks. An Overall Perspective for the Innovation Management*, 2018 IEEE 4th International Forum on Research and Technology for Society and Industry (RTSI), 2018, Figure 1.

[12] Francois Rancy, *IMT for 2020 and Beyond*, Radiocommunication Bureau, International Telecommunication Union, 2019, Figure 6.3.

[13] <https://techblog.comsoc.org/2019/05/22/tech-mahindra-india-needs-to-begin-5g-spectrum-auction-now/>

[14] Félix Pablo CANO PAÍNO, *Implementation and evaluation of FD-MIMO beamforming schemes for highway scenarios*, Diploma Thesis, Vienna University of Technology, 2017, Figure 1.2, pp.3.

[15] M. Kamel, W. Hamouda, A. Youssef, *Ultra-Dense Networks: A Survey*, IEEE Communications Surveys & Tutorials, 2016, vol. 18, no. 4, pp.2522-2545.

[16] C. C. Coskun, E. Ayanoglu, *Energy-spectral efficiency tradeoff for heterogeneous networks with QoS constraints*, Proc. IEEE International Conference on Communications(ICC), 2017, pp. 1-7.

[17] Zhiyong Liu, Xin Chen, Ying Chen, Zhuo Li, *Deep Reinforcement Learning Based Dynamic Resource Allocation in 5G Ultra-Dense Networks*, 2019 IEEE International Conference on Smart Internet of Things (SmartIoT), 2019, Fig 1.1, pp.169.

[18] Zhiyong Liu, Xin Chen, Ying Chen, Zhuo Li, *Deep Reinforcement Learning Based Dynamic Resource Allocation in 5G Ultra-Dense Networks*, 2019 IEEE International Conference on Smart Internet of Things (SmartIoT), 2019, pp.168-169.

[19] Zhiyong Liu, Xin Chen, Ying Chen, Zhuo Li, *Deep Reinforcement Learning Based Dynamic Resource Allocation in 5G Ultra-Dense Networks*, 2019 IEEE International Conference on Smart Internet of Things (SmartIoT), 2019, pp.170-172.

[20] Zhiyong Liu, Xin Chen, Ying Chen, Zhuo Li, *Deep Reinforcement Learning Based Dynamic Resource Allocation in 5G Ultra-Dense Networks*, 2019 IEEE International Conference on Smart Internet of Things (SmartIoT), 2019, algorithm 1, pp.171.

[21] Zhiyong Liu, Xin Chen, Ying Chen, Zhuo Li, *Deep Reinforcement Learning Based Dynamic Resource Allocation in 5G Ultra-Dense Networks*, 2019 IEEE International Conference on Smart Internet of Things (SmartIoT), 2019, Table 1, pp.172.

[22] <https://www.researchgate.net/figure/Agent-Environment-interaction-cycle-in-RL_fig2_221309044>

[23] Chris Yoon. *Dueling Deep Q Networks*, towardsdatascience.com. 2019, Architecture

[24] Ziyu Wang, Tom Schaul, Matteo Hessel, Hado van Hasselt, Marc Lanctot, Nando de Freitas, *Dueling Network Architectures for Deep Reinforcement Learning*, 2016, Chapter 3.

[25] Ziyu Wang, Tom Schaul, Matteo Hessel, Hado van Hasselt, Marc Lanctot, Nando de Freitas, Dueling Network Architectures for Deep Reinforcement Learning, 2016, Figure 1.

[26] <https://www.electronics-notes.com/articles/basic_concepts/electronic-rf-noise/thermal-noise-calculations-calculator-formulas.php>

[27] <https://en.wikipedia.org/wiki/Fading>

[28] <https://datascience.stackexchange.com/questions/25317/what-are-the-pros-and-cons-of-keras-and-tflearn>

[29] <https://www.researchgate.net/figure/Bandwidth-scalability-in-3GPP-LTE-LTE-A_tbl1_272413342>

[30] <https://www.ibm.com/cloud/learn/machine-learning>

[31] <https://cvxopt.org/>

[32] James Montantes. *What You Need to Know About Deep Reinforcement Learning*, towardsdatascience.com. 2020, The challenge of Model-Free RL.