Machine Learning Technique for Uplink Link Adaptation in 5G NR RAN at Millimeter Wave Frequencies

Master's Thesis

By

Hazem Elgabroun

Department of Electrical and Information Technology
Faculty of Engineering, LTH, Lund University
SE-221 00 Lund, Sweden

Aug-2019

Popular Science Summary

In 1959, Arthur Samuel defined Machine Learning (ML) as the “field of study that gives computers the ability to learn without being explicitly programmed”. Since then, computers had major improvements, and now can handle extremely large number of processes in a trivial fraction of time. This highlighted the new benefits ML can provide in different areas, especially where large raw data is handled, as it can be challenging to visualize and process all data by human individuals. 5G networks is one of these fields, where ML is expected to be present in many applications to enhance the overall performance.

Link Adaptation (LA) is one of the Random Access Network (RAN) techniques used in 5G, it determines the signals modulation order and coding scheme used in the transmission. Therefore, it has a direct impact on the throughput and robustness of the transmitted signal. In LA, the scheduler adjusts the throughput mainly according to channel conditions and the Block Error Rate (BLER). This is done in two steps called, Inner-Loop Link Adaptation (ILLA) and Outer-Loop Link Adaptation (OLLA).

In 5G New Radio (NR) networks, a new high frequency spectrum was introduced as a key feature to meet 5G standards. This high frequency spectrum -known as millimeter Wave (mmWave) spectrum- provided a new unutilized frequency bands, with more intricate channel conditions compared to sub 6 GHz frequency spectrum used in 5G predecessors. However, many new techniques were developed to enable a reliable communication in mmWave frequency environment. In this matter, ML was presented as one of the techniques that can help tolerating these difficult conditions in many applications. This thesis focused on using ML techniques to improve the performance of Uplink (UL) OLLA, while considering 5G standards at mmWave frequency environment.

Reinforcement Learning (RL) is one of the popular ML techniques in dynamic environments. Therefore, RL was considered a promising approach for this case, as it allows the schedular to learn directly from the feedback obtained from a previously chosen action, using mathematical algorithms in the process.

In this thesis, a simulation tool was used to demonstrate the results of implementing an RL algorithm on UL OLLA in 5G at mmWave frequency environment. As a result, the ML algorithm was able to fine-tune the scheduler’s Modulation and Coding Scheme (MCS) decisions, using the feedback of previous actions, leading to a decreased BLER while achieving higher throughput, when compared to a non-ML system under the same conditions.

Abstract

The demands on wireless communications are continuously growing, due to the fact that when higher network capabilities are delivered, new features and applications are created, calling for even higher requirements. To keep pace with these demands and to allow new applications to rise, the limits of mobile networks must be pushed regularly. Therefore, the International Telecommunication Union targets on achieving new milestones almost every decade. 5G is the fifth-generation standard for wireless cellular networks, which was planned to push the limits once more to a new level.

To achieve the standards of 5G, a new frequency spectrum of mmWave was introduced. This spectrum was unutilized in earlier generations due to its complex environment relatively to sub 6 GHz spectrum. However, since then, new techniques were introduced helped to overcome these challenges.

This thesis is investigating on the possibility of improving UL Link Adaptation using ML technique on mmWave frequency environment.

After studying previous related work and different ML techniques, a reinforcement learning algorithm was suggested. The algorithm uses the feedback of previous actions in consideration when taking future decisions. The system was implemented on a professional simulation tool provided by Ericsson. The results showed an improvement on both throughput and BLER performance when compared to a non-ML system.

Acknowledgments

With greatest pleasure, I would like to address my sincere gratitude to my supervisor at Ericsson, Irfan Biag, and Ericsson office in Lund for giving me this chance and providing a great environment with all needed support for this thesis. I would also like to thank my supervisor at Lund University, Ove Edfors, for all his support and guidance on this work. Furthermore, my sincere acknowledgement to all individuals who helped and encouraged me during my study journey, starting with all teachers and fellow colleagues and ending with family and friends.

Also special thanks to my wife Jinan for her endless support during the period of this work. Last but not least, I cannot express enough thanks to my father and mother for their encouragement and support throughout all of my study journey, and therefore, I will forever be thankful.

Table of Contents

List of Figures

List of Tables

List of Acronyms

- **3GPP** - Third Generation Partnership Project
- **5G** - Fifth generation
- **ACK** – Acknowledgement
- **AI** - Artificial Intelligence
- **AWGN** - Additive White Gaussian Noise
- **BLER** - Block Error Rate
- **BS** - Base Station
- **CA** - Carrier Aggregation
- **CF** - Correction Factor
- **CRI** - CSI-RS Resource Indicator
- **CSI** - Channel State Information
- **CSI-RS** - Channel State Information Reference Signal
- **CQI** - Channel Quality indicator
- **DC** - Dual Connectivity
- **DCI** - Downlink Control Information
- **DL** - Downlink
- **DSA** - Dynamic Spectrum Access
- **EB** - Exabytes
- **EE** - Energy Efficiency
- **eMBB** - Enhanced Mobile Broadband
- **FD** - Full-Duplex
- **FR** - Frequency range
- **gNB** - Next Generation Node Base Station
- **HARQ** - Hybrid Automatic Repeat Request
- **ILLA** - Inner-Loop Link Adaptation
- **IMT** – 2020 - International Mobile Telecommunication system 2020
- **IoT** - Internet of Things
- **ITU** - International Telecommunication Union
- **LA** - Link Adaptation
- **LI** - Layer Indicator
- **LOS** - Line of Sight
- **LTE** - Long Term Evolution

- **MAB** - Multi-Armed Bandits
- **MCS** - Modulation and Coding Scheme
- **MDP** - Markov Decision Process
- **mMIMO** - Massive MIMO
- **mMTC** - Massive Machine-type Communication
- **mmWave** - Millimeter Wave
- **NACK** - Negative Acknowledgement
- **NOMA** - Non-Orthogonal Multiple Access
- **NR** - New Radio
- **OLLA** - Outer-Loop Link Adaptation
- **OWC** - Optical Wireless Communication
- **PMI** - Precoding Matrix Indicator
- **QAM** - Quadrature Amplitude Modulation
- **QoS** - Quality of Service
- **QPSK** - Quadrature Phase Shift Keying
- **RAN** - Radio Access Network
- **RF** - Radio Frequency
- **RI** - Rank Indicator
- **RRM** - Radio Resource Management
- **RV** - Redundancy Version
- **SE** - Spectral Efficiency
- **SINR** - Signal to Interference and Noise Ratio
- **SIR** - Signal to Interference Ratio
- **SNR** - Signal to Noise Ratio
- **SVM** - Support-Vector Machine
- **TBS** - Transfer Block Size
- **UE** - User Equipment
- **UL** - Uplink
- **URLLC** - Ultra-Reliable and Low Latency Communications
- **UDenseNets** - Network Ultra-Densification
- **V2I** - Vehicle to Infrastructure communication
- **V2V** - Vehicle to Vehicle communication

CHAPTER 1

1 Introduction

In 5G New Radio (NR) networks, the transmission techniques and protocols have strong similarity to Long Term Evolution (LTE), its predecessor. The scheduler is responsible for allocating the resources of transmission between different User Equipment (UE). One of the crucial roles of the scheduler is to regulate the Link Adaptation (LA) procedure, since it has a direct influence on the link throughput and Block Error Rate (BLER). The LA procedure has been receiving close review, as more efficient approaches might be proposed in 5G networks.

On the other hand, Machine Learning (ML) is a rising technique that is undergoing intense study, as it has shown the capability to improve the system overall performance in numerous applications and services [1].

This thesis focuses on the benefits that could be gained by applying ML techniques to improve the scheduler performance. Particularly, in the uplink LA scenario for 5G NR systems, at millimeter wave (mmWave) frequencies.

1.1 Background and Motivation

The 5G NR Radio Access network (RAN) is expected to provide very high user throughput (>1Gbps) under Enhanced Mobile Broadband (eMBB) use case [2]. Subsequently, meeting higher user throughput demand requires large frequency bandwidth in RAN. In this context, the mmWave band has been considered an enabler of the 5G NR bandwidth requirements, due to availability of wider bandwidths in the mmWave band. However, there are several challenges associated with mmWave frequencies:

- The radio propagation loss is more profound for mmWave when compared to lower frequencies as the wavelength is very small, making the signals more prone to propagation attenuation such as absorption by water vapor [3].
- Wireless channel conditions and link quality can change more drastically compared to low frequencies during slight movement of users.

All these challenges call for enhanced scheduler decisions and LA in both Uplink (UL) and Downlink (DL) to maintain reliable connectivity and Quality of Service (QoS) for users. Therefore, an ML algorithm can be proposed to adopt better decisions by schedulers for LA, leading to an increased throughput while maintaining a decent level of Block Error Rate (BLER).

1.2 Project Aims and Main Challenges

This thesis work will propose an ML model for UL LA which shall dynamically adjust the UL modulation and coding scheme. The ML algorithm shall be built using several decision parameters, targeting improvement of the Outer-Loop LA (OLLA) of the network. The substantial thesis goals will span over comparison of the relative performance and robustness of evaluated ML techniques to the traditional technique for UL LA as described below. Thus, the thesis work will involve the following:

- Investigate existing UL LA techniques, and their potential drawbacks.
- Explore previous related work and propose an ML technique for UL LA.
- Develope a simulation model to evaluate and compare prior art and the proposed ML technique, as well as simulating various possible scenarios.

1.3 Approach and Methodology

LA in current 5G NR systems depends on look-up tables to decide the suitable Modulation and Coding Scheme (MCS). These tables are built depending on several simulations, which in average results the highest performance of the transmission link. The scheduler chooses the MCS value that corresponds to given measured inputs, mainly Signal to Interference and Noise Ratio (SINR), while satisfying the constraint of keeping the BLER below a certain threshold (10%). However, use of static look-up tables leads to ignoring of significant information related to each cell environment and UE parameters, as well as being highly dependent on the estimated SINR values. Therefore, the application of an ML algorithm can enhance the performance and robustness of the scheduler LA decisions, as it can be dedicated to each user and each cell alone. Moreover, it will automatically adapt to any new changes in the environment, as it will always seek to improve the performance using the feedback from previous actions.

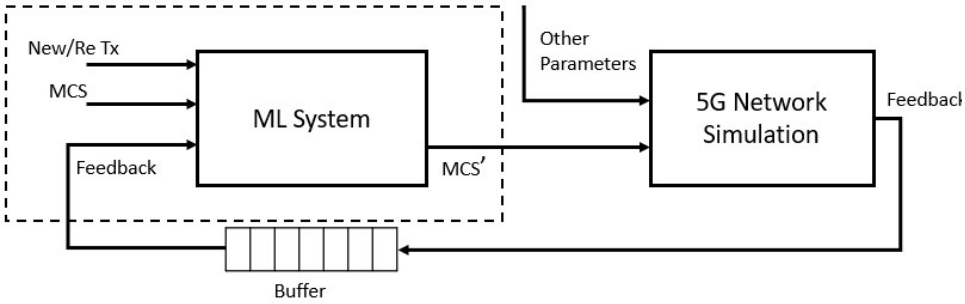


Figure 1.1 *The general block diagram of the proposed ML system for UL link adaptation.*

This thesis will evaluate existing and proposed approaches for UL LA, and further suggests an ML algorithm suitable for 5G NR systems at mmWave frequencies. This will be done using a professional simulation tool provided by Ericsson with needed adjustments and possible improvements. Figure 1.1 shows the general block diagram of the ML system proposed for UL LA.

1.4 Thesis outline

This thesis work was organized into 6 chapters. Chapter 1 is an introduction to the thesis topic. Chapter 2 reviews the general theoretical background of 5G NR networks, the mmWave environment, and LA. Chapter 3 briefly discusses the concept behind different ML techniques. Chapter 4 provides an overview of prior art, followed by a detailed description of the proposed ML algorithm. Chapter 5 evaluates the obtained results of the implemented system. Finally, Chapter 6 summarizes the main conclusions and outlines future work based on the results.

CHAPTER 2

2 Theoretical Background

The demands on wireless communications are growing rapidly. New applications are rising every day, adding new substantial services to the users, thus driving for increased requirements. As an example, Facebook is one of the applications launched only in the previous 15 years. Today it is considered an essential source for entertainment, information, advertisement, as well as many other services. Moreover, it has more than 2 billion monthly active users [4], with over 1 billion of those users considered as mobile-only users [5].

In this matter, mobile cellular networks are expected to absorb a massive amount of the wireless data expansion. In order to keep pace with these demands, the International Telecommunication Union (ITU) initiated the process of evolving towards 5G networks in 2015. New scenarios with improved performance were introduced in the International Mobile Telecommunication system 2020 (IMT-2020). According to the Ericsson's mobility report, published in June 2019, 5G networks will carry up to 35 percent of mobile data traffic globally by 2024. The expected growth on the global mobile data traffic in exabytes (EB) per month for this decade is demonstrated in Figure 2.1 [6].

