T.C.
ISTANBUL OKAN UNIVERSITY
INSTITUTE OF GRADUATE SCIENCES

THESIS

FOR THE DEGREE OF

MASTER OF SCIENCE

IN ADVANCED ELECTRONICS AND

COMMUNICATION TECHNOLOGIES PROGRAM

Wameedh Naseer Taha AL-BDOOR

REINFORCEMENT LEARNING AND Q LEARNING FOR RESOURCE ALLOCATION IN MIMO NETWORK WITH INTELLIGENT REFLECTIVE SURFACES

ADVISOR

Dr. Öğr. Üyesi Didem KIVANÇ TÜRELİ

ISTANBUL, January 2024

T.C.
ISTANBUL OKAN UNIVERSITY
INSTITUTE OF GRADUATE SCIENCES

THESIS

FOR THE DEGREE OF

MASTER OF SCIENCE

IN ADVANCED ELECTRONICS AND

COMMUNICATION TECHNOLOGIES PROGRAM

Wameedh Naseer Taha AL-BDOOR
(203006006)

REINFORCEMENT LEARNING AND Q LEARNING FOR RESOURCE ALLOCATION IN MIMO NETWORK WITH INTELLIGENT REFLECTIVE SURFACES

ADVISOR

Dr. Öğr. Üyesi Didem KIVANÇ TÜRELİ

ISTANBUL, January 2024

T.C.
ISTANBUL OKAN UNIVERSITY
INSTITUTE OF GRADUATE SCIENCES

THESIS FOR THE DEGREE OF MASTER OF SCIENCE
IN ADVANCED ELECTRONICS AND COMMUNICATION
TECHNOLOGIES PROGRAM

Wameedh Naseer Taha AL-BDOOR
(203006006)

REINFORCEMENT LEARNING AND Q LEARNING FOR RESOURCE ALLOCATION IN MIMO NETWORK WITH INTELLIGENT REFLECTIVE SURFACES

Date Thesis Delivered to Institute : January 24, 2024

Date of Thesis Defense : January 24, 2024

Thesis Advisor: Dr. Öğr. Üy. Didem KIVANÇ TÜRELİ ________________

Jury Members: Doç. Dr. Ömer Cihan KIVANÇ________________________

Prof. Dr. Mehmet Serdar Ufuk TÜRELİ ________________

ISTANBUL, January 2024

ABSTRACT

REINFORCEMENT LEARNING AND Q LEARNING FOR RESOURCE ALLOCATION IN MIMO NETWORK WITH INTELLIGENT REFLECTIVE SURFACES

In the last decade, modern wireless communication system is widely developed to enhance the channel performance and overcome the issue of fading. However, with new infrastructures of modern cites, the issue of multipath reflections from buildings and obstacles has become a significant problem that reduce the quality of the signal and the bit error rate. In such circumstances, many paths to the signal between the user and the node could be generated that limit the accuracy of modulation/ de-modulation process. Therefore, it is a subject of this research to invoke a hybrid technique based on combing two intelligent algorithms of R-learning and Q-learning processes. Such technique could be applicable to modern wireless communication systems based on IRS and MIMO networks. In this matter the use of R-learning starts the action to determine wither is desired according the criteria of signal quality or not. The turn of the Q-learning comes to localize the user with best action through generating a criterion based on minimum BER response. It is observed from such combination, a significant reduction is achieved in the BER of the received signal that almost realizes an encloses resonance to the ideal case without noise or fading effects. Later, an analytical study is introduced by increasing the number of IRS elements as hardware solution to be compared with the achieved results from the previous solution. It is found a significant enhancement is accomplished in the signal quality with increasing the number of IRS elements, however, the system complexity is increased rapidly. This

gives an indication on how much such algorithms is an effective solution to maintain minimum hardware complexity with low cost.

Keywords: Q learning, for resource allocation, MIMO, intelligent reflective surfaces (IRS).

KISA ÖZET

AKILLI YANSITICI YÜZEYLERE SAHİP MIMO AĞINDA KAYNAK TAHSİSİ İÇİN GÜÇLENDİRME ÖĞRENME VE Q ÖĞRENME

Son on yılda, modern kablosuz iletişim sistemi, kanal performansını artırmak ve solma sorununu aşmak için yaygın olarak geliştirildi. Ancak modern şehirlerin yeni altyapılarıyla birlikte binalardan ve engellerden gelen çok yollu yansımalar sorunu, sinyal kalitesini ve bit hata oranını düşüren önemli bir sorun haline gelmiştir. Bu gibi durumlarda, kullanıcı ile düğüm arasındaki sinyale giden, modülasyon/de-modülasyon işleminin doğruluğunu sınırlayan birçok yol oluşturulabilir. Bu nedenle, R-öğrenme ve Q-öğrenme süreçlerinin iki akıllı algoritmasının birleştirilmesine dayanan hibrit bir tekniğin kullanılması bu araştırmanın konusudur. Bu teknik, IRS ve MIMO ağlarına dayanan modern kablosuz iletişim sistemlerine uygulanabilir. Bu konuda R-öğrenmenin kullanılması, sinyal kalitesi kriterlerine göre soldurmanın istenip istenmediğini belirleme işlemini başlatır. Q-öğrenmenin sırası, minimum BER yanıtına dayalı bir kriter oluşturarak kullanıcıyı en iyi eylemle yerelleştirmeye gelir. Bu tür bir kombinasyondan, alınan sinyalin BER'inde, gürültü veya sönümleme etkileri olmadan neredeyse ideal duruma yakın bir rezonans gerçekleştiren önemli bir azalma elde edildiği gözlemlenmektedir. Daha sonra donanım çözümü olarak IRS elemanlarının sayısı arttırılarak önceki çözümden elde edilen sonuçlarla karşılaştırılacak analitik bir çalışma ortaya konmuştur. IRS elemanlarının sayısının artmasıyla sinyal kalitesinde önemli bir iyileşme sağlandığı ancak sistem

karmaşıklığının hızla arttığı bulunmuştur. Bu, bu tür algoritmaların minimum donanım karmaşıklığını düşük maliyetle sürdürmek için ne kadar etkili bir çözüm olduğunun bir göstergesidir.

Anahtar Kelimeler: Q öğrenme, kaynak tahsisi için, MIMO, akıllı yansıtıcı yüzeyler (IRS).

To My Family

ACKNOWLEDGMENT

I would like to extend my sincere thanks and gratitude to my supervisor, without whose advice and assistance at all times, I would not have achieved this thesis.

Also, I want to thank my father and mother for their sacrifices, their constant giving, and their support for me all the time and finally to my daughter Eileen which is the best think in my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SYMBOLS

d_1 distance between BS and the IRS

d_2 distance between IRS and the USER

L_m the product distance

α the path loss exponent

t_1 is the fading parameter

$f(x)$ Probability density function of the elements

F phase-shift values at each element

y_k received signal at $user_m$

$\boldsymbol{\Phi}$ the reflection-coefficient matrix

$\boldsymbol{\phi}_L$ the phase shift of IRS element

Z_K independent and identically distributed additive white Gaussian noise

β here is the amplitude coefficient of IRS elements

W_k transmit precoding vector

S_K the information-bearing symbols of users

P transmit power in the base station

ABBREVIATIONS

MIMO	multiple input multiple output
IRS	intelligent reflective surfaces
BS	base station
RL	Reinforcement learning
EE	energy efficiency
SE	spectral efficiency
QoS	quality of service
CSI	channel state information
LoS	line of sight
FBGA	field-programmable gate array
IoE	Internet of Everything
LIS	large intelligent surface
SDM	software-controlled meta surfaces
SDS	software-defined surfaces
PIS	passive intelligent mirrors
SINR	signal-to-interference-plus-noise ratio
NOMA	non-orthogonal multiple access
AWGN	additive white Gaussian noise
i.id.	independent and identically distributed

I. INTRODUCTION

1.1. Overview

One of today's technologies that is quickly evolving and expanding is wireless networking, with spectacular new services and goods appearing practically daily. As the need for additional wireless bandwidth develops rapidly, these advancements provide huge hurdles for communications engineers. In fact, designers face a number of problems in the field of wireless communications. this is because the physical medium is hard to work with and the network dynamics are complicated. the most important technical difficulty in wireless communications network is multipath induced fading, which refers to random variations in channel gain caused by scattering of transmitted signals from objects between the transmitter and the receiver. As a result, multipath scattering is often seen as a detriment to wireless communication. However, it may now be seen as a chance to dramatically increase the capacity and dependability of such systems[1] , many advanced wireless technologies, such as intelligent reflective surfaces (IRS). and multiple input multiple output (MIMO) systems, suggested to use in next-generation Wireless networks In accordance with the new radio standard for the fifth generation (5G), whenever the frequency is increased above 6GHz, Because the high-frequency signals with many are very sensitive effects that may block the signal from reaching in high resolution such as trees and buildings [2], but with the help of a reflective surface that can be included inside the infrastructure of the city, for example, it will be very effective and also at a lower cost[3] , For an IRS to work, the incident signals must pass through reflective elements, each of which might alter the incident signals' phase shifts and perhaps their amplitude adjusting the global channel state information allows for either the constructive or destructive superposition of the user received signals and base station (BS) by aligning the signals reflected by the IRS [4].

1.2. Reinforcement learning (RL) and Q-learning

If we look at the nature of learning, the concept that we pick up new information through interacting with the world around us is usually the first thing that comes to mind for most of us. An infant does not have a teacher in the traditional sense while it is playing, waving its arms, or looking about, but it does have a direct sensorimotor link to the world around it. Utilizing this relationship yields a plethora of knowledge about cause and effect, regarding the repercussions of acts, and what to do in order to accomplish one's objectives. These kinds of encounters throughout our lives are, without question, one of the most important sources of information about both ourselves and the world around us. Whether we are learning how to drive a car or how to have a conversation, we are acutely aware of how our environment reacts to what we do, and we strive to influence what occurs through the behaviors that we exhibit. This is true whether we are learning how to hold a conversation or how to drive a car. The concept that humans acquire knowledge primarily via social contact serves as a cornerstone for almost all educational and intellectual frameworks.[5]

Reinforcement learning (RL) has been getting a lot of attention lately. It has been used successfully in so many fields, like operations research, game theory, simulation-based optimization, information theory, control theory, and statistics. (RL) is a type of machine-learning which is become an important tool in machine learning field. This is a strategy in which devices make their own decisions in a given environment without knowing anything about the past or being told what to do [6] Artificial intelligence will improve capabilities in all areas and will give better and faster solutions. Enhanced education will continue to evolve and introduce new variables [7]. As Q-Learning progresses, is the best score for each action in each state, The procedure begins with no information, hence a blank value should be entered into the database automatically. Since this is the case, the agent gets to know the environment in which he wants to work through several random actions and the reactions he gets. If the action is good, he gets a reward, and if it is not good, he gets a penalty, based on the experience or information obtained from previous procedures and the state of the environment, the next action will be taken. it's very similar to how human behavior, we shall choose anything at random and see what happens. Yes, that is the way Q-Learning works.

