

**REPUBLIC OF TURKEY
ISTANBUL GELISIM UNIVERSITY
INSTITUTE OF GRADUATE STUDIES**

Department of Electrical and Electronics Engineering

**RECONFIGURABLE INTELLIGENT SURFACE (RIS)
FOR UAV COMMUNICATION LINKS IMPROVEMENT**

Master Thesis

Nameer Mufeed SALIH AL-GBURI

Supervisor

Asst. Prof. Dr. Mahmoud H.K. ALDABABSA

Istanbul – 2024

THESIS INTRODUCTION FORM

Name and Surname : Nameer Mufeed Salih AL-GBURI

Language of the Thesis : English

Name of the Thesis : Reconfigurable Intelligent Surface (Ris) For Uav
Communication Links Improvement

Institute : Istanbul Gelisim University Institute of Graduate Studies

Department : Electrical and Electronics Engineering

Thesis Type : Master

Date of the Thesis : 25/01/2024

Page Number : 62

Thesis Supervisors : Asst. Prof. Dr. Mahmoud H.K. ALDABABSA

Index Terms : Deep Deterministic Policy Gradient, Reconfigurable
intelligent surface, UAV

Turkish Anstract : İHA İletişim Bağlantılarının İyileştirilmesi İçin Yeniden
Yapılandırılabilir Akıllı Yüzey (Ris)

Distribution List : 1. To the Institute of Graduate Studies of Istanbul
Gelism University
2. To the National Thesis Center of YÖK (Higher
Education Council)

Signature

Nameer Mufeed Salih AL-GBURI

**REPUBLIC OF TURKEY
ISTANBUL GELISIM UNIVERSITY
INSTITUTE OF GRADUATE STUDIES**

Department of Electrical and Electronics Engineering

**RECONFIGURABLE INTELLIGENT SURFACE (RIS)
FOR UAV COMMUNICATION LINKS IMPROVEMENT**

Master Thesis

Nameer Mufeed SALIH AL-GBURI

Supervisor

Asst. Prof. Dr. Mahmoud H.K. ALDABABSA

Istanbul – 2024

DECLARATION

I hereby declare that in the preparation of this thesis, scientific ethical rules have been followed, the works of other persons have been referenced in accordance with the scientific norms if used, there is no falsification in the used data, any part of the thesis has not been submitted to this university or any other university as another thesis.

Nameer Mufeed Salih AL-GBURI

.../.../2024



TO ISTANBUL GELISIM UNIVERSITY
THE DIRECTORATE OF GRADUATE EDUCATION INSTITUTE

The thesis study of Nameer Mufeed Salih AL-GBURI titled as Reconfigurable Intelligent Surface (RIS) For Uav Communication Links Improvement has been accepted as MASTER in the department of Electrical-Electronic Engineering by out jury.

Director

Asst. Prof. Dr. Mahmoud H. K. ALDABABSA

(Supervisor)

Member

Asst. Prof. Dr. Halit YAHYA

Member

Asst. Prof. Dr. Mohammed SALEM

APPROVAL

I approve that the signatures above signatures belong to the aforementioned faculty members.

... / ... / 2024

Prof. Dr. Izzet GUMUS

Director of the Institute

SUMMARY

Reconfigurable intelligent surface (RIS) is a potential technology for energy-efficient wireless communications that has recently received a lot of interest. In dynamic circumstances, combining an unmanned aerial vehicle with RIS (UAV-RIS) can provide on-demand deployment services. The energy of the battery-powered UAV, however, will limit the benefits of UAV-RIS. We propose an innovative energy harvesting strategy for Simultaneous Wireless Information and Power Transfer (SWIPT) in UAV-RIS (Unmanned Aerial Vehicle-Reflecting Intelligent Surface) systems. Unlike previous studies, our approach involves segmenting passive reflected arrays in geometric space, enabling the concurrent transfer of information and energy harvesting from impinging radio-frequency (RF) signals. Additionally, we are developing a Deep Deterministic Policy Gradient (DDPG) technique to dynamically allocate resources of UAV-RIS in both time and space domains, aiming to maximize the overall harvested energy while maintaining communication quality for individual users. Our simulations demonstrate a significant performance enhancement of the proposed UAV-RIS SWIPT system compared to existing benchmarks.

Key Words: Deep Deterministic Policy Gradient, Reconfigurable intelligent surface, UAV

ÖZET

Yeniden yapılandırılabilir akıllı yüzey (RIS), enerji açısından verimli kablosuz iletişim için gelecek vaat eden bir teknolojidir ve son zamanlarda büyük ilgi görmüştür. İnsansız hava aracını RIS (UAV-RIS) ile birleştirmek, dinamik senaryolarda isteğe bağlı konuşlandırma hizmetleri sağlayabilir. Bununla birlikte, İHA-RIS'in faydalarından yararlanmak, pille çalışan İHA'nın enerjisi ile sınırlı olacaktır. İHA-RIS'lerin dayanıklılığını artırmak için, eşzamanlı kablosuz bilgi ve güç aktarımı (SWIPT), kaynak tahsisi ve çarpan radyo frekansı (RF) sinyallerinden enerji hasadı için yeni bir enerji toplama planı geliştiriyoruz. Mevcut çalışmalardan farklı olarak, önerilen şema, pasif yansıyan dizileri, aynı anda bilgi taşımak ve enerji toplamak için geometrik uzayda yaratıcı bir şekilde böler. Ayrıca, her kullanıcı için iletişim kalitesini garanti ederken toplanan toplam enerjiyi en üst düzeye çıkarmak için UAV-RIS'in kaynaklarını hem zaman hem de uzay alanlarında sürekli olarak tahsis etmek için derin bir deterministik politika gradyanı (DDPG) şeması tasarlanmıştır. Simülasyon sonuçlarımızın da gösterdiği gibi, önerilen UAV-RIS SWIPT sistemi, kıyaslamamızın üzerinde performansı önemli ölçüde artırıyor.

Anahtar kelimeler: Derin Deterministik Politika Gradyanı, Yeniden yapılandırılabilir akıllı yüzey, İHA

TABLE OF CONTENTS

SUMMARY	i
ÖZET.....	ii
TABLE OF CONTENTS.....	iii
ABBREVIATIONS	v
LIST OF TABLES	vi
LIST OF GRAPHICS	vii
LIST OF FIGURES	viii
INTRODUCTION.....	1

CHAPTER ONE

BACKGROUND

1.1. General Context.....	2
1.2. Introduction to Reconfigurable Intelligent Surfaces.....	3
1.2.1. Meta-surfaces.....	3
1.2.2. Reconfigurable Intelligent Surfaces	5
1.3. Benefits of RIS	7
1.4. Applications of RIS	8
1.5. Research Motivation.....	11
1.6. Aims and Objectives.....	2
1.7. Outline	12

CHAPTER TWO

LITERATURE REVIEW

2.1. RIS for Wireless Communication.....	13
2.1.1. RIS based communications.....	13
2.1.2. RIS-assisted UAV communications.....	15
2.2. Application of UAV Assisted RIS.....	18
2.3. Wireless Power Transfer.....	20
2.4. Localization and Mapping	21
2.5. Application of RIS for Localization	23
2.6. Reconfigurable Intelligent Surfaces-Aided Wireless Networks.....	24
2.7. Deep Reinforcement Learning in Wireless Networks	26
2.7.1. An Overview of Markov Decision Processes	26
2.7.2. A Brief Overview of Deep Reinforcement Learning	27
2.8. Related Works	32
2.9. Contributions	36

CHAPTER THREE

METHODOLOGY

3.1. IRS-Aided UAV Communications	37
3.2. The DDPG Method.....	37
3.3. Proposed Algorithms	42
3.3.1. DQN-based Algorithm for Discrete Cases	42
3.3.2. Algorithm for Continuous Cases Based on DDPG	44

CHAPTER FOUR

RESULTS AND DISCUSSION

4.1. Simulation Setting	46
4.2. Simulation Results and Discussions	46
4.2.1. Parametric analysis	46
4.2.2. Performance Comparison	49
4.3. Conclusion	54
CONCLUSION AND FUTURE WORKS	55
REFERENCES.....	57

ABBREVIATIONS

AI	:	Artificial Intelligence
FSK	:	Frequency Shift Keying
IoT	:	Internet of Things
PAN	:	Personal Area Network
OFDM	:	Orthogonal Frequency Division Multiplexing
TDMA	:	Time Division Multiple Access
IRS	:	Intelligent Reflecting Surfaces
RIS	:	Reconfigurable Intelligent Surfaces
DQL	:	Deep Q-Learning
DDPG	:	Deep Deterministic Policy Gradients

LIST OF TABLES

Table 1. A comparison with existing literature.....36

Table 2. Simulation parameters 46



LIST OF GRAPHICS

Graphic 1. The overall performance of the DDPG-based algorithm at various learning rates.	47
Graphic 2. Performance of convergence with various discount factors.....	48
Graphic 3. (a) DQN and (b) DDPG average reward versus number of training Episodes...	49
Graphic 4. DQN, DDPG, FF, and RR with varying numbers of reflecting elements attain average energy efficiency.	50
Graphic 5. DQN, DDPG, FF, and RR with shifting quantities of reflecting components achieve normal energy productivity.....	51
Graphic 6. Performance without state normalization or behavior noise	52
Graphic 7. (a) DQN and DDPG performance under varying UE computational capabilities. (b) DQN and DDPG offloading ratios under varied UE computational capabilities. (c) A comparison of DQN and DDPG	53

LIST OF FIGURES

Figure 1: Example of static and dynamic meta-surfaces made of meta-atoms (Zhang, Song, Seddik, Han, & Li, 05 May 2020).....	4
Figure 2. Representation of elementary functions supported by meta-surfaces.....	5
Figure 3. NTT DOCOMO's prototype, example of a programmable glass (Basar, et al., 2019).....	7
Figure 4. General use of RIS in outdoor environment (Basar, et al., 2019).	8
Figure 5. General use of RIS in outdoor environment (Basar, et al., 2019).	8
Figure 6. Usage of RIS for vehicular communication in urban areas (Basar, et al., 2019). ...	9
Figure 7. Usage of RIS in indoor environment (Basar, et al., 2019).....	10
Figure 8. Reinforcement learning model in wireless networks.....	12
Figure 9. Examples of application of RIS (Balakrishnan, De, & Wang, 2020).....	14
Figure 10. RIS-assisted UAV-aided communications (Mei, Yang, Shen, & Liu).	16
Figure 11. RIS-assisted UAV for extended coverage (Mei, Yang, Shen, & Liu).....	17
Figure 12. RIS-UAV for spectrum sharing (Mei, Yang, Shen, & Liu).	17
Figure 13. RIS-UAV for physical layer security (Mei, Yang, Shen, & Liu).	18
Figure 14. Illustration of the relation between frequency range and radio channel propagation effects (Al-Hourani, Kandeepan, & Jamalipour, 2014).....	19
Figure 15. RIS-aided WPT system conceptual application scenario.....	20
Figure 16. RIS-assisted UAV for SWIPT (Mei, Yang, Shen, & Liu).	21
Figure 17. Application examples of RIS-based localization and mapping services (Ju & Zhang, 2013).....	24
Figure 18. Deep Q-learning	31
Figure 19. Actor-Critic architecture	32
Figure 20. UAV training using deep reinforcement agent.	41
Figure 21. Design of DQN calculation	42
Figure 22: DQN based Algorithm.....	44
Figure 23. Architecture of DDPG algorithm	45

INTRODUCTION

Unmanned Aerial Vehicles (UAVs), commonly known as drones, represent a promising yet intricate platform for the study of autonomous and cooperative control. A UAV is an electromechanical system capable of autonomous operation, remote control, or a combination of both. Among the diverse UAV types available, fixed-wing UAVs have garnered significant attention and uptake in various domains, spanning from academic research to commercial applications. This is largely attributed to their energy-efficient performance, especially when tasked with transporting payloads (Zhang, Song, Seddik, Han, & Li, 05 May 2020) (Basar, et al., 2019). A fixed-wing UAV regularly has an extensively easier construction than a rotating wing UAV (quadrotor, helicopter, slant rotor airplane, etc). This less difficult development typically requires a more straightforward upkeep and fix process, considering a more drawn out functional length at a diminished expense. In particular, this clear plan considers more compelling streamlined features, taking into consideration longer flight periods at quicker speeds with bigger payloads. The elements, incitation, take-off, and arriving of fixed-wing UAVs, then again, include complex optimal design. Moreover, from the stance of control hypothesis, the advancement of agreeable independent route frameworks fit for managing significant mission vulnerabilities (both in the UAV elements and in the general climate) is one of the essential examination issues for fixed-wing UAVs (Feng, Wang, Li, & Wen, 2020) (Peng, Tsai, Wang, & Han, 2021). Thus, the objective of this proposal is to give novel versatile and dispersed versatile methods that can address mission vulnerabilities in a methodical way at both the single fixed-wing UAV and numerous UAV levels. The significance of this exploration is straightforwardly connected with the advancement of further developed independent route frameworks for mechanical independent specialists performing different positions with a wide unique working reach. With this vision in mind, in the following we will explain the research goals in more detail.

CHAPTER ONE

BACKGROUND

Which is a prologue to the issue and portrays the issue explanation, points and targets, and examination questions that are proposed to arrive at the review's objective.

1.1. Aims and Objectives

Albeit existing DRL calculations have been generally utilized in remote organizations, they are challenging to adjust to current conditions because of thorough framework imperatives, for example, UAV flying length, transmission distance, and versatile clients. We discretize the flying way into a network, similarly as with the DQL and dueling DQL calculations, and the UAV just has to conclude the activity in a limited activity space. The brain organizations can be immediately prepared and conveyed for the web-based stage because of their limited state and activity space. We tried several existing RL algorithms and discovered that some of them are ineffective in tackling our stated task. In the interim, the constant solver RL calculations are ineffectual and hard to apply to the compromise issue. Accordingly, we propose the DQL and dueling DQL calculations in this part to accomplish the best compromise altogether reachable aggregate rate and direction. Therefore, we may proficiently change a certifiable application into a computerized climate for streamlining.

1.2. General Context

As wireless systems continue to evolve, there are ongoing challenges related to power consumption for communication and the unpredictable nature of the wireless environment. One persistent hurdle is the lack of control over this environment, both indoors and outdoors. To overcome this challenge, it is crucial to effectively manage the interaction between electromagnetic waves and the objects in the surrounding environment. The goal is to mitigate the negative effects that can arise from reflections, refractions, path-loss, and fading phenomena, such as unmanageable

interference. By addressing these issues, we can improve the performance and reliability of wireless systems, ensuring better connectivity and communication in various settings.

The emergence of Reconfigurable Intelligent Surfaces (RIS) in recent years has garnered attention as a promising technology for addressing the challenges posed by the unpredictable wireless environment. RIS is composed of low-power integrated electronic circuits, empowering these surfaces to actively manipulate the wireless environment and augment the capacity and coverage of wireless networks. This innovative technology has the ability to seamlessly integrate with a variety of existing wireless technologies, thereby offering the potential to greatly enhance their performance. By harnessing the capabilities of RIS, we can effectively overcome the limitations imposed by interference, reflections, and other phenomena, ultimately paving the way for more reliable and efficient wireless communication systems.

1.3. Introduction to Reconfigurable Intelligent Surfaces

The wireless environment is inherently unpredictable, and the quality of communication within it is greatly influenced by the presence of various objects. In outdoor settings, urban structures are prevalent and have a significant impact on wireless communication. Conversely, indoor environments are often crowded with multiple devices that can potentially cause interference. Additionally, at lower frequencies (sub-6 GHz), the surrounding structures act as scatterers of electromagnetic waves, further affecting communication. As frequencies increase into the millimeter wave or terahertz range and beyond, even smaller objects can have a substantial influence as scatterers, adding to the complexity of wireless communication dynamics. These factors highlight the intricate nature of wireless communication and the challenges posed by the environment in which it operates

1.3.1. Meta-surfaces

In recent years, the design concept of Reconfigurable Intelligent Surfaces (RIS) has undergone development, building upon the foundational idea of meta-surfaces. Meta-surfaces are planar structures composed of repeated conductive elements called meta-atoms, which are placed on a dielectric substrate (as shown in Figure 1). These meta-surfaces have the capability to manipulate electromagnetic waves. In the

context of RIS, meta-surfaces play a crucial role as they are equipped with interconnected switch components known as meta-iota. These components act as control elements that influence the flow of surface currents. By effectively combining these components, precise control over the overall electromagnetic response of the surface can be achieved, ultimately determining the aggregate radiated field generated by each surface current. This design approach allows for the customization and adaptation of the surface's electromagnetic behavior to optimize wireless communication and overcome the challenges posed by the wireless environment (Zhang, Song, Seddik, Han, & Li, 05 May 2020).