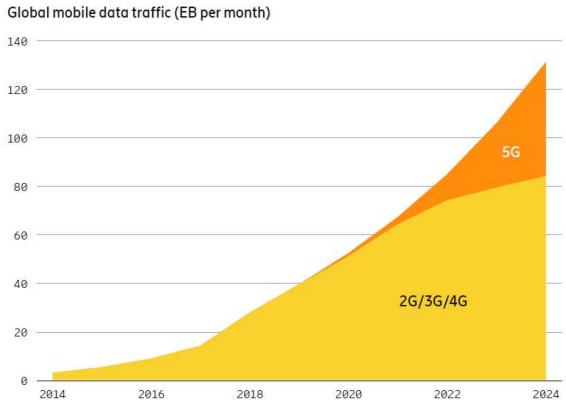


Figure 2.1: The expected growth of global mobile data traffic at the present decade. [6]

2.1 Introduction to 5G New Radio (NR)

In 2016, the Third Generation Partnership Project (3GPP) initiated the standardization process for 5G NR [7], which should fulfil the specifications of achieving 1000 times higher network spectral efficiency (SE), low cost, guaranteed QoS, mobility supporting up to 500 km/h, ultra-reliable & low- latency communication, and 100 times better energy efficiency (EE) compared to its predecessor [7]. To achieve this target, the unutilized mmWave spectrum was introduced, adding new capabilities to the network, though in reverse requiring new techniques to overcome its intricate environment. In Figure 2.2, a comparison between 4G and 5G network requirement, illustrated through a "spider web" diagram.

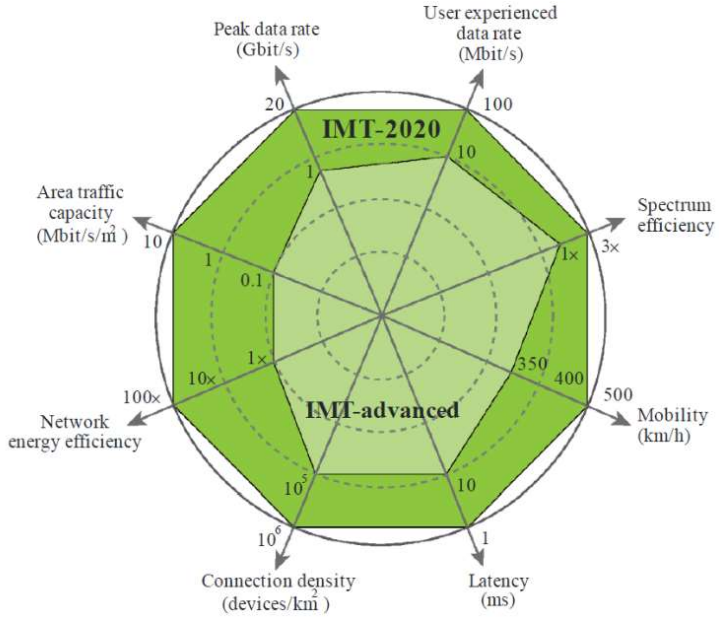


Figure 2.2 *A comparison between IMT-advanced and IMT-2020 specifications.* [2]

2.1.1 5G Mobile Networks New Concept

To meet the high standards of 5G networks, a new concept was introduced in IMT-2020. This new concept was to split the desired 5G specifications into multiple scenarios, each with different requirement and different services, as shown in Figure 2.3 [2]. The first scenario is eMBB, where the users require increased data throughput, with high mobility and network capacity.

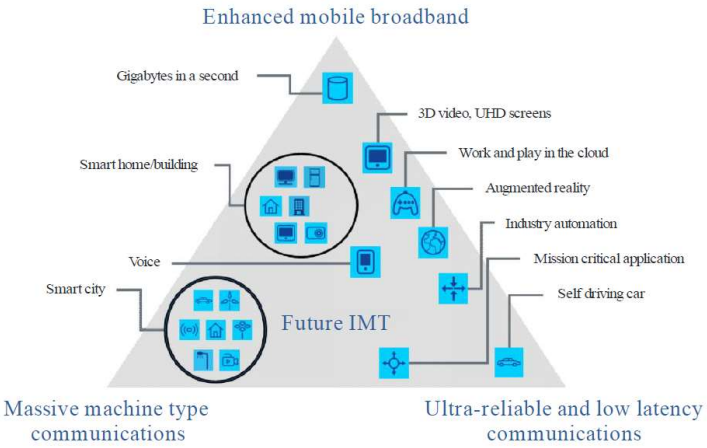


Figure 2.3 *IMT-2020 different services and use cases.* [2]

On the other hand, Massive Machine-type Communication (mMTC) is more focused on lowering power consumption, and greatly increasing the number of connected devices to the network, as this scenario is mostly dedicated to the Internet of Things (IoT) devices, that have low data traffic and will be connected to the wireless network. Finally, ultra-reliable and low latency communications (URLLC) is introduced, providing communication services such as Vehicle to Vehicle (V2V) and Vehicle to Infrastructure (V2I) communication, which requires high reliability and low latency for low data flows. Figure 2.4 illustrates the key capabilities of the three different scenarios.

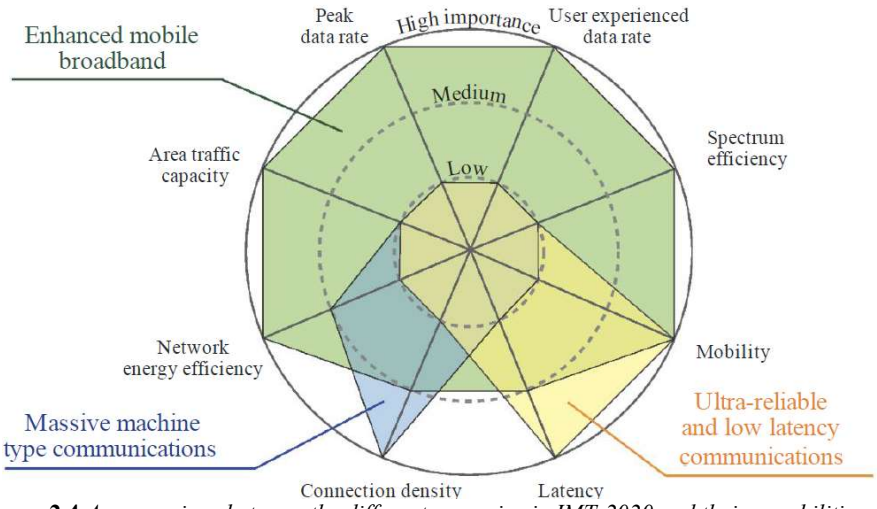


Figure 2.4 *A comparison between the different scenarios in IMT-2020 and their capabilities.* [2]

As a result, splitting the use cases into different categories was tremendously beneficial and essential to meet 5G high standards. This allowed for introducing new techniques, which enhanced performance in one desired aspect, even if it leads to a lower performance in other less relevant aspect.

2.1.2 5G NR New Key Features

In 5G NR, several new key technologies are introduced. These technologies can be combined to meet the requirements of IMT-2020 for 5G systems. Examples of these new features are: Network Ultra-Densification (UDenseNets), mmWave, Optical Wireless Communication (OWC), Massive MIMO (mMIMO), Full-Duplex technology (FD), Dynamic Spectrum Access (DSA), and Non-Orthogonal Multiple Access (NOMA). Figure 2.5 illustrates the uses of these new key technologies in a dense 5G macrocell environment.

The new technologies can be very beneficial to each other, whether adding new advantages, or helping to overcome the trade-offs of other techniques. For example, mmWave can add a lot of new spectra to the system, and significantly increase the data throughput. However, due to high penetration losses introduced from the atmosphere in mmWave range, the coverage area can be limited. Nevertheless, if combined with mMIMO, the coverage area can be extended using beamforming techniques, meanwhile the small wavelength of mmWaves can make it easier to implement small size antennas of mMIMO.

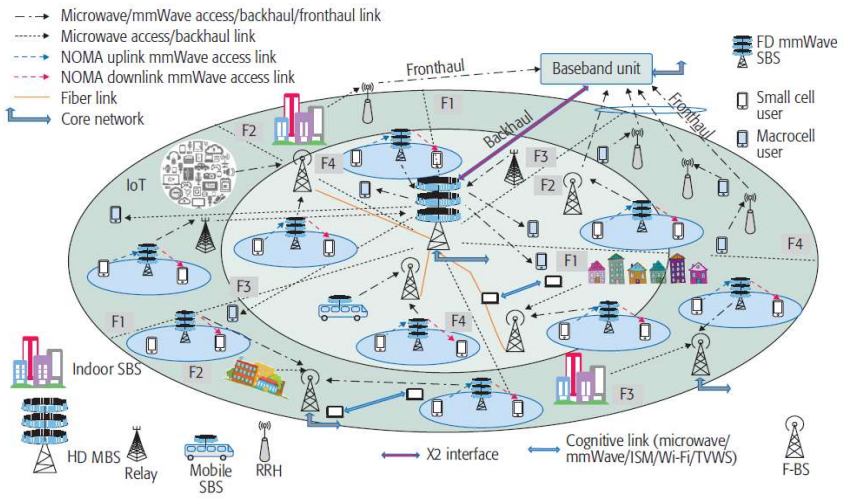


Figure 2.5 *An illustration of 5G NR macrocell with the use of its new technologies.* [10]

On the other hand, mmWave and mMIMO integration requires large number of radio frequency (RF) chains, which can be responsible for approximately 70% of the total transceiver energy consumption [8]. To reduce this disadvantage, hybrid beamforming (analog and digital) is considered, which will affect the system performance. Another proposed solution is using beam-space MIMO, which can significantly reduce both the number of RF chains and energy consumption [9]. However, each RF chain can support only one user in the time frequency domain. The important role of different new technologies to back up one another can be highlighted here, as NOMA can be introduced to help overcome this dilemma through its multiplexing capability when users are on the same beam range. Along with FD, one RF chain can serve users in both downlink (DL) & uplink (UL) channels. [10]

Similarly, many emerging technologies are suggested to join the above features for improved overall performance. One of these techniques which can be used in the 5G network is ML, which might introduce new advantages to overcome many challenges. In this thesis, we will investigate the possibility of improving the scheduler's decisions for UL LA, in mmWave environment.

2.1.3 mmWave Frequencies, Capabilities and Limitations

Considering the congestion occurring in different radio frequency ranges in previous generations, it was just a matter of time for the idea to include the unlicensed higher frequencies. The name of mmWave frequencies was originally derived from the small wavelengths measured in millimeters at this spectrum [11]. As frequencies ranging between 30 and 300 GHz lead to wavelengths of 10 down to 1 millimeter. However, in 5G terminology, it is common to indicate frequencies ranging between 24.5 and 52.6 GHz as mmWave frequencies, since only this portion of high frequencies is used in 5G NR. Figure 2.6 demonstrates the operating bands specified in 3GPP release 15 for frequencies above 6 GHz [12].

In release 15, 3GPP divided the operating 5G NR frequency band into two frequency ranges [13]:

- Frequency range 1 (FR1) includes all bands below 6 GHz.
- Frequency range 2 (FR2) includes the new added spectra in the range 24.25-52.6 GHz. Table 2.1 shows the operating bands defined by 3GPP in Frequency Range 2 [12].

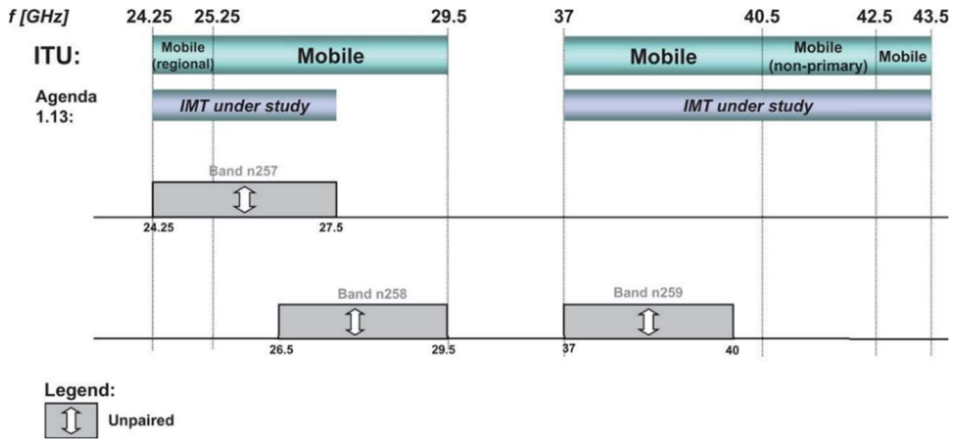


Figure 2.6 *Operating bands specified by 3GPP for frequencies above than 6 GHz (FR2).* [12]

In 5G Networks, mmWave frequencies can be used in all three scenarios, eMBB, URLLC, and mMTC. For example, in URLLC, mmWave signals are considered as the main resource in V2V and V2I communication, as it enables low latency connections, and provides the required frequency spectrum to allocate the data generated from the sensors in each vehicle. Meanwhile, in the cellular networks, it can enable extremely high data rates, and add a huge spectrum to the network. Finally, due to its large capacity, many IoT devices are expected to work in FR2, such as smart city sensors, smart phones, wearable devices, wireless headsets, augmented reality applications, etc.