Based on the knowledge gained from decisions or experiences that will be chosen randomly, appropriate decisions will be made, as we are human beings, our experience increases based on the experiences that pass us, and our decisions are based on previous experience more than we rely on fate or luck [8].

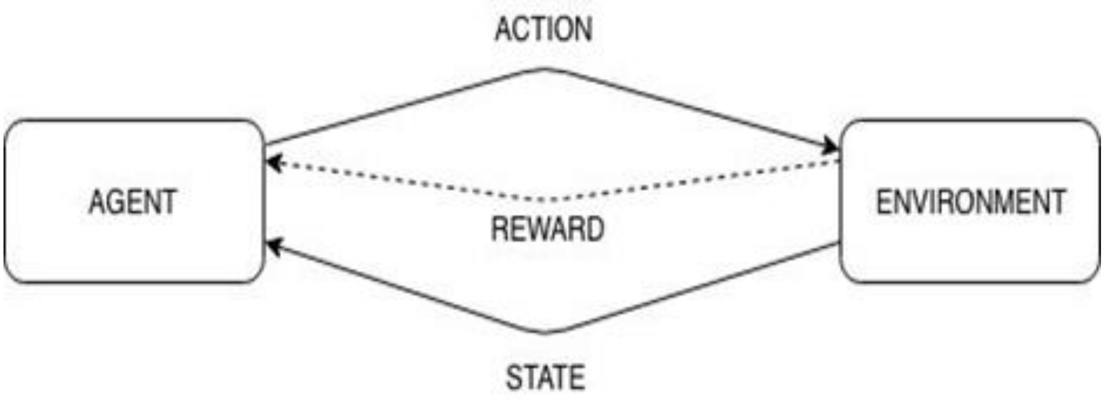

Figure I.1. How the Q-learning work in environment [7].

1.3. Multiple input multiple output (MIMO)

The usage of wireless communication applications and devices has expanded significantly during last decades. The widespread use of devices put high requirements on current wireless network infrastructure in terms of coverage, resilience and capacity in order to fulfill the fast-expanding demand for mobile services.

Presently, wireless network enhancements need the aggressive installation of numerous access points in order to raise energy efficiency (EE), spectral efficiency (SE) [9]. To reach great successes in power, hardware efficiency and spectral in wireless communication systems, the deployment of (MIMO) technology present huge potential in terms as a leading candidate. Towards this objective, suggestions for increasing spectral efficiency through cellular networks and accompanying examples are presented in [10], Important topics including spatial, channel estimation, signal processing spatial, spatial resource allocation and power are all discussed in depth throughout the book. Massive MIMO's potential for boosting both spectrum and energy efficiency was also highlighted. Multiple-input Multiple-output (MIMO) antennas have also been offered as a potential solution by multiple studies. This is due to the fact that MIMO technology shows promise in attaining the needed extremely high network dependability and high data rates, hence improving the quality of service (QoS) for mobile users and allowing them to continue using their massive data-hungry apps [11].

iMIMO is a wireless method that makes use of multiplexing in order to expand wireless bandwidth and increase wireless range. MIMO is able to transmit information over two or more antennas by using the algorithms present in a radio chip. When traveling from an access point to a wireless card, a wireless signal will cause additional reflections of the original signal as it is reflected off of walls, ceilings, and any other obstacles along its path. Also, this means that the original signal will get to its destination through a number of different paths [12]. The signals arrive at varying moments because they traveled different distances rather than the same distance, and interference will occur if we use only one antenna so we use the MIMO multi-antenna technology, and we can improve the signal by having multiple copies of them. There are many signals that we can choose from. MIMO also allows access to larger and better ranges, so the signal strength is much faster and better [13].

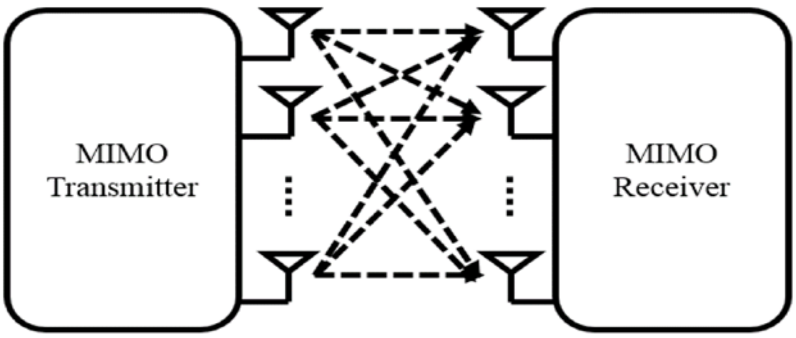

Figure I.2. MIMO [12].

MIMO is a big success for two main reasons: the first is the increase in data rate that was achieved through spatial multiplexing, and the second is the improvement in reliability that was achieved by exploiting the diversity in the channel. Both of these factors contributed to MIMO's rise to prominence [14].

1.4. Intelligent Reflecting Surface (IRS)

[IRS] is a modern hardware technology that has recently emerged and gotten a lot of attention in recent years with the goals of expanding signal range, decreasing power consumption, and decreasing the overall cost of deployment [15], IRS is Two-dimensional surface Composed of a large number of sub wavelength reflecting components (small antennas), each of which is coupled to a tunable chip (PIN diode or varactor, for example) to adjust its load impedance. Depending on whether the PIN diodes are on or off, the resulting change in load impedance causes a phase shift of [16], Each element may be controlled to change its reflective properties, such as its angle of reflection and transmissivity. When using a mirror to reflect light, the angle of incidence is also the angle of reflection. The angle of reflection may be adjusted arbitrarily, allowing RISs to redirect the reflected radio signal in any desired direction [17], By making the necessary adjustments to the global channel state information (CSI) in the correct way, the received signal of both the base station (BS) and of the users can be either constructively or destructively overlaid [4].

In Figure I.3 the main use of the IRS as it is an obstacle between the base station (BS) and the user, in this situation, having an IRS with clear connections to both the user and the base station (BS) aids in avoiding the obstruction through clever signal reflection, resulting in the establishment of a virtual line of sight (LoS) link between them. This really is extremely beneficial for the coverage expansion of mmWave communications, as they are more sensitive to being blocked [18].

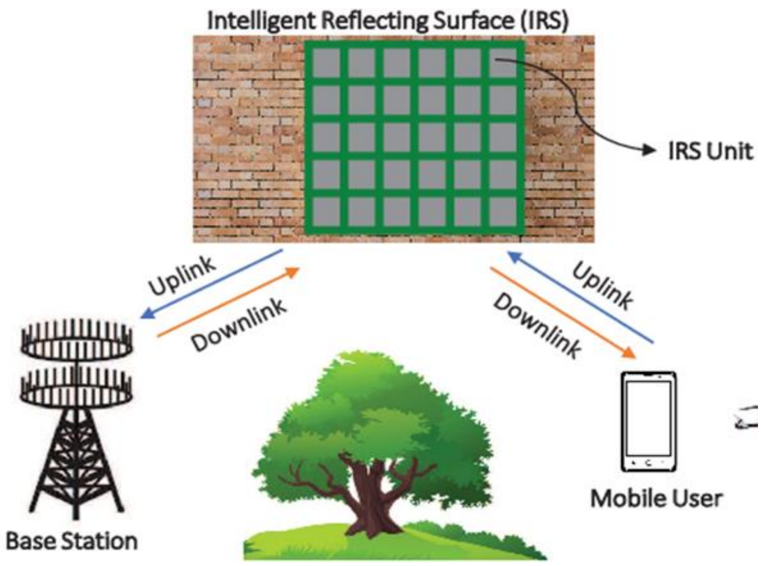

Figure I.3. An intelligent reflecting surface is used to improve communication between a base station and a user. [18].

1.4.1. IRS architecture

A typical IRS architecture consists of three different layers, all of them link up to smart controller as shown in Figure I.4.

The first layer of the device is made up of a substantial number of reprogrammable metal patches that have been printed on a substrate in order to intelligently regulate the incident signal [19], Copper is used for the second layer to decrease energy loss during the reflection phase, In the third layer, there is a Control circuit board that keeps changing the phase shift and amplitude of each element's reflection when told to do so by a smart controller that is linked to the IRS. In practice, the controller can be made out of a field-programmable gate array (FPGA) [20]

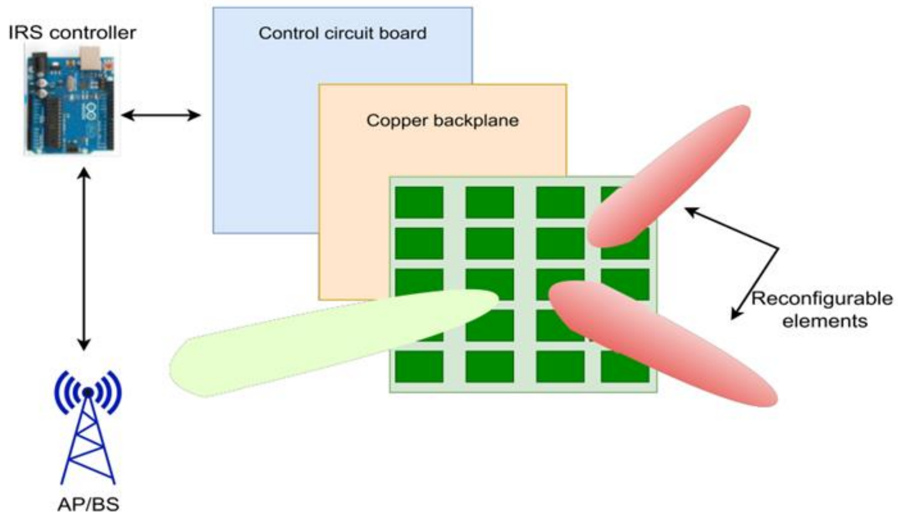

Figure I.4. Architecture of IRS [21].

1.4.2. IRS benefits

By properly changing the reflection coefficient, the IRS is projected to do a major role in meeting the needs of energy efficiency and spectrum efficiency to the5G and beyond wireless networks (i.e., phase shifts). To produce full duplex communication, an IRS doesn't need radio frequency (RF) chains., which enables it to fulfill the function of passive signal reflection. Also, an IRS can change the propagation environment by rearranging its reflection components. This makes the signal beamed toward the receiver, which helps the signal of interest and reduces interference from other users (MU interference).

The meaning of a smart radio environment, the idea that existing radio signals can be reused through the design and deployment of an IRS, and the possibility of using an IRS in future wireless networks (FWNs) are developed in.