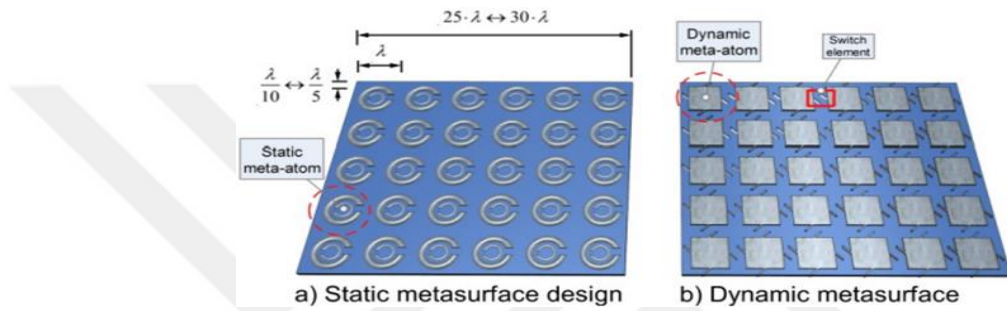


Figure 1.Example of static and dynamic meta-surfaces made of meta-atoms (Zhang, Song, Seddik, Han, & Li, 05 May 2020).

Meta-surfaces possess a wide range of electromagnetic (EM) functionalities, each serving distinct purposes (Basar, et al., 2019). The first functionality is reflection, which involves redirecting an incoming wave from its original direction towards a specific custom direction. Another function is refraction, where the trajectory of an incoming radio wave is altered towards a predetermined direction, even if it does not align with the wave's natural course. Absorption is yet another capability, which ensures that minimal power is reflected or refracted by the impinging waves. Lastly, polarization, a fundamental function, involves modifying the orientation of the wave's electric and magnetic fields. These various functionalities make meta-surfaces highly versatile and applicable in a wide range of scenarios that involve manipulating electromagnetic waves.

Meta-surfaces offer an array of advanced functionalities, including essential operations like beamforming, which involves redirecting an incoming wave in the same direction. In addition to the mentioned functionalities, meta-surfaces also enable beam splitting, which means they can create multiple reflected or refracted

radio waves from a single incident radio wave. Furthermore, meta-surfaces have the ability to support complex manipulations at the electromagnetic (EM) level. For example, a sophisticated meta-surface can refract a radio wave in a way that represents the first-order derivative or integral of the impinging radio wave. This showcases the capability of meta-surfaces to perform intricate EM-level computations and transformations, expanding their potential for advanced applications in various fields.

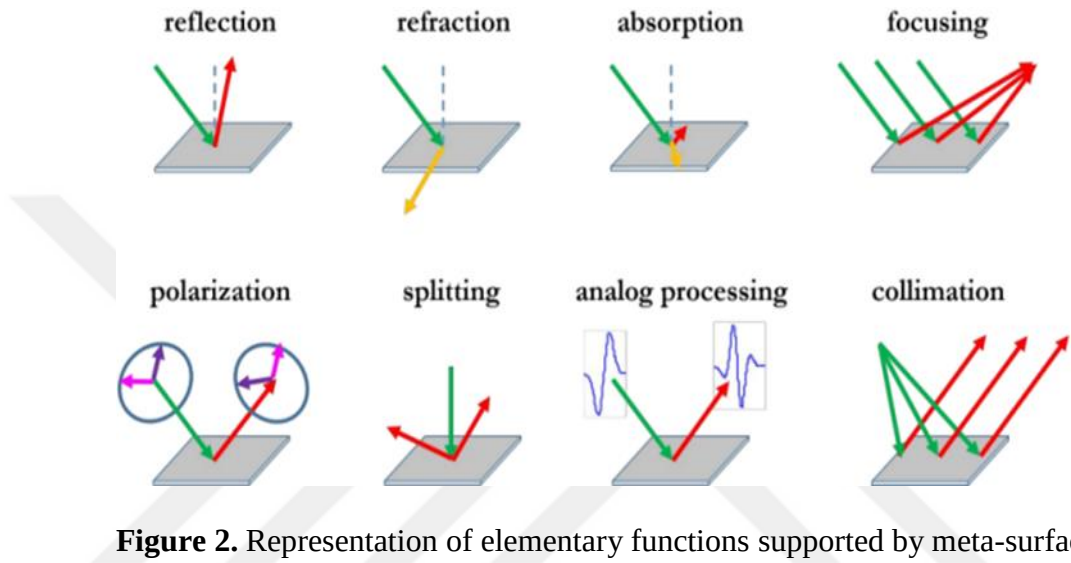


Figure 2. Representation of elementary functions supported by meta-surfaces.

1.3.2. Reconfigurable Intelligent Surfaces

In recent years, significant research efforts have been focused on exploring Beyond 5G networks, with the aim of developing technologies that can make the wireless environment programmable and manageable. This concept is often referred to as Smart Radio Environment (SRE) or Intelligent Radio Environment (IRE). A notable advancement in the design of these SREs is the integration of Reconfigurable Intelligent Surfaces (RIS). RIS technology allows for the dynamic control and manipulation of electromagnetic waves by employing meta-surface structures with interconnected switch components. By incorporating RIS into the wireless network architecture, it becomes possible to enhance the performance, efficiency, and flexibility of communication systems. The integration of RIS in SREs represents a promising avenue for achieving higher data rates, improved coverage, reduced interference, and enhanced energy efficiency in future wireless networks. These

surfaces, also referred to by other terms, embody a transformative approach to wireless communication, enabling control and adaptability within the wireless medium, paving the way for unprecedented advancements in wireless networking technologies:

Large Intelligent Surfaces (LIS): LIS refers to surfaces designed to advance the current Massive Multiple Input, Multiple Output (MIMO) technology, catering to extensive data and multiple output requirements.

Intelligent Reflecting Surfaces (IRS) refer to surfaces that serve as reflectors and are made up of individually tunable unit elements. These unit elements enable precise adjustments in phase response, thereby enabling functionalities such as beam steering, focusing, and other similar operations. On the other hand, Digitally Controllable Scatterers (DCS) is a term closely related to RIS, emphasizing the ability to digitally control the behavior of objects and devices, specifically those that are integrated or composed of smart surfaces. This highlights the capability of manipulating and dynamically controlling the scattering properties of these surfaces, allowing for advanced signal manipulation and optimization in wireless communication systems. The integration of IRS and DCS technologies opens up new possibilities for enhancing the performance and efficiency of wireless networks by leveraging the unique capabilities of these intelligent and controllable surfaces.

Reconfigurable savvy surfaces are meta-surfaces that are basically evolved as a versatile slender composite material sheet equipped for controlling radio waves. There are right now two fundamental plans for understanding these surfaces: programmable dainty backdrop and programmable slim glass. Both of these designs capability in an almost detached and dynamic way, and that implies that the surface doesn't produce extra radio waves. Additionally, Reconfigurable Intelligent Surfaces (RIS) possess several key characteristics that make them highly advantageous. Firstly, RIS operates without the need for power amplification during standard operational phases, reducing energy consumption. Secondly, minimal digital signal processing capabilities are required for configuring the surface, simplifying the overall system design. Moreover, RIS demands minimal power consumption for both configuration and operation, contributing to its energy efficiency. These characteristics enable RIS to be constructed using low-power electronic circuits and

allow for dynamic configuration to control the direction of outgoing waves. One practical implementation of RIS is the use of a transparent artificially engineered layer, which can be discreetly deployed due to its inconspicuous nature. This intelligent glass consists of numerous sub-wavelength components, with each unit cell significantly smaller than the wavelength, arranged on a flat surface and coated with a glass substrate. Such implementations highlight the versatility and potential of RIS in various applications.

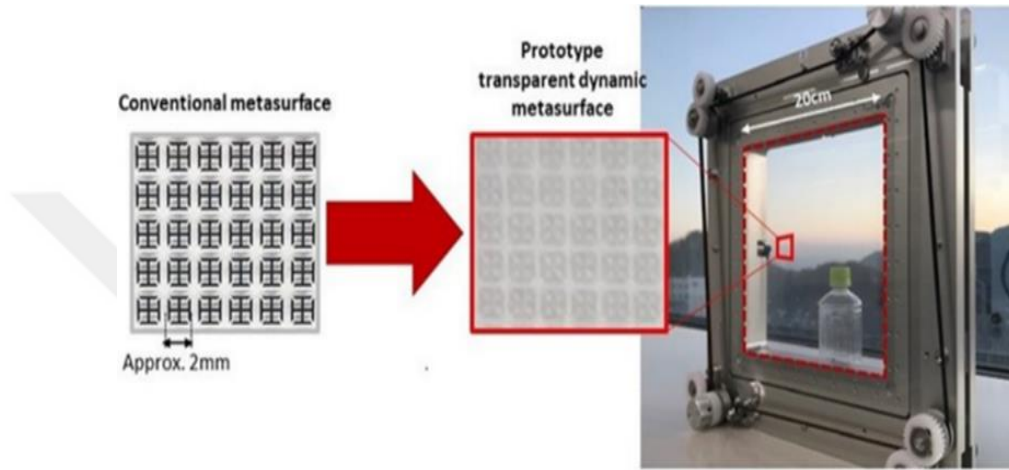


Figure 3. NTT DOCOMO's prototype, example of a programmable glass (Basar, et al., 2019).

1.4. Benefits of RIS

The integration of reconfigurable surfaces has proven to significantly improve performance in both indoor and outdoor wireless environments. The appealing characteristics of reconfigurable intelligent surfaces (RIS) include cost-effectiveness in terms of materials, minimal power consumption, and ease of deployment on various structures such as indoor walls, airborne platforms, roadside billboards, highway checkpoints, vehicle windows, and even clothing worn by pedestrians. Furthermore, RIS technology is environmentally friendly and stands out from conventional relaying systems due to its passive nature. This means that RIS does not require active power sources and operates by manipulating the incident waves, resulting in reduced energy consumption and a more sustainable wireless communication solution.

1.5. Applications of RIS

As detailed in the preceding subsection, the seamless integration of RIS into the existing wireless environment opens up a plethora of potential applications where its benefits can be harnesses. The versatility of RIS deployment extends to a vast array of applications, showcasing its potential to enhance wireless communication across different domains.

- Smart cities. Since the sending of RISs might diminish how much organization gear required as shown in figure 4, RISs can be mounted on structures to further develop inclusion, increment ghastly effectiveness, and lessen openness to EM radiation in outside regions (Basar, et al., 2019).



Figure 4.General use of RIS in outdoor environment (Basar, et al., 2019).

- Intelligent Buildings: RIS finds valuable application in smart buildings where it can be utilized to craft enormous windows using specialized glass that facilitates seamless indoor-outdoor and outdoor-indoor connectivity (See figure 5).



Figure 5. General use of RIS in outdoor environment (Basar, et al., 2019).

- Smart vehicles. RIS can be executed in vehicles, trains, and airplane. For instance, car windows and rooftops can be covered with RISs and utilized as moving almost aloof transfers to further develop vehicle-to-vehicle and vehicle-to-foundation interchanges. The inside of the trains can be covered with RISs to work on signal inclusion and breaking point traveler openness to EM fields as shown in figure 6. At long last, the carriers' above receptacles can be covered with RISs, which can serve travelers with high velocity Web while decreasing their EM field openness (Basar, et al., 2019).



Figure 6. Usage of RIS for vehicular communication in urban areas (Basar, et al., 2019).

- Smart Homes: Within smart homes, RIS can be strategically installed on interior walls to enhance device connectivity. Additionally, homes can incorporate smart glass solutions, illustrated in Figure 7, which significantly contribute to improving local connectivity.



Figure 7. Usage of RIS in indoor environment (Basar, et al., 2019).

- **Smart Clothing:** Utilizing metamaterials and integrating intelligent sensors, clothing can be designed to create wearable body networks that monitor individuals' health and well-being. Subsequently, RIS is one of the most encouraging innovations for the up and coming age of interchanges, and it stays a hot region for current review. The reconfigurable intelligent surface (RIS) boasts distinctive features that set it apart in Wireless communication benefits from the integration of reconfigurable intelligent surfaces (RIS) in several ways. Firstly, RIS is environmentally friendly due to its passive nature, allowing for the use of eco-friendly materials in constructing sustainable wireless networks. Secondly, RIS employs a unique scatterer deployment strategy, utilizing sub-wavelength scattering elements, which is uncommon in traditional wireless communication systems. This approach enhances the performance and efficiency of wireless networks. Additionally, RIS finds effective applications in the near-field, enabling the concentration of electromagnetic power in specific regions. This capability enables precise radio localization and facilitates wireless power transfer to recharge low-power device batteries, further enhancing the versatility and utility of RIS in wireless communication scenarios.

- **Channel Capacity Enhancement:** With proper configuration, RIS can augment channel capacity by leveraging a high-rank channel matrix. Conversely, it can employ destructive interference to degrade the signal, safeguarding information from unauthorized users.

1.6. Research Motivation

UAV and RIS provide various advantages in terms of network coverage, network quality, and energy efficiency. Due to severe flight duration and power consumption constraints, there are still certain obstacles in implementing UAV and RIS in real-world applications. Customary ways to deal with asset allotment, direction plan, and RIS stage shift improvement are proposed. Nonetheless, earlier examinations generally ignore the trouble in channel evaluation, versatile clients, power utilization, and data transmission delay. Besides, some troublesome joint enhancement errands are as often as possible viewed as sub-issues. They are irrelevant progressively and huge scope streamlining.

Reinforcement learning (RL) computation is fundamentally designed to empower experts in making a sequence of decisions. The expert depicted in Figure 8 can represent any entity equipped with a processing unit in wireless networks, such as UAVs, servers, ground clients, and more. Through reinforcement learning, these experts interact with the environment to optimize rewards. Consequently, Deep Reinforcement Learning (DRL) approaches have been employed to grant each node the autonomy to enhance the intelligence of the wireless network. The initial step in transforming wireless environments involves creating a digital representation that includes various components such as IoT locations, parameters related to UAV flight (speed, angle, position), and channel models between network elements. To train the experts, Deep Reinforcement Learning (DRL) methodologies are employed. For discrete problems, techniques like Deep Q-Learning (DQL), Double Q-Learning (DDQL), Dueling Deep Q-Learning (Dueling DQL), and Proximal Policy Optimization (PPO) are utilized. On the other hand, for continuous design, Deep Deterministic Policy Gradients (DDPG) and Proximal Policy Optimization (PPO) are applied. Furthermore, to minimize transmission delay between the centralized processing unit and users, multi-agent learning is implemented. This approach aims to enhance the efficiency and effectiveness of wireless communication systems by optimizing the decision-making process and reducing latency.

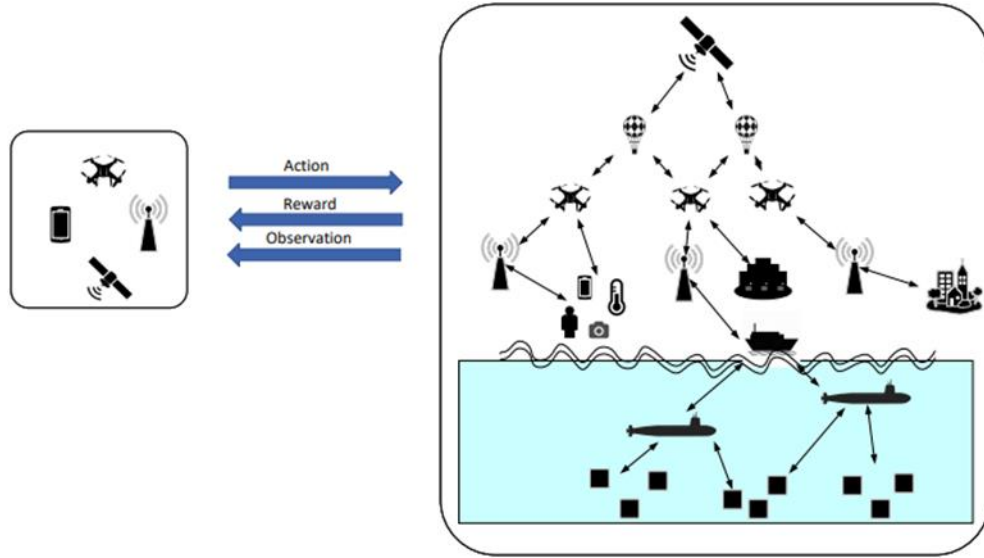


Figure 8. Reinforcement learning model in wireless networks.

1.7. Outline

The thesis study begins with

- Chapter 1, which is the introduction to the topic and it describes the problems statement, aims and objectives, and the research questions that are proposed to achieve the goal of the study.
- In Chapter 2, a thorough background of all the necessary knowledge required to understand and conduct the research study is stated.
- In the next chapter, i.e. the Method, a detailed overview of the complete research design is presented. The method is divided into two parts, namely, the Literature review and the Experiment for answering the research questions.
- In Chapter 5, the results of the literature review and the results of all the experiments performed in the study are presented.

CHAPTER TWO

LITERATURE REVIEW

In Chapter 2, a thorough background of all the necessary knowledge required to understand and conduct the research study is stated.