Table 2.1 *The operating bands defined by 3GPP for NR in FR2.* [12]

NR Band	Uplink and Downlink Range (MHz)	Duplex Mode	Main Regions
n257	26,500 - 29,500	TDD	Asia, Americas (global)
n258	24,250 - 27,500	TDD	Europe, Asia (global)
n259	37,000 - 40,000	TDD	US (global)

Adding FR2 to the frequency spectrum approximately doubles the network total bandwidth. However, there is very strong and valid reasons why this huge unlicensed spectrum was not used previously. As a matter of fact, for long time, it was argued that it might not be possible to use high frequency environment for reliable mobile communications. However, the arrival of many new technologies has helped to overcome its intricate environment and enabled the possibility to take advantage of this spectrum. Some mmWave challenges and solutions are demonstrated below.

- **Challenges**
  - In addition to the high isotropic free space losses, mmWave signals suffer from excessive penetration loss. Also, high propagation attenuation is introduced due to the atmosphere absorption of oxygen molecules and water vapor [3]. Therefore, smaller cell size must be used to decrease the attenuation, at the expense of increasing number of handovers. This also results in increasing the cost and complexity of the network, which is a side effect of raising the number of access points and base stations to achieve good coverage.

- Raindrops can highly affect the availability of the connection [11].

- Doppler spread as well as frequency and phase errors can have a greater effect while receiving mmWave signals.

- The power efficiency of the electronics, especially power amplifiers, decrease when operating on higher frequencies [14].

To reduce and overcome these challenges, several solutions can be applied. Some of these solutions are listed below:

- **Solutions**

  - Network Ultra-Densification (UDenseNets) can be considered as one of the offered solutions, as mmWaves can be used as hot spots in the highly populated dense areas, where the user speed is limited and Line of Sight (LOS) signals are present. This will limit the signal attenuation, and therefore ensure extremely high data rates, while maintaining good reliability, and greatly increasing network capacity.

  - Dual Connectivity (DC) of 4G and mmWave (5G), can be used to improve the link handover and reliability [12] [15].

  - Using massive beam forming at high frequencies can increase the coverage area along with reasonable antenna size.

  - To increase the robustness against doppler spread, frequency, and phase errors; higher numerology of sub-carrier spacing are used. This, in addition to smaller cell size, lead to lower latency, which can be very beneficial in delay-sensitive applications.

To take maximum advantage of mmWave spectrum, 5G NR networks also allow using Carrier Aggregation (CA) between FR1 and FR2. Applying all these solutions enables the network to benefit from the currently under-utilized mmWave spectrum, resulting in adding new features such as, extremely high data rates, increased network capacity, and lower latency.

2.2 Link Adaptation in Wireless Networks

In wireless networks, LA is the process where the robustness of the transmitted signal is determined. Signal immunity against noise and interference is traded against link throughput. This is due to the fact that to increase the robustness of a signal, higher code rates are needed, and/or less bits are represented by each transmitted signal using lower modulation orders.

2.2.1 The Principle of Wireless Link Adaptation

For a better understanding of the link adaptation procedure, lets imagine the scenario in Figure 2.7. In this scenario, multiple speakers are transmitting information using different languages, for specific listeners (lets denote them here as users). Each speaker has scalable power levels (from 1 to 10). The goal is to determine the optimum power level for a successful transmission. For simplicity, all speakers will transmit using the same power level:

- First, starting at the lowest power level of the speakers (level 1), the users would have a lot of troubles in receiving the information, as the voice will be very low, and most of the information will be lost.

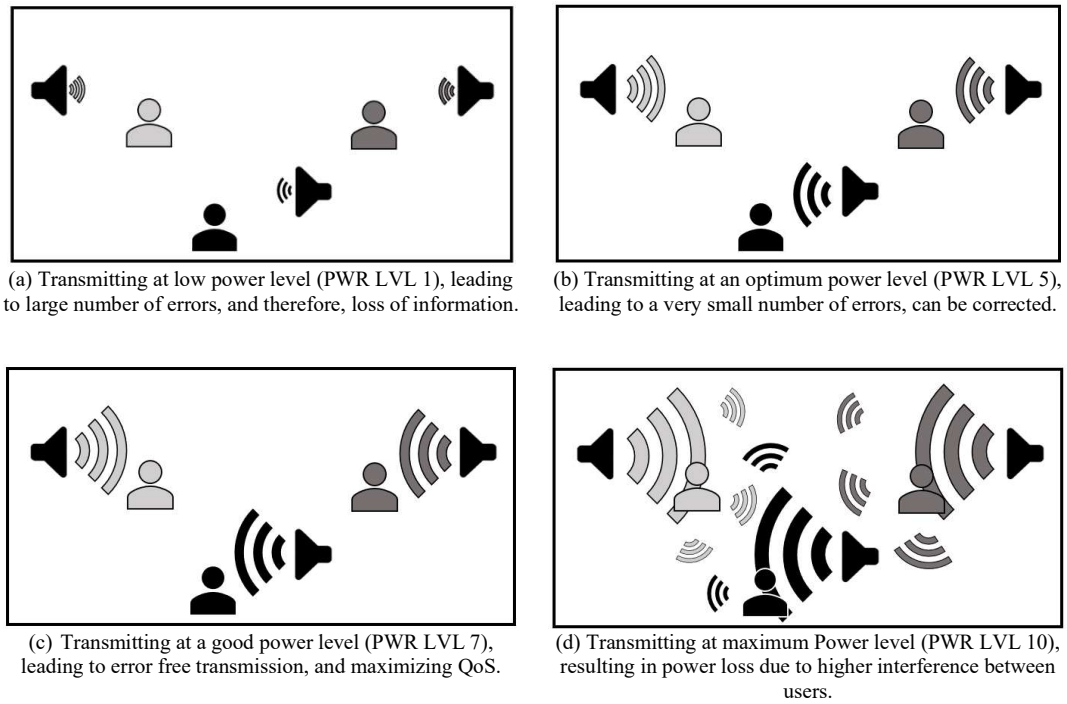
- By increasing the power level, the lost information will be decreasing, until the point that all information is successfully received.

- If the power level was set to an optimum value (let us assume 5), the audio signal will reach to the users with a suitable power level, and most of the information will be successfully received. Even while some words were not clearly heard, it was still possible for the users to fully understand all the information from the context, and therefore, it is not needed to increase the power level more than that.

- When the power level reaches level 7, the audio signal will be clearer to the users, and they will not need to correct many information from the context, since less information was lost compared to power level 5. Note that in this case audio speakers have very low interference level on each other.

- If the power level was set to the maximum value (10), the audio signal would still be heard clearly by the users. Although, this time increasing the power will not enhance the audio signal quality, this is due to higher interference introduced to the users by the other speakers. Moreover, some energy will be lost in the process, since the information would still be successfully received if lower power levels were used.

A similar example of this scenario is what so called, the cocktail party effect, which is the phenomenon of the human's ability to extract information from one audio signal, while filtering out a range of other less important audio signals, as when a person can focus on a single conversation in a noisy room [16]. The cocktail party effect states that higher audio power levels will not necessarily improve the situation, unless it introduces an increase to the Signal-to-Noise Ratio (SNR) [17]. After reaching a certain threshold -power level 7 in this case-, the SINR will saturate and will be the same for all higher power levels.



(a) Transmitting at low power level (PWR LVL 1), leading to large number of errors, and therefore, loss of information.

(b) Transmitting at an optimum power level (PWR LVL 5), leading to a very small number of errors, can be corrected.

(c) Transmitting at a good power level (PWR LVL 7), leading to error free transmission, and maximizing QoS.

(d) Transmitting at maximum Power level (PWR LVL 10), resulting in power loss due to higher interference between users.

Figure 2.7 *Adjusting the power of the transmittied audio signals, to determine the optimum power level.*

Now, after demonstrating the above scenario, it is time to relate it to the link adaptation procedure in wireless communication networks, starting by defining the below:

- **Signal to Interference and Noise Ratio** can simply be described as a measure of how strong the information signal is, relative to the unwanted signals (noise and interference). SINR or SNR, when the interference is neglected- is a critical parameter for any LA procedure, as it has a direct effect on the link capacity, as given in equation (2.1) below [18]:

$$C = W * \log_2(1 + SNR), \qquad \text{[b/s]}. \tag{2.1}$$

  Typically, when the transmitted signals are relatively low, the noise signals can have the dominant effect on decreasing the SINR. While -as in the scenario above-, by increasing the transmitting power in a network, the interference signals will have a higher effect than noise on the received signal quality, which will eventually cause the SINR to saturate on a certain threshold no matter how much the transmitting power is increased.

- **Code Rate** can be defined as the ratio between the number of information bits and the total number of transmitted bits. Code rate always ranges between zero and one, since the useful number of bits can never be more than the total number of transmitted bits for a successful transmission. Decreasing code rate can increase the robustness of the transmission. Coding creates a higher number of bits -in total- to be transmitted, but it allows the receiver to detect and correct errors even at lower signal power levels. Though, this is on the expense of decreasing the data throughput. Coding is similar to the process of correctly detecting the information from the context as in the scenario above. In relation to this fact, "the father of information theory" Claude Shannon proved that English prose has a redundancy of more than 50% [19], enabling the possibility of error detection and correction in written texts. However, in real wireless systems, more complex coding facilitates reliable transmission while using more efficient coding than the redundancy present in English language.

❖ **Modulation** is the phase where the digital bits are converted to analog signals, that can be represented and transmitted through the wireless channel as an electromagnetic signal. By increasing the order of the modulation scheme used, more bits are represented as one analog signal at a certain frequency-time resource. Higher order modulation schemes have more points allocated in the constellation diagram, leading to each signal representing a greater number of information bits. However, this means that the Euclidian distance between each point will be decreased since there are more points, and therefore will lead to increased number of errors for a fixed SINR value. Figure 2.8 [20] below shows the constellation diagram for different modulation schemes supported in 5G NR, where Quadrature Phase Shift Keying (QPSK) represents two bits per symbol, and 16-Quadrature Amplitude Modulation (QAM) and 64-QAM represent four and six bits per symbol, respectively.

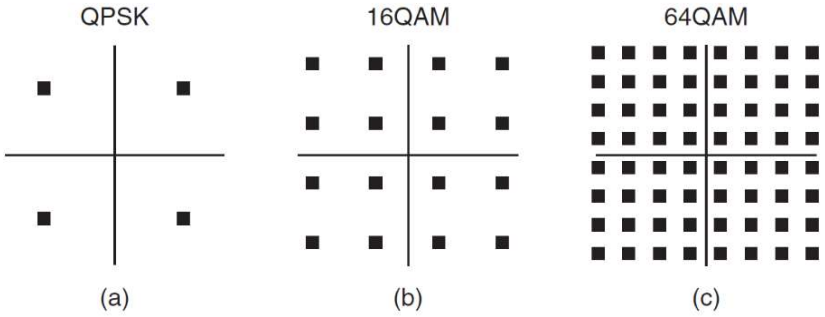
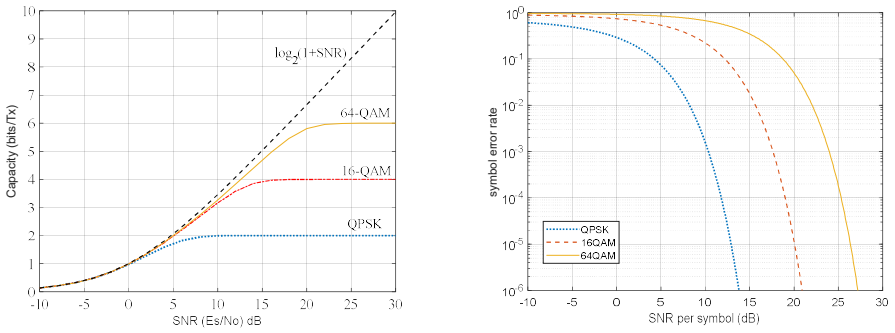


Figure 2.8 *Signal constellations for (a) QPSK, (b) 16-QAM and (c) 64-QAM.* [20]

Figure 2.9 (a) and (b) where generated by Malab to demonstrate the performance variation between QPSK, 16-QAM, and 64-QAM modulation schemes, over an Additive White Gaussian Noise (AWGN) channel [21]. The two charts illustrate the trade-off of using higher modulation schemes that is, increasing the capacity of the link from one hand, and decreasing the performance versus symbol error rate on the other.



(a) *Capacity vs SNR (E_s/N_o) for QPSK, 16-QAM, and 64-QAM.*

(b) *Symbol Error Rate vs SNR (E_s/N_o) for QPSK, 16-QAM, and 64-QAM.*

Figure 2.9 *The trade-off between channel capacity and symbol error rate VS SNR using different modulation schemes.*

Finally, as a conclusion of this scenario, it can be proposed that 5 is the optimum power level can be used, which insures successful information transmission with lowest possible power consumption. On the other hand, it can also be argued that the power level 7 is the ideal value for this case, since it provides a higher received signal quality, or what can be called as a higher QoS. Also, power level 6 can be suggested which can provide a little bit of both. Therefore, the optimum power level can only be determined after deciding the acceptable level of audio signal quality. In relation to this concept, 3GPP decided to set a BLER target in 5G standards as one of QoS measures, where BLER should be targeted to be below a certain threshold of 10%.

