This electronic thesis or dissertation has been downloaded from the King's Research Portal at https://kclpure.kcl.ac.uk/portal/

Integrated UAVs Communications in Cellular Network: Deployment and Optimization via Deep Reinforcement Learning Technique

Burhanuddin, Liyana

Awarding institution:
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT

Unless another licence is stated on the immediately following page this work is licensed

under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

licence. https://creativecommons.org/licenses/by-nc-nd/4.0/

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 13. Jan. 2025

Integrated UAVs Communications in Cellular Network: Deployment and Optimization via Deep Reinforcement Learning Technique

Liyana Adilla binti Burhanuddin

A Thesis Submitted for the Degree of

Doctor of Philosophy at

King's College London

July 2023

Dedicated to my family, teachers and friends.
Alhamdulillah, without you, I cannot be here. ...

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Monday 24$^{\text{th}}$ July, 2023

Abstract

The thesis focuses on optimizing the coverage and capacity of wireless networks through the use of unmanned aerial vehicles (UAVs) in a cellular-connected environment. The limitations of UAVs in capturing large areas necessitate the deployment of multiple UAVs to support the system, with the UAV Base Station (UAV-BS) communicating with other UAVs. The positioning of UAVs is crucial to maximize data communication rate and meet real-time requirements. To ensure quality of experience (QoE) in real-time video streaming, a coordination co-design between UAVs and the UAV-BS is implemented to capture dynamic firefighting areas, by optimal bit-rate and power allocation to ensure smoothness quality in multiple locations. The study also addresses the challenge of downlink interference to terrestrial users (TUEs) when BSs serve both TUEs and UAVs. An interference coordination mechanism is proposed to mitigate inter-cell interference and maximize radio connectivity for TUEs. Dynamic cell muting interference and resource allocation scheduling schemes (MOSDS-DQN) are introduced, leading to a significant improvement in throughput and satisfactory level for both UAVs and TUEs. Conventional beam-sweeping approaches face challenges due to the high mobility and autonomous operation of UAVs. To address this, the deep reinforcement learning (DRL)-based framework using hierarchical Deep Q-Network (hDQN) is proposed for UAV-BS beam alignment in a mmWave radio setting. The framework utilizes location information to maximize beamforming gain during communication requests. To improve convergence time, the convolution neural network radio mapping and hDQN-based framework (hDRM) are employed. Simulation results showed that QoE is improved 12% compared to the non-learning algorithm with 41% improvement of the long-term video smoothness. The proposed MOSDS-DQN showed 18% improvement compared to the DQN algorithm. The proposed hDRM framework improved 63% over the converging time compared to vanilla hDQN approaches under real-time conditions. Overall, the thesis contributes to the optimal positioning of UAVs and BSs, dynamic bit-rate selection, interference mitigation, and efficient beam alignment

using advanced techniques such as coordination co-design, dynamic scheduling, and deep reinforcement learning. These approaches enhance the performance and coverage of UAV-UEs, mitigate interference, and improve the overall efficiency of wireless networks in dynamic environments.

Publications

Throughout this PhD career, TWO publications have been accepted and published of which the following are related to this thesis:

- L. A. b. Burhanuddin, X. Liu, Y. Deng, U. Challita and A. Zahemszky, (2022). QoE Optimization for Live Video Streaming in UAV-to-UAV Communications via Deep Reinforcement Learning, in IEEE Transactions on Vehicular Technology, doi: 10.1109/TVT.2022.3152146.
- L. A. b. Burhanuddin., Liu, X., Deng, Y. (2023). Inter-cell Interference Mitigation for Cellular-connected UAVs using Deep Reinforcement Learning in IEEE Transactions on Cognitive Communications and Networking.