2.1. RIS for Wireless Communication

In recent years, there has been a growing interest in the utilization of Reconfigurable Intelligent Surfaces (RIS) across various communication settings, both in research and industry. RIS offers numerous advantages, including its straightforward design and adaptability to different technologies, making it a crucial technology for the development of 6G wireless networks. When comparing RIS to the conventional antenna arrays commonly used in 5G, a significant distinction arises from the fact that RIS is not just a static component of the transmitter or receiver. Instead, RIS serves as a modifiable element of the wireless environment itself. This unique characteristic sets RIS apart, enabling it to actively manipulate the wireless signals and optimize the communication performance in ways that traditional antenna arrays cannot achieve. By harnessing the potential of RIS, 6G wireless networks can unlock new levels of flexibility and efficiency, paving the way for advanced and transformative communication technologies of the future (Lee, Jung, Kasgari, Saad, & Bennis, 2020).

2.1.1. RIS based communications.

Reconfigurable Intelligent Surfaces (RIS) have emerged as a promising technology with the potential to revolutionize communication systems. One of the key advantages of RIS lies in its ability to enhance communication by amplifying signal reception at specific destinations and improving the Signal-to-Noise Ratio (SNR) at the receiver. By strategically configuring the surface elements, RIS can manipulate the propagation environment, effectively steering and focusing the wireless signals towards their intended targets. This targeted signal enhancement significantly improves the overall quality of communication, allowing for stronger and more reliable connections. With its ability to actively shape the wireless channel, RIS opens up new possibilities for optimizing signal reception and transmission,

ultimately leading to enhanced communication performance in various applications and scenarios. Beyond that, RIS finds application in localization, sensing, and bolstering physical layer security through the strategic introduction of artificial noise (AN) to unintended receivers. Furthermore, RIS can play a supportive role in existing communication systems, acting as a reflective surface and shaping the reflected signal in a tailored manner to optimize the communication process (Lei, Zhang, Yu, Fowler, & Yu, 2021).

A prime example of the application of a Radiology Information System (RIS) is within a hospital's radiology department . RIS serves as a comprehensive solution for managing and organizing radiological data, including patient information, imaging orders, scheduling, and reporting. When a physician requests a diagnostic imaging procedure, the RIS facilitates the scheduling of appointments, tracks patient data, and ensures seamless communication between different stakeholders involved in the process. Radiologists use the system to access patient histories, interpret images, and generate detailed reports (see figure 9).

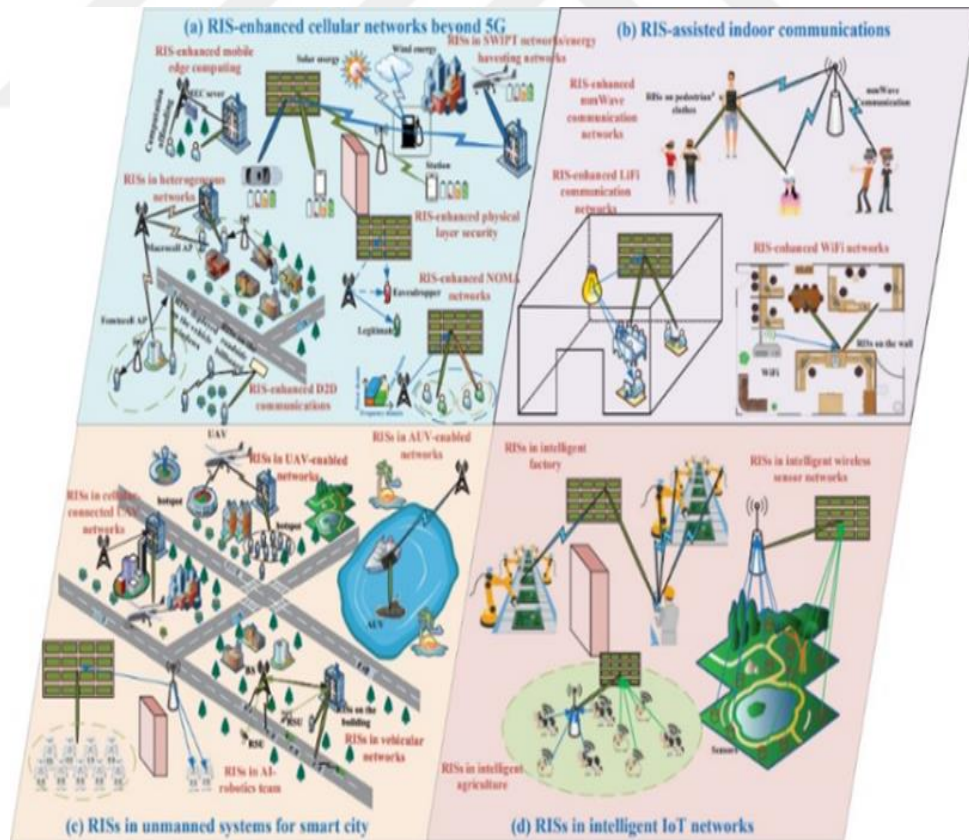


Figure 9. Examples of application of RIS (Balakrishnan, De, & Wang, 2020).

Ambient Backscatter communication, an intriguing application, harnesses ambient radio frequency (RF) signals to facilitate device-to-device communication and power devices without relying on batteries. Leveraging the nearly-passive nature of RIS and its capacity to manipulate incoming waves, there's potential to repurpose existing waves. This innovative approach could play a pivotal role in reducing electromagnetic (EM) pollution, particularly in EM-sensitive environments like hospitals.

2.1.2. RIS-assisted UAV communications

The next generation of communications is set to witness a significant breakthrough with the integration of unmanned aerial vehicles (UAVs) into the network. UAVs offer a range of distinct advantages that make them a crucial advancement in this field. One such advantage is their heightened flexibility in deployment and control, allowing them to be easily maneuvered and positioned in different locations according to the communication requirements. Additionally, UAVs boast minimal operational and maintenance expenses compared to traditional infrastructure, making them a cost-effective solution. Furthermore, UAVs have the unique ability to establish line-of-sight (LoS) connections with various nodes, enabling direct and efficient communication between different points in the network. With these advantageous features, UAVs hold immense potential for revolutionizing the future of communication systems, providing enhanced connectivity, improved coverage, and increased reliability in a variety of applications and scenarios (Mei, Yang, Shen, & Liu).



Figure 10. RIS-assisted UAV-aided communications (Mei, Yang, Shen, & Liu).

UAV-assisted RIS presents various strategies for enhancing communication systems:

- **Enhanced coverage:** Utilizing a specialized intelligent surface known as an omni-surface (IOS), this approach employs antennas on both sides capable of reflecting incoming signals from opposing directions. One of the key features that sets the Intelligent Operating System (IOS) apart is its unique capability to process incoming signals from all directions, enabling comprehensive 360-degree coverage and enhancing spectral efficiency. Unlike traditional systems, the IOS can intelligently analyze and utilize signals from both sides, allowing for a more effective utilization of the wireless environment. This means that the IOS can dynamically adjust and optimize the outgoing signal from the Reconfigurable Intelligent Surface (RIS) to direct it towards potential recipients without creating any blind spots. By intelligently steering the outgoing signal, the IOS ensures that the communication is not limited to a specific direction or location, but rather provides seamless coverage and connectivity in all directions. This breakthrough technology holds great promise for improving wireless communication systems, enabling enhanced connectivity, and maximizing the utilization of available resources.

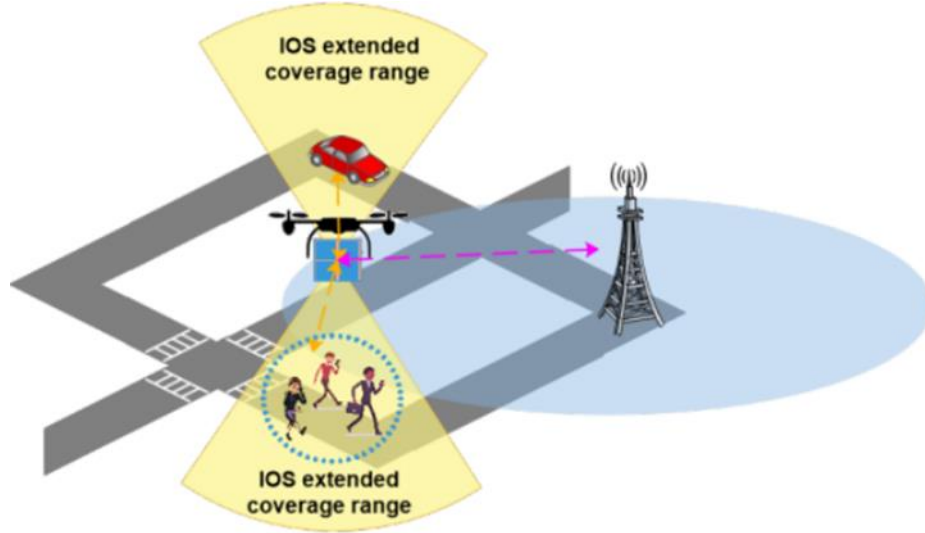


Figure 11. RIS-assisted UAV for extended coverage (Mei, Yang, Shen, & Liu).

- **Spectrum sharing:** Range sharing can be empowered by means of RIS. Range sharing is a technique for improving the utilization of the wireless transmissions, or remote interchanges channels, by permitting many sorts of clients to have a similar recurrence groups securely. Thus, RIS can lessen impedance between gadgets communicating in a similar recurrence band simultaneously. Besides, by using range sharing, RIS-helped UAV frameworks can increment framework limit (Wu & Zhang, 2020). The inherently passive nature of RIS results in significantly lower energy consumption compared to conventional spectrum sharing techniques.

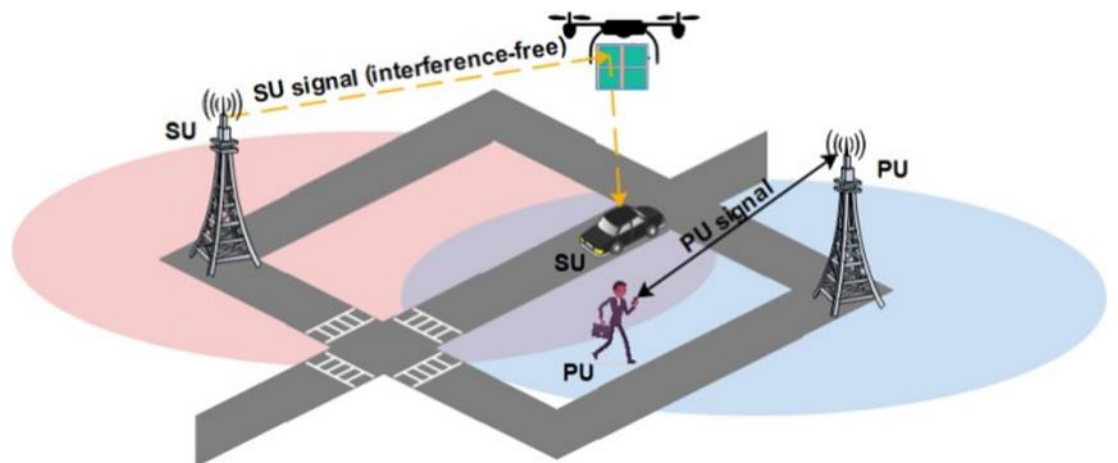


Figure 12. RIS-UAV for spectrum sharing (Mei, Yang, Shen, & Liu).

- **Enhanced capacity and throughput** are key benefits of deploying RIS with UAVs. The passive nature of RIS, when combined with UAVs, overcomes

challenges related to antenna noise amplification and self-interference, leading to reduced computational needs and lower power consumption compared to active full-duplex (FD) transmissions. Additionally, integrating RIS with UAVs enhances the antenna elements' phase dynamics, resulting in abundant line-of-sight (LoS) communications and further optimizing capacity and throughput.

- **RIS-Assisted UAV Communications for physical layer security (PLS):** Various types of PLS can be given by RIS-helped UAVs. It can, for instance, be utilized to produce bogus commotion to confound possible aggressors while safeguarding clients' protection and information. Moreover, by expanding the communicate power, it very well may be utilized as an aeronautical transfer among approved and illegal clients. Moreover, RIS-helped UAVs can be used to diminish the information rate for snoops and to battle complex aggressors (Mei, Yang, Shen, & Liu). It was illustrated, specifically, that the mystery pace of real clients develops as the distance between correspondence peers diminishes. Because of UAVs' opportunity of versatility, it is feasible to diminish the distance between the transmission source and a solitary client.

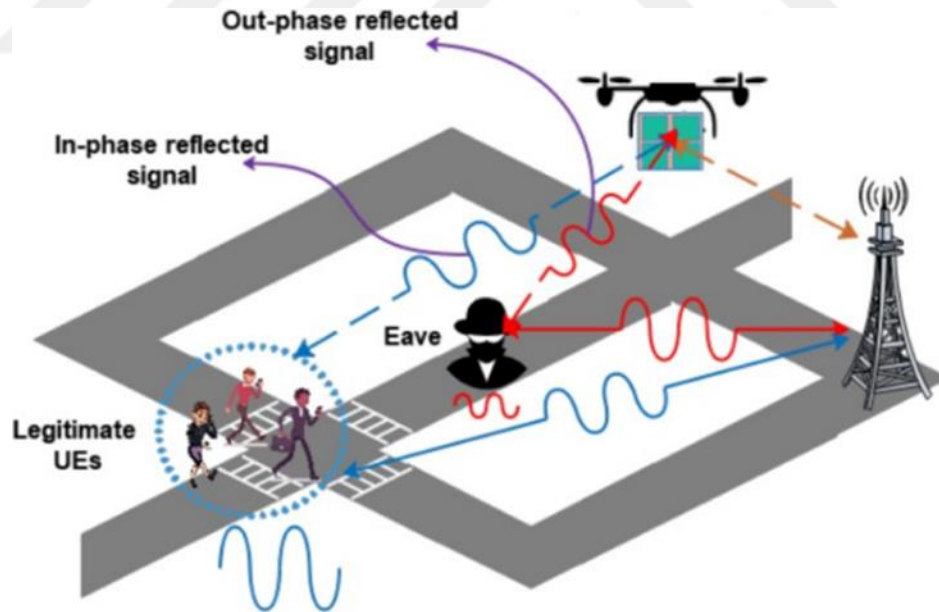


Figure 13. RIS-UAV for physical layer security (Mei, Yang, Shen, & Liu).

2.2. Application of UAV Assisted RIS

Current research is actively exploring new technologies to fulfill the demands of upcoming 6G wireless networks. Integration of RIS-assisted UAVs with these

technologies is a focal point. Various approaches are being studied to seamlessly integrate RIS-aided UAVs into the fabric of 6G networks, ensuring they play a pivotal role in enhancing communication, coverage, capacity, and overall performance of these advanced networks. The goal is to leverage the unique capabilities of RIS and UAVs to meet the evolving needs and expectations of the 6G era, optimizing network efficiency and enabling a wide array of groundbreaking applications. The RIS-UAV correspondence framework, which utilizes AI procedures (ML) and man-made brainpower, is one of the essential thoughts for future organizations. This sort of headway will empower UAVs to pursue independent choices, remove information, and figure results. ML calculations can be utilized to further develop RIS channel gauges, otherworldly proficiency, balance different tradeoffs, and expand the extent of existing streamlining methods.

The use of recurrence groups over 100 GHz is one more component of expected 6G organizations. The expanded blockage pace of sent signals, as well as the steadfastness and accessibility of remote correspondences administrations, will affect mm-Wave and Terahertz (THz) interchanges. The utilization of RIS-helped UAVs might be a feasible answer for these issues.

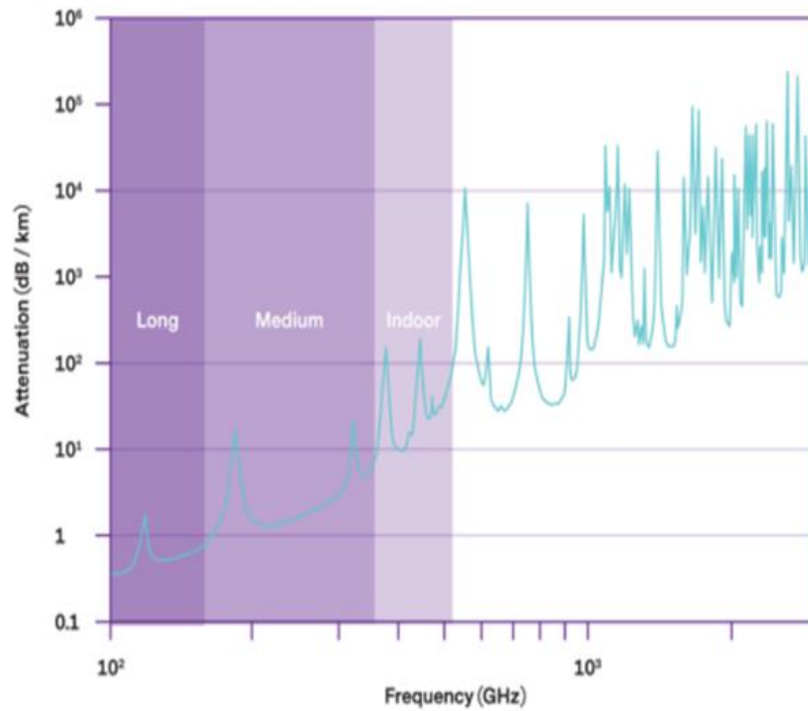


Figure 14. Illustration of the relation between frequency range and radio channel propagation effects (Al-Hourani, Kandeepan, & Jamalipour, 2014).