And we can list some of IRS benefits.

It is a two-dimensional surface made of negative elements, and these elements have a low cost, which gives freedom to place a number of elements and they are combined into one metasurface, allowing for simple deployment on walls, buildings, underground tunnels and ceilings with a clear (LoS) to the base station, as well as long-term, sustainable operation (BS). And since it has no RF links, an IRS has very low power requirements.

Higher data speeds need a more densely deployed 5G network, which is necessitated by the short transmission range of mmWave bands. However, the SINR (signal to interference + noise) and throughput (throughput divided by interference) both suffer from an increase in interference brought on by a dense BS deployment.

In this case, an IRS is invaluable because to its ability to boost system capacity with no additional expense by increasing signal strength and decreasing interference strength at the receiver through intelligent beamforming.

Passive beamforming allows for flexible reconfiguration by concurrently optimizing the phase shift of all scattering elements. By using the abundant reflecting surfaces, the incident signal may be focused on the desired receiver while being cancelled out in all other directions, boosting the wireless network's throughput.

Decreased cellular edge loss Signal strength drops down and interference levels rise for users towards the cell's periphery. In this scenario as well, an IRS may enhance the overall signal quality for the cell-edge users by reducing interference. Spreading components in MU wireless networks may spread the signal and help with data. As a result, an IRS boosts the performance of the sum-rate and provides higher quality of service with less power.

In order to support developing technologies, like virtual reality, holographic communication, and other Internet of think applications, an IRS will be a crucial component in the near future. This will allow these technologies to meet their extremely high data rate requirements [21].

1.5. Background

The significance of wireless communications in providing worldwide Internet access is crucial. There has been remarkable technological development in the satellite sector during the past 60 years. Since the size and power of satellites have increased, Satellite communications has expanded beyond its original purpose of data transmission. Television and radio transmission, high-speed Internet access, maritime and other mobile applications, and the backbone of military communications are all areas where Satellite communications has proven its worth [14]. Mobile communications in today's world include a wider scope than the simple exchange of untethered audio messages between two individuals. It encompasses a wide range of services, such as ubiquitous access to multimedia communications, mobile edge computing for the Internet of Everything (IoE), wireless power transmission, and many more besides. Additionally, it is anticipated that cellular vehicle-to-everything in fifth generation (5G) would change the automobile sector for many years to come. Regardless of the path that mobile network evolution has taken, there has always been a focus on improving both spectral efficiency and energy efficiency. The next generation of wireless technology, known as 6G, will attempt to enable a wide variety of novel applications and more advanced key performance indicators [22].

reinforcement learning history may be broken down into two primary strands, each of which is a lengthy and fruitful journey that was taken in isolation before being intertwined in contemporary reinforcement learning. One line of inquiry focuses on the process of learning by trial and error, which originated in the study of the psychology of animal learning. Reinforcement learning saw a renaissance in the early 1980s as a direct result of this line of inquiry, which goes through some of the first work in artificial intelligence. The second line of inquiry focuses on the issue of optimum control and how value functions and dynamic programming may be used to find a solution to the problem. This conversation didn't really entail any sort of education for the most part. The temporal-difference approaches, such like as used in the tic-tac-toe examples, are the subject of a third, less distinguishable thread. Although the two threads have remained essentially independent of one another, the exceptions center around this third thread. The present area of reinforcement learning was established in the late 1980s as a result of the convergence of these three strands. In the late 1950s, the word "optimal control" was used to characterize the challenge of developing a regulator that minimizes some performance metric for a dynamical system. In the 1950s, Richard Bellman and colleagues discovered a solution to this issue by building on a theory first proposed by Hamilton and Jacobi in the 1800s. The state of the dynamical system and the value function, also known as the "optimal return function," are used in this method to establish a functional equation, which is now commonly known as the Bellman equation. This equation led to the development many of techniques known as dynamic programming, which may be used to resolve optimum control issues (Bellman, 1957a). “Markovian decision processes (MDPs) are a stochastic discrete variant of the optimum control problem that were first proposed by Bellman (1957b), and the policy iteration approach for MDPs was developed by Ronald Howard (1960). Each of them is crucial to the theory and algorithms that power contemporary reinforcement learning. Dynamic programming is the main approach to solve broad stochastic optimum control problems. It suffers from "the curse of dimensionality," meaning its processing needs expand exponentially with the number of state variables, yet it's still more efficient and extensively applicable than any other generic approach. Since the late 1950s, dynamic programming has now been expanded to include partially observable MDPs (Lovejoy, 1991), various applications (White, 1985, 1988, 1993), approximation techniques (Rust, 1996), and asynchronous methods (Bertsekas, 1982, 1983). Bertsekas, 2005, 2012; Puterman, 1994; Ross, 1983; and Whittle, 1982, 1983) offer good recent approaches of dynamic programming. Bryson (1996) explains optimum control [23]."

also, MIMO has an intriguing history. Originally employed in electric circuit and filter theory in the 1950s, it's currently utilized to define communication approaches in this book. "MIMO" originally meant circuits with multiple input and output ports. Information theorists and communication system researchers developed this phrase in the 1990s to refer to innovative signal processing techniques for multi-antenna communication systems. In this modern definition, the communications channel is the reference point, and multiple input refers to signals from various transmit antennas "entering" or "being input" to the channel. Multiple output refers to signals coming from multiple receiver antennas, or "exiting" the channel. In 1999, Peter Driessen and Gerry Foschini released a study analyzing the theoretical communications capacity of a system having numerous transmit and receive antennas [24].

As the central element of a fully intelligent and collaborative wireless environment, reconfigurable MTSs have emerged as a crucially important enabling technology for the next generation of sensing and communications. This environment is now often referred to as an "intelligent reflecting surface" (IRS)[15]. MTSs have had and will continue to have a strong influence on antenna and microwave application technologies. Between the years 2000 and 2010, MTSs were virtually the same all across the area they occupied and were made up of printed components arranged in periodic patterns. The MTSs of this generation were the first ever produced. During that time period, a number of novel uses for microwaves and antennas came into being. MTSs were redesigned for the second generation (2010–2020) with the intention of altering the boundary conditions (BCs) in space in order to manage the field that was initiated by an in-plane or external feed. Today, we are on the cusp of a transition to a third generation, which will leverage the achievements made during the second generation and will include MTSs changing their BCs in space and time while also becoming intelligent and controlled. Utilizing electronics, time-changing materials, or several switchable feed points dispersed throughout the MTS are all viable options for accomplishing reconfigurability of the MTS[25]. The ability to control each individual element individually gave birth to the idea of "digital-coding MTSs," which later developed into "programmable MTSs" following the introduction of a field-programmable gate array to control various MTS capabilities depending on various digital states of the elements. This idea may also be applied to self-adaptive and cognitive MTSs as a further extension. Because of its reconfigurability, the MTS is capable of supporting huge MIMO and is able to redirect signals in nonspecific directions. “Because of this, reconfigurable MTSs have emerged as an essential component of a fully intelligent and cooperative wireless environment, making them an essential enabling technology for the next generation of sensing and communications systems.”

Moreover, the intelligent reflecting surface (IRS) has garnered an increasing amount of interest over the past few years. IRS is an artificial structure that is two-dimensional and contains a large number of passive reflective elements. The electromagnetic properties of these elements, such as scattering, reflection, and refraction, can be controlled electronically and independently in real-time by applying a variety of control signals. As a consequence of this, the phase and amplitude of electromagnetic waves that are impinged upon may be rearranged and reflected in a manner that is software-defined. In addition, the path that the reflected signal takes may be carefully directed to the receiver of one's choice, which enables a radio environment that can be totally programmed. This never-before-seen capacity to program the radio spectrum has enormous possibilities for wireless communications [26].

1.6. Aims of research

We seek in this research to deliver the signal from the main station to the user with higher efficiency and lower cost and without any problems in it, using many modern methods and techniques, and these techniques are

1. The agent chooses the correct procedure by using Reinforcement learning (RL) and Q-learning, which is work Based on the experience or

information obtained from previous procedures and the state of the environment,

2. Intelligent reflecting surface (IRS) It works to reverse the signal reception from the main station (BS) and reflect it in the direction of the USER through the elements that make up the reflective surface.

II. LITERATURE REVIVEW

In this part, we present a summary of the literature on IRS-assisted telecommunications, and we will also provide a summary of the MIMO literature and resource allocation and how it has been studied and developed throughout history.

IRS is the most recent change in the field of wireless communication. works of literature, they called a "large intelligent surface."(LIS), software-controlled meta surfaces (SDM), software-defined surfaces (SDS), reconfigurable intelligent surfaces, passive intelligent surfaces and passive intelligent mirrors (PIS) [27].

Figure 2 shows a quick overview of the IRS historical point of view. Based on “frequency selective surfaces (FSS), intelligent walls (IW) can change the way electromagnetic (EM) waves move, how they interact with their surroundings, and how well they work by switching”. After that, 2D metamaterials were made to work at different frequencies as an alternative to FSS [28], Di Renzo et al.. [28]the authors talked about the different ways 2D metasurfaces can be used. They also talked about how wave guides can be used to trap and guide EM energy between two metasurfaces, and how terahertz devices can be used to control metasurfaces so that they work better at terahertz frequencies. The benefits and uses of 2D metasurfaces in wireless communication were also discussed. Compared to 3D surfaces, 2D metasurfaces take up less physical space and have lower loss, and they can work at different frequency bands. Spatial microwave modulators (SMM) based on tunable metasurfaces were made by putting the SMM on the walls to boost the power or range of the transmission signal. More recently, coding metamaterials (MM) with the ability to change the EM properties by replacing the phases 0 and p with a binary 0 and 1 was introduced”. “This helped even more with the implementation of the software program described in. Also, in and references therein, programmable metasurfaces based on a PIN diode were introduced. In 2016, passive elements were added to reconfigurable reflect arrays, and a large intelligent surface was suggested as a way to go beyond mMIMO.