Contents

List of Figures

List of Tables

List of Abbreviations

3D three-dimensional

3GPP 3rd generation partnership project

5G fifth generation

6G sixth generation

A2C actor-critic

AoD angle of departure

AoA angle of arrival

ANN artificial neural networks

AWGN additive white gaussian channel

BB broad beam

BS base station

BSs base stations

CNN convolutional neural network

CSI channel state information

DNN deep neural networks

DRL deep reinforcement learning

hDRL hierarchical deep reinforcement learning

DQN deep Q-network

hDQN hierarchical deep Q-network

ESN echo state network

FML fast machine learning

fsp free space path loss

gNB 5G NR base station

GLOBECOM Global Communication Conference

GPS global positioning system

IID independent and identical distribution

LoS line-of-sight

MAB multi-armed bandit

MIMO multiple input multiple output

ML machine learning

MSE mean squared error

mmWave Millimeter wave

MDP markov decision process

NB narrow beam

nLoS non-line-of-sight

NN neural networks

NR new radio

OFDM orthogonal frequency division multiple access

POMDP partially observable Markov decision process

QoS quality of service

RBD receiver beam direction

ReLU rectifier linear units

RB resource block

RX received

RL reinforcement learning

RF radio frequency

RIM RSSI Information Matrix

RSS received signal strength

SGD stocastic gradient descent

SNR signal-to-noise ratio

SINR signal-to-interference-plus-noise ratio

SSB subframe structure block

SS subframe structure

SCS sub-carrier spacing

TX transmit

TTU travel time unit

TDMA time domain multiple access

UAV unmanned aerial vehicle

UAVs unmanned aerial vehicles

UE user equipment

UEs user equipments

ULA uniform linear array

UMa urban macro-cellular

UPA uniform planar array

V2X vehicle-to-everything

Chapter 1

Introduction

The market for Unmanned Aerial Vehicles (UAVs) is rapidly growing and has become useful for improving efficiency, especially for streaming video [30, 113, 111, 119], agricultural use cases [185], streamline operations, and military applications. The study in [105] shown that in coming year 2029, the worldwide commercial UAV market will triple in size by cost of 14 billion dollars and these will lead to traffic congestion of communication between UAVs and cellular-connected ground users. In particular, UAVs also may give a vast benefit to the communication community to provide reliable and cost-effective wireless communication solutions in a real-world scenarios.

UAVs commonly communicate using a direct link and function as aerial user equipment (UEs), known as cellular-connected UAVs, in coexistence with ground users (e.g., video streaming or packet delivery) [113]. Despite the promising advantages of cellular-enabled UAV communications, there are still scenarios where the cellular services are limited reach in remote or rural regions, as well as in disaster-stricken areas or areas with challenging terrain. As an example, the numerous wildfires have caused challenges for firefighters to control and monitor fire in remote areas [114, 131]. To support such scenarios, the technologies such as the flying ad hoc network or namely unmanned aerial vehicle base station (UAV-BS) will support cellular communications beyond the terrestrial coverage of cellular networks [131].

On the one hand, UAV-BS can deliver reliable, cost-effective, and on-demand wireless communications to desired areas [68, 113]. Furthermore, the adjustable altitude of UAV-BSs enables them to effectively establish line-of-sight (LoS) communication links thus it mitigating signal blockage and shadowing [116]. Due to potential advantages, UAVs admit many potential use cases in wireless networks such as, UAV-BS can be deployed to enhance the wireless capacity and cover-

age at ad-hoc events or hotspots such as sports stadiums and disasters [73, 141]. Moreover, UAVs can be used in public safety scenarios to support disaster relief activities and to enable communications when conventional terrestrial networks are damaged [30, 33, 83, 101, 125, 176, 120, 126]. Therefore, it is envisioned that the future wireless network for supporting large-scale UAV communications will have an integrated 3D architecture consisting of direct link UAV (UAV to Network), UAV-to-ground communications (UAV to Device) and UAV-to-UAV, as shown in Fig. 1, where each UAV may be enabled with one or more communication technologies to exploit the rich connectivity diversity in such a hybrid network. However, in cellular networks, the base station (BS)'s inter-site distance (ISD) is designed according to ground level channel models is not optimized for UAVs in different propagation environments. As a result, the transmission performance of UAVs and TUEs is severely affected by interference among them, when BSs serving them in the same frequency simultaneously [23].