2.3. Wireless Power Transfer

In the foreseeable future, the number of IoT devices is anticipated to escalate, reaching tens of billions by 2030. Traditional methods of managing this vast array of devices, such as power cords and batteries, are no longer adequate to cater to their burgeoning numbers and the associated complexities of maintenance and operation. Hence, novel and more sophisticated approaches are imperative to ensure effective control and optimal functioning of this rapidly expanding IoT ecosystem. The critical answer for this issue is Radio Recurrence (RF) remote power move (WPT). The transmission of electrical energy without the utilization of an actual connection is what remote power move is about. RF Wireless Power Transfer (WPT) presents a notable advantage with its capability for long charging distances. However, its efficiency diminishes as the distance increases. RIS comes as an efficient solution to this drawback, providing the ability for beamforming and beam focusing without the need for active components. This characteristic makes RIS a cost-effective and energy-efficient alternative. Illustrated in Figure 2.7, RIS integration into building walls or spaces can significantly enhance the efficiency of power transfer.



Figure 15. RIS-aided WPT system conceptual application scenario.

The RIS is placed in the roof of a mechanized production line to help the WPT framework in this situation. The RIS mirrors an EM pillar from a power reference point and afterward focuses the considered wave the ground gadgets. RIS can be

utilized in remote data and power move simultaneously (SWIPT). The sign is used to move both energy and data in these frameworks. SWIPT can be executed in two ways:

Power splitting, an essential technique, involves dividing the received signal into two distinct parts within the power domain. Each part serves a specific purpose—one for energy harvesting and the other for information transfer. This approach efficiently utilizes the received signal for both extracting power and transmitting essential information, optimizing the overall communication and energy harvesting process.

- Time switching: In particular time windows, it changes between energy gathering and data move. Because of its one of a kind versatility, simplicity of sending, and minimal expense, UAV innovation could likewise be utilized to help the SWIPT.

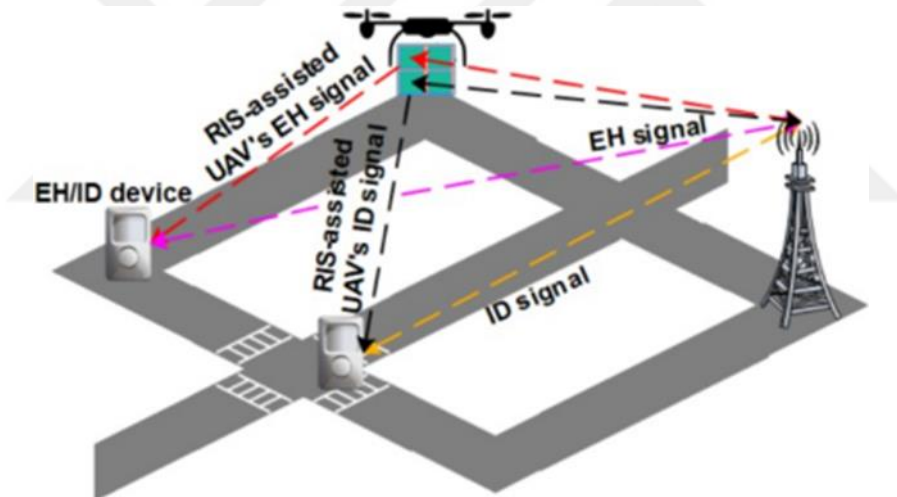


Figure 16. RIS-assisted UAV for SWIPT (Mei, Yang, Shen, & Liu).

2.4. Localization and Mapping

Localization refers to a device's capability to ascertain its own position or the positions of other devices in its operational surroundings. Various technologies are employed in this research domain, with radio-based methods like radar, WiFi, satellite positioning, and cellular localization being prominent choices. Radio-based technologies are favored due to their resilience to environmental influences and their dual functionality in sensing and communication. Starting from the era of 4G, signals dedicated to localization achieved accuracies of about 10 meters. The advent of 5G,

with its larger bandwidths and higher carrier frequencies, further improved accuracy to approximately 1 meter. The forthcoming generation of systems holds the promise of refining accuracy even further. However, propagation at higher frequencies is significantly impacted by obstacles along the path between the transmitter and the receiver, resulting in high path loss for these frequencies.

The essence of radio localization and mapping revolves around three fundamental principles: measurements, the reference system, and inference algorithms. Measurements involve collecting data from the radio signals, which are crucial for determining positions and creating maps. The reference system provides a standardized framework or set of known reference points against which the collected measurements can be compared and interpreted accurately. Finally, inference algorithms analyze the measurements in relation to the reference system to estimate positions, distances, or other relevant parameters, ultimately contributing to the localization and mapping process. They are as follows:

- **Measurements:** The radio waves between a transmitter and a recipient are estimated quickly from the channel gauge technique. These actions are recognized by two qualities: goal and exactness. The ability to recognize two signs in light of their estimations is alluded to as goal, and not set in stone by the sign transfer speed and term, transporter recurrence, number of radio wires, and rationality combination time. Then again, exactness alludes to the sum to which we can decide the boundary of interest.

- **Reference system:** The reference system plays a critical role in ensuring consistency, where all measurements are taken in relation to anchor points with precisely known states. These anchor points provide a reliable foundation for making accurate measurements. Additionally, the geometric arrangement and placement of the reference points are essential factors, significantly impacting the overall precision and accuracy of a localization system.

- **Inference algorithms:** In the realm of localization, algorithms are distinct from communication algorithms in that they primarily utilize Line of Sight (LoS) signals. Unlike communication algorithms that leverage multipath signals, localization algorithms focus on LoS signals for measurements, as these signals offer direct and precise information regarding the user's location.

2.5. Application of RIS for Localization

The integration of RIS into the process of localization significantly impacts all three fundamental principles mentioned earlier. In the context of localization, two key factors come into play. Firstly, the measurements are influenced by the geometric channel model, which allows for the establishment of a mapping between position space and measurement space based on the locations and electromagnetic (EM) properties of the surrounding environment. This mapping helps in accurately determining the position of objects or users. Secondly, the references for localization can be established through the interaction between Base Stations (BS) or Reconfigurable Intelligent Surfaces (RIS) acting as anchor points, and users or passive objects that lack known location and orientation information. The integration of RIS in localization has attracted significant research attention due to its potential benefits. However, this integration also brings forth various challenges that need to be addressed. These challenges include issues related to signal interference, channel estimation, synchronization, and the need for efficient algorithms to handle the complexity introduced by RIS. Overcoming these challenges is crucial for unlocking the full potential of RIS in localization applications:

- **Near-field propagation:** Signal propagation is significantly influenced by the RF region in which users are situated. While localization algorithms are readily applicable with RIS in far-field scenarios, near-field cases present challenges due to the impact of wavefront curvature on propagation. This necessitates the need for novel signal design approaches to effectively utilize near-field properties in the context of localization algorithms.

- **Channel modeling:** Creating accurate channel models for RIS poses a significant challenge, primarily because RIS behavior is highly dependent on operational frequencies. Models will likely vary based on the frequency band, necessitating different approaches for each band. However, for multi-band operations, it is advantageous to develop models that maintain consistency and reliability over a broad frequency range.

- **Localization and mapping algorithms** can greatly benefit from the utilization of Reconfigurable Intelligent Surfaces (RIS) due to their known area and direction. This makes RIS particularly valuable for confinement and planning calculations, as it

enables the accurate recovery of the user's location. To successfully determine the user's state, which includes their 3D position, 3D direction, and clock bias, a sufficient number of resolvable signal paths must be available. By effectively managing the RIS, it can provide significant advantages, leading to a reduction in data associations and improving the overall efficiency of localization and mapping algorithms. The integration of RIS in these algorithms holds immense potential for enhancing localization accuracy and streamlining the processing of location-related data.



Figure 17. Application examples of RIS-based localization and mapping services (Ju & Zhang, 2013).

An introduction to meta-surfaces and their application in designing Reconfigurable Intelligent Surfaces (RIS) is provided. RIS presents significant advancements to meet the demanding needs of future wireless networks due to its nearly passive nature and unique physical characteristics. Notably, RIS has versatile implementations; it can act as a covering material for creating smart cities, vehicles, clothing, and primarily to enhance communication. One key approach is utilizing RIS to aid Unmanned Aerial Vehicles (UAVs). Another application involves Wireless Power Transfer (WPT), enabling energy delivery to challenging areas. Moreover, the technology can extend to include information transfer for powering devices in such areas. Overall, considering its diverse applications, RIS is relatively cost-effective, with manageable complexity, making it suitable for deployment in the current wireless environment while promoting sustainability.

2.6. Reconfigurable Intelligent Surfaces-Aided Wireless Networks

Reconfigurable Intelligent Surfaces (RIS) have recently emerged as a highly promising technology for revolutionizing connectivity and inclusivity in communication systems. The fundamental principle behind RIS is the ability to

manipulate the phase shift mechanism and active components to reflect incoming signals towards the intended receivers. This intelligent manipulation allows for the enhancement of received signals, resulting in improved signal strength and quality. Additionally, by precisely controlling the reflection properties of the RIS, interference from unexpected sources can be minimized, thereby optimizing the overall communication performance. With its transformative potential, RIS technology holds the key to unlocking new possibilities in communication services, ensuring seamless connectivity and enabling a more inclusive and reliable communication experience for users. Moreover, the RIS's classification and zero-delay are key benefits in working with trustworthy and minimal expense distant correspondences.

RISs have arisen as another strategy in light of the fact that to their capability to change far off correspondences. RISs have different benefits, including straightforward arrangement and long haul exercises, expanded limit, unpleasant adequacy and energy efficiency, versatile reconfiguration, and similitude. Due to the minimal expense of the inactive scattering parts, RISs can be effectively embedded and supplanted. They can be associated with high designs, rooftops, vehicles, and different plans, as well as UAVs, to give better cell sorts of help and to extend network consideration for a minimal price. The raised areas of the RISs can help with the work on the gained sign and snag decrease. Accordingly, limit, energy capability, and max-min sensibility among clients are extraordinarily further developed in RIS-helped remote organizations. Moreover, the RIS's stage shift design can be arranged to reflect the sign in different headings. The compelling movement of stage shift can altogether further develop network execution. Moreover, the RISs may be joined with other creating advancements, for instance, UAV, to furnish clients with additional reliable and speedy organizations.

RISs are utilized in different courses in far off organizations. RISs can be utilized in cell organizations to conquer snags and further develop accessibility from the base station to clients. Moreover, RISs can be utilized to work on signal strength and decrease impedance in gadget to-device joint efforts. RISs, then again, can act as trustworthy focus layers in genuine layer security by dropping undesirable signs. Moreover, the RIS is appended to the UAV to furnish ground clients with an on-

request and rapid raised network. The RIS can likewise be utilized to help with the advancement of savvy urban areas. Independent vehicles (Abedin, Munir, Tran, Han, & Hong, 2021), and insightful far off sensor organizations.

2.7. Deep Reinforcement Learning in Wireless Networks

As of late, AI has been proposed for keen help of UAVs and other organization gadgets. Support learning (RL) can utilize experimentation figuring out how to track down the smartest idea. In any case, getting an ideal procedure while considering a wide state and activity space is challenging for without model RL calculations, for example, the Q-learning strategy. Luckily, the refined mix of RL and profound learning, named profound support learning (DRL), is perfectly reasonable for taking care of high-layered issues utilizing developing brain organizations. Therefore, DRL calculations have found far reaching use in disciplines, for example, advanced mechanics (Polvara, , Sharma, , Wan, , Manning, , & SUTTON, 2019). Business the executives (Lee, et al., 2019), and gaming (Wang, Wang, & Zhang, A deep reinforcement learning approach to flocking and navigation of uavs in large-scale complex environments., 2019). DRL has as of late acquired ubiquity in remote organization critical thinking because of its thinking skill and adaptable collaboration with the climate. DRL, for instance, has been used to address hardships in asset allotment, route and impedance the board (Peake, , McCalmon,, Zhang, , & Raiford, , 2020). The Markov choice cycle (MDP) (Moon, , Papaioannou, , Laoudias, , & Kolios, , 2020) and the RL calculation are presented in this segment.

2.7.1. An Overview of Markov Decision Processes

The Markov Decision Process (MDP) is a mathematical framework that describes a decision-making problem with uncertain outcomes. It is represented by a 4-tuple, denoted as $\langle S, A, P, R \rangle$, where S represents the finite set of states, A represents the finite set of actions, P is the transition probability function, and R is the reward function. The transition probability function, $P_{ss_0}(a)$, specifies the probability of transitioning from state s to state s_0 after taking action a . The reward function, R , determines the immediate reward received after executing an action. The objective of an MDP is to find an optimal policy, denoted as π^* , that maximizes the cumulative reward over time. In other words, the goal is to find the policy that leads

to the highest possible total reward. By determining the optimal policy, one can make informed decisions and take actions that maximize the expected cumulative reward in a given MDP:

$$\max E_{\pi} [\sum_{t=1}^T \gamma R(s', s, a)] \quad (1)$$

Where $E[\cdot]$ is the expectation function, γ is the discounting factor.

- $\max$: This represents the maximum operator, indicating that we are taking the maximum value.
- $E(\pi)$: This is the expectation operator with respect to a policy denoted by π . It calculates the expected value.
- $\sum_{t=1}^T \gamma R(s', s, a)$: This is a summation symbol, where t ranges from 1 to T . It represents the sum over a certain number of time steps.
- γ : This is the discount factor, which determines the importance of future rewards. It is a value between 0 and 1.
- $R(s', s, a)$: This represents the immediate reward obtained when taking action a in state s that leads to state ' s' '.

2.7.2. A Brief Overview of Deep Reinforcement Learning

At least one specialists associate with the climate and learns through cooperation in the RL calculation. The pith of RL is experimentation learning, in which the specialist notices the state and plays out an activity toward the climate to change its conduct because of remunerations obtained. Profound learning permits RL to take care of additional intricate issues with a high-layered state and activity space.

When it comes to solving Reinforcement Learning (RL) problems, there are two primary approaches: value-based methods and policy search methods. Value-based methods, such as deep Q-learning, SARSA, and double deep Q-learning, focus on estimating the value function associated with each state-action pair. By iteratively updating these value estimates, these methods aim to find the optimal policy. On the other hand, policy search methods, like vanilla policy gradient and proximal policy optimization, directly search for the optimal policy by iteratively adjusting the policy parameters based on the observed rewards. These methods aim to optimize the policy's performance without explicitly estimating the value function. Additionally,

there exists a hybrid approach called the actor-critic method, which combines elements of both value-based and policy search methods. This approach, exemplified by the deep deterministic policy gradient algorithm, utilizes both value function approximation and policy optimization to find the optimal policy in RL problems. By leveraging the strengths of both approaches, the actor-critic method provides a powerful framework for solving complex RL tasks. Profound brain networks are utilized in DRL to surmise the worth capability V , the activity esteem capability Q , the benefit capability A , and the strategy. To modify the worth capability in the worth pursuit technique, we dissect the distinction between the got prizes in two examples. We straightforwardly find the approach for the issues utilizing the arrangement search strategy.

This chapter provides an overview of the reinforcement learning frameworks employed in the suggested approach. To model our situation, we propose Markov Decision Processes (MDP). Then we go over the two DRL calculations that were utilized to handle this issue. We dissect a limited skyline MDP characterized by the tuple (S, A, R, P) , where S is the state space, A is the activity space, R is the prize capability, and P is the state change likelihood. The specialist communicates with its environmental factors by executing exercises in discrete time steps and afterward getting a prize. At each time step t , the specialist notices state s_t , acts a_t , gets a prize r_t in view of s_t and a_t , and afterward moves to the following state s_{t+1} with the likelihood $P(s_{t+1}|s_t, a_t)$.

The objective of RL is to plan a strategy that expands potential compensations while not realizing the state change likelihood, which is addressed by the state progress tuple (s_t, a_t, r_t, s_{t+1}) . Contingent upon whether it is deterministic or stochastic, an approach (s) is depicted as a planning from state s to an activity or a dispersion of activities. Truth be told, the strategy oversees the specialist's exercises and, accordingly, the prizes it gets from its environmental elements. The return R_t is characterized as the limited all out of potential compensations from the current state s_t to the terminal state at time T :

$$R_t = \sum_{r=t}^T \gamma^{r-t} r_r, \quad (2)$$

Where γ is the rebate factor, which can go from 0 to 1.

- R_t : This is the cumulative discounted reward starting from time step t .
- $\sum$: This is a summation symbol, where r ranges from t to T . It represents the sum over time steps from t to T .
- γ^{r-t} : This is the discount factor raised to the power of $(r-t)$. The discount factor γ determines the importance of future rewards, and raising it to the power of $(r-t)$ ensures that rewards further in the future are given less weight.
- r_r : This represents the reward obtained at time step r .