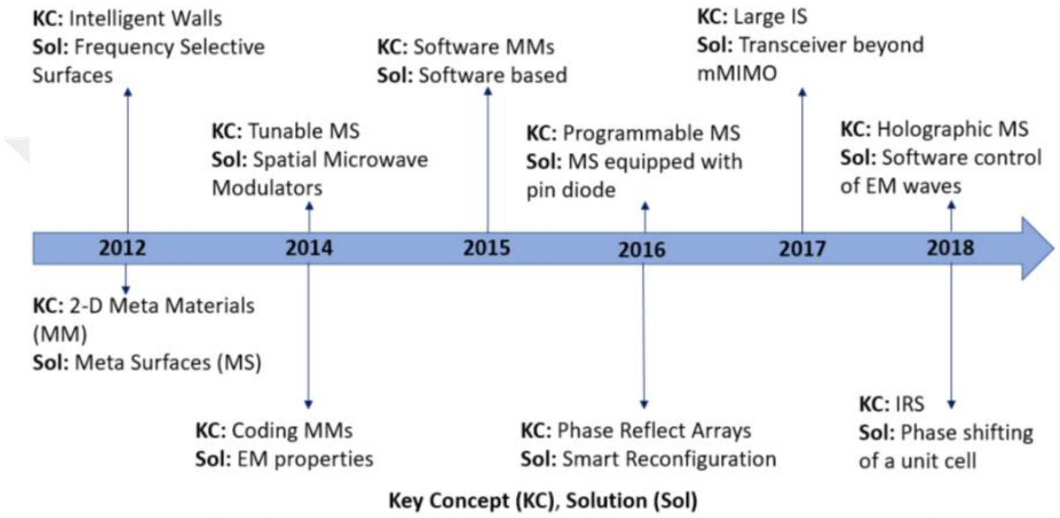

Figure II.1. Multicarrier systems.

Zhang et al. [29]proposed that the IRS should be run by software. Each IRS element's main job is to individually control the signal's amplitude, polarization and phase frequency. This makes it possible to make a very focused beam. It has already been shown that very direct and limited-gain beamforming improves signal quality or can be used to get rid of interference. This additional control in the environment gives a higher level of freedom that will improve performance across the network. Additionally, the IRS offers a number of benefits including reduced energy usage and cost-effectiveness.

Joseph M. Jornet et al. [30] proved that an IRS may reflect the incident signal toward the receiver in order to construct a stable virtual connection between transceivers, hence overcoming the NLoS problems in mm Wave communication. Two kinds of performance measures may be used to quantify the advantages of utilizing an IRS as a signal reflector: probabilistic metric and ergodic metric.

The development of MIMO technology has made a big difference in how well networks use energy and spectrum. The environment can affect how well a communication system works. Signal blockage, different frequency effects, and a complicated system all lead to a higher energy demand. This means that the network needs to use little power and not be too complicated. To solve these problems, the IRS becomes a beacon of hope by being the first to meet the needs of the next generation of communications systems. Using IRS and MIMO together will improve spectral efficiency with little extra work and can be used for many things, such as IRS-assisted MIMO communication, multicell communication, SWIPT MIMO, the Cognitive Radio system and point to point communication [31].

Bera et al. [27] offered a comprehensive evaluation of the intelligent reflecting surface (IRS) aided (MIMO) system They demonstrated that the network's performance is enhanced by the use of IRS and by combining MIMO with intelligent reflecting surfaces (IRS) is seen as a front-runner for use in 5G and 6G networks and beyond. They went through the fundamental ideas and provided the important problems that needed to be solved, such as IRS positioning strategies, channel estimation protocols, algorithms and applications.

X. Zhang et al. [32] looked at the possibility of wireless signal transmission in difficult environments. Within a MIMO network, the spectral efficiency as well as the power control were both optimized. In a MIMO network, they have taken into consideration both the uplink and the downlink transmission. The ensemble approach is utilized to improve power regulation and spectral efficiency, and the performance of this method is compared to the gradient descent method, the genetic algorithm and the Newton method, and. The ensemble technique is an application of artificial intelligence, which is use a neighborhood field optimization technique, also they showed that the suggested algorithm considerably increased the spectral efficiency in the multiple input multiple output (MIMO) network in comparison to traditional techniques by comparing the results to benchmark schemes and demonstrating the improvement.

Z. He et al. [33] present a practical channel estimation technique based on an actual codebook design restriction for enhancing the IRS reflecting elements in a MU-MIMO scenario. This procedure is designed to maximize the IRS's performance. To be more explicit, they provide two original methods for improving the reflective elements of a standard passive IRS using machine learning, as well as a safe channel estimation method for IRS. In addition to this, we discuss a technique for estimating the channel in an IRS.

When training the network, the deep supervised network employs exhaustive search to explore every possible reflection pattern, whereas the reinforcement network uses Q-learning to determine how to achieve the highest possible reward. The imperfect estimated channel knowledge can be utilized by our two methods in order to maximize the IRS in terms of the minimum rate or the sum rate across all users. Simulations show that the real-world algorithms can achieve sum rate and minimum rate performances that are similar to what theory says they should be able to do.

M. Jung et al. [34] describe a new algorithm for allocating resources that takes user scheduling and power control into account, this algorithm is designed with the passive beamforming and modulation approaches proposed in mind. Simulation results indicate that recommended strategies are near to their upper limits. when a lot of RIS-reflecting elements are present. This proves that the rate that can be reached in real RISs meets asymptotic optimality.

Y. Hu et al. [35] explored the design of a resource allocation algorithm for a system that uses inductive power transfer (IRS) to simultaneously send both information and energy wirelessly. They used an IRS design based on physical principles, which calculates the incidence and reflection angles. They solved the challenge of combined optimization of transmitted power and reflection coefficients of IRS elements by separating the IRS into a series of tiles and then designing the phase shift of each tile into multiple transmission modes. Using a combination of a penalty-based technique, sequential approximation, and a semi-definite relaxation algorithm, they finally minimized the transmitted power.

Navneet Garg et al. [36] In order to maximize the long-term energy efficiency of the multi-user systems, they consider into account per-antenna power allocation with a finite set of power levels, all the while satisfying the quality of service (QoS) constraints at the end users in terms of required signal-to-interference-plus-noise ratio (SINRs), which in turn depends on channel information. After modeling the constraint problem as an unconstraint problem under the assumption that channel states change according to a Markov process, the Q-learning technique is used to determine how much power should be allocated to each channel. It is shown through simulation findings that it is possible to minimize power usage while still meeting the SINR criterion at consumers' devices.

Lee rt al. [37] they take reinforcement learning (RL) into account with the MIMO-NOMA (non-orthogonal multiple access) system. Spectrum-efficient communication methods, such as NOMA, have been intensively explored for use in 5G wireless networks. With MIMO added to NOMA, spectral efficiency may be increased even more. Power allocation and user pairing are other crucial NOMA strategies. High computational complexity owing to dynamic radio channels is a major drawback of NOMA, though. This restriction hampers effective channel use and radio resource allocation. They provide a simultaneous user pairing and power distribution approach based on reinforcement learning to lessen the computational burden. Using Q-learning, we can minimize the computational complexity of user pairing and power allocation by doing both simultaneously. Simulation results (ES) show that the suggested method gets a sum rate that is about the same as that of the exhaustive search.

III. SYSTEM MODEL WITH METHODOLOGY AND ANALYSIS

In this paper we present the MIMO system, where the connection is made between the user and the base station (BS) by many antennas, where the user is in a dead zone where the signal is very weak and the user’s demands are not met correctly, so we use an intelligent reflective surface (IRS) to work to reflect the signal to the user with higher efficiency.

In the networks that have been suggested, a base station (BS) would make use of M transmitter antennas (TAs), while users would make use of K reception antennas (RAs). It is important to note that L IRS elements are co-located on an IRS array that is placed on the same building that is positioned in the center of the disc. This allows M users to be served simultaneously by L IRS components. It is possible to manipulate the electromagnetic signal in a way that is to one's advantage by adjusting the phase shifts and amplitude coefficients of the IRS components in a suitable manner. The system model of the proposed IRS-aided network is shown in Figure III.1.

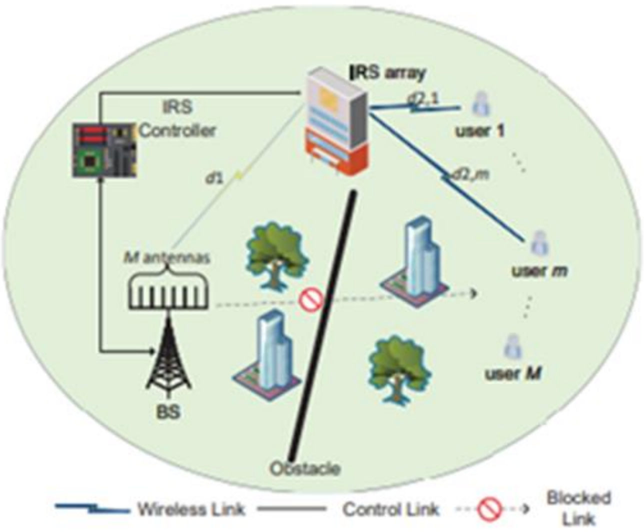

Figure III.1. Illustration of the process of sending a signal with the help of the IRS [37].

The IRS and BS are fixed and unchanged because they are infrastructure elements on the other side of the user, which is mobile and whose location is not fixed. The distance between BS and the IRS is referred to as d_1 and the distance between the IRS and USER is referred to as $d_{2,m}$. Most of the time, the distance is greater than 1, and the reason for this is to simplify the analytical results [38].

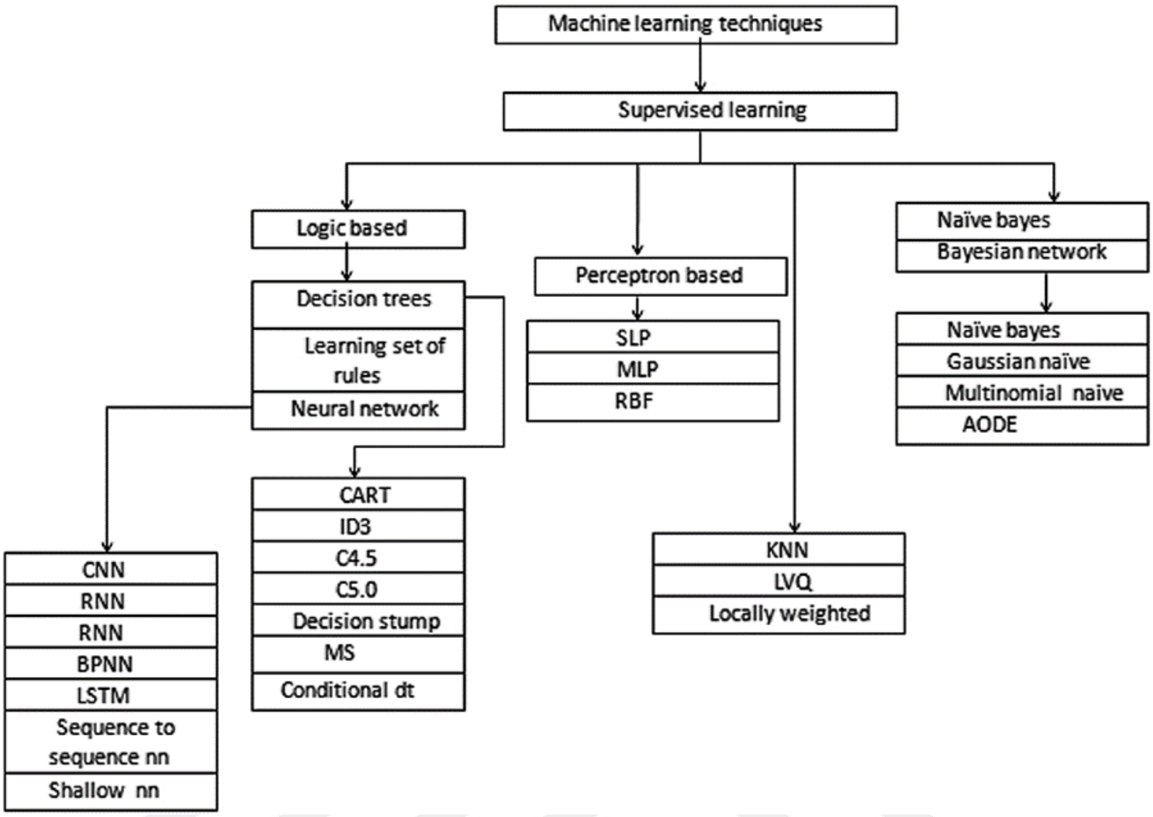

Figure III.2. Machine-learning algorithms classifications.