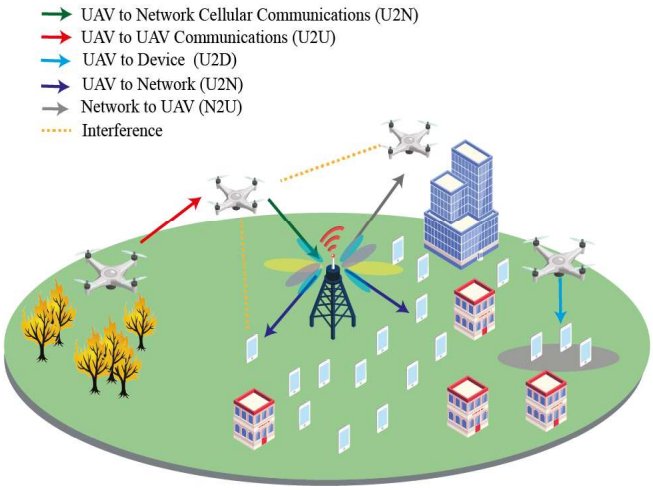

Figure 1.1: Supporting UAV communications with integrated network architecture

However, in the fifth generation (5G) and beyond, the Millimeter wave (mmWave) frequencies (30 GHz to 300 GHz) together with multiple input multiple output (MIMO) beam-forming are capable to provide high capacities and line-of-sight (LoS) dominant connectivity where mmWave have high-capacity and inexpensive, could improve the signal directivity and reduce the co-channel interference between users. The study focuses to optimize the mmWave beam alignment for efficient control of unmanned aerial vehicles (UAVs) in beyond 5G communication, the high deployment of cellular-enabled UAVs-user equipment together with unique features in mobility in three-dimensional (3D) space and autonomous operations to the aerial-ground communications between BS and UAVs to increase high reliability data rate.

Thus, the challenge of attaining high reliability and low-latency communications, while simultaneously improving Quality of Experience (QoE) and Quality of Service (QoS), as well as minimizing interference among users, arises.

Situation: High diversity UAV connectivity in hybrid network.

1.1 Overview

The current practice in the dynamic control and communication model in urban and sub-urban of the 5G cellular-connected unmanned aerial vehicle (UAV) is focused on this study. UAV will be an integral part of the next generation of wireless communication networks and adopt in various communication-based applications is expected to improve coverage and spectral efficiency as compared to traditional ground-based solutions. However, this new degree of freedom that will be included in the network will add new challenges. Generally, UAVs need wireless communication infrastructure to control based on UAV application categories. Therefore, UAV wireless communication can be divided into four (4) types: (i) Terrestrial Base Station (BS) to UAV (ii) UAV to BS (iii) UAV-BS to Terrestrial User (i.e., mobile user) and (iv) UAV-to-UAV communications.

1.1.1 UAV Communications

UAV communication refers to the exchange of information between UAV and other systems, such as ground control stations. This communication enables the control, navigation, and monitoring of UAVs, and can be obtained through various means. The studies found that the UAV's performance can be improved by using high-speed wireless communication channels [27]. However, due to the extensive coverage area and high mobility, channel communication may fluctuate, especially in rural and sub-urban areas where it is difficult to get a full-coverage signal for autonomous purposes. Existing wireless technologies, such as WiFi, Bluetooth, and radio wave, can only support UAVs' communication with a short transmission range, making UAV-cellular collaboration inefficient.

1.1.2 UAV act as Flying User

The type of communication used for a UAV depends on several factors, including the size and range of the UAV, the mission requirements, and the environment in which the UAV is operating. The use of flying UAVs user can help in many

sectors and can continuously move to provide full coverage to targeted areas within a minimum possible time. As the UAV needs continuously moving, it required a network to received and transmit control command. The issues need to be highlighted when deploy the UAV user is whether current cellular network is suitable for UAV communication, how inclusive UAV with the environment including pathloss, power and network efficiency [38, 39, 55] and how UAV can self-positioning themselves [55, 141]. In addition, how UAVs will impact to ground user and the overall network efficiency should take into consideration [22, 119].

1.1.3 UAV act as Flying Base Station

However, geographical constraints caused many areas to suffer from poor network connectivity and difficulty to deploy ad-hoc communications, i.e. search and rescue. Thus, UAV can act as emergency cellular networks that support the users [87]. For example, Emergency Cellular Network was introduced to adapt and adjust the drone small scale [87] while static truck BS was placed in a similar position as the original BS position to help the urban surveillance areas. The authors in [166] considered the moving aerial base station (ABS) to give fair coverage probability to all users and found that the average fade duration is reduced compared to the static drone. Moreover, the use of UAVs is quite natural due to their agility, mobility, flexibility, and dynamic altitude which enable to provide a booster of the performance of the existing ground wireless networks in terms of coverage, capacity, delay and overall quality which required real-time control to support it [113].