The markdown factor gauges the significance of quick and future additions. Higher upsides of show that the specialist is more worried about future advantages, while lower values show the inverse. We need to perceive a technique that supports the ordinary return from the hidden time step, which is imparted as:

$$Q^\pi(s, a) = E[R_t | s_t = s, a_t = a] \quad (3)$$

The expression 3 represents the expected cumulative reward, conditioned on starting in state s , taking action a , and following policy π . Let's break down each symbol:

- $Q^\pi(s, a)$: This is the action-value function, commonly known as the Q-function, for a given state s and action a under policy π . It represents the expected cumulative reward for taking action a in state s and then following policy π .
- $E[\cdot]$: This denotes the expectation operator, representing the average or expected value of the quantity inside the brackets.
- R_t : This is the cumulative reward, i.e., the sum of rewards obtained from time step t onwards.
- $||$: This vertical bar denotes conditioning. In this context, it signifies that the expectation is conditioned on the event $st=s$ and $at=a$, meaning that the calculation considers only scenarios where the current state is s and the chosen action is a .

Q-learning (QL) (Wang, Wang, Shen, & Zhang, 2019) is a notable RL calculation that tries to distinguish the ideal Q-values for each state-activity pair (s, a) by iteratively expanding the assessed Q-capability with the accompanying update.

$$Q(S_t, \alpha_t) = Q(S_t, \alpha_t) + \alpha \left[r_t + \gamma \max_{a_{t+1}} Q(S_{t+1}, a_{t+1}) - Q(S_t, \alpha_t) \right] \quad (4)$$

- $Q(S_t, \alpha_t)$: This represents the current estimate of the action-value (Q) function for state S_t and action α_t .
- α : This is the learning rate, a small positive constant that determines the step size of each update. It controls the impact of new information on the current estimate.
- r_t : This is the immediate reward received after taking action α_t in state S_t .
- γ : This is the discount factor, a value between 0 and 1 that determines the importance of future rewards. It discounts future rewards, giving more weight to immediate rewards.
- $r_t + \gamma \max_{a_{t+1}} Q(S_{t+1}, a_{t+1}) - Q(S_t, \alpha_t)$: This represents the maximum estimated future reward for the next state S_{t+1} , considering all possible actions a_{t+1} .
- $Q(S_{t+1}, a_{t+1})$: This is the current estimate of the action-value function for the next state S_{t+1} and the chosen action a_{t+1} .

To guarantee that different state-activity pairings are investigated during preparing, the specialist utilizes a-covetous strategy that plays out an irregular activity with chance and the eager activity $\arg \max_a Q(s, a)$. In the wake of deciding all that Q-esteems, the ideal approach can be acknowledged by taking part in avaricious way of behaving, as communicated in:

$$\pi^*(s) = \arg \max_a Q^*(s, a) \quad (5)$$

In any case, QL is wasteful when the state and activity spaces are enormous since the Q-an incentive for each state-activity pair should be put away in a table. DRL approaches can be utilized to address the scourge of dimensionality by combining profound brain organizations (DNNs) and RL. Presently, we will take a gander at the two DRL calculations utilized in this proposition, DQL and DDPG.

DQL was presented in, which utilizes two DNNs, the web-based network and the objective organization, to rough the Q() capability by diminishing the misfortune L:

$$L(\theta^Q) = \mathbb{E} \left[(Q(s_t, a_t | \theta^Q) - y_t)^2 \right], \quad (6)$$

Where θ^Q denotes the weights of the online DNN, and y_t known as the target is expressed as:

$$y_t = r_t + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1} | \theta^{Q'}) \quad (7)$$

Where $\theta^{Q'}$ is the objective organization's loads, which have a similar construction as the web-based DNN. Be that as it may, its loads are imitated from the web-based network consistently and are held in the middle between to guarantee the objective qualities' security. Moreover, preparing tests in RL are corresponded, which doesn't match the autonomous and indistinguishable appropriation prerequisite for DNN preparing. To resolve this issue, the experience replay technique was executed. We explicitly utilize a replay cradle to hold the state progress tuples (s_t, a_t, r_t, s_{t+1}) .

To refresh the DNN, we utilize a small clump haphazardly picked from the replay support. In light of the haphazardness of the examples, experience replay can diminish the relationship between successively created tests and further develop the DNN's assembly conduct. Moreover, gaining from tests a few times increments information effectiveness. Figure 18 portrays a prepared Q-organization, where the brain network yields the Q-values for each activity.

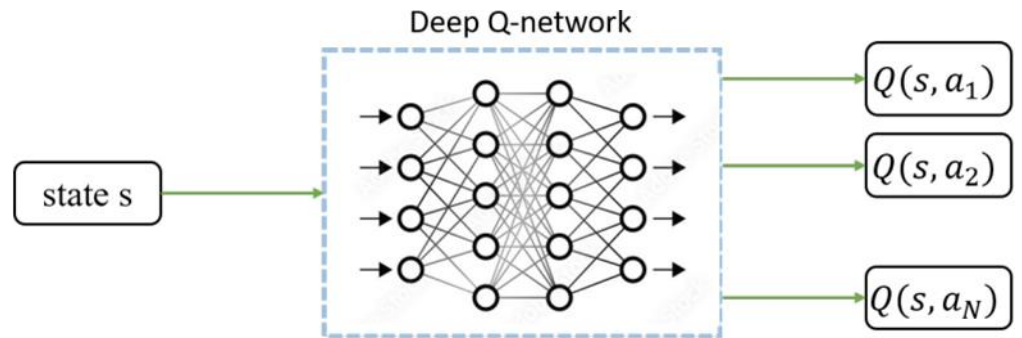


Figure 18. Deep Q-learning

The constraint of DQL is that it is simply material to control issues with a discrete and low-layered activity space. It is hard to apply DQL to ceaseless control

since it should decide the activity that enhances the $Q()$ capability in (3.5), which is a difficult errand. UAV direction arranging and sensor power control, then again, are both ceaseless control issues.

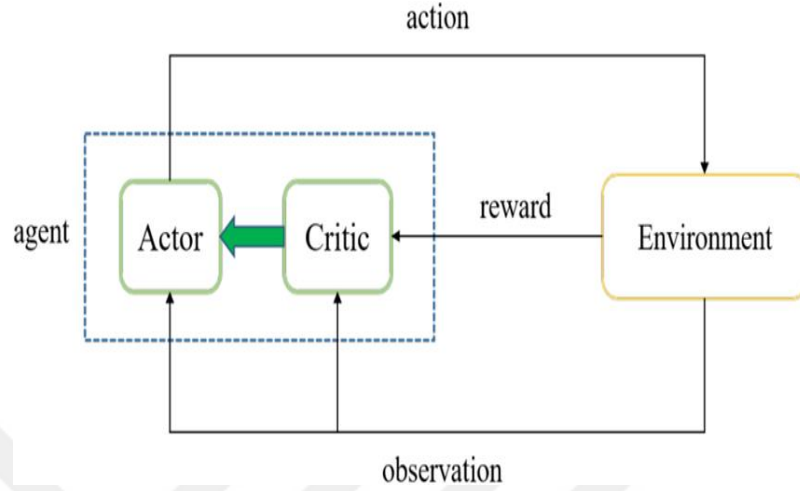


Figure 19. Actor-Critic architecture

DDPG (Samir, Elhattab, Assi, Sharafeddine, & Ghrayeb, 2021). was made as an entertainer pundit approach consequently. Figure 19 portrays the entertainer pundit engineering, in which both the entertainer and the pundit are executed by DNNs. The pundit is gone about as $Q(s_t, a_t | Q)$ with boundaries Q and trains involving a similar update as DQL. The result of the pundit is given to the entertainer network as input on the first activity. To make an activity given state s_t , the entertainer utilizes a deterministic defined arrangement $(s_t |)$ with boundaries.

2.8. Related Works

In remote power transmission (Zou, et al., 2020). And versatile edge figuring (GUO, et al., 2021)[58], the proficiency of RIS is additionally analyzed. The creators of made a period exchanging convention for a RIS that incorporates an energy gathering stage to charge the RIS capacitor and a sign reflecting stage to help transmission from the passageway (AP) to the recipients. The send beamforming of the AP, the stage booking of the RIS, and the detached beamforming were completely improved to amplify the data rate. The standard semi-clear unwinding approach and monotonic improvement were utilized to take care of two sub-issues. The send precoding frameworks of the BS and the uninvolved stage shift lattice of the RIS were enhanced in to boost the weight aggregate pace of all data recipients in

power move situations. The RIS is likewise used to help with D2D interchanges (Camci, & Kayacan, , 2019). The makers of (Zhang, et al., 2020) saw two sub-issues with fixed idle beamforming vectors at the transmitters and fixed stage shift networks at the RIS. To take care of these sub-issues, the slope approach and a nearby pursuit calculation were introduced.

DRL calculations have been used to give every hub some independence by depending on their dynamic limit (Bellman, , 1957). (Lee, & Kim, , 2020) Inspected an ideal DRL-based channel access way to deal with amplify aggregate rate and -reasonableness. DRL methods were utilized in to further develop the energy proficiency of D2D correspondences. Reserving concerns were thought about further in (Bouhamed, Ghazzai, Besbes, & Massoud, 2020) to amplify store achievement hit rate and limit transmission time. The creators made a concentrated and a decentralized framework model and used an entertainer pundit calculation to decide the smartest strategy.

DRL algorithms have shown significant success in addressing resource management challenges in UAV-assisted wireless networks (Wang, , et al., 2019). In (Wang, Liu, Liu, Thompson, & Kadoch, 2019), multiple UAVs were deployed to assist the cellular network, aiming to enhance the network's overall throughput. An iterative approach was employed involving two stages: leveraging Q-learning computation and a variant of convex optimization to optimize UAV locations, implement beamforming, and manage UAV-user interactions. Another application highlighted in (Ye, Li, & Juang, 2019) involved using UAVs for charging ground equipment and data collection through the Wireless Power Transfer (WPT) approach.. To accomplish the three objectives of aggregate rate augmentation, reaped energy expansion, and energy utilization minimization, the multi-objective profound deterministic strategy angle approach was introduced.

Deep Reinforcement Learning (DRL) algorithms have not only been applied to various domains but also to path planning in UAV-assisted wireless communications. This application has been discussed in several research papers, including references. In one particular study, the authors presented a DRL algorithm that tackles the problem of determining optimal flight paths, transmission power levels, and the corresponding cell assignments in UAV-powered wireless networks. The algorithm

builds upon the resonance state matrix approach described in a previous work (Shin, Kang, & Kim, 2019). By leveraging DRL techniques, this algorithm offers a promising solution for optimizing the performance and efficiency of UAV-assisted wireless communication systems, ultimately enhancing their overall capabilities and effectiveness. In (Wang, Fan, Han, & Pan, 2020), the deterministic approach angle calculation of (Wang, Wang, Zhang, & Zhang, Autonomous navigation of UAV in large-scale unknown complex environment with deep reinforcement learning, 2017) was utilized for UAV-helped cell organizations. The direction of the UAV was intended to expand the uplink aggregate rate accomplished without information on the client's situation or send power. In addition to the aforementioned study, there are several other studies that have applied Deep Q-Learning (DQL) algorithms to address various challenges in UAV-assisted wireless communications. For instance, in reference (Li, Li, Sanyal, Wang, & Qiu, 2019), the authors utilized the DQL algorithm to optimize UAV routing decisions based on received signal powers estimated using a massive MIMO (Multiple-Input Multiple-Output) approach. Another study (Akhroufi, , Arola, , & Bonnet, , 2019) employed Q-learning in the context of controlling the movement of multiple UAVs in two different scenarios: one with static client locations and another with dynamic client locations following a random walk model. Additionally, the authors of reference (B., Z., , W., & B., 2018) proposed a DQL-based approach to mitigate data packet loss in UAV-assisted power transfer and data gathering systems. Lastly, in reference, a multi-agent Deep Reinforcement Learning (DRL) approach was applied for trajectory planning and model selection within a UAV cellular network. These studies collectively demonstrate the versatility and potential of DQL and DRL techniques in addressing various challenges and optimizing performance in UAV-assisted wireless communication systems. In any case, the previously mentioned commitments have not tended to the extreme examination errand of upgrading the joint direction and information get-together of UAV-helped networks. Moreover, these past exploration to a great extent overlooked obstruction, 3D direction, and dynamic climate.

The prerequisite for a procedure that is adaptable and versatile to changes in the climate while meeting genuine cutoff points is expanding, and DRL calculations are among the most encouraging ways to deal with address these issues in remote organizations. DRL calculations have as of late been applied with promising

outcomes in RIS-helped remote organizations (Cui & Wang, 2021). In (Li, K., Ni, W., Duan, L., , Abolhasan,, & Niu, , 2017), the power designation and stage shift enhancement were improved to augment the aggregate rate. A RIS-helped UAV was utilized to serve ground clients in (Chen, Zhang, Liang, Kang, & Zhang, 2019). The direction and stage shift streamlining in view of DRL was proposed to expand the aggregate rate and decency, everything being equal. The producers of used a RIS to help with safeguarding correspondences against sneaks around. Deep Reinforcement Learning (DRL) algorithms have been employed to optimize beamforming techniques in base stations (BS), with the inclusion of reflecting beamforming in reconfigurable intelligent surfaces (RIS) being identified as a means to improve secrecy rates and enhance quality-of-service satisfaction. An article referenced as discusses the development of a deep deterministic policy gradient algorithm that aims to find the optimal phase shift grid configuration for the RIS. The objective is to maximize the received signal-to-noise ratio (SNR) in a multiple-input single-output (MISO) system. Another study referenced as (Basar, et al., 2019) . explores the potential benefits of power enhancement and coordinated phase shift optimization in a multi-UAV-assisted network. By leveraging DRL techniques, these studies contribute to the advancement of beamforming strategies, particularly in the context of RIS, to improve wireless communication performance and meet the increasing demands of modern networks.

Since DRL is a superb strategy for managing a powerful climate with ceaseless moving, a few late examinations have investigated the productivity of DRL procedures for RIS-helped remote organizations. The maker of used the DRL estimation to update the send beamforming vector and the RIS stage shift model to increase indisputably the total rate. In the MISO correspondences, a significant Q-learning and significant deterministic procedure slant were made and shown extraordinary results. The makers of (Imanberdiyev, , Kayacan, , & Chen, , 2016) proposed a DRL estimation to change the UAV's level and the RIS stage shift to restrict the total season of-information.

Table 1. A comparison with existing literature

	[5]	[29]	[46]	[77]	[91]	[32]	[92]	[124]	[93]	[94]
3D trajectory				✓			✓			✓
Sum-rate maximisation	✓		✓	✓				✓	✓	✓
Time minimisation		✓	✓				✓		✓	
Dynamic environment				✓	✓			✓		
Unknown users										
Reinforcement learning			✓	✓		✓		✓		
Deep neural networks			✓			✓		✓		

2.9. Contributions

As recently expressed, it is basic to upgrade dynamic time to fulfill the severe guidelines of UAV-helped remote organizations. In Section 3, we propose new methodologies in view of the DRL calculation for planning the UAV direction and allotting assets to amplify the EE and organization aggregate rate. For a compromise issue, we offer a procedure in view of the DQL and the dueling DQL technique. Our proposed approaches give the most conservative direction and total rate while meeting the entirety of the serious distance, flight time, and correspondence prerequisites. In Table 1, we strikingly and straightforwardly contrast our proposed arrangement with the cutting edge. Our proposed approach might work in a powerful climate with clients who are haphazardly dispersed and versatile. The assigned capability is versatile to various missions of quick sending and aggregate rate boost. Moreover, the previously mentioned works commonly accept amazing CSI and static client conditions.

CHAPTER THREE

METHODOLOGY

The methods proposed for this thesis work, as well as the overall conceptual mechanism, are described in the first section of this chapter. The subsequent segment adjusts chosen designs and proposes new models, as well as calibrating and hyperparameter setting. The third segment is worried about the strategy's execution viewpoints, and it depicts calculation plans for imagining the result of the models planned, as well as customization to prepare on the proposal dataset

3.1. IRS-Aided UAV Communications

IRS mix in UAV-supported correspondence frameworks has as of late been a hot exploration issue. In (Chen, González-Prelcic, & W. Heath, 2020), the creators, for instance, examined a downlink correspondence framework contained a turning. The creators recommended in this paper a consecutive curved guess (SCA)- based calculation to work on the IRS's direction and latent beamforming. The utilization of IRS in UAV-helped correspondence frameworks was analyzed in (Grando, Jesus, & Drews-Jr, 2020). The researchers reasoned that the utilization of IRS can bring about critical execution gains in UAV-helped cell organizations. Different benefits of the IRS-helped technique can likewise be recognized in the writing. Most existing calculations depend on raised streamlining hypothesis, which might bring about horrible showing and is tedious because of how much cycles expected for the calculation's combination. How much reflecting components builds their intricacy.

3.2. The DDPG Method

The DDPG calculation is a blend of the worth capability and strategy search techniques. Subsequently, the DDPG calculation is fitting for enormous scope activity and state spaces. The entertainer capability (π) maps the states to a particular activity in light of the ongoing arrangement, with being the entertainer network boundaries, and the pundit capability $Q(s, a)$ examines the nature of the activity executed. To help combination speed and forestall pointless computations, we apply an encounter replay cushion and an objective organization system in the DDPG calculation.