Another point should be highlighted here, that is, in this scenario human languages were used, where even when having a higher received signal quality, the obtained data throughput cannot be increased. But unlike human languages, in modern wireless communication networks, the increased signal quality -SINR- can lead to a higher data throughput. This can be achieved by using higher modulation orders and lower code rates. To take a full advantage of the high received signal quality, optimal modulation scheme and code rate must be chosen before the transmission, at the transmitter side. This requires knowledge about channel conditions in advance.

However, in real wireless networks, multiple variables add much more complexity to the procedure of predicting the optimal modulation scheme and code rate. As an example, users can move in random directions with random speeds, causing rapid changes on channel conditions. Also, other users can introduce interference signals that might hugely decrease SINR. Therefore, it must be a periodic procedure that adapts with these changes and optimizes the modulation order and code rate frequently. In 5G networks, this procedure is done through two steps as demonstrated in the next section (Section 2.2.2).

2.2.2 Inner-Loop and Outer-Loop Link Adaptation

LA in current 5G NR systems use look-up tables to determine the modulation and coding scheme. Each MCS value on the table represents a certain combination of modulation order and code rate. The scheduler chooses between 28 possible MCS candidates, mainly depending on the estimated SINR. The chosen MCS value should result on the highest possible throughput, while satisfying the constraint of maintaining the BLER below a certain threshold (10%). However, keeping up with this constraint can be very challenging, since suitable channel conditions and acceptable accuracy of the estimated SINR are needed.

Moreover, the demand in keeping the BLER below a certain threshold requires more conservative MCS selection after failure transmissions. The feedback from the receiver side of the transmission success or failure is referred to as Acknowledgement and Negative Acknowledgement (ACK/NACK). Those failures can be a result of optimistic scheduler MCS decisions, which are mainly caused by the inaccurate estimation of SINR. The process of LA depending on the estimated SINR is called Inner-Loop Link Adaptation (ILLA), while the process of adjusting the MCS value depending on the ACK/NACK of previous transmissions is called Outer-Loop Link Adaptation (OLLA).

In this thesis work, we will focus on improving the OLLA performance using a reinforcement learning algorithm. This algorithm will use ACK/NACK, as well as the scheduler MCS decision as input, to predict a new and more accurate MCS value.

2.2.3 Downlink and Uplink Link Adaptation in 5G NR

The process of LA follow the same concept in both UL and DL scenarios. However, a few differences are present due to the fact that the channel quality is measured in opposite places, while the scheduler decisions for both cases are made at the Base Station (BS) side, which is denoted as Next Generation Node B (gNB) in 5G terminology. Below is a brief description of the DL and UL LA processes, to avoid any confusion between the two cases.

❖ Downlink Link Adaptation

In DL case, the channel is estimated at UE side, the reference signals sent from gNB are used for this estimation, these refence signals are called Channel State Information Reference Signals (CSI-RS). Since the SINR is a continuous (non-discrete) value, an integer value ranging between 0 to 15 (4 Bits) is used as a reference of channel quality. This value is defined as Channel Quality indicator (CQI). The CQI is sent by the UE in control channel reports called Channel State Information (CSI), which might also include other information, such as Precoding Matrix Indicator (PMI), CSI-RS Resource Indicator (CRI), Layer Indicator (LI) and Rank Indicator (RI) [22]. Table 2.2 shows the 3GPP CQI 4 Bit table and its corresponding modulation, code rate, and efficiency in one of the DL scenarios.

Once the CQI is received at the gNB, the scheduler maps the CQI value to a suitable corresponding MCS value. However, since CQI is a 4-bit integer ranging between 0 to 15 and MCS is a 5-bit integer ranging between 0 to 31, there will be more than one MCS value that corresponds to a certain CQI value in many cases. Therefore, the scheduler must determine which MCS value corresponds to the received CQI value. This procedure was not standardized by 3GPP and it was left for the different vendors to optimize individually.

Table 2.2 *4-bit CQI table defined by 3GPP as Table 2.* [22]

CQI index	modulation	code rate x 1024	efficiency
0	out of range		
1	QPSK	78	0.1523
2	QPSK	193	0.3770
3	QPSK	449	0.8770
4	16QAM	378	1.4766
5	16QAM	490	1.9141
6	16QAM	616	2.4063
7	64QAM	466	2.7305
8	64QAM	567	3.3223
9	64QAM	666	3.9023
10	64QAM	772	4.5234
11	64QAM	873	5.1152
12	256QAM	711	5.5547
13	256QAM	797	6.2266
14	256QAM	885	6.9141
15	256QAM	948	7.4063

The chosen modulation order and code rate will be used for transmission on the Physical Downlink Shared Channel (PDSCH). Then the MCS value will be sent among the Downlink Control Information (DCI), to enable the UE for correct demodulation. Table 2.3 shows the 3GPP MCS 5 Bit table and its corresponding modulation, target code rate, and efficiency in one of the DL scenarios.

There are 28 possible values, and 4 reserved values, as seen in Table 2.3. By increasing the MCS value, the Transfer Block Size (TBS) of the transmitted subframe is increased, which represent the useful bits in a subframe. As a consequence, higher data rates can be achieved on the expense of signal robustness against noise and interference.

Table 2.3 *MCS index Table 2 for PDSCH [22].*

MCS Index I_{MCS}	Modulation Order Q_m	Target code Rate R x 1024	Spectral efficiency
0	2	120	0.2344
1	2	193	0.3770
2	2	308	0.6016
3	2	449	0.8770
4	2	602	1.1758
5	4	378	1.4766
6	4	434	1.6953
7	4	490	1.9141
8	4	553	2.1602
9	4	616	2.4063
10	4	658	2.5703
11	6	466	2.7305
12	6	517	3.0293
13	6	567	3.3223
14	6	616	3.6094
15	6	666	3.9023
16	6	719	4.2129
17	6	772	4.5234
18	6	822	4.8164
19	6	873	5.1152
20	8	682.5	5.3320
21	8	711	5.5547
22	8	754	5.8906
23	8	797	6.2266
24	8	841	6.5703
25	8	885	6.9141
26	8	916.5	7.1602
27	8	948	7.4063
28	q	Reserved	
29	2	Reserved	
30	4	Reserved	
31	6	Reserved	

The reserved values are used for retransmissions, where each value corresponds to a different Redundancy Version (RV). The RV is a two-bit integer that represents different bit combinations, and it is an important input at the receiver for correct detection. As in 5G NR systems, Hybrid Automatic Repeat Request (HARQ) uses the previous erroneously received packets, which have different RV, to increase the probability of correctly detecting the packet in the retransmission procedure. Each erroneously received packet is combined with the recently retransmitted packets, to obtain a single, combined packet that is more reliable for detection [23].

❖ Uplink Link Adaptation

In the UL case, the gNB estimates the SINR using the reference signals transmitted by the UE. By applying this estimation, the MCS value can be mapped using look-up tables, which were optimized using several simulations with different scenarios. Then the chosen MCS value is transmitted to the UE to enable correct modulation. The UE uses the look-up tables to determine the Modulation Order (Q_m) and Code Rate (R).
The MCS value is transmitted in the DCI reports, and it ranges between 0 and 31 (5 Bits). Table 2.4 below demonstrates one of these tables used by 3GPP to map the Modulation Order and Code Rate.

Table 2.4 *MCS index table for UL with transform precoding and 64QAM [22].*

MCS Index I_{MCS}	**Modulation Order** Q_m	**Target code Rate R x 1024**	**Spectral efficiency**
0	q	240/q	0.2344
1	q	314/q	0.3066
2	2	193	0.3770
3	2	251	0.4902
4	2	308	0.6016
5	2	379	0.7402
6	2	449	0.8770
7	2	526	1.0273
8	2	602	1.1758
9	2	679	1.3262
10	4	340	1.3281
11	4	378	1.4766
12	4	434	1.6953
13	4	490	1.9141
14	4	553	2.1602
15	4	616	2.4063
16	4	658	2.5703
17	6	466	2.7305
18	6	517	3.0293
19	6	567	3.3223
20	6	616	3.6094
21	6	666	3.9023
22	6	719	4.2129
23	6	772	4.5234
24	6	822	4.8164
25	6	873	5.1152
26	6	910	5.3320
27	6	948	5.5547
28	q	Reserved	
29	2	Reserved	
30	4	Reserved	
31	6	Reserved	

CHAPTER 3

3 Introduction to Machine Learning

As wireless networks evolved, more and more data were generated along with the new rising applications and services. Taking full advantage by processing the generated data can improve human life significantly. However, this requires new techniques since human capability in processing large data is limited. Therefore, an increased attention is driven towards research on automatically processing these large data sets using Artificial Intelligence (AI). The processed information can be used in many major areas such as optimization, statistics, data mining, and many others.

ML is considered as a sub-category of AI. It mainly depends on processing data sets to predict future actions or suitable classifications. This can be done using mathematical models based on sample data sets.

3.1 Machine Learning Basic Concept

In 1959, Arthur Samuel defined ML as the "field of study that gives computers the ability to learn without being explicitly programmed" [24]. This is done by enabling computers to learn from experience, which will eliminate the need of detailed and complex programming. This will also allow computers to learn directly from data, without the need for the human interference. Since humans has a limited ability of learning from large data compared to computer, surpassing this limitation will introduce many new features and applications were not possible previously.

ML is a very hot research area that is being widely invested in to improve numerous life applications. These includes search engines, market prediction, social media applications, health care, and speech recognition, to mention a few. 5G networks as well are expected to take advantage of ML. Many workshops are being held by the ITU that discusses the important role of ML in 5G networks. For example, in the context of 5G RAN, ML offers enhanced scheduler performance, Radio Resource

Management (RRM) optimization, Beam pattern optimization, indoor positioning, and countless other applications [1].

The ML algorithms differ based on the input data and the system requirement. The most popular algorithms are: supervised learning, unsupervised learning, and reinforcement learning.

- **Supervised Learning**

In supervised learning, the aim is to build a mathematical model that can predict the output from a certain input data. This is done by exploiting the knowledge of a previously observed set of data, this data contains the input data and its actual output, this is referred as the training data set. The mathematical model should be able to predict the output of new inputs were not included in the training sets. In supervised learning it is critical to have large enough and accurate training sets that allow for an accurate mathematical model to be built.

Supervised learning can also be used to classify an input into a certain class, based on previous inputs or previous classification attempts and the success or failure through supervision.

- **Unsupervised Learning**

Unlike supervised learning, this ML algorithm does not use training sets. Instead, unsupervised learning seeks to find hidden similarities and patterns in a required set of data. Based on these similarities the data set can be categorized and grouped into different subsets. This clustering is used in various applications, where it is necessary to automatically separate data that contains similar unique features. For example, social media applications use unsupervised learning to label a group of users depending on their activities, which helps in selecting more suitable advertisements to display for these users.

- **Reinforcement Learning**

A system that uses Reinforcement Learning can be described as “an AI system that learns from its own mistakes”. RL uses the experience gained from its previous actions for better future decisions. In the beginning, the system will not have enough experience to take the correct action. After a while of training and making incorrect actions, the system will gain more experience by storing the result of each action using a mathematical model. The stored experience allows the system to take better decisions in the future. Reinforcement Learning includes two main conditions

known as: exploration and exploitation. These two conditions -or policies- determine how the system takes an action. In exploitation, the system depends on the previous experience as described above. In exploration, it will try to explore new outcomes, mostly depending on randomness. Since this thesis discusses the use Reinforcement Learning in 5G networks, more detailed description can be found in the next section (Section 3.2).

3.2 Introduction to Reinforcement Learning

In Reinforcement Learning, an agent in a certain environment must choose a suitable decision depending on its state (S). Each action or decision taken by the agent will have some effect on the environment. After each action a feedback is returned to the agent by observing the environment. The feedback is used as an input to calculate a reward (R) for the action (a), using a predefined reward function. This reward will lead to adjusting the policy of the agent accordingly. If the reward is positive, the agent will increase the chances of using the same action under similar conditions -state- and vice versa. The decisions are then determined using a mathematical model relying on the stored experience. If the actions taken can affect the next state, the mathematical model will take in consideration the cumulative expected reward when making a decision. Figure 3.1 shows the general procedure of a RL system.

In summary, to build an RL system, a mathematical calculation is needed to profit from and store previous experience. However, it is important to first define the input variables below:

- The set of States $S = \{s_1, s_2, \ldots, s_n\}$ of n possible states of the agent in the required environment.

- The set of Actions $A(t) = \{a_1(t), a_2(t), \ldots, a_m(t)\}$ of m possible actions by the agent, at a certain time in the environment.

- The Reward function R which defines the reward resulting of taking an action at a certain state. The reward represents the instant measure of the eligibility of choosing this action for that state, and it will affect the possibility of selecting this action again in future.