Reinforcement learning (RL) and Q-learning can be applied to resource allocation problems in a Multiple-Input Multiple-Output (MIMO) network with intelligent reflective surfaces (IRS). The goal of resource allocation in this context is to optimize the utilization of the available resources, such as the transmit power, the beamforming weights, and the phase shifts of the IRS elements, to maximize system performance. Here's an overview of how RL and Q-learning can be used for resource allocation in an MIMO network with intelligent reflective surfaces:

1. Define the problem: First, you need to define the specific resource allocation problem you want to solve. For example, it could be

maximizing the signal-to-interference-plus-noise ratio (SINR) at the receiver by optimizing the transmit power, beamforming weights, and phase shifts of the IRS elements.

2. State and action representation: Next, you need to define the state and action spaces. The state represents the current environment or system conditions, which could include information about the channel conditions, interference levels, and previous actions taken. The action represents the resource allocation decisions, such as the transmit power levels, beamforming weights, and phase shifts of the IRS elements.
3. Reward function: Design a reward function that reflects the performance objective of the system. For example, you can define the reward as the achieved SINR or the system throughput. The reward function guides the RL agent towards actions that improve system performance.
4. Q-learning algorithm: Q-learning is a model-free RL algorithm that learns an action-value function, called the Q-function, to estimate the expected cumulative reward for each state-action pair. The Q-function is typically represented as a lookup table or a neural network. The Q-learning algorithm iteratively updates the Q-values based on the observed rewards and the agent's experience.
5. Exploration and exploitation: During the learning process, the RL agent needs to balance exploration and exploitation. Exploration allows the agent to discover new, potentially better actions, while exploitation exploits the current knowledge to select actions with higher expected

rewards. Techniques like epsilon-greedy or softmax exploration policies can be used to control this trade-off.

6. Training: Train the RL agent using a simulator or by interacting with the real environment. The agent receives the current state, selects an action based on its current policy (e.g., using an epsilon-greedy strategy), and observes the reward and the next state. The agent then updates its Q-values based on the observed reward and the Q-learning update rule.
7. Testing and evaluation: Once the RL agent is trained, evaluate its performance on unseen scenarios. Measure key performance metrics such as SINR, throughput, or fairness to assess the effectiveness of the resource allocation strategy.
8. Fine-tuning and optimization: Depending on the results obtained, you may need to fine-tune the RL algorithm or adjust the reward function to further improve the resource allocation performance.

By applying RL and Q-learning to resource allocation in an MIMO network with intelligent reflective surfaces, you can develop intelligent algorithms that adaptively optimize the resource allocation decisions based on the current network conditions, leading to improved system performance and efficient resource utilization.

Considering that the horizontal distance between the IRS and the BS is much greater than the vertical distance and the length of the IRS matrix, so the vertical length and the length of the IRS will be ignored, also the user M will be randomly distributed to a thickness of radius R according to the binomial point process, and there will be attenuation caused by the path, which is subject to the product distance law and it can be expressed as follows:

$$L_m = (d_1 d_{2.m})^{-\alpha} \tag{III.1}$$

The α here mean the path loss exponent. The Nakagami fading channel theory is used as a matrix to connect the BS and the IRS and can be expressed as follows:

The G here mean K × L matrix It is a channel connecting the antennas in the BS and the element in the IRS. The Probability density function of the elements is given by:

$$f_1(x) = \frac{t_1^{t_1} x^{t_1 - 1}}{\Gamma(t_1)} e^{-t_1 x} \tag{III.2}$$

t_1 is the fading parameter. Also, the Nakagami fading channel theory will be applied to the channel that exists between the IRS elements and the user, and this theory may be expressed by:

$$H_m = \begin{bmatrix} h_{m,1,1} & \cdots & h_{m,1,L} \\ \vdots & \vdots & \vdots \\ h_{m,K,1} & \cdots & h_{m,K,L} \end{bmatrix} \tag{III.3}$$

H_{m9} here mean K x L which is the channel between IRS elements(L) and the antenna at user (K) in this channel it will be associated by fading parameter (t_2) [39]. In addition, we take into account the fact that there is a limited set of discrete values that may be used for the phase shift at each element of the IRS, which simplifies practical implementation. Let's call the number of bits (b) utilized to represent the number of bits (L_m) of phase shifts ($L_m = 2^b$). Assuming for the purpose of argument that such quantized phase-shift values are created by uniformly quantizing the interval [0, 2], we may focus on the latter assumption. The discrete phase-shift values at each element are given by:

$$F = \{0 \cdot \Delta\theta, \cdots, (L_m - 1)\Delta\theta\} \tag{III.4}$$

Where, the $\Delta\theta = \frac{2\pi}{L_m}$. The number of possible beam patterns is $(L_m)^N$ for an IRS with N elements, where each element has a phase shift level of L_m. After that the signal will be received at USER m and that can be given by:

$$y_k = (h_m + \Phi G) \sum_{j=1}^{k} \omega_j S_j + z_k \tag{III.5}$$

The matrix of reflection coefficients is a diagonal matrix The same as the commonly agreed beliefs, Where $\boldsymbol{\Phi}$ is the reflection-coefficient matrix which is $diag$ $[\beta_2\boldsymbol{\phi}_1, \beta_2\boldsymbol{\phi}_2, \ldots, \beta_L\boldsymbol{\phi}_2]$, β here is the amplitude coefficient of IRS elements which is Є (0, 1], we assumed that β = 1, and the $\boldsymbol{\phi}_L$ is the phase shift of IRS element which is = exp ($j\theta_L$), j = $\sqrt{-1}$, $\theta_L \in [0, 2\pi)$. We made the assumption that both the amplitude coefficients and the phase shifts of the IRS components are continuous, which allows for them to be controlled in an ideal manner. $\sqrt{L_m}$ is the path loss induced amplitude attenuation of user, Z_K represents independent and identically distributed additive white Gaussian noise (AWGN) with zero mean and variance σ_k^2 at the receiver of user K. We will analyze the conventional continuous linear precoding at the base station (BS), where $w_k \in C^{m\times1}$will stand in for the transmitted precoding vector for user K. The complex baseband signal that is sent from the base station (BS) may thus be written as:

$$x = \sum_{k=1}^{k} w_k s_k \tag{III.6}$$

where S_K which signify the information-bearing symbols of users. These symbols are modeled as independent and identically distributed (i.i.d.) random variables that have zero mean and unit variance [40]. As a result, the overall transmit power that was used in the base station (BS) may be calculated as.

$$P = \sum_{k=1}^{K} \|\omega_k\|^2 \tag{III.7}$$

Also, the SINR at the user k is given by:

$$\text{SINR}_k = \frac{\left|\left(h_m \Phi G\right)\omega_k\right|^2}{\sum_{j \neq k}^{k} \left|(h_m \Phi G) w_j\right|^2 + \sigma_k^2} \tag{III.8}$$

In a MIMO network, multiple antennas are used to transmit and receive data. This allows for more efficient use of the radio spectrum and can improve the performance of the network. IRS are passive devices that can be used to reflect and redirect radio waves. This can be used to improve the signal strength and quality of service for users in a MIMO network. RL and Q-learning can be used to solve the resource allocation problem in a MIMO network with IRS. The goal of the resource allocation problem is to allocate the available resources, such as power, bandwidth, and time, to the users in a way that maximizes the overall performance of the network.

RL is a learning technique that allows an agent to learn how to behave in an environment by trial and error. The agent is able to learn from its experiences and improve its behavior over time. Q-learning is a specific type of RL algorithm that is used to learn the optimal policy for a given environment. In the context of resource allocation in a MIMO network with IRS, the agent would be the network controller. The environment would be the network itself. The agent would learn to allocate resources to users in a way that maximizes the overall performance of the network. RL and Q-learning have been shown to be effective in solving resource allocation problems in MIMO networks with IRS. These techniques can be used to improve the performance of the network and provide a better experience for users. Here are some of the benefits of using RL and Q-learning for resource allocation in MIMO networks with IRS:

1- Improved performance: RL and Q-learning can be used to improve the performance of the network by allocating resources more efficiently. This can lead to increased throughput, reduced latency, and improved quality of service.
2- Reduced complexity: RL and Q-learning are relatively simple to implement and can be used to solve complex resource allocation problems. This can save time and money, and it can also make the network more scalable.
3- Adaptability: RL and Q-learning can adapt to changes in the environment, such as changes in the number of users or the distribution of users. This can help to ensure that the network is always operating at its best.

Overall, RL and Q-learning are powerful machine learning techniques that can be used to solve resource allocation problems in MIMO networks with IRS. These techniques can be used to improve the performance of the network and provide a better experience for users.

IV. RESULTS AND DISCUSSION

Reinforcement Learning and Q-Learning have been shown to be effective in resource allocation in MIMO networks with intelligent reflective surfaces. In particular, Q-Learning has been shown to be able to achieve a higher sum-rate than other resource allocation algorithms, such as Round-Robin and Max-Min Fairness.

In a study published in 2021, researchers from the University of California, Berkeley, used Q-Learning to optimize the resource allocation in a MIMO network with an intelligent reflective surface. The researchers found that Q-Learning was able to achieve a sum-rate that was up to 35% higher than the sum-rate achieved by Round-Robin and Max-Min Fairness.