In an iterative process, the agent engages with the environment by selecting and executing an action, denoted as a_t , and subsequently receives an immediate reward, represented as r_t , along with the resulting state. This tuple, consisting of the current state, action, reward, and next state, is then stored in a replay buffer, referred to as D . The purpose of this replay buffer is to accumulate a collection of these tuples for training the actor and critic networks. As new samples are obtained, they are added to the buffer, while the oldest samples are removed to maintain the buffer's finite size. By continually updating the replay buffer, the agent can learn from a diverse range of experiences and improve its decision-making capabilities over time. Once a sufficient number of samples are accumulated, the agent performs G updates in a batch to train the network. We expressly train the entertainer and pundit networks utilizing stochastic slope plunge (SGD) on a little clump G tests.

$$L = \frac{1}{G} \sum_i^G \left(y^i - Q(s^i, a^i; \theta_q) \right)^2, \quad (8)$$

With

$$y^i = r^i(s^i, a^i) + \zeta Q'(s^{i+1}, a^{i+1}; \theta_{q'})|_{a^{i+1}=\mu'(s^{i+1}; \theta_{\mu'})}, \quad (9)$$

In the context where the action at time step $i + 1$ needs to be determined, the target actor network, denoted as θ' , can be utilized. The target actor network is a separate network with its own set of parameters (θ') and plays a crucial role in this process. By providing the state S^{i+1} as input to the target actor network, the desired action can be obtained. It is important to note that the parameters of the target actor network act as boundaries or limits for the decision-making process. These boundaries are periodically updated or refreshed to ensure that the target actor network remains aligned with the current state of the environment and the most recent knowledge acquired by the agent. This updating process helps the agent make accurate and informed decisions based on the latest information available, thus improving the overall performance and effectiveness of the actor-critic framework:

$$\nabla_{\theta_{\mu}} J \approx \frac{1}{G} \sum_i^G \nabla_{a^i} Q(s^i, a^i; \theta_q)|_{a^i=\mu(s^i)} \nabla_{\theta_{\mu}} \mu(s^i; \theta_{\mu}). \quad (10)$$

Moreover, after a certain number of episodes, we replicate the actor and critic networks to create target actor and target critic networks. Utilizing a single network to predict the goal value aids in minimizing the computational workload. We define the Markov Decision Process (MDP) (Liu, Ma, Gao, & Tang, 2019) using a 4-tuple $\langle S, A, P, R \rangle$ for the hovering scenario. Subsequently, we devise a gaming approach to address the problem (3.4).

Agent: The brought together processor of the UAV will act as a specialist. The specialist collaborates with the climate to decide the smartest idea for boosting the general aggregate rate. Following preparation, the activity arranging plans will be sent to the UAV to expect the ideal EH time booking, and the RIS will actually want to choose the stage shift framework..

- State space: The channel is comprised of both immediate and reflecting connections. Subsequently, we characterize the state space as:

$$\mathcal{S} = \{h_1 + H\Phi g_1, h_2 + H\Phi g_2, \dots, h_N + H\Phi g_N\}, (11)$$

- Action space: Because the UAV hovers in a fixed position, we optimize the EH time and the RIS's phase shift. The action space is defined as follows

$$\mathcal{A} = \{\tau, \theta_1, \theta_2, \dots, \theta_K\} \quad (12)$$

At the state s_t , the UAV takes the action $a_t = \{\tau_t, \theta_{t1}, \theta_{t2}, \dots, \theta_{tK}\}$ and move to the next state s_{t+1} .

- Reward function: The UAV collaborates with its environmental factors to augment its return. In our work, we plan the motivating force capability to accomplish the most noteworthy all out aggregate rate execution as could really be expected:

$$\mathcal{R} = \sum_{n=1}^N (1 - \tau^t) B \log_2(1 + \gamma_n^t) \quad (13)$$

The UAV is drifting over XUAV and chooses an activity relying upon the channel status data acquired (CSI). The UAV then conveys energy to the IoT gadgets while the RIS regulator changes the stage shift for every component. While the IoT devices provide data in the uplink to the UAV for the extended time, the RIS will not affect the stage shift (1). It is infuriating because the RIS is critical in reducing impediments. As a result, we should support a clever system for the RIS to improve

network execution in the downlink power move and uplink data transfer exhibits. We propose a DRL, more precisely a DDPG calculation, to find the best strategy for the UAV and the RIS.

At time step t , the specialist has channel data and uses the entertainer organization to choose the activity a_t . Via completing the activity a_t , the specialist gets a response from the climate in view of the prize capability. The pundit activity adds to the effectiveness of the picked activity a_t . In the wake of putting away an adequate number of information in cradle D , the specialist utilizes the SGD with Adam optimiser to prepare the organizations over a small bunch G . To restrict flight energy utilization, we expect that the UAV is floating at a decent spot in this part. It involves a game of trade-offs between energy and maximizing the potential aggregate rate. In the upcoming section, we propose a unified optimization approach involving direction, EH (Energy Harvesting) time, and phase shift to enhance network performance within a limited operational duration.

Given the UAV's short flight time, we propose a cooperative improvement method including the UAV's direction, EH time planning of IoT, and the RIS's stage shift to boost complete conceivable total rate. As in Area 4.3, we characterize the state space and the prize capability. The activity space is adjusted as follows.

$$\mathcal{A} = \{v, \varsigma, \tau, \theta_1, \theta_2, \dots, \theta_K\} \quad (14)$$

We want to find the best methodology for amplifying the normal prize R . To get the prize, the specialist utilizes neighborhood information and communicates with the climate. The neighborhood data is used to make the accompanying state (14) and afterward the activity is chosen in light of it (14). The specialist refreshes the strategy and executes another activity at another state in view of the acquired prize. The tedious collaborations permit the specialist to foster a superior strategy with a more prominent prize. Following each activity, the UAV will make a trip to another position and get reactions from the environmental elements. The specialist can pick the right speed and flight course for the UAV in each time step in view of the accomplished CSI by associating iteratively with the climate. Simultaneously, the EH planning and the stage shift lattice are streamlined to further develop network execution. M and T are the greatest number of episodes and time steps, separately.

To prepare UAVs to investigate the climate independently, different Support Learning (RL) methods are applied. RL can address different issues by going about as a human master in the point. The specialist communicates with the climate by examining its condition, playing out an activity, and getting a prize. UAV cameras and sensors gather information from the climate to portray the state. As displayed in Figure 20, the specialist assesses the caught state and results an activity that characterizes the heading of the UAV development or manages the push of the propellers.

Established researchers explored numerous UAV route difficulties, including Visual UAV route, UAV Running and Way Arranging (Peng, Tsai, Wang, & Han, 2021). In any case, to the best of the creators' information, no review on RL applications in UAV route exists. Subsequently, the reason for this study is to give an exhaustive and methodical outline of the utilization of a few RL calculations to different independent UAV route difficulties. The accompanying individuals added to this overview:

- Assist practitioners and researchers in selecting the best algorithm for the job at hand based on the application area and environment type.
- Make sense of the center ideas and attributes of various RL calculations, as well as their connections and arrangement relying upon the climate type.

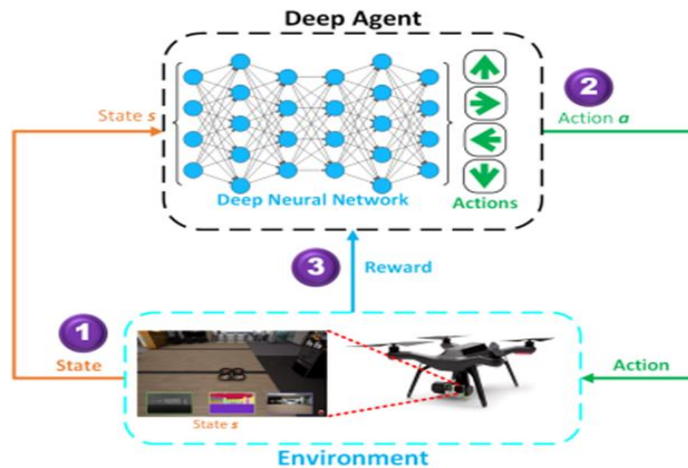


Figure 20. UAV training using deep reinforcement agent.

3.3. Proposed Algorithms

3.3.1. DQN-based Algorithm for Discrete Cases

This portion shows the DQN-based estimation. In any case, we describe the state, movement, and grant. From that point onward, the full IRS-helped UAV correspondence framework is displayed as a climate. It is imagined that the specialist will cooperate with the climate to choose the suitable activities to do to expand the amassed rewards. Inside a state grouping, $RT = PT \ t' = t \ t' \ RT'$, where is the markdown factor.

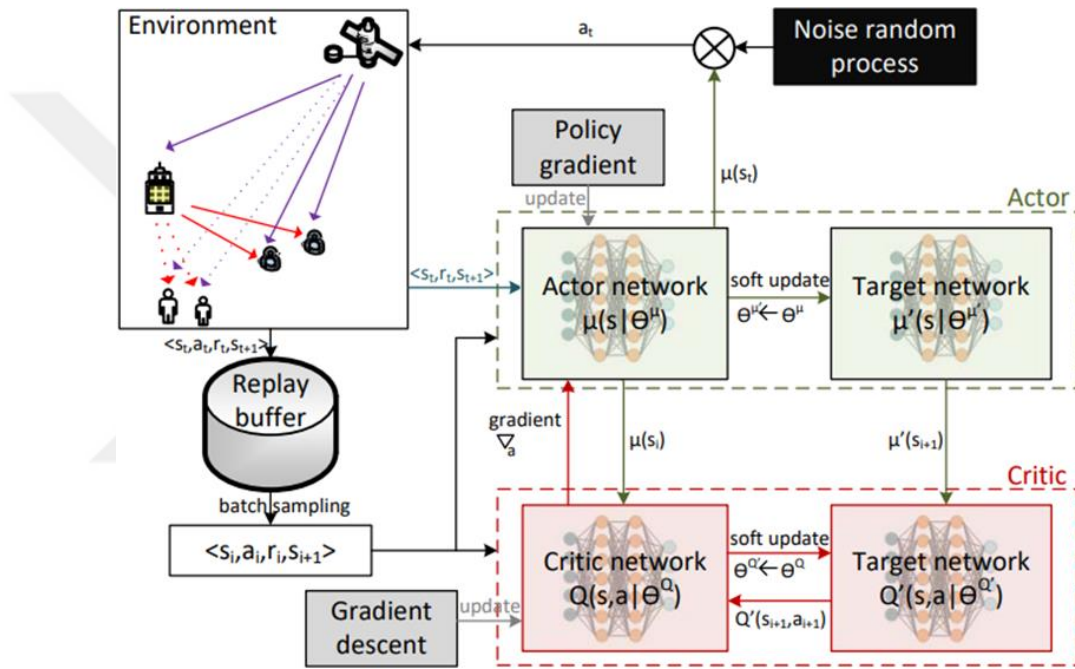


Figure 21. Design of DQN calculation

Figure 21 depicts the overall architecture of the system. In the Deep Q-Network (DQN) framework, an agent is responsible for governing the interactions between the Unmanned Aerial Vehicle (UAV) and the environment. Two Deep Neural Networks (DNNs) are employed: the evaluation network and the target network. The target network has a structure that is identical to the evaluation network, but its updates are less frequent. The action at a given time step is determined based on the Q-values computed using an ϵ -greedy policy. Subsequently, the reward r_t is obtained from the environment. It is important to emphasize that the proposed DQN-based algorithm improves the UAV's direction within a discretely limited action space. To this end, the action space in each time slot is defined as A ,

which encompasses a set of specific actions. By utilizing this approach, the algorithm effectively navigates the UAV, leading to improved performance and efficient decision-making in the given environment:

$$\mathcal{A} = \begin{cases} [x^{\max}, 0, 0], \\ [-x^{\max}, 0, 0], \\ [\frac{x^{\max}}{2}, 0, 0], \\ [-\frac{x^{\max}}{2}, 0, 0], \\ [0, y^{\max}, 0], \\ [0, -y^{\max}, 0], \\ [0, \frac{y^{\max}}{2}, 0], \\ [0, -\frac{y^{\max}}{2}, 0], \\ [0, 0, z^{\max}], \\ [0, 0, -z^{\max}], \\ [0, 0, 0]. \end{cases} \quad (15)$$

The transition, consisting of st , at , rt , $st+1$, is then stored in an experience replay memory. Once a sufficient number of transitions accumulate in the experience replay memory, the learning process commences. To train the evaluation network, a mini-batch is randomly sampled from the experience replay memory, and the loss function for updating the evaluation network is calculated. This loss function can be represented as:

$$L_i(\delta_i) = \mathbb{E}_{s,a} \left[\left(r + \gamma \max_{a'} Q(s', a' | \delta_{i-1}) - Q(s, a | \delta_i) \right)^2 \right], \quad (16)$$

The general pseudo code of the proposed DQN calculation is given in figure 22. Lines 1-2 are utilized to introduce the assessment, target organizations, and experience replay memory. We start every episode by introducing the state st . The specialist then follows a - ravenous strategy to make at in every TS. In particular, the specialist chooses with likelihood that has the most noteworthy Q-value, or haphazardly chooses with likelihood 1 from A . Line 13 stores the progress into experience replay memory. Line 14 starts the growing experience by arbitrarily picking M advances from memory for preparing the assessment organization, the

boundary of which is refreshed by Eq (16). At last, the objective organization is refreshed consistently.

Algorithm 1 DQN-based algorithm

```

1: Initialize evaluation, target networks with parameters  $\delta$ ;
2: Initialize experience replay memory;
3: for Episode = 1,2,..., $N^{\text{eps}}$  do
4:   Initialize state  $s_t$ ;
5:   for TS  $t = 1, 2, \dots T$  do
6:     Obtain  $s_t$ ;
7:     Select  $a_t = \operatorname{argmax}_{a_t \in \mathcal{A}} Q(s_t, a_t)$  with probability  $\epsilon$ ;
8:     Randomly select  $a_t$  from  $\mathcal{A}$  with probability  $1 - \epsilon$ ;
9:     Execute  $a_t$ ;
10:    Calculate the energy consumption of UAV  $e_t$  from Eq. (4);
11:    Obtain the optimized phase shifts of selected IRS  $k$  according to Section IV-C;
12:    Calculate  $r_t$  according to Eq. (16);
13:    Store transition  $\{s_t, a_t, r_t, s_{t+1}\}$  into experience replay memory;
14:    if the learning process starts then
15:      Randomly sample  $M$  transitions from experience replay memory;
16:      Update evaluation network from Eq. (18);
17:      Update target network periodically;
18:    end if
19:  end for
20: end for

```

Figure 22: DQN based Algorithm

3.3.2. Algorithm for Continuous Cases Based on DDPG

In this section, we present a DDPG-based approach for monitoring and optimizing the UAV's trajectory using the well-known actor-critic methodology. Figure 22 outlines the architecture of the DDPG algorithm. It involves two DNNs: the actor network, which determines the action $a = \pi(s)$, and the critic network, which estimates $Q(s, a|Q)$, respectively.

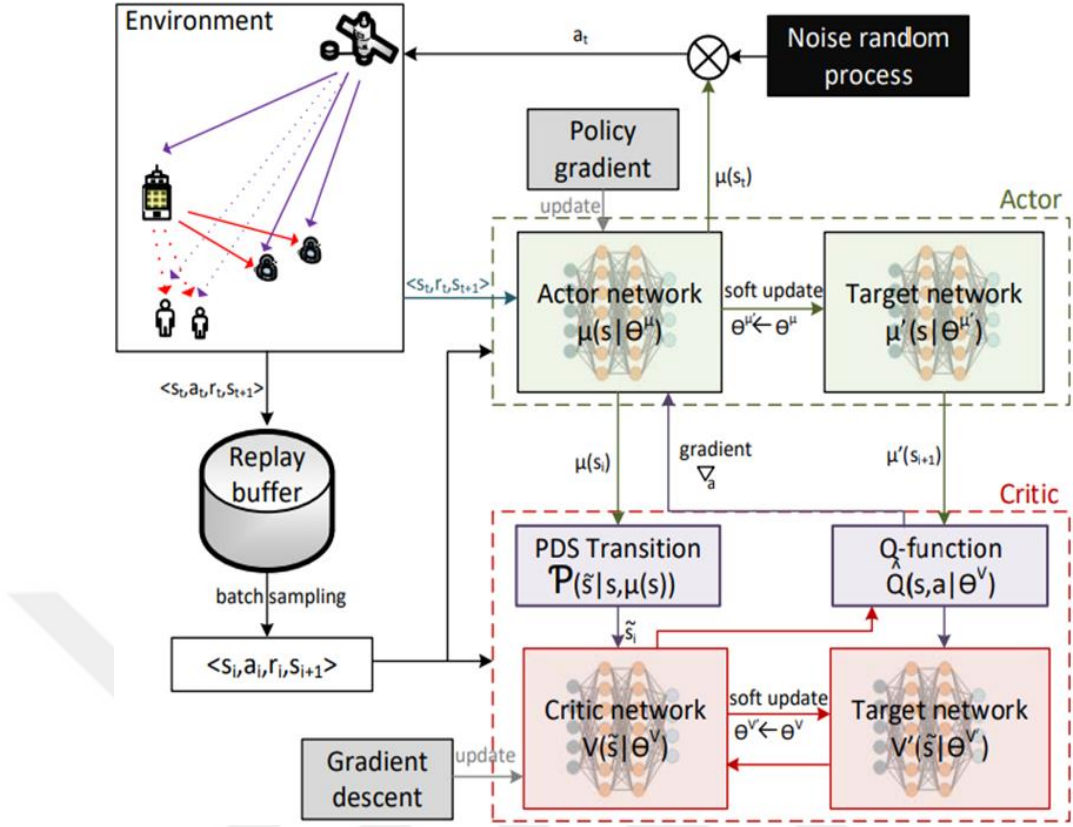


Figure 23. Architecture of DDPG algorithm

The agent receives the state of the environment, s_t , and produces the action, a_t , using its actor network. This transition is then stored in the experience replay memory. When there are sufficient transitions in the memory, the actor and critic networks are trained. In this process, the critic network computes the Q-values, $Q(s, a)$, to determine the policy gradient based on the states, s , and actions, a , as explained in (Çetin, Barrado, Muñoz, Macias, & Pastor, 2019).