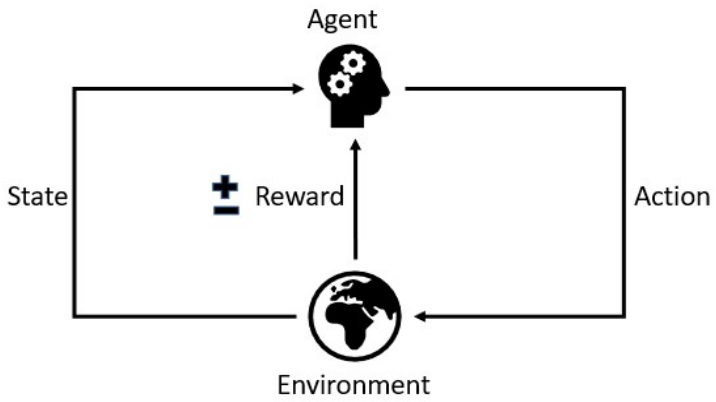


Figure 3.1 *Reinforcement Learning general procedure.*

3.2.1 Reinforcement Learning Concept

The idea behind Reinforcement Learning is to imitate one of the human learning behaviors that is, to learn by interacting with a certain environment. This can be demonstrated by the practical example below:

- Walking requires a complex mixture of muscles and body movement; however, humans learn it at a considerably early age. Since it is difficult for us to learn by language at that time, we tempt to learn by interacting, that is, to take an action and observe the consequences of that action. After many wrong decisions -actions- and falls, a few steps can be made for the first time. These first steps will result on a positive reward, and therefore allow for the repetition of these consecutive actions.

- At this point, it is possible to move these certain muscles with the correct amount of power to keep balance while moving. All movements that lead to a fall will result in a negative reward, and therefore, will be avoided in future. Repetition of this procedure for a suitable amount of time will result in storing the experience of walking in our brains. This allows us to use this experience for walking in future. The time needed for the repetition mainly depends on the complexity of the required procedure.

- When our brains store all the needed experience of walking, most of the decisions will be taken based on the previous experience. However, at early stages, it is also important to frequently explore new ways to optimize our movement. By exploring more and more actions, this allows us to enhance our balance and learn new techniques such as jogging, running, and jumping.

- As we grow older, more actions are taken to exploit previous experience and less actions depend on exploration, this is due to the fact that exploration decisions normally have higher probability of negative rewards. It is very complex to determine when the system should depend on previous experience and when it should try to explore new better techniques, this is what so called the exploration-exploitation dilemma, which will be discussed in more detail in the coming sections. However, in general, the RL system should depend more on exploration at the beginning, since there is not enough experience. By gaining more and more experience the agent should decrease the exploration decisions and start exploiting the previously gained experience.

3.2.2 Exploration vs Exploitation

One of the most controversial topics in RL is the exploration/exploitation trade-off. As discussed in the previous example, it can be quite challenging to determine an optimal strategy that switches between exploration and exploitation, with the target of maximizing the cumulative reward.

To better demonstrate the trade-off, let us consider the scenario in Figure 3.2, where the agent is required to find the best route to reach his goal -market-, and the actions are the movement around the environment shown in the figure. The state of the agent changes according to the actions taken, where the state (S) is the position of the agent. In this case, a small reward (+1) is given each time the agent gets closer to its destination -the market-, and a negative reward (-1) is given each time it gets farther. If the agent reaches its goal, it will receive a huge reward (+100). The positive reward will ensure that the agent seeks its goal, while the negative reward is to ensure that the shortest path is found. If the agent only focuses on the immediate rewards, it will get stuck on the closed end of the maze. If so, the agent will not be able to reach its goal, and therefore, the total cumulative reward will not be maximized.

For the agent to be able to reach the destination, it needs to go to the opposite direction for a while (4 steps in this example), even when getting negative rewards. This can be achieved by enabling the agent to explore for some time in the beginning. During the exploring time, the agent's actions will not depend on the rewards. However, the real challenge is to determine the time needed for exploration, especially in other environments where it can be more complex than this example. Also, unlike this example, the environment in many cases can change with time. For example, if the walls of the maze are moving or an obstacle appears

every now and then. This requires continuous exploration to adapt with any new changes on the environment, even if a suitable time of learning the maze has previously passed.

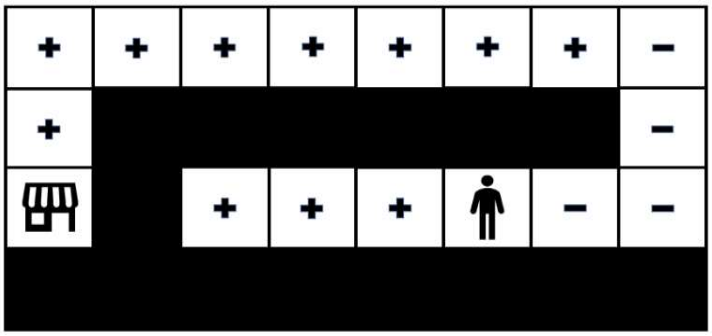


Figure 3.2 *An example of an RL environment, where the agent (human) is required to reach a goal (market) using the mathematical expression depending on the rewards gained during the process.*

In fact, this practical example faces us many times in our lives. Such as if we want to go to a new place for the first time without using our phones of course. If we were walking when we see the building of the destination, we try to go directly towards it. The positive reward is when we see it getting closer and closer. If we face an obstacle on our way like a fence, we try to go around it for a while, even if we go in the opposite direction. Going in the opposite direction leads to a negative reward, as we see the building getting farther and farther. Although, if the obstacle -fence- is a bit wide, it is always difficult to decide when we should turn back and try for another direction.

If we succeed to reach our destination, we can remember this route and use it in future. However, it can also be confusing on our second time, as we will hesitate if we should exploit our previous experience and go on the same long direction, or if we should try to look for another way. After exploring and exploiting multiple times, the optimal route will eventually be found. Even so, it can always be a good idea to explore every now and then, since a new shorter route might have been opened.

One of the strategies used is to include randomness in this procedure, where a random variable is generated before each action. This random variable controls the policy of the decision, where if it is below a certain threshold, the random process of exploration is generated, and vice versa. The threshold is called epsilon, and it should decrease as the agent gains more experience. This is known as a Markov Decision Process (MDP), where the output of the mathematical model is partly random and partly controlled by a decision maker. After a suitable amount of actions, the agent will gain a good experience and epsilon is decreased down to a determined minimum value, which allows for the exploitation of the gained knowledge, as well as to explore every once in a while.

3.2.3 Mathematical Representation

The main goal of the RL algorithm is to maximize the cumulative reward of a sequence of actions. To reach this aim the mathematical expression responsible for making the decisions should take in consideration expected future rewards. Thus, it is obvious that future rewards should have lower affect than immediate rewards, this is due to the fact that future rewards have lower certainty in normal cases. The discount value also helps on reaching the maximum value with the least actions possible. Then the discounted return can be calculated as

$$G_t = \sum_{k=0}^{\infty} \gamma^k * R_{t+k+1} = R_{t+1} + \gamma * R_{t+2} + \gamma^2 * R_{t+3} + \cdots, \quad \textbf{(3.1)}$$

where G_t is the discounted cumulative expected reward, R_t is the reward resulting from an action at a certain state and time instant, and γ is the discount factor that weights the future rewards, where $0 \leq \gamma \leq 1$. The higher the discount factor is, the less is the discount and therefore the future rewards are considered more important. While low discount factor leads to decreasing the impact of the future rewards on the decisions. For example, when the discount factor reaches zero, the agent will only consider current rewards, and future rewards will not have any influence.

Equation (3.1) formulates the general mathematical representation of RL systems. However, it does not include the policy used to update the probability of choosing an action depending on the resulting reward. This is an essential demand since it is unattainable in real scenarios to have a complete knowledge of the environment behavior.

This also brings the question of how important the new experience is, compared to the old one, and when to discard the old experience and depend on the new one. In

other words, if the same action lead to different reward than previously experienced, how can the agent decide which experience is more reliable in future. Knowing that the new experience has the advantage of being up to date, while the old experience has been accumulated along several previous actions and therefore should also be reliable for future decisions.

In the next section, we will discuss one of the model-free learning methods, known as the Q-Learning method, that have a continuous update on the decision-making strategy, depending on the observed rewards.

3.2.4 Introduction to Q-Learning

As discussed earlier, RL uses previous experience to perform an action in a certain state. This can be applied by using many methods and mathematical models. However, Q-learning is considered one of the most popular RL methods, that does not require a strong knowledge of the environment, where it seeks to find the optimal policy for maximizing the expected cumulative reward. To achieve this goal, a reward function is needed, which accurately reflects the agent's final target.

Q-learning depends on a table that stores all previous experience, known as the Q-table. The Q stands for Quality, as it describes the quality of an action at a certain state. This is done by following Bellman's optimality principle, through an iterative process. In other words, after building the Q-table that describes all actions at all states, the agent follows a greedy policy, where it chooses the action that corresponds to the highest Q-value. Figure 3.3 demonstrates the general procedure of Q-learning.

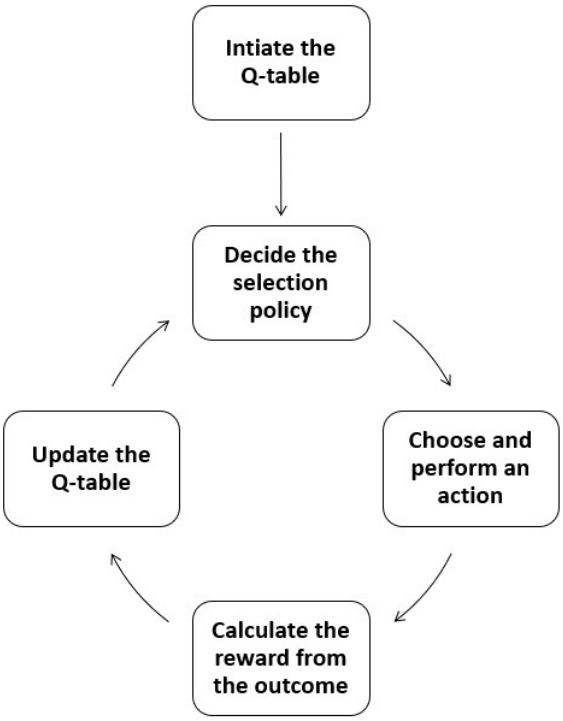


Figure 3.3 *Q-Learning general procedure.*

Q-leaning procedure

- First, the Q-table is initiated, normally by setting its values to zero for all possible actions. In cases of impossible and unwanted actions in the environment, the Q-value of that action is set to negative infinity, to ensure that the selection of this action is avoided. The Q-table size depends on the number of states (n) and possible actions (m).
- Before choosing an action, the agent must decide the action selection policy (exploration or exploitation), which can be controlled by the value of epsilon. The policy determines whether the action selection follows a random process, or if it will exploit the previous experience. As discussed in sections 3.2.1 & 3.2.2, the agent tends to explore more at the beginning, when the Q-table is not optimized, and after gaining more experience, more decisions are made depending on previous knowledge.

- After the agent decides the selection policy, the action can be chosen. The random policy is quite straightforward, where the action is randomly selected out from the set of all actions. On the other hand, the exploitation process depends on a greedy policy, by choosing the action that leads to the highest Q-value.
- The agent then observes the outcome of the performed action at that certain state, and calculates the reward depending on a predefined reward function.
- The most important step in this procedure is updating the Q-table, as it is responsible for storing the gained experience of the chosen action at this state. The Q-value is updated by using Bellman's equation,

$$New\ Q(s,a) = Q(s,a) + \alpha * [R(s,a) + \gamma * \max(nextQ'(s',a') - Q(s,a)], \quad \textbf{(3.2)}$$

where Q(s,a) is the Q value of action (a) at state (s), γ is the discount factor, R(s,a) is the reward for action (a) at state (s), and α is the learning rate, where $0 \leq \alpha \leq 1$ and weights the importance of the new experience compared to the old experience. In other words, the learning rate weights how quickly the agent abandons former information and replaces it with the newly observed experience, that is, if $\alpha = 1$, the new Q-value will completely replace the former value and, therefore, the agent is not concerned with keeping older observations.