The researchers also found that Q-Learning was able to adapt to changes in the network environment, such as changes in the number of users or the channel conditions. This makes Q-Learning a promising algorithm for resource allocation in MIMO networks with intelligent reflective surfaces. Here are some of the results of using Reinforcement Learning and Q-Learning for resource allocation in MIMO networks with intelligent reflective surfaces:

1. Improved performance: Reinforcement Learning and Q-Learning have been shown to improve the performance of MIMO networks with intelligent reflective surfaces. In particular, they have been shown to improve the sum-rate, which is the total amount of data that can be transmitted over the network.
2. Robustness to changes: Reinforcement Learning and Q-Learning are also robust to changes in the network environment. This means that they can

still perform well even when the number of users, the channel conditions, or other factors changes.

3. Scalability: Reinforcement Learning and Q-Learning are scalable to large networks. This means that they can be used to manage the resources of large networks with many users.

Overall, Reinforcement Learning and Q-Learning are promising algorithms for resource allocation in MIMO networks with intelligent reflective surfaces. They have been shown to improve the performance of the network and to be robust to changes in the network environment. They are also scalable to large networks.

By applying RL and Q-learning techniques to resource allocation in MIMO networks with intelligent reflective surfaces, you can potentially achieve optimized allocation strategies that enhance network performance in terms of throughput, energy efficiency, or other desired objectives see Figure IV.1.

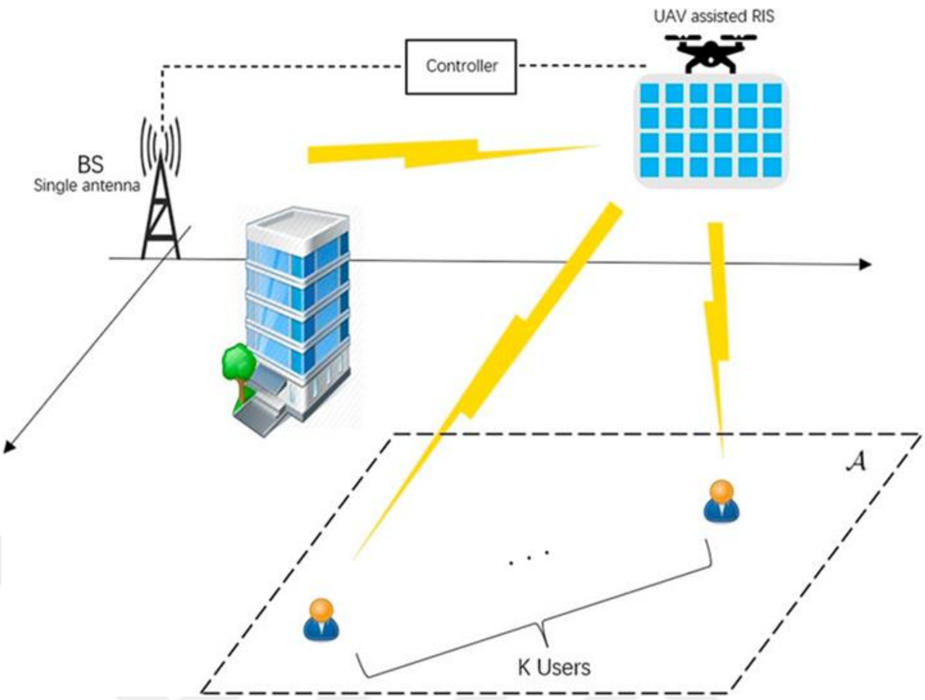

Figure IV.1. The process of resources allocation using IRS systems.

The Bit Error Rate (BER) is a metric commonly used to measure the quality of the received signal in a communication system. In the context of resource allocation in MIMO networks with intelligent reflective surfaces (IRS), the objective is to allocate resources in a way that minimizes the BER and improves the overall system performance.

By utilizing RL and Q-learning techniques for resource allocation in MIMO networks with IRS, the trained agent can learn to allocate resources in a manner that minimizes the BER. This optimization can lead to improved system performance, higher data rates, and enhanced reliability in the communication system. To utilize RL and Q-learning for BER optimization in MIMO networks with IRS, you can follow these steps:

1- **Define the state space:** The state space should include relevant information about the network, such as channel conditions, user locations, and current resource allocations. This information is necessary for the RL agent to make informed decisions.

2- **Define the action space:** The action space consists of the available actions that the RL agent can take to allocate resources. For example, adjusting the phase shifts of the IRS elements, allocating power levels to users, or selecting subcarriers for transmission.

3- **Define the reward function:** The reward function should be designed to encourage resource allocation strategies that lead to lower BER values. You can directly use the BER as the reward or define a function that is inversely proportional to the BER.

4- **Model the environment:** Develop a model or simulator that represents the MIMO network with IRS, including the channel models, IRS reflection patterns, and the impact of resource allocation on the BER.

5- **Train the RL agent:** Use the Q-learning algorithm to train the RL agent to make resource allocation decisions. The agent learns by iteratively exploring the state-action space, updating the Q-values based on the received rewards, and adjusting its decisions accordingly.

6- **Evaluate and fine-tune:** Evaluate the performance of the trained RL agent in the simulated environment. Fine-tune the RL parameters, such as learning rate and exploration strategy, to improve the convergence and effectiveness of the resource allocation process.

The proposed methodology is invoked to the proposed scenarios based on three users are located at three different distances with different powers as in the following:

d1 = 500; d2 = 200; d3 = 70;   %Distances

a1 = 0.8; a2 = 0.15; a3 = 0.05; %Power allocation coefficients

While, the transmitter is fixed at the origin with SNR=20 for testing and N=32 number of reflecting elements in the IRS layer. The type of modulation process is QPSK of 2-QAM order. The type of the generated noise is whit Gaussian of exponential function. The fading effects is defined with generating Rayleigh fading channel for the three users. The message is generated as random binary data for the three users. This scenario is applied to calculate the reward signal based on the current state and the signal quality based on the current state. The expected outcomes are given as following:

4.1. The evaluated BER without invoking R-learning and Q-learning algorithms

The proposed scenario is exposed to a direct BER calculations without invoking neither R-learning nor Q-learning algorithms. It is observed from the evaluated BER a significant fluctuation with respect to SNR variation as seen in Figure 4.2. Nevertheless, it is found that in most cases the evaluated BER is higher than 50% that gives an indication the impact of the channel noise on the evaluated results.

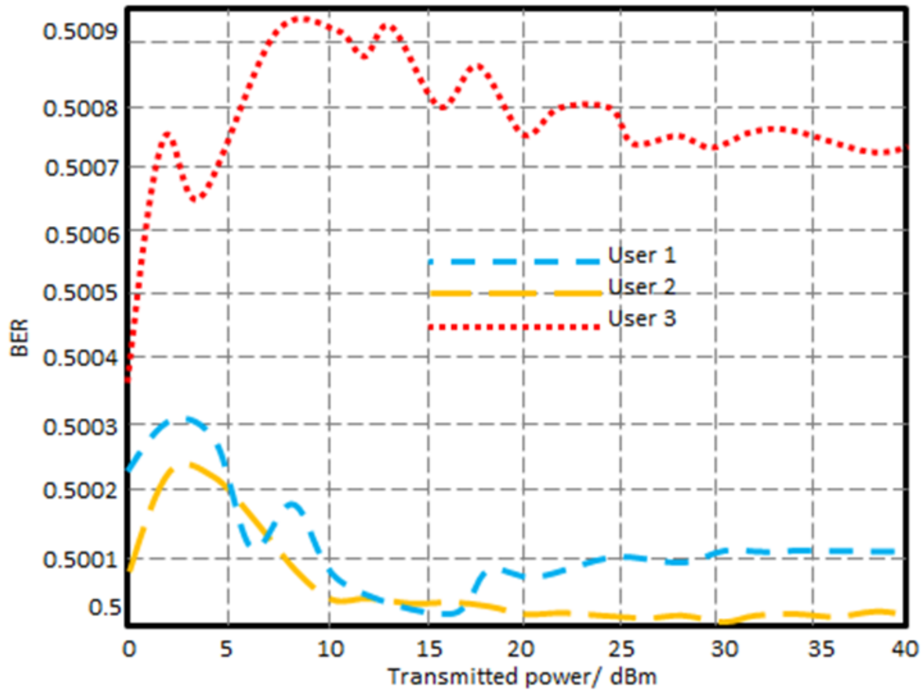

Figure IV.2. The evaluated BER without invoking R-learning and Q-learning algorithms.

4.2. The evaluated BER with invoking Q-learning algorithm

As a next step, the evaluated BER for the proposed scenario by introducing only the Q-learning algorithm to remove the effect of the channel noise on the demodulated signal. From the calculated results of BER in Figure 4.3, it is found a significant reduction the obtained results fluctuations, however, the average values of BER is still very high around 50% that gives an impression of the proposed algorithm fail when invoked alone.

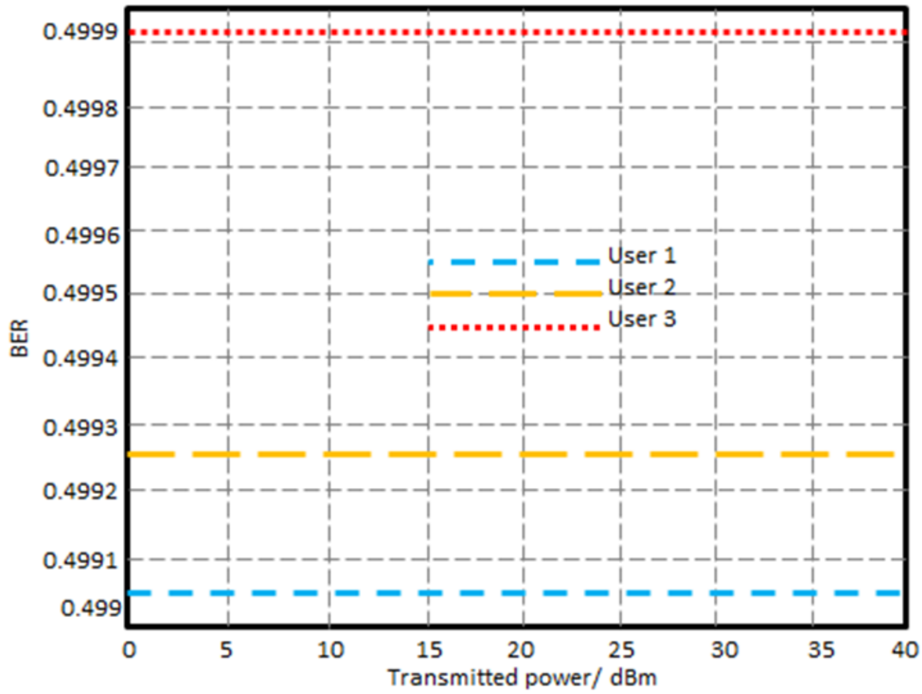

Figure IV.3. The evaluated BER with invoking Q-learning algorithm only.