CHAPTER FOUR

RESULTS AND DISCUSSION

In this segment, we employ mathematical simulations to illustrate the proposed DDPG-based architecture for process offloading in the UAV-assisted MEC framework. Initially, we present the setup of the simulation parameters. Subsequently, we evaluate the performance of the DDPG-based system under various conditions and compare it with alternative estimation methods.

4.1. Simulation Setting

The detailed simulation parameters are outlined in Table 2, unless specified otherwise. For our experiments, the average reward for performance evaluation is derived from multiple runs of Algorithm 3 with identical settings. We have considered four standard methods as benchmarks for comparison:

Table 2. Simulation parameters

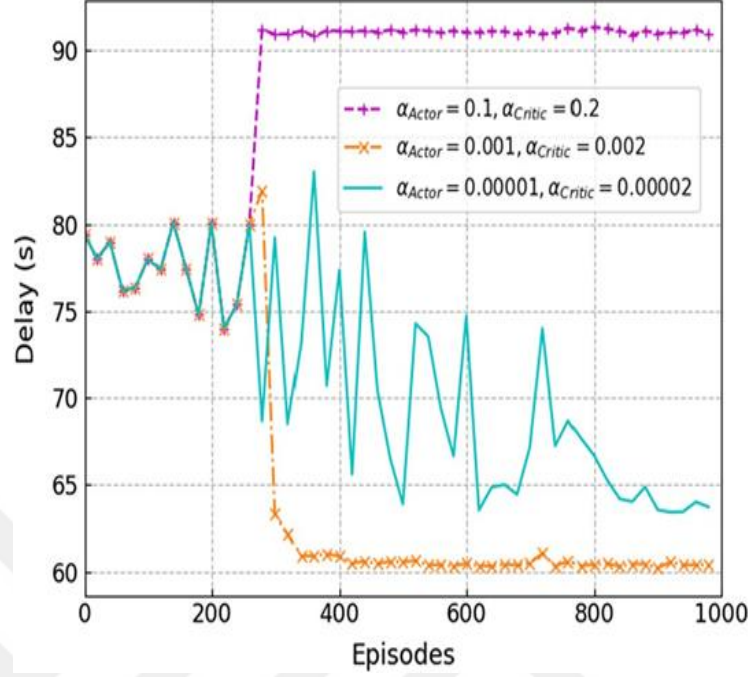
Parameter	Default value	Parameter	Default value
K	4	L, W, H	100 m
MU_{AV}	9.65 kg	T	400 s
I	40	v_{max}	50 m/s
t_{fly}	1 s	a_0	-50 dB
B	1 MHz	r^2	-100 dBm
P	20 dB	P	0.1 W
E_b	500 kJ	S	1000 cycles/bit
f_{UE}	0.6 GHz	f_{UAV}	1.2 GHz
c	5×10	$c ; c$	100

4.2. Simulation Results and Discussions

4.2.1. Parametric analysis

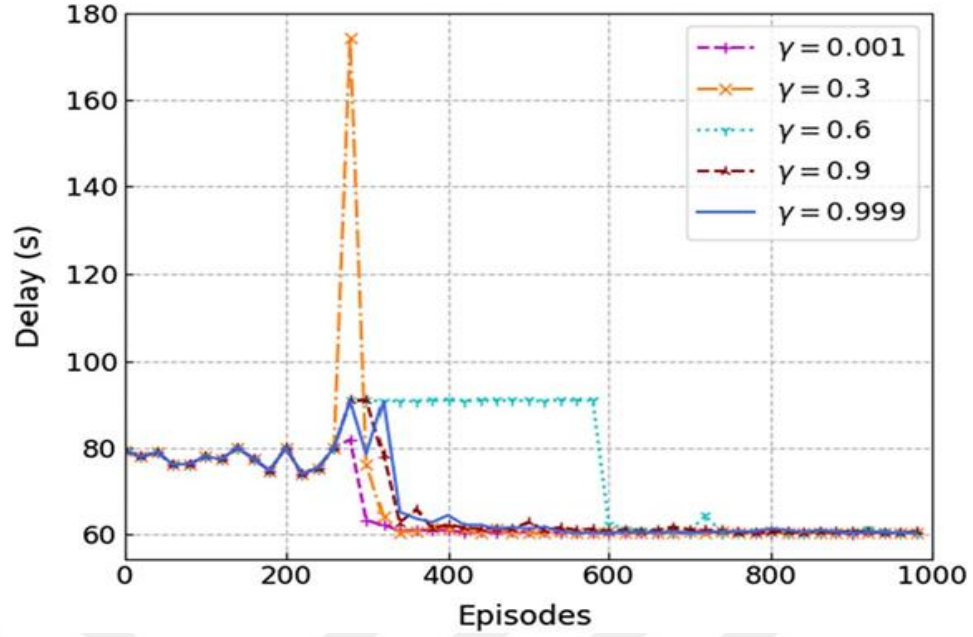
We initiate by conducting a series of trials to ascertain the optimal values for key hyperparameters employed in the correlation computations. The collective performance of the proposed algorithm with varying learning rates is depicted in

Graphic 1. We hypothesize that the critic network and actor network utilize distinct learning rates.



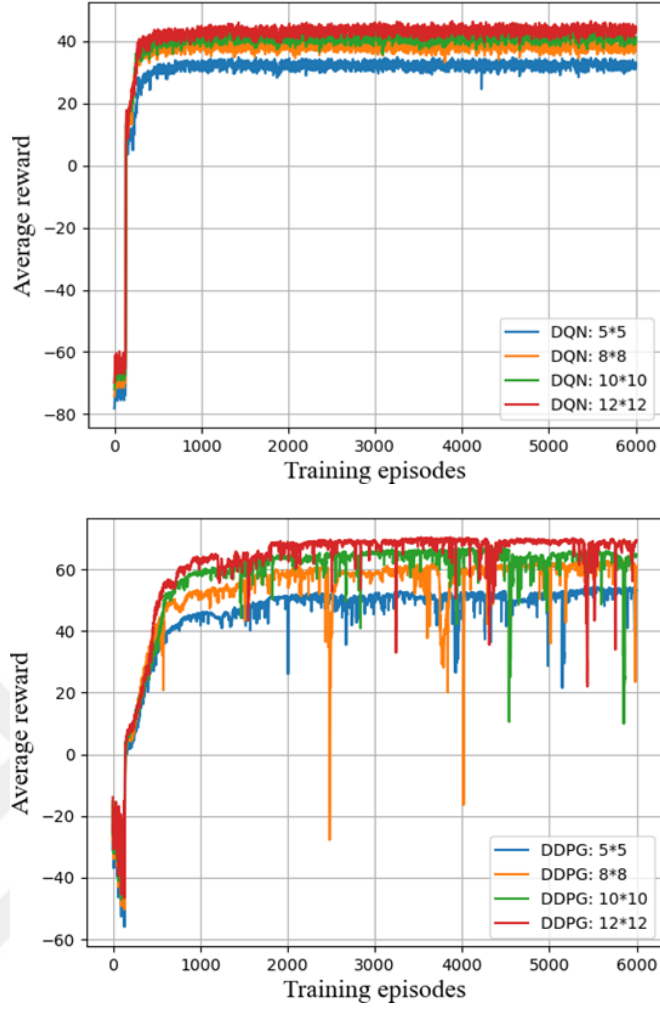
Graphic 1. The overall performance of the DDPG-based algorithm at various learning rates.

In Graphic 2, we analyze the influence of different discount factors (γ) on the collective performance of the proposed algorithm. Optimal performance of the learned computation offloading policy is observed when the discount factor γ is set to 0.001. This is attributed to the significant environmental changes over time, where information from a single time span may not adequately represent long-term behavior, resulting in suboptimal policy across time spans. Therefore, an appropriate value of γ enhances the potential performance of our training methods. In the subsequent experiments, we set the discount factor γ to 0.001.



Graphic 2. Performance of convergence with various discount factors

In Graphic 3 (a), the learning curves of average rewards are depicted, showing initial negative trends for different numbers of reflection elements. This decline in average rewards can be attributed to suboptimal UAV performance, such as flying beyond the target area, resulting in negative rewards. However, as the UAVs start to converge, the average rewards increase and eventually stabilize, indicating that the system has achieved optimal performance. Additionally, it is evident that increasing the number of reflecting elements leads to higher average rewards. Moving on to Graphic 3 (b), it presents the average reward of the proposed Deep Deterministic Policy Gradient (DDPG)-based policy compared to the number of training episodes. The trajectory of the average rewards follows a similar pattern to the DQN-based approach shown in Graphic 3 (a). However, when the number of reflecting elements is the same, the DDPG-based approach consistently yields superior rewards compared to the DQN-based approach, as expected. This is because DQN-based algorithms are limited to a fixed number of actions, whereas DDPG-based systems continuously optimize the variables, allowing for more flexible and adaptive decision-making. Overall, these results highlight the effectiveness of the DDPG-based approach in achieving higher rewards and improved performance in the given scenario.

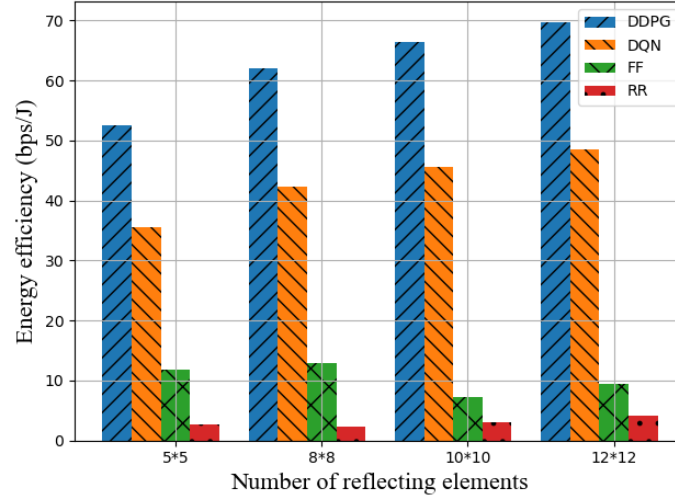


Graphic 3. (a) DQN and (b) DDPG average reward versus number of training Episodes

4.2.2. Performance Comparison

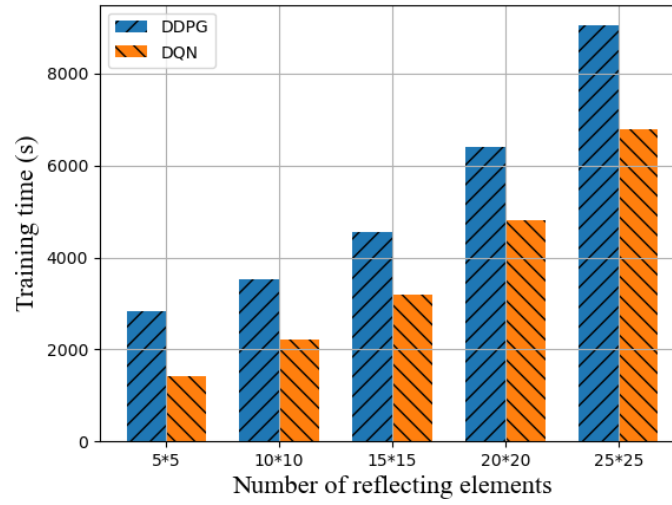
Graphic 4 portrays an exhibition correlation of different strategies. This graph demonstrates the way that the air conditioner calculation can't merge as the quantity of emphasis develops, while both the DQN and DDPG strategies can. This is because of the air conditioner calculation's issue with simultaneous refreshing of the entertainer organization and the pundit organization. The entertainer organization's activity not entirely set in stone by the worth capability of the pundit organization, however the pundit network itself is challenging to join. Subsequently, the air conditioner calculation might neglect to join in certain examples. DQN and DDPG, then again, benefit from a double organization structure comprising of an assessment

organization and an objective organization that might be utilized to remove the relationship between preparing information to find the best activity system.



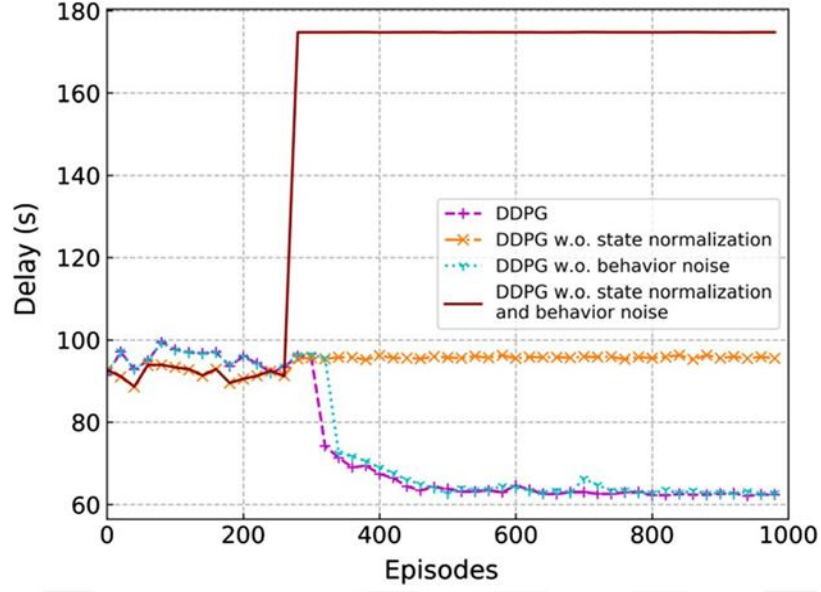
Graphic 4. DQN, DDPG, FF, and RR with varying numbers of reflecting elements attain average energy efficiency.

The DQN and DDPG-based algorithms are evaluated in terms of the standard energy efficiency of UAVs, as shown in Graphic 5. The energy efficiency achieved by the DDPG algorithm steadily improves over time, ranging from 52 bps/J to 70 bps/J. Moreover, across different numbers of reflecting elements, DDPG consistently achieves superior energy efficiency compared to the other algorithms. DQN performs slightly worse than DDPG but still outperforms the FF (Feedforward) and RR (Random) algorithms in terms of energy efficiency. These results demonstrate the effectiveness of the DDPG algorithm in optimizing energy efficiency for UAVs and its superiority over other baseline algorithms in this regard.



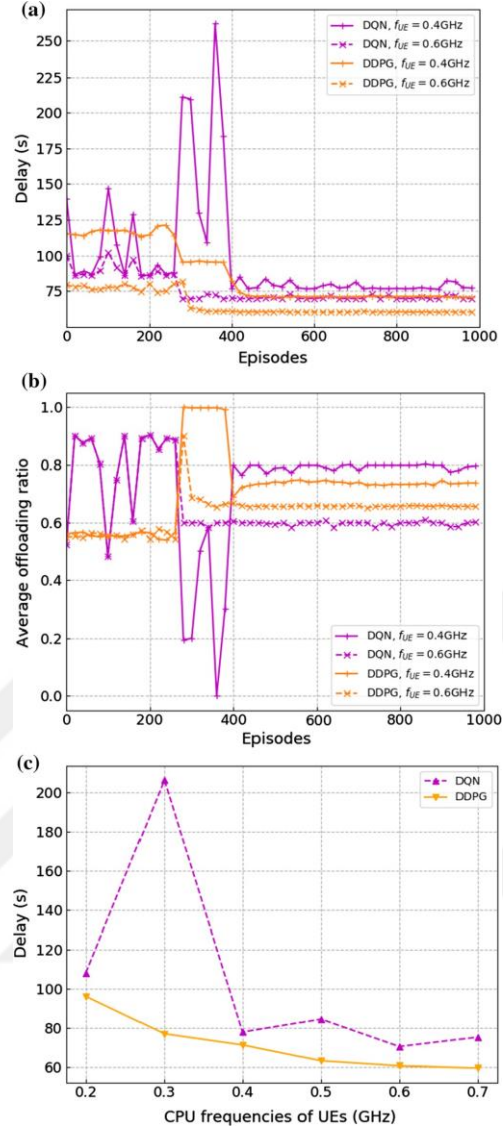
Graphic 5. DQN, DDPG, FF, and RR with shifting quantities of reflecting components achieve normal energy productivity.