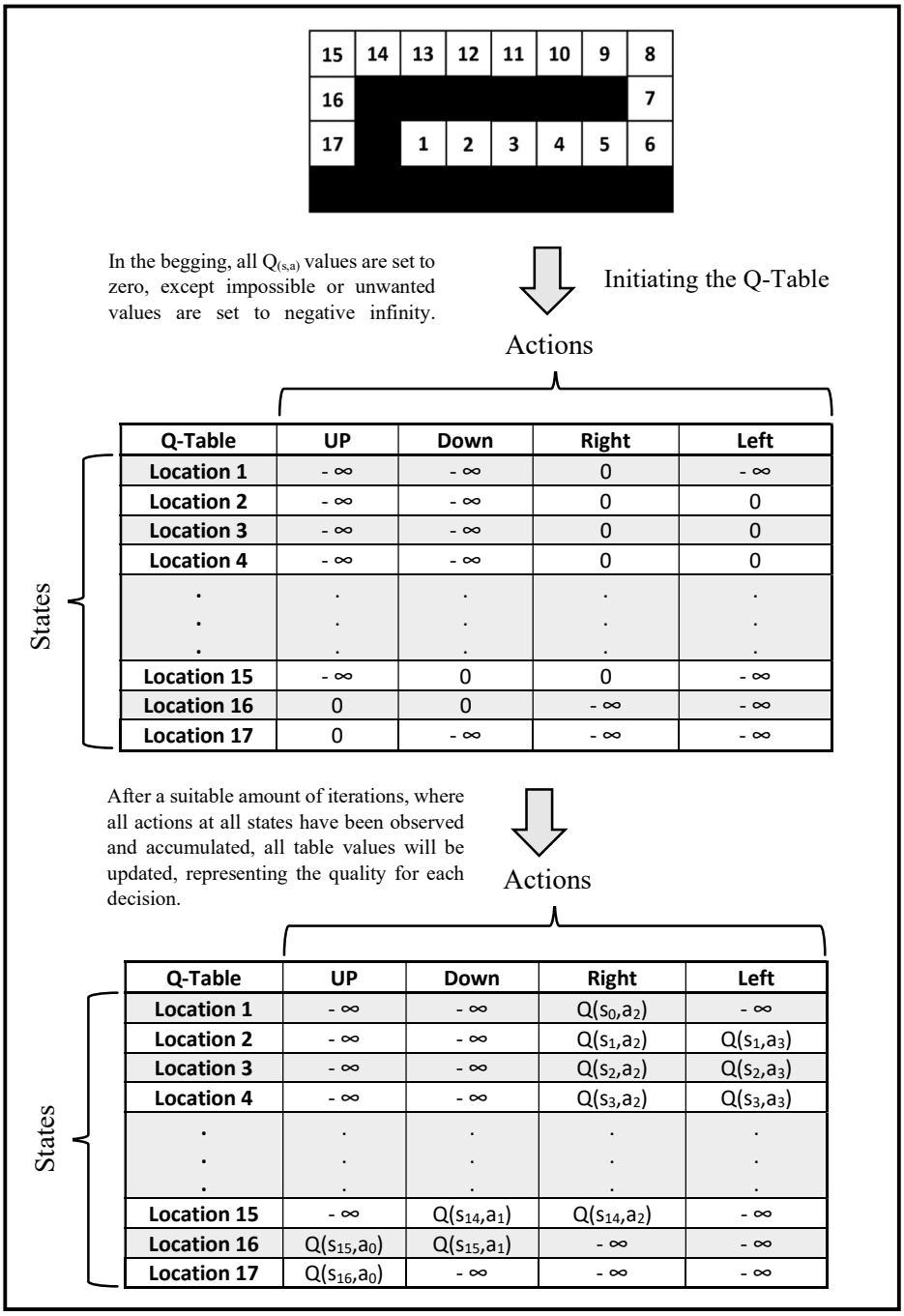


Figure 3.4 *An example for the process of initiating and training the Q-Table*

After a suitable amount of iterations, the Q-table will be optimized to select the actions that lead to the highest expected cumulative reward, bearing in mind that this optimization is according only to previously observed outputs. From there, the agent will be able to perform its goal in its environment. It will, however, be possible for the agent to automatically adapt with any future changes on the environment, since the Q-table depends on the observation of the performed actions. Figure 3.4 illustrates the Q-table of the example in Figure 3.2, where location 1 is at the dead end, and location 17 is at the destination (market). Therefore, the number of states is m = 17, and the number of possible actions is n = 4, which are up, down, right, and left.

CHAPTER 4

4 Machine Learning for Link Adaptation in Wireless Networks

LA is considered one of the popular research areas in wireless networks. This is arising from the fact that 3GPP association sets the standard of only the general LA procedure, which does not include all aspects. Therefore, it left the door wide open for different vendors to investigate individually and compete on achieving highest possible performance, leading to many researches in this field. Furthermore, since growing attention is driven towards ML techniques recently, many papers are focusing on the utilization of ML in 5G networks, including in LA procedures. In the next section, we will review some of the related research work. Then a theoretical approach for the proposed system is described, before finally describing the system implementation.

4.1 Previous Related Work

The main goal of this master thesis is to propose a ML model that enhance uplink link adaptation in 5G networks. The ML system should dynamically adjust UL modulation and coding scheme in the mmWave environment. As previously mentioned, several studies [25]-[29] regarding the utilization of ML in wireless LA can be found. Although, most previous works were examining different independent problems, with multiple ML techniques, and mainly focusing on the DL LA. However, no previous work could be found that discussed the use of ML in the mmWave environment, or for UL scenario in 5G networks.

In [25],[27], and [28], the authors depend on supervised and unsupervised learning by treating this problem as classification and Support-Vector Machine (SVM) models. The main weakness of this solutions is that it requires large sets of training data to build a model of the wireless channel dynamics. Also, in [25] and [27], the system does not keep pace closely with changes occurring in the environment, as it does not rely on the feedback of the system.

In [26] and [29], a reinforcement learning was proposed, based on Q-learning algorithm to avoid the use of model training phases. However, in [29] the state space is defined over the continuous value of received SINR leading to a large number of states after discretization, thus a large Q-table needs to be handled by the learning algorithm, which requires long exploration to be filled and optimized.

In [26], the system is designed for the LA DL procedure in LTE networks. The target of this paper is slightly different from the goal of the thesis since the DL procedure depends on CQI values, which are the main input of its algorithm. Moreover, the system focuses on ILLA, since all UE use the same Q-table and, since it uses CQI values which are direct outputs derived from the estimated SINR. The disadvantage of this algorithm is that it needs a long time for the Q-table to be filled and optimized. This can be quite challenging, since some MCS values are more common than others in real channel conditions. This leads to unoptimized Q-values and therefore the actions can be unreliable even after long time have passed, which also increases the time needed to adapt to new changes on environment. Thus, a long exploration time is needed to optimize the large Q-table, but even so, this will not guarantee optimizing uncommon Q-values. Also, it does not separate between new transmissions and retransmissions cases, which have a clear difference since the HARQ procedure is used. The retransmitted signals will have higher probability of successful reception, due to the HARQ procedure use of previous erroneously received packets with different RV in detecting the retransmitted packets. The proposed algorithm in [26] also results in a lower performance than the standard look-up table currently used when the estimated SINR is accurate, which can be the case most of the time. It also has higher retransmission occurrences when trying to achieve better throughput.

In [30], the authors provide a solution for DL LA in 4G networks, by using the MAB algorithm for RL. The main weakness of this solution is that the mathematical model was built on the BLER, which in some cases can be a deceiving value, and does not reflect the accuracy of the chosen action. The algorithm always seeks to push the BLER to a certain target by controlling MCS value. However, in rough channel conditions, the BLER can be very high and therefore the algorithm will be too conservative in future, even if a good channel conditions appear. For instance, it can happen when an obstacle interferes the signal for a while, then the BLER will increase significantly, and therefore, even if the obstacle is removed and the channel conditions become good, the system will still choose very low MCS values until the BLER recovers. This may not be suitable in mmWave rough environments, where the signal can be blocked easily. Instead, the optimal algorithm should always seek

for the lowest possible BLER value while maximizing the throughput. This can happen even if the BLER is above the threshold due to out of control channel blockage.

Also, the algorithm used in [30] contains complex calculations. This can cause some delay due to processing time, which is vital in for scheduling decisions, and may result in outdated actions, especially since each UE will have its own procedure, and therefore, large processing is needed at the gNB.

4.2 Reinforcement Learning for mmWave Uplink Link Adaptation

In this section, the reasons behind choosing the proposed RL solution for OLLA will be discussed, followed by the description of the system implementation.

4.2.1 Theoretical Approach

Link Adaptation (LA) is a crucial procedure in current mobile cellular networks, as it has a direct influence on both the throughput and BLER. In this thesis, the goal is to improve the scheduler decisions at mmWave frequencies, while obligating the standards of 5G NR networks. In these standards, LA procedure uses an integer value that ranges between 0-28 to represent the most suitable modulation and coding scheme that should be used in the estimated channel conditions.

In the UL scenario, the SINR is measured at the BS (gNB), and eventually mapped to the MCS value, using a predefined look-up tables that were built depending on previous simulations. The MCS value is then sent to the UE in the DCI report to be used for transmission. This LA procedure that relies on the look-up tables and estimated SINR is known as ILLA. However, it can be noticed that it can be vastly vulnerable to outdated and inaccurate SINR estimated values.

To diminish and overcome this issue, an OLLA is used. The OLLA can reduce the MCS subordination to the estimated SINR, by tuning its value when consecutive ACKs/NACKs are being received. In case of receiving numerous consecutive ACKs -while the maximum MCS value is not achieved-, the system can be assumed as being conservative, and therefore, higher throughput might be achieved by increasing the MCS value. Contrarily, if the system is being aggressive, the OLLA should decrease the MCS in case of receiving multiple NACKs, in order to get a successful transmission and reduce the BLER. However, this brings in the demand for optimizing a system that is able to define when to intervene, and what is the optimal value that the MCS should be attuned to.

The necessity for the OLLA is more acute in the intricate mmWave environment, as the channel is expected to experience higher variations, as well as an increment in sudden blockages due to moving obstacles in the environment. Moreover, the SINR estimated value can be outdated or inaccurately measured in many cases.

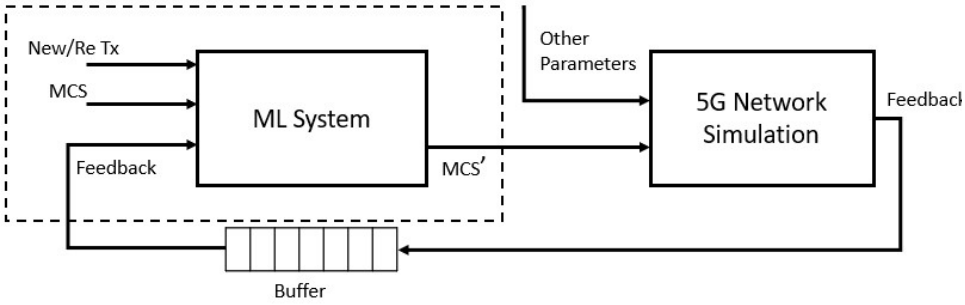


Figure 4.1 *a block diagram of the proposed RL system for UL link adaptation*

In this thesis, a reinforcement learning system is proposed to control this procedure. The system will constantly observe the outcome of previous correction factors and MCS values used in each transmission. Consequently, it can notice after a suitable number of transmissions if the estimated MCS value can be defined as conservative or aggressive based on these previous observations.

The Q-learning method is one of RL techniques, that provide a simple solution to store and exploit the previous experience of the system. As it is mainly based on Bellman's equation to store the quality of each action at a certain state. This information is used in future decisions and updated frequently in a table known as the Q-table.

By using this method, the system can automatically adapt to any changes and inaccuracy in the estimated SINR values, which is done in a relatively simple procedure that does not need high processing. However, the Q-learning system depend on a predefined reward function, that can reflect for the system a quality measure of the chosen action. This allows the system to learn if this action was favorable or not in this exact condition, and therefore, if it should be chosen again in similar circumstances in future. -Figure (1) illustrates a block diagram of the proposed ML system for UL link adaptation-.

The system -in Figure 4.1- was proposed depending on results from previous related work [26] for LTE systems at lower frequencies. As it added robustness to DL LA decisions with the use of ML techniques. Even in cases of having inaccurate measurements as an input, the system could still be able to predict a suitable MCS value. This can be a good addition to improve the LA in mmWave environment, especially that the channel is changing rapidly, and the scheduler is required to make fast decisions. However, due to disadvantages mentioned in section 4.1, and since this thesis focuses on OLLA rather than ILLA, many enhancements were added to allow the system to adapt in the mmWave environment. In the next section, we will discuss the input and output parameters, as well as the reward function of the system.

4.2.2 System Implementation

The aim of the ML system is to predict the suitable MCS value which results on highest possible link throughput while simultaneously decreasing the BLER. The MCS value ranges between 28 integers, starting from 0 with the lowest possible throughput, while the 27th leads to the highest possible throughput.

For this case, RL can be used, which uses a reward function that enables the system to automatically achieve the optimal performance based on positive and negative feedback of previous actions. These actions are chosen using the Q-learning method, where it chooses the action with the maximized future reward values, depending on the factors below:

- **System states (s) / Input parameters**

  the system states can be considered as the combination of two factors. First is the MCS value chosen by the scheduler for the UL transmission. The second input is a flag identifier for new transmissions that is, if the packet sent is a not sent previously the flag is set to one, while if packet is being retransmitted, the flag is set to zero. It can be considered that this flag allows the ML system to use two different Q-tables, where one is used in the case of new transmissions and the other is used in the case of retransmissions.
  retransmission of the packet with HARQ process ID (1).

  This flag is used as an input to reflect the different properties of the received packets introduced by using HARQ procedure, since a retransmitted packet is more robust and has higher probability for detection when using HARQ procedure. As well as it reflects that the previous action and/or the channel conditions are not ideal.

Note that this flag is not equal to the ACK/NACK flag will be used in the reward function, since ACK/NACK are sent by the receiver depending on the condition of the latest transmission. However, this flag follows the UL HARQ Process ID, where each packet is assigned to a process ID, and retransmitted in case of unsuccessful transmission according to the process ID. For example, if the packets of HARQ process ID (0, 1, 2, and 3) are sent respectively, and only (0, 2, and 3) were successfully received, and the scheduler assigns the next transmission for the packet with HARQ processes ID (1), while using different Redundancy Version (RD) as mentioned in section 2.2.3. In this case, the ACK/NACK will refer to the latest transmission, which is process ID (3), and therefore, will be ACK. While the flag of the new transmission will be set to zero, since it is a retransmission.

- **Possible Actions (a) /Output Parameters:**

In Q-learning technique, it is desirable to decrease the number of system actions as much as possible. This is to reduce the resulting Q-table and therefore the complexity of the system, as well as decreasing the exploration time needed to reach the optimal performance.