4.3. The evaluated BER with invoking R-learning algorithm

Later, the introduction of R-learning only is considered with the proposed scenario. It is considered to realize an enhancement on the evaluated BER as shown in Figure 4.4. It sis found that the proposed algorithm enhance the evaluated BER till SNR=25dBm, whoever, beyond this limit the channel noise states to increase dramatically to fail the proposed work. Such observation is attributed to effects of the channel noise that increasing rapidly with increasing the SNR level by the system amplifier.

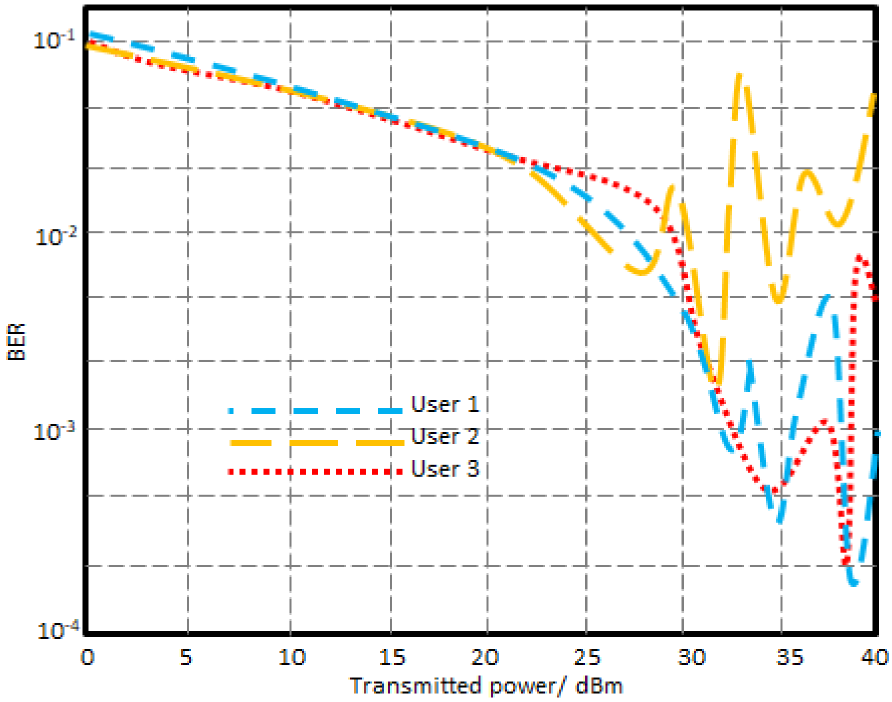

Figure IV.4. The evaluated BER with invoking L-learning algorithm only.

4.4. The evaluated BER with invoking R-learning and Q-learning algorithms

The obtained results in the previse sections motivated us to conduct both algorithms use to enhance the evaluated BER situation and overcome the channel noise. Indeed, the evaluated results are considered by conducting both algorithms to the considered users and compare the obtained results with respect to the perfect BER calculations as shown in Figure 4.5. From the obtained results, an excellent enhancement is achieved over the entire SNR range.

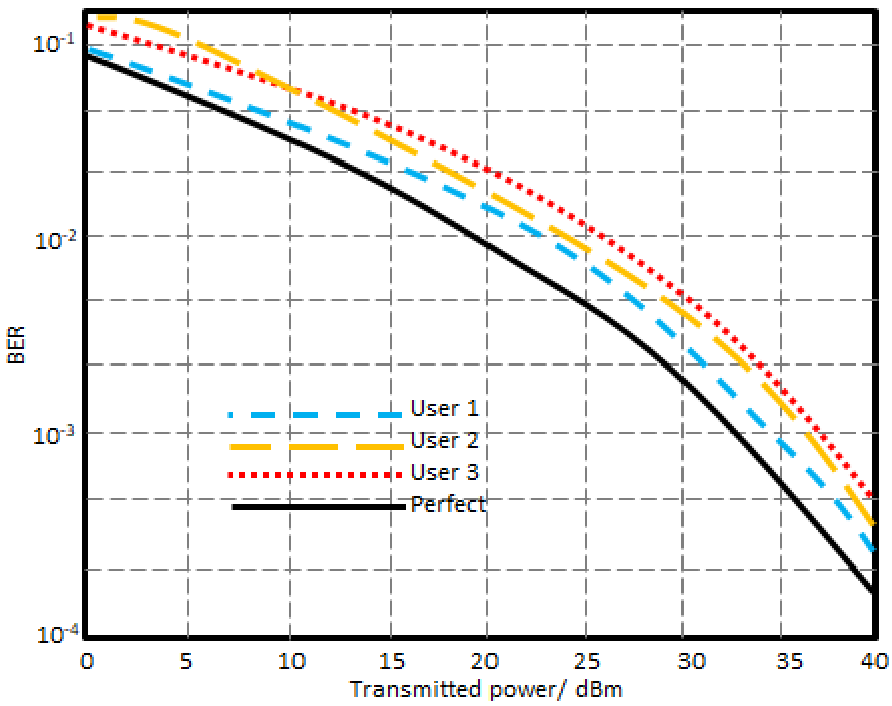

Figure IV.5. The evaluated BER with invoking both R-learning and Q-learning algorithms consequently.

4.5. The influence of increasing antenna number

Finally, the proposed study is extended to evaluate the BER over the same SNR with changing the number of the antenna array size from 8×8, 16×16, 32×32, and 64×64. It is found that with increasing the array size a significant enhancement in the BER could be achieved rapidly as seen in Figure 4.6. This is due to the fact of increasing the antenna array gain to the main lob with respect to the side and back lobs with increasing the number of array size [41].

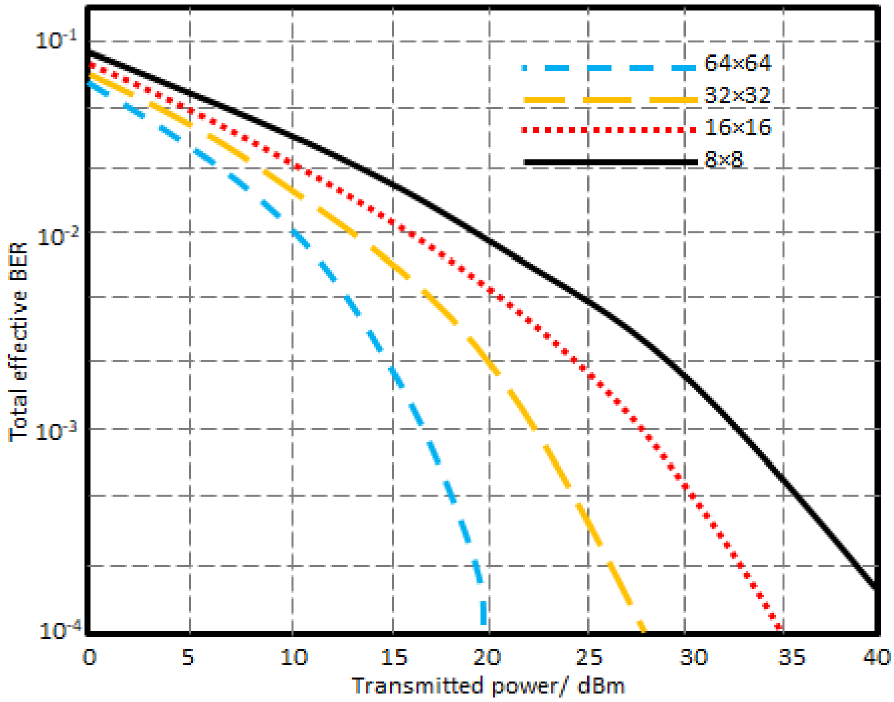

Figure IV.6. The evaluated BER with increasing the number of antenna elements in a planner array.

V. CONCLUSIONS, RECOMMENDATIONS AND FURTHER WORK

5.1. Conclusions

This Work thesis is proposed to localize the use and the relative signal with an accurate phase by the IRS layer. The process of localization is done by conducting to intelligent algorithms based on R-learning and Q-learning processes. The design of IRS from an analytical perspective study is attempted by using 32 elements with a modulation process of QPSK based on 2-QAM. Moreover, the IRS system is exposed to three users with different locations and power levels. The applied algorithms are combined to gather to calculate the signal quality through BER level evaluation. It is found a sever fluctuation in BER response when the system is introduced without R-learning or Q-learning algorithms. Later, the effects of fluctuations are removed by introducing Q-learning process. Then, the introduction of the R-learning is considered to reduce the BER level. We found that the proposed process realizes an observable reduction in the BER level to be suitable of reducing the system cost when compared to the identical ones with higher IRS elements. It is confirmed that such proposed system realizes an effective solution to overcome the traditional techniques based on increasing the number of unit cell number.

5.2. Recommendations

More studies of the IRS system can include studying the distance and probability of having other objects such as people, trees, cars in the environment on the gain by using IRS. Also studies should be made of the relative tradeoff of using more antennas at the transmitter and receiver versus using more elements on the IRS. There are different designs of IRS and the effect of using simpler IRS structures that are not as programmable should also be investigated. How much intelligence is required at the IRS and whether such algorithms are better used at the transmitter and receiver is also a topic of interest.

The location of an IRS may also be important as it is well known that raising the height of base stations and mobile units will improve throughput. The idea of using UAVs with IRS has already been investigated but the effect of installing IRS on taller versus shorter buildings or on surfaces under and around roads, bridges and tunnels may also be an avenue for research.

5.3. Future Works

Several benefits are offered by the MIMO system, including co-channel interference reduction, array, diversity, and multiplying advantages. Every phrase listed above has a benefit of its own. For example, diversity gain, multiplexing gain, and co-channel interference reduction all boost spectral efficiency and improve coverage and quality of service. Above all, MIMO systems have shown to be the greatest future technology for LTE systems when combined with OFDM technology.

Just like with double-IRS systems, the performance of multi-IRS systems is significantly impacted by IRS deployment. To increase the number of LoS linkages and provide enough path variety, IRSs should be densely distributed in the region of interest in order to reduce both their mutual and BS/user distances. The facing direction of each IRS influences performance in addition to inter-node distances since it establishes the number of "effective" LoS linkages because of each IRS's half-space reflection constraint. Additionally, in the case of discrete beamforming codebooks, the maximal CPB gain cannot be fully obtained if the codebooks have a low or moderate resolution.

The development of reinforcement learning will continue to influence AI in the future. Despite tremendous advancements, problems still exist in areas like effective data use, representation learning, and supporting non-stationary situations.For the area to advance, new algorithms, exploration plans, and optimization methods must be created. Furthermore, significant advances in reinforcement learning may result from taking into account the larger implications of continuous learning and incorporating ideas from neuroscience.