Graphic 6 illustrates the impact of training strategies that do not incorporate state normalization or behavior noise in DDPG (Deep Deterministic Policy Gradient) training algorithms. When the DDPG algorithm is trained without behavior noise, it results in slower convergence speed. Conversely, if the policy is trained without state normalization, meaning the absence of a scale factor in the state normalization process, the training algorithm fails to function properly. This is due to the large values of variables such as $E_{battery}(i)$, $D_{remian}(i)$, and $D_k(i)$ when state normalization is not applied. Consequently, the random initialization of Deep Neural Networks (DNNs) leads to the generation of larger output values. Therefore, without the state normalization strategy proposed in this study, the DDPG algorithm eventually behaves like a greedy algorithm. It is important to note that the absence of state normalization and behavior noise significantly impacts the training process and the overall performance of the DDPG algorithm.



Graphic 6. Performance without state normalization or behavior noise

Graphic 7 (a and b) illustrate the performance in terms of delay and offloading ratio for the same set of experiments. Graphic 6a depicts the convergence performance of the DQN scheme and the DDPG scheme with varying UE computing capabilities. The proposed scheme is not compared to the AC scheme as the AC scheme is not yet concurrent. It is evident that with a modest UE computing capability ($f_{UE} = 0.4$ GHz), both optimization methods significantly improve the processing delay. Conversely, when the UE possesses high computational capabilities ($f_{UE} = 0.6$ GHz), the average offloading ratio in the system is low, as depicted in Graphic 6b, indicating a preference for local task execution by the UE. Lower computational capabilities in the UE result in slower data processing within the system, leading to a larger maximum delay between local execution and offloading. Graphic 6c presents a refined delay comparison of the proposed framework with the DQN scheme under various central processing unit (CPU) frequency conditions. It shows that the proposed DDPG framework achieves reduced delay compared to the DQN approach across different UE computing capabilities. This improvement is attributed to the DDPG scheme's ability to generate a higher number of continuous actions compared to the limited discrete actions generated by the DQN scheme. Consequently, DDPG can identify a precise component the offloading ratio—that significantly impacts delay in the continuous action control system.



Graphic 7. (a) DQN and DDPG performance under varying UE computational capabilities. (b) DQN and DDPG offloading ratios under varied UE computational capabilities. (c) A comparison of DQN and DDPG

4.3. Conclusion

In this review, we explored the joint improvement of UAV direction and uninvolved stage movements of reflection components in an IRS-supported UAV correspondence framework, considering UE development and IRS determination. We want to augment the framework's energy productivity, which offers a planning time advantage anyway a show setback and may be suitable for circumstances when getting ready time is fundamental. Then, at that point, to further develop execution, we utilized a DDPG-based calculation that can consistently change the framework's factors. The exploratory outcomes exhibit that the proposed calculations beat other conventional strategies.



CONCLUSION AND FUTURE WORKS

Conclusion

During the past two decades, conventional optimization methods, such as convex optimization and AO, have been extensively applied to design resource allocation algorithms for wireless systems. While the conventional optimization methods solve the resource allocation problems by exploiting the (hidden) convexity, the resulting computational complexity can increase significantly with respect to the DoF in wireless systems. For the B5G wireless systems with high DoF, the computational complexity of conventional optimization methods can be prohibitively high, making the algorithms difficult to implement in practical systems. In this thesis, we investigated the DRL-based resource allocation algorithms design for wireless systems, with the objective of providing both IoT devices and mobile users with satisfactory QoS. In the proposed algorithms, the policies for solving the resource allocation problems in B5G wireless systems are learned by exploiting various model-free DRL techniques, with the help of the DNN structures that we customized specifically for each resource allocation problem. Using the proposed algorithms, we were able to explore the potential of three specific physical layer and medium access control techniques, namely, IRS, and RIS, to improve the spectral efficiency of wireless systems.

Move learning and meta learning have colossal commitment for further developing the growing experience in remote correspondences. The quintessence of move learning is the utilization of a formerly scholarly model and information to another climate. Move learning enjoys various upper hands over conventional methodologies in remote correspondences, including 1) expanding the quality and amount of preparing information, 2) accelerating the growing experience, 3) lessening registering requests in the preparation cycle, 4) relieving correspondence above, and 5) safeguarding information protection. One more conceivable method is meta-realizing, which alludes to gaining calculations that gain from the result of other learning calculations to adjust to another climate with few preparation tests rapidly. Both exchange learning and meta learning can possibly be powerful in helping UAV-empowered remote correspondences since the UAV will immediately

know about ideal asset the executives, arrangement, and direction arrangements in new settings. They are likewise very successful in RIS-helped networks when the climate changes because of another client or RIS. The pre-prepared model or other model result will help the RIS in rapidly changing the stage shift grid to accomplish the best exhibition.

Future Works

In the following, we discuss the limitations of the proposed DRL-based algorithms design and present the potential directions for future research work. In this thesis, we assumed that the learning agent in DRL can obtain perfect information of the state of the underlying MDP. In particular, in Chapters 3 and 4, we proposed DRL based algorithms for B5G wireless systems with perfect CSI. However, perfect CSI may be difficult to obtain in practical wireless systems.

Hence, it is worthy to study the potential methods to tackle the lack of perfect CSI in DRL-based algorithms design. Recently, several robust DRL algorithms (e.g., have been proposed for policy learning under noisy state observation. Existing works on channel estimation have shown that the channel estimation error in practical systems can be modeled using complex Gaussian distributions (Lee, Jung, Kasgari, Saad, & Bennis, 2020). Therefore, it is interesting to investigate how to apply the robust DRL algorithms to tackle the channel estimation error in B5G wireless systems. In addition, the training algorithm and DNN structure design may need to be revisited to accommodate the imperfect CSI during the policy learning process.

REFERENCES

- Al-Hourani, A., Kandeepan, S., & Jamalipour, A. (2014). Modeling air-to-ground path loss for low altitude platforms in urban environments. doi:<https://doi.org/10.1109/GLOCOM.2014.7037248>
- Abedin, S., Munir, M., Tran, N., Han, Z., & Hong, C. (2021). Data Freshness and Energy-Efficient UAV Navigation Optimization: A Deep Reinforcement Learning Approach. (pp. 5994 - 6006). IEEE. doi:<https://doi.org/10.1109/TITS.2020.3039617>
- Akhloufi, , M. A., Arola, , S., & Bonnet, , A. (2019). Drones chasing drones: Reinforcement learning and deep search area proposal. Drones, . (pp. 3(3), 58.). IEEE.
- B., D., Z., , W., W., w., & B., Z. (2018). Neural Q learning algorithm based UAV obstacle avoidance. In 2018 IEEE CSAA Guidance, Navigation and Control Conference (CGNCC) (pp. 1-6). IEEE.
- Balakrishnan, A., De, S., & Wang, L.-C. (2020). Traffic Skewness-Aware Performance Analysis of Dual-Powered Green Cellular Networks. IEEE. doi:<https://doi.org/10.1109/GLOBECOM42002.2020.9322397>
- Basar, E., Renzo, M., Rosny, J., Debbah, M., Alouini, M.-S., & Zhang, R. (2019). Wireless Communications Through Reconfigurable Intelligent Surfaces. 116753 - 116773. doi:<https://doi.org/10.1109/ACCESS.2019.2935192>
- Bouhamed, O., Ghazzai, H., Besbes, H., & Massoud, Y. (2020). Autonomous UAV Navigation: A DDPG-Based Deep Reinforcement Learning Approach. doi:<https://doi.org/10.1109/ISCAS45731.2020.9181245>
- Bellman, , R. (1957). (1957). A Markovian decision process. Journal of mathematics and mechanics, 679-684. 6(5), 679-684. doi:<https://www.jstor.org/stable/24900506>
- Carlet, J., & Abayowa, B. (2017, sep 25). Fast Vehicle Detection in Aerial Imagery. doi:<https://doi.org/10.48550/arXiv.1709.08666>
- Çetin, E., Barrado, C., Muñoz, G., Macias, M., & Pastor, E. (2019). Drone Navigation and Avoidance of Obstacles Through Deep Reinforcement Learning. doi:<https://doi.org/10.1109/DASC43569.2019.9081749>
- Chen, Y., González-Prelcic, N., & W. Heath, R. (2020). Collision-Free UAV Navigation with a Monocular Camera Using Deep Reinforcement Learning. doi:<https://doi.org/10.1109/MLSP49062.2020.9231577>

- Chen, J., Zhang, L., Liang, Y.-C., Kang, X., & Zhang, R. (2019). Resource Allocation for Wireless-Powered IoT Networks With Short Packet Communication. 18, pp. 1447 - 1461. IEEE. doi:<https://doi.org/10.1109/TWC.2019.2893335>
- Cui, Z., & Wang, Y. (2021). UAV Path Planning Based on Multi-Layer Reinforcement Learning Technique. 9, 59486 - 59497. doi:<https://doi.org/10.1109/ACCESS.2021.3073704>
- Feng, K., Wang, Q., Li, X., & Wen, C.-K. (2020). Deep Reinforcement Learning Based Intelligent Reflecting Surface Optimization for MISO Communication Systems. 745 - 749. doi:<https://doi.org/10.1109/LWC.2020.2969167>
- Grando, R., Jesus, J., & Drews-Jr, P. (2020). Deep Reinforcement Learning for Mapless Navigation of Unmanned Aerial Vehicles. doi:<https://doi.org/10.1109/LARS/SBR/WRE51543.2020.9307015>
- GUO, T., JIANG, N., LI, B., ZHU, X., WANG, Y., & DU, W. (2021). UAV navigation in high dynamic environments: A deep reinforcement learning approach. 34(2), 479-489. doi:<https://doi.org/10.1016/j.cja.2020.05.011>
- Ju, H., & Zhang, R. (2013). Throughput Maximization in Wireless Powered Communication Networks. 13(1), 418 - 428. doi:<https://doi.org/10.1109/TWC.2013.112513.130760> 12369 - 12373. IEEE. doi:<https://doi.org/10.1109/TVT.2020.3011942>
- Imanberdiyev, , N., Kayacan, , E., & Chen, , I. (2016). Autonomous navigation of UAV by using real-time model-based reinforcement learning. In 2016 14th international conference on control, automation, robotics and vision (ICARCV) (pp. 1-6). IEEE. doi:<https://doi.org/10.1109/ICARCV.2016.7838739>
- Lee, S., Shim, T., Kim, S., Park, J., Hong, K., & Bang, H. (2019). Vision-Based Autonomous Landing of a Multi-Copter Unmanned Aerial Vehicle using Reinforcement Learning. IEEE. doi:<https://doi.org/10.1109/ICUAS.2018.8453315>
- Lei, M., Zhang, X., Yu, B., Fowler, S., & Yu, B. (2021). Throughput maximization for UAV-assisted wireless powered D2D communication networks with a hybrid time division duplex/frequency division duplex scheme. IEEE. doi:<https://link.springer.com/article/10.1007/s11276-021-02562-w>
- Li, Y., Li, M., Sanyal, A., Wang, Y., & Qiu, Q. (2019). Autonomous UAV with Learned Trajectory Generation and Control. doi:<https://doi.org/10.1109/SiPS47522.2019.9020508>

- Liu, C., Ma, X., Gao, X., & Tang, J. (2019). Distributed Energy-Efficient Multi-UAV Navigation for Long-Term Communication Coverage by Deep Reinforcement Learning. doi:<https://doi.org/10.1109/TMC.2019.2908171>
- Mei, H., Yang, K., Shen, J., & Liu, Q. (n.d.). Joint Trajectory-Task-Cache Optimization With Phase-Shift Design of RIS-Assisted UAV for MEC. *10*, 1586 - 1590. doi:<https://doi.org/10.1109/LWC.2021.3074990>
- Moon, J., Papaioannou, S., Laoudias, C., & Kolios, P. (2020). Deep reinforcement learning multi-UAV trajectory control for target tracking. *8(20)*, pp. 15441-15455. IEEE.
- Lee, G. T., & Kim, O. C. (2020). Autonomous control of combat unmanned aerial vehicles to evade surface-to-air missiles using deep reinforcement learning. *IEEE Access*, *8*, pp. 226724-226736. IEEE.
- Lee, G., Jung, M., Kasgari, A., Saad, W., & Bennis, M. (2020). Deep Reinforcement Learning for Energy-Efficient Networking with Reconfigurable Intelligent Surfaces. *IEEE*. doi:<https://doi.org/10.1109/ICC40277.2020.9149380>
- Li, K., Ni, W., Duan, L., Abolhasan, M., & Niu, J. (2017). (2017). Wireless power transfer and data collection in wireless sensor networks. *IEEE Transactions on Vehicular Technology*, *67(3)*, pp. 2686-2697. IEEE.
- Peng, H., Tsai, A.-H., Wang, L.-C., & Han, Z. (2021). LEOPARD: Parallel Optimal Deep Echo State Network Prediction Improves Service Coverage for UAV-Assisted Outdoor Hotspots. doi:<https://doi.org/10.1109/TCCN.2021.3115765>
- Peake, A., McCalmon, J., Zhang, Y., & Raiford, B. (2020). Wilderness search and rescue missions using deep reinforcement learning. In *2020 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)* (pp. 102-107). IEEE.
- Polvara, R., Sharma, S., Wan, J., Manning, A., & SUTTON, R. (2019). Autonomous vehicular landings on the deck of an unmanned surface vehicle using deep reinforcement learning. *Robotica*, *37(11)*, 1867-1882. IEEE.
- Samir, M., Elhattab, M., Assi, C., Sharafeddine, S., & Ghrayeb, A. (2021). Optimizing Age of Information Through Aerial Reconfigurable Intelligent Surfaces: A Deep Reinforcement Learning Approach. *70(4)*, 3978 - 3983. doi:<https://doi.org/10.1109/TVT.2021.3063953>
- Shin, S.-Y., Kang, Y.-W., & Kim, Y.-G. (2019). Automatic Drone Navigation in Realistic 3D Landscapes using Deep Reinforcement Learning. doi:<https://doi.org/10.1109/CoDIT.2019.8820322>

- Wang, C., Wang, J., Shen, Y., & Zhang, X. (2019). Autonomous Navigation of UAVs in Large-Scale Complex Environments: A Deep Reinforcement Learning Approach. 2124 - 2136. doi:<https://doi.org/10.1109/TVT.2018.2890773>
- Wang, C., Wang, J., & Zhang, X. (2019). A deep reinforcement learning approach to flocking and navigation of uavs in large-scale complex environments. IEEE. doi:<https://doi.org/10.1109/GlobalSIP.2018.8646428>
- Wang, C., Wang, J., Zhang, X., & Zhang, X. (2017). Autonomous navigation of UAV in large-scale unknown complex environment with deep reinforcement learning. doi:<https://doi.org/10.1109/GlobalSIP.2017.8309082>
- Wang, Z., Liu, R., Liu, Q., Thompson, J., & Kadoch, M. (2019). Energy-Efficient Data Collection and Device Positioning in UAV-Assisted IoT. 7, pp. 1122 - 1139. IEEE. doi:<https://doi.org/10.1109/JIOT.2019.2952364>
- Wang, D., Fan, T., Han, T., & Pan, J. (2020). A Two-Stage Reinforcement Learning Approach for Multi-UAV Collision Avoidance Under Imperfect Sensing., 5, pp. 3098 - 3105. doi:<https://doi.org/10.1109/LRA.2020.2974648>
- Wang, , H., Wang, , J., Ding, , G., Chen, , J., Gao, , F., & Han,, Z. (2019). Completion time minimization with path planning for fixed-wing UAV communications. IEEE
- Wu, Q., & Zhang, R. (2020). Joint Active and Passive Beamforming Optimization for Intelligent Reflecting Surface Assisted SWIPT Under QoS Constraints. 1735 - 1748. doi:<https://doi.org/10.1109/JSAC.2020.3000807>
- Transactions on Wireless Communications. 18(7), , pp. , 3485-3499. IEEE.
- Ye, H., Li, G., & Juang, B.-H. (2019). Deep reinforcement learning based resource allocation for V2V communications. IEEE Transactions on Vehicular Technology, 68(4),. 68(4), 3163-3173. doi:<https://doi.org/10.1109/TVT.2019.2897134>
- Zhang, H., Song, L., Seddik, K., Han, Z., & Li, G. (05 May 2020). Reconfigurable Intelligent Surfaces for Wireless Communications: Principles, Challenges, and Opportunities. 6, pp. 990 - 1002. IEEE. doi:<https://doi.org/10.1109/TCCN.2020.2992604>
- Zhang, J., Yu, Z., Mao, S., Periaswamy, S., Patton, J., & Xia, X. (2020). IADRL: Imitation Augmented Deep Reinforcement Learning Enabled UGV-UAV Coalition for Tasking in Complex Environments. 8, 102335 - 102347. doi:<https://doi.org/10.1109/ACCESS.2020.2997304>

Zou, Y., Gong, S., Xu, J., Cheng, W., Hoang, D., & Niyato, D. (2020). Wireless Powered Intelligent Reflecting Surfaces for Enhancing Wireless Communications. *69(10)*, pp.