Therefore, to avoid dedicating an action for each possible MCS value, the output can be considered as a correction factor (CF) added to the actual MCS, and ranges between positive and negative constant value. This constant value determines the number of possible actions, and represent the threshold of the correction factor, and will be referred to as the correction factor margin (k), such as CF = {-k ..., -1, 0, 1, …, k}.

However, in the exploitation phase, adding CF might lead to an MCS value less than zero, or greater than the maximum MCS value (27). These impossible actions are set to negative infinity in the Q-table, so the agent will avoid choosing them in future. While in the exploration phase, the random values that lead to impossible actions are discarded. Then the below condition is applied:

$$newMCS = max[\,0\,, min[MCS + \Delta MCS\,, 27\,]\,]. \quad (4.1)$$

❖ **Reward function (R):**

as described previously, reward function allows the system to automatically achieve the optimal goal based on positive and negative feedback of previous actions. In this case, the reward of a certain action at a certain state can be calculated as below:

$$R(s_t, a_t) = \text{TBS} * \text{ACK}, \tag{4.2}$$

Where TBS is the Transfer Block Size of the transmitted subframe, and ACK is the acknowledgment flag. TBS represents the number of useful bits transmitted using the chosen MCS value. The TBS are used to reflect the usefulness of the chosen MCS value. Hence, the TBS is calculated assuming one-layer UL channel with full buffer state, so there are no padded bits that might influence the TBS value and therefore the reward function. On the other hand, the ACK is an indicator of the success of previous action that is, if the transmission was failed, the ACK will be equal to zero, and therefore the agent will not gain any reward from this action. However, a buffer is used to combine between each HARQ process ID and its ACK/NACK, since multiple transmissions can be performed before receiving their ACK/NACK.
Using this simple reward function can reflect to the system the quality of the decision, and therefore, lead for better decisions in future.

❖ **Learning rate (α):**

Learning rate ranges (α) between [0,1]. As α increases, the importance of new information against old information is increased. In this system, the learning rate was chosen relatively high (α = 0.85), to allow the system to quickly adapt to the changes occurring in the channel. Using high learning rate also decreases the required time to reach its optimized performance.

❖ **Discount rate (Γ):**

ranges between [0,1], and as the discount rate increases, the highly the system evaluates the long-term rewards. However, in LA, the current MCS value have no direct influence on the next MCS value, as it mainly depends on the variation of the channel conditions. Therefore, the discount rate was set to zero.

❖ **Epsilon:**

The switching between exploration and exploitation states is controlled by the exploration rate (epsilon). As mentioned in section 3.2.2, epsilon value decreases progressively to decrease the exploration events in the expense of increasing exploitation, until it reaches a minimum value. The minimum value allows the agent to explore every once in a while. In this case, the minimum value of epsilon should be below the constraint value of BLER (10%), and therefore it was chosen to be (0.05). This means that the random exploration action occurs once every 20 iterations to look for a better reward.

However, in new transmissions, better rewards can only be achieved by increasing the MCS value. Hence, the random action is set to only result in one direction increment of the MCS value. Contrarily, in retransmissions, the exploration event will only result in decreasing the MCS value, since the channel conditions were not suitable for the previously chosen MCS value.

Finally, Bellman's equation is used, where all factors are considered to calculate the new Q values that is, the expected reward for taking certain action and in a certain state, to help in the prediction of future rewards for all states and for all possible actions, as below:

$$New\ Q(s,a) = Q(s,a) + \alpha\,.[R(s,a) + \gamma\,.\max(nextQ'(s',a') - Q(s,a)], \qquad (4.3)$$

After a considerable amount of iterations, the system will be able to choose the optimal action that leads to the highest expected cumulative reward, by utilizing the obtained Q-table, and selecting the action that corresponds to the maximum Q-value.

CHAPTER 5

5 Simulation & Results

In this chapter, the simulation environment is briefly described, followed by an evaluation of the obtained results.

5.1 Simulation Environment

The main goal of simulation is to evaluate the proposed ML algorithm performance and compare it with the currently used LA procedure. Therefore, a professional simulation tool is used to simulate the mmWave environment in 5G networks. The simulation tool was developed by Ericsson, and with the efforts of experts in the field. All needed support, devices, tools and materials was provided by Ericsson Lund AB. The main ML algorithm was coded in Java, then Matlab was used to display the obtained results.

For simplicity of the implementation, the simulation scenario included only one tagged user in the network, randomly moving with walking speed of 1.3 m/s (almost 5 Km/h). The simulation was repeated with 20 different seeds, each with random user direction and starting position, and with simulation period of 10 seconds. Also, the tagged user is assumed of having full buffer all the time, to avoid cases with padded bits, which might have an influence on the resulted throughput. Table 5.1 demonstrates the parameters used in the simulation.

Table 5.1 *Parameter used in the simulation with 20 different seeds.*

Parameter	Value
Scenario	5G NR Network with full buffer user
Simulatin Period	10 s
User Speed	Average Walking speed (1.3 mps)
Carrier Frequency	28 GHz
Bandwidth	200 MHz
No. Of Users	1
Cell Radius	100 m
BS Antenna Height	25.0
UE Antenna Height	1.5 m
Antenna Tilt Angle	15

5.2 Results

The aim of LA procedure is to maximize the throughput while maintaining BLER below 10%. Therefore, to evaluate the scheduler decisions for the UL LA procedure, these two main outputs should be taken in consideration.

After running the simulation, UL throughput and BLER were obtained for 20 different seeds and using 15 different correction factor margin (K). Thus, to facilitate the demonstration of these results, all seed's outputs were averaged. In relevance to this matter, Matlab was used to average the outputs, and present the obtained results in clear figures.

The Correction Factor Margin (K) can limit the actions of the ML system and therefore it's interventions on the MCS value. Such that if K was set to zero, the ML algorithm will be unable to modify the MCS value, as the highest CF possible will be zero, while having K equal to 27 allows the ML system to have full control on the LA procedure. Since the Correction Factor (CF) = {-K ..., -1, 0, 1, …, K}.

Although, by increasing K, the number of actions is increased, leading to higher complexity and larger Q-tables. Therefore, setting K to a small number can be more reasonable, especially that OLLA is expected to correct the estimated MCS value predicted by ILLA. In addition, by using lower K value, it is expected to result in a better performance, due to the decrement in exploration time.

The 15 values of K used in this simulation are [0 (non-ML), 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 24, 27]. The simulation is run for each value with the period of 10 seconds and for 20 different random seed.

❖ UL User Throughput

As mentioned earlier, the output of 20 different seeds is averaged to get a reliable result from the simulation. To simplify the demonstration of system performance, the mean of all the 10 seconds period is calculated, in accordance to the UL user throughput. The obtained results for the total averaged UL throughput showed an improvement when using K = {1, 2, 3, 4, 5}. While the performance kept degrading as K increased. This can be expected, since K value increases the possible actions, and lead to decreasing the dependence on the ILLA. The optimal K value (K = 2) increased the UL user throughput by 1.9 Mbps compared to the non-ML throughput. Figure 5.1 demonstrates the bar chart of the total averaged UL user throughput for 15 different values of K over the simulation period.

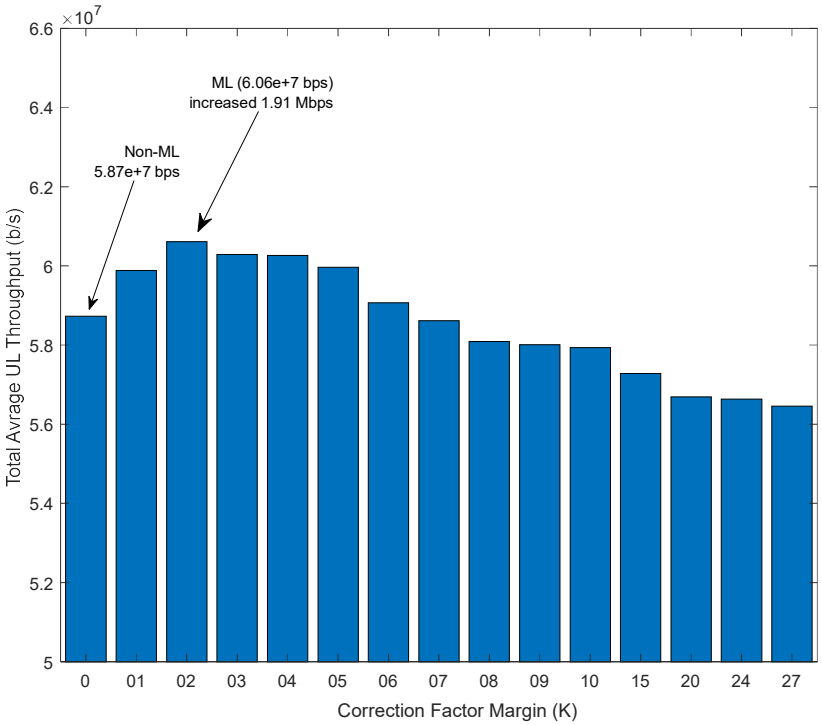


Figure 5.1 *Bar chart of the total UL user throughput averaged over the simulation period, while using different values of K.*

For a better demonstration of the results, and for close view of the improvement introduced by ML on the obtained averaged UL user throughput, the K values of (1, 2, and 3) are plotted along with non-ML curve, over the duration of 10 seconds. The result illustrates that these K values result in increment on UL user throughput over the entire period, as in all three ML cases, the throughput curve was above the non-ML at all the 10 seconds duration. The optimal performance was achieved in the case of using K = 2, while K = 1 performed slightly lower. Figure 5.2 shows the UL average user throughput of k = [0 (non-ML), 1, 2, 3] during the simulation period of 10 seconds.

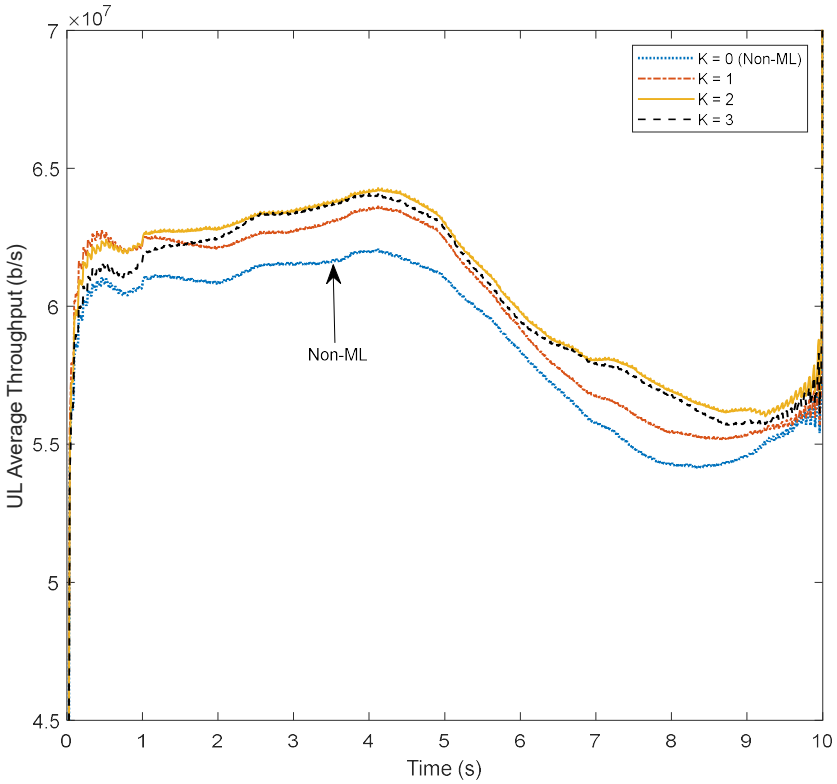


Figure 5.2 *UL average user throughput of different K values during the simulation period.*

❖ BLER

Similar to the procedure followed to present the throughput, UL BLER was averaged for all 20 used seeds, then the mean was calculated over all the simulation period. Figure 5.3 shows the bar graph of the total average BLER for 15 different K value. However, in BLER only two K values (1 and 2) resulted in an enhanced performance compared to non-ML system. Furthermore, despite the fact that BLER for K = 3 was increased, it was able to keep the constraint with being exactly at 10%. However, all other values of K failed to achieve this constraint, and kept degrading as K increased, with comparable performance to the obtained UL throughput.