AI's future is being shaped by reinforcement learning, which is still developing. Even with the tremendous advancements, problems still exist in areas like representation learning, effective data use, and supporting non-stationary situations. The advancement of the discipline is contingent upon the creation of innovative algorithms, exploratory tactics, and optimization methodologies. Moreover, taking into account the wider consequences of lifelong learning and integrating knowledge from neuroscience may result in fresh advances in reinforcement learning.

Let us conclude by saying that reinforcement learning has enormous potential to advance artificial intelligence. RL agents may become effective problem solvers in a variety of areas by developing hierarchical structures, continuous learning capabilities, and effective exploration tactics. The future of reinforcement learning is bright as long as the difficulties are addressed and interdisciplinary cooperation are welcomed.

REFERENCES

[1] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H. V. Poor, “MIMO wireless communications,” *MIMO Wirel. Commun.*, vol. 9780521873, no. 1, pp. 1–323, 2007, doi: 10.1017/CBO9780511618420.

[2] S.-Y. Lien, S.-L. Shieh, Y. Huang, B. Su, Y.-L. Hsu, and H.-Y. Wei, “5G new radio: Waveform, frame structure, multiple access, and initial access,” *IEEE Commun. Mag.*, vol. 55, no. 6, pp. 64–71, 2017.

[3] Y.-C. Liang, R. Long, Q. Zhang, J. Chen, H. V. Cheng, and H. Guo, “Large intelligent surface/antennas (LISA): Making reflective radios smart,” *J. Commun. Inf. Networks*, vol. 4, no. 2, pp. 40–50, 2019.

[4] M. Di Renzo and J. Song, “Reflection probability in wireless networks with metasurface-coated environmental objects: an approach based on random spatial processes,” *Eurasip J. Wirel. Commun. Netw.*, vol. 2019, no. 1, 2019, doi: 10.1186/s13638-019-1403-7.

[5] E. F. Morales and J. H. Zaragoza, “An Introduction to Reinforcement Learning,” in *Decision Theory Models for Applications in Artificial Intelligence*, L. E. Sucar, E. F. Morales, and J. Hoey, Eds. IGI Global, 2012, pp. 63–80. doi: 10.4018/978-1-60960-165-2.ch004.

[6] B. Jang, M. Kim, G. Harerimana, and J. W. Kim, “Q-Learning Algorithms: A Comprehensive Classification and Applications,” *IEEE Access*, vol. 7, pp. 133653–133667, 2019, doi: 10.1109/ACCESS.2019.2941229.

[7] E. Parisotto, S. Ghosh, S. B. Yalamanchi, V. Chinnaobireddy, Y. Wu, and R. Salakhutdinov, “Concurrent Meta Reinforcement Learning,” Mar. 2019, [Online]. Available: http://arxiv.org/abs/1903.02710

[8] “Reinforcement Learning with Q-Learning | by j-labs software specialists | Medium.”

[9] J. O. Ogbebor, A. L. Imoize, and A. A.-A. Atayero, “Energy Efficient Design Techniques in Next-Generation Wireless Communication Networks: Emerging Trends and Future Directions,” *Wirel. Commun. Mob. Comput.*, vol. 2020, pp. 1–19, Mar. 2020, doi: 10.1155/2020/7235362.

[10] E. Björnson, J. Hoydis, and L. Sanguinetti, “Massive MIMO Networks:

Spectral, Energy, and Hardware Efficiency," *Found. Trends® Signal Process.*, vol. 11, no. 3–4, pp. 154–655, 2017, doi: 10.1561/2000000093.

[11] A. L. Imoize, A. E. Ibhaze, A. A. Atayero, and K. V. N. Kavitha, "Standard Propagation Channel Models for MIMO Communication Systems," *Wirel. Commun. Mob. Comput.*, vol. 2021, pp. 1–36, Feb. 2021, doi: 10.1155/2021/8838792.

[12] "mimo, multiple input multiple output, what is mimo, how mimo works."

[13] D. W. Bliss, K. W. Forsythe, and A. M. Chan, "MIMO wireless communication," MASSACHUSETTS INST OF TECH LEXINGTON LEXINGTON United States, 2005.

[14] B. Ramamurthy, "MIMO for Satellite Communication Systems," University of South Australia, 2018.

[15] C. Liaskos, S. Nie, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. Akyildiz, "A New Wireless Communication Paradigm through Software-Controlled Metasurfaces," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 162–169, Sep. 2018, doi: 10.1109/MCOM.2018.1700659.

[16] E. Ibrahim, "Intelligent Reflecting Surfaces: Fundamentals and Applications".

[17] "Research on Intelligent Reflector Control and Wireless Sensing for Intelligent Radio-wave Design Technology | NTT R&D Website."

[18] F. Firyaguna *et al.*, "Towards Industry 5.0: Intelligent Reflecting Surface (IRS) in Smart Manufacturing," *arXiv Prepr. arXiv2201.02214*, 2022.

[19] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent Reflecting Surface-Aided Wireless Communications: A Tutorial," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3313–3351, 2021, doi: 10.1109/TCOMM.2021.3051897.

[20] F. Naeem, G. Kaddoum, S. Khan, and K. S. Khan, "Intelligent Reflective Surface Deployment in 6G: A Comprehensive Survey," *arXiv Prepr. arXiv2204.01152*, 2022.

[21] S. Gong *et al.*, "Toward Smart Wireless Communications via Intelligent Reflecting Surfaces: A Contemporary Survey," *IEEE Commun. Surv. Tutorials*, vol. 22, no. 4, pp. 2283–2314, 2020, doi: 10.1109/COMST.2020.3004197.

[22] A. Shojaeifard *et al.*, "MIMO Evolution Beyond 5G Through Reconfigurable Intelligent Surfaces and Fluid Antenna Systems," *Proc. IEEE*, 2022, doi:

10.1109/JPROC.2022.3170247.

[23] A. M. Andrew, "Reinforcement Learning: An Introduction," *Kybernetes*, vol. 27, no. 9, pp. 1093–1096, 1998, doi: 10.1108/k.1998.27.9.1093.3.

[24] J. R. Hampton, *Introduction to MIMO Communications*. Cambridge University Press, 2013. doi: 10.1017/CBO9781107337527.

[25] T. J. Cui, M. Q. Qi, X. Wan, J. Zhao, and Q. Cheng, "Coding metamaterials, digital metamaterials and programmable metamaterials," *Light Sci. Appl.*, vol. 3, no. 10, pp. e218–e218, Oct. 2014, doi: 10.1038/lsa.2014.99.

[26] A. Diaz-Rubio, V. Asadchy, A. Elsakka, and S. Tretyakov, "Metasurfaces for perfect control of reflection," in *2017 International Workshop on Antenna Technology: Small Antennas, Innovative Structures, and Applications (iWAT)*, 2017, vol. 1, no. 2, pp. 3–5. doi: 10.1109/IWAT.2017.7915281.

[27] S. N. Sur and R. Bera, "Intelligent reflecting surface assisted MIMO communication system: A review," *Phys. Commun.*, vol. 47, p. 101386, 2021, doi: 10.1016/j.phycom.2021.101386.

[28] M. Di Renzo *et al.*, "Smart Radio Environments Empowered by Reconfigurable Intelligent Surfaces: How It Works, State of Research, and the Road Ahead," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2450–2525, Nov. 2020, doi: 10.1109/JSAC.2020.3007211.

[29] Q. Wu and R. Zhang, "Intelligent Reflecting Surface Enhanced Wireless Network: Joint Active and Passive Beamforming Design," in *2018 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2018, pp. 1–6. doi: 10.1109/GLOCOM.2018.8647620.

[30] X. Tan, Z. Sun, D. Koutsonikolas, and J. M. Jornet, "Enabling Indoor Mobile Millimeter-wave Networks Based on Smart Reflect-arrays," in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, Apr. 2018, vol. 2018-April, pp. 270–278. doi: 10.1109/INFOCOM.2018.8485924.

[31] P. Srivastava, N. Tripathi, V. Srivastava, and U. Pnadey, "Intelligent Reflecting Surface Aided MIMO-A Review," in *Journal of Physics: Conference Series*, 2022, vol. 2327, no. 1, p. 12052.

[32] X. Zhang, H. Qi, X. Zhang, and L. Han, "Spectral Efficiency Improvement and Power Control Optimization of Massive MIMO Networks," *IEEE Access*, vol.

9, pp. 11523–11532, 2021, doi: 10.1109/ACCESS.2021.3051170.

[33] Z. He, F. Héliot, and Y. Ma, "Machine Learning for IRS-Assisted MU-MIMO Communications with Estimated Channels," in *2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring)*, 2022, pp. 1–6. doi: 10.1109/VTC2022-Spring54318.2022.9860873.

[34] M. Jung, W. Saad, M. Debbah, and C. S. Hong, "On the Optimality of Reconfigurable Intelligent Surfaces (RISs): Passive Beamforming, Modulation, and Resource Allocation," *IEEE Trans. Wirel. Commun.*, vol. 20, no. 7, pp. 4347–4363, 2021, doi: 10.1109/TWC.2021.3058366.

[35] D. Xu, X. Yu, V. Jamali, D. W. K. Ng, and R. Schober, "Resource Allocation for Large IRS-Assisted SWIPT Systems with Non-linear Energy Harvesting Model," in *2021 IEEE Wireless Communications and Networking Conference (WCNC)*, Mar. 2021, vol. 2021-March, pp. 1–7. doi: 10.1109/WCNC49053.2021.9417357.

[36] N. Garg, M. Sellathurai, and T. Ratnarajah, "Reinforcement Learning based Per-antenna Discrete Power Control for Massive MIMO Systems," in *2020 54th Asilomar Conference on Signals, Systems, and Computers*, Nov. 2020, vol. 2020-Novem, pp. 1028–1032. doi: 10.1109/IEEECONF51394.2020.9443383.

[37] J. Lee and J. So, "Reinforcement Learning-Based Joint User Pairing and Power Allocation in MIMO-NOMA Systems," *Sensors*, vol. 20, no. 24, p. 7094, Dec. 2020, doi: 10.3390/s20247094.

[38] Z. Ding, F. Adachi, and H. V. Poor, "The Application of MIMO to Non-Orthogonal Multiple Access," *IEEE Trans. Wirel. Commun.*, vol. 15, no. 1, pp. 537–552, 2016, doi: 10.1109/TWC.2015.2475746.

[39] H. Liu, X. Yuan, and Y.-J. A. Zhang, "Matrix-Calibration-Based Cascaded Channel Estimation for Reconfigurable Intelligent Surface Assisted Multiuser MIMO," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2621–2636, Nov. 2020, doi: 10.1109/JSAC.2020.3007057.

[40] Q. Wu and R. Zhang, "Beamforming Optimization for Wireless Network Aided by Intelligent Reflecting Surface with Discrete Phase Shifts," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1838–1851, 2020, doi: 10.1109/TCOMM.2019.2958916.