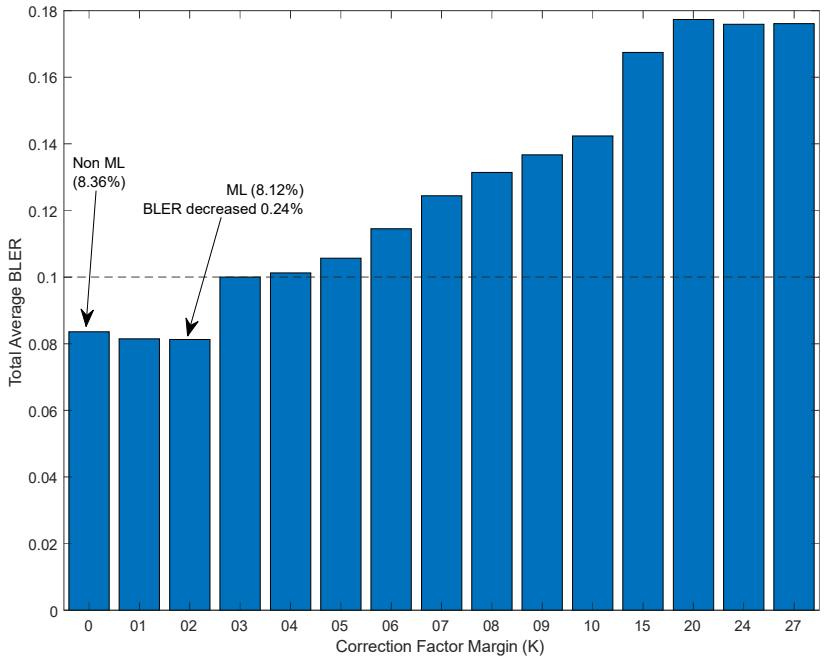


Figure 5.3 *Bar chart of the UL total BLER averaged over the simulation period, while using different values of K.*

By observing the BLER curves obtained from K values that managed to improve the system performance against BLER, and comparing it with non-ML system, it is shown that unlike the throughput case, the different K values replaced places couple of times during the 10 second period. That might reflect that there is also more space of improvement on the current system. Figure 5.4 illustrates the average UL BLER of k values (0 non-ML, 1, 2) during the period of 10 seconds.

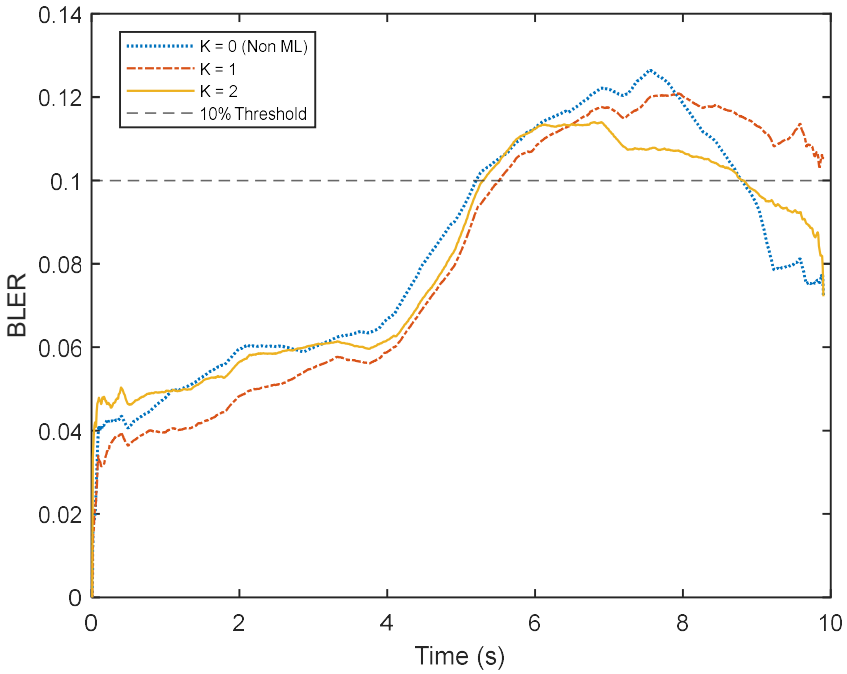


Figure 5.4 *Average UL BLER of different K values during the simulation period.*

❖ Close Review of the obtained results (Throughput and BLER)

Finally, to compare the performance of both the throughput and BLER, only the K values that resulted in at least one type of improvement were considered. However, a normalization process is needed, since the throughput is measured in bps and have a very high value, while the BLER only ranges between 0 and 1. Therefore, the non-ML learning value was considered as the normalization factor for both the throughput (5.87e+7 bps) and BLER (8.36%). As a result, it can be seen that only two values of K (1 and 2) achieved better performance in both aspects, while other values struggled to maintain the constraint of the BLER. Furthermore, the best improvement in both aspects was achieved when K = 2, and therefore it can be considered as the optimal K value in this exact scenario. Figure 5.5 illustrate the normalized UL BLER and Throughput bar chart for different K values.

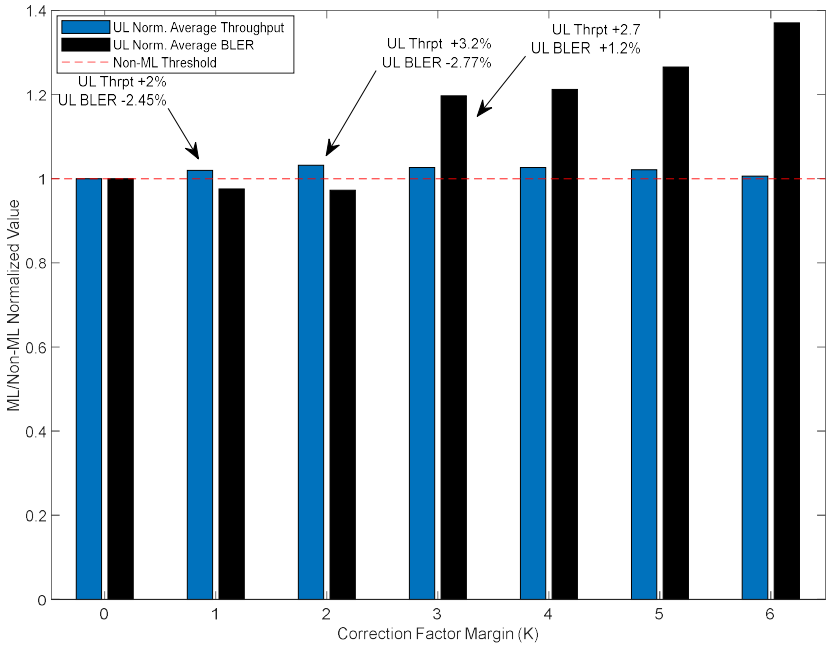


Figure 5.5 *The bar chart of the normalized UL BLER and user Throughput for different K values.*

CHAPTER 6

6 Conclusion and Future Work

6.1 Conclusion

As a conclusion of the obtained results, the RL technique can be considered as a strong candidate to enhance the performance of LA in wireless networks. As it can compensate for inaccuracies in the selected MCS value. As well as it is able to automatically adapt to the changes occurring in the environment that causes a fluctuation in estimated SINR, by observing the outcome of each decision. The Q-learning method can be used to regulate the scheduler decisions in UL LA. The proposed Q-learning system is adaptable to 5G networks and at mmWave environment.

After running the simulation of the implemented system on 20 different seeds, the adjusted MCS values resulted in an improvement in both aspects of BLER and the throughput. By comparing the results with the currently used procedure, that depends on look-up tables, the ML system increased the total average throughput by 1.91 Mbps, while also decreasing the total average BLER from 8.36% to 8.12%, which represents a decrement of 2.89% of BLER value. This result reflects the improvement of applying the ML algorithm for only one user, and therefore when the implemented system is used for multiple users in the network, the total network throughput is expected to gain considerable increasement.

6.2 Future Work

In future work, key system parameters can be optimized to reach better performance, such as: the MCS correction margin (K), Learning rate (α), and epsilon. These values can be adaptable depending on the environment, and changing automatically as well, using optimality theories, to provide the optimal performance. Also, other users can be introduced to the network to evaluate the total performance of the network and monitor the effect of increasing the number of users and therefore, the network interference. Moreover, in future work, more investigation is needed on adding new inputs to the Q-learning system, which might enhance the performance, such as: channel bandwidth, the tendency in MCS value, and CQI value. The idea of including CQI value is that it reflects the quality of the DL channel, and therefore, might have relevant information regarding the UL channel, even while being in opposite locations, since the UE and gNB normally has the Tx and Rx antennas implemented in close positions relatively to the transmission distance.

References

[1] Agrawal, R. (2018). Machine Learning for 5G RAN. Algorithm Innovations, Mobile Networks ATF, Tech. Rep, 8.

[2] Series, M. (2015). IMT Vision–Framework and overall objectives of the future development of IMT for 2020 and beyond. Recommendation ITU, 2083, 0.

[3] Alejos, A. V., Sanchez, M. G., & Cuinas, I. (2008). Measurement and analysis of propagation mechanisms at 40 GHz: Viability of site shielding forced by obstacles. IEEE Transactions on Vehicular Technology, 57(6), 3369-3380.

[4] Facebook, Inc. (Nasdaq: FB). (2019). Facebook Reports First Quarter 2019 Results. California.

[5] Protalinski, E. (2016, November 2). Facebook Passes 1 Billion Mobile-Only Monthly Users. VentureBeat. https://venturebeat.com/social/facebook-passes-1-billion-mobile-only-monthly-users/.

[6] Ericsson, A. B. (2019). Ericsson mobility report 5G switched on. Ericsson, Sweden.

[7] Parkvall, S., Dahlman, E., Furuskar, A., & Frenne, M. (2017). NR: The new 5G radio access technology. IEEE Communications Standards Magazine, 1(4), 24-30.

[8] Rusek, F., Persson, D., Lau, B. K., Larsson, E. G., Marzetta, T. L., Edfors, O., & Tufvesson, F. (2012). Scaling up MIMO: Opportunities and challenges with very large arrays. IEEE signal processing magazine, 30(1), 40-60.

[9] Brady, J., Behdad, N., & Sayeed, A. M. (2013). Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements. IEEE Transactions on Antennas and Propagation, 61(7), 3814-3827.

[10] Yadav, A., & Dobre, O. A. (2018). All technologies work together for good: A glance at future mobile networks. IEEE Wireless Communications, 25(4), 10-16.

[11] Federal Communications Commission. (1997). Millimeter wave propagation: spectrum management implications. Bulletin, 70, 1-24.

[12] Dahlman, E., Parkvall, S., & Skold, J. (2020). 5G NR: The next generation wireless access technology. Academic Press.

[13] 3GPP. (2018). New frequency range for NR (24.25-29.5 GHz). 3rd Generation Partnership Project (3GPP), Technical Report (TR) 38.815.

[14] Li, Y., Pateromichelakis, E., Vucic, N., Luo, J., Xu, W., & Caire, G. (2017). Radio resource management considerations for 5G millimeter wave backhaul and access networks. IEEE Communications Magazine, 55(6), 86-92.

[15] Polese, M., Giordani, M., Mezzavilla, M., Rangan, S., & Zorzi, M. (2017). Improved handover through dual connectivity in 5G mmWave mobile networks. IEEE Journal on Selected Areas in Communications, 35(9), 2069-2084.

[16] Getzmann, S., Jasny, J., & Falkenstein, M. (2017). Switching of auditory attention in “cocktail-party” listening: ERP evidence of cueing effects in younger and older adults. Brain and cognition, 111, 1-12.

[17] Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. Acta Acustica united with Acustica, 86(1), 117-128.

[18] Höst, S. (2017). Information Theory and Communication Engineering, compendium. Department of Electrical and Information Technology, Faculty of Engineering, LTH, Lund University.[17] http://www. oberhumer. com/opensource/lzo.

[19] Shannon, C. E. (1951). The redundancy of English. In Cybernetics; Transactions of the 7th Conference, New York: Josiah Macy, Jr. Foundation (pp. 248-272).

[20] Dahlman, E., Parkvall, S., Skold, J., & Beming, P. (2010). 3G evolution: HSPA and LTE for mobile broadband. Academic press.

[21] Proakis, J. G. (1998). Digital communications fourth edition, 2001. McGraw-Hill Companies, Inc., New York, NY.

[22] 3GPP. (2019). Technical Specification Group Radio Access Network; NR; Physical layer procedures for data (Release 15). 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 38.214.

[23] Dahlman, E., Parkvall, S., & Skold, J. (2016). 4G, LTE-advanced Pro and the Road to 5G. Academic Press.

[24] Awad, M., & Khanna, R. (2015). Efficient learning machines: theories, concepts, and applications for engineers and system designers (p. 268). Springer nature.

[25] Xu, G., & Lu, Y. (2006, September). Channel and modulation selection based on support vector machines for cognitive radio. In 2006 International Conference on Wireless Communications, Networking and Mobile Computing (pp. 1-4). IEEE.

[26] Bruno, R., Masaracchia, A., & Passarella, A. (2014, September). Robust adaptive modulation and coding (AMC) selection in LTE systems using reinforcement learning. In 2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall) (pp. 1-6). IEEE.

[27] Daniels, R. C., Caramanis, C. M., & Heath, R. W. (2009). Adaptation in convolutionally coded MIMO-OFDM wireless systems through supervised learning and SNR ordering. IEEE Transactions on vehicular Technology, 59(1), 114-126.

[28] Daniels, R., & Heath, R. W. (2010, April). Online adaptive modulation and coding with support vector machines. In 2010 European Wireless Conference (EW) (pp. 718-724). IEEE.

[29] Leite, J. P., de Carvalho, P. H. P., & Vieira, R. D. (2012, April). A flexible framework based on reinforcement learning for adaptive modulation and coding in OFDM wireless systems. In 2012 IEEE Wireless Communications and Networking Conference (WCNC) (pp. 809-814). IEEE.

[30] Pulliyakode, S. K., & Kalyani, S. (2017). Reinforcement learning techniques for outer loop link adaptation in 4G/5G systems. arXiv preprint arXiv:1708.00994.