1.1.4 UAV-to-UAV Communication

The reliability of the proposed UAV-BSs and relay to help other UAVs transmit to the nearby terrestrial BS with a low signal-to-noise ratio (SNR) shows good performance where the communication between UAV-BS can be extended to flying users. Study shown that when the distance of UAV communication decreases, the SNR of the transmission increases and the transmission performance becomes better. Furthermore, by using the power control policy, it will help to improve the performance of ground users. In evaluating the power control, the height dependency, and interference between spectrum sharing of U2U communication can also be matrices in evaluating the U2U performance.

1.1.5 High UAV number

As mentioned earlier, the growing number of the worldwide commercial UAV market will triple in size [105] and this leads the traffic growth of cellular-connected of flying UAV and ground user become crucial to minimize interference. There are important requirements to provide coexistence and optimal performance for both aerial and terrestrial users to ensure reliable performance. However, today's cellular networks aren't built for aerial coverage, and deployments are primarily focused on providing excellent service to terrestrial users. Therefore, it cause UAVs and terrestrial users face problem to get high performance when there are existing user in t-th time slot. Besides that, from the factors mentioned above, the combined with stringent regulatory requirements resulted in extensive research and standardization efforts to ensure that current cellular networks can reliably operate aerial vehicles in a variety of deployment scenarios.

1.2 What motivates this study?

UAV deployment offers several advantages that contribute to more effective and safer firefighting efforts. UAVs equipped with cameras and sensors provide real-time aerial surveillance, allowing firefighters to gather crucial situational awareness and make informed decisions [65]. They can quickly assess the extent of the fire, identify hotspots, and monitor its progression. Furthermore, UAVs can reach inaccessible or dangerous areas, providing valuable information without risking human lives. With the advancement of cameras such as thermal imaging capabilities, they can detect hidden fires and locate trapped individuals. Overall, UAVs are able to enhance firefighting capabilities, improve response times, aid in minimizing damage, preserve lives, and improve network reliability and latency.

The motivation behind this research is to solutions to these problems: (i) dynamic wireless communication (ii) high interference, (iii) dynamic location and video resolution in dynamic environments, (iv) emergency situations Quality of experience for high-resolution video streaming (v) high volume of terrestrial and flying users and (vii) overhead beam sweeping in mmWave.

1.2.1 Dynamic Wireless Communication

Wireless communication is tending to fading and interfere with other wireless devices. A wireless signal may fade in time, space, and phase, causing temporary connection failures and loss of packets. Fading along with signal attenuation in

mobile wireless system makes bandwidth limited compared to wired networks [115]. The networking and communication constraints associated to an ad-hoc network of UAVs include limited transmission range, high mobility of nodes and frequent topology change that may require route changes and multi-hop communication. The terrain, environmental changes, and obstacles in space can cause higher and bursty bit errors [115]. Even if direct communication link is available, the achieved throughput may not be sufficient for an acceptable quality of multimedia transmission. In such case, some UAVs might be required as relay nodes to provide connectivity and to increase the communication range. High mobility is another constraint to be considered. High mobility makes the network topology change frequently.

In long-term connection performance, the connectivity between the UAV and the destination node may not be maintained due to a fluctuated change in the network topology. QoE is important matrix to evaluate the user satisfaction. As an example of video streaming situation, when then the network is unstable, the quality of video will fluctuate and caused the user to have bad experience. If the connection is poor or lost, UAVs cannot coordinate as a system and the mission objectives might be jeopardized.

1.2.2 High interferences

Another challenge in providing connectivity to the multiple UAVs through the existing cellular network arises due to the increased interference in the network. The increased altitude, distance and favourable propagation condition cause UAVs to generate more interference to the neighbouring users. The uplink interference problem may result in degraded performance, whereas the downlink interference problem may make it challenging for UAV to maintain connection with the network.

Generally, cellular-connected UAVs are integrated into cellular networks to support many applications, as it provide higher probability of line of sight (LoS) transmission to BSs. However, the presence of stronger link between UAV and its associated BS and inter-cell interference (ICI) from neighbouring BSs cause severe downlink interference to terrestrial users (TUEs) and other UAVs, especially when the network has heavy loaded.

1.2.3 Dynamic location and video resolution in dynamic environments

UAV can help to improve the quality of the data captured by the UAV and make it easier to use this data for a variety of purposes. The reception conditions vary in time, space and number of users, therefore the achievable throughput vary for different members of the UAVs group. The distance, location, power control, video resolution and reception condition of the wireless members of the UAVs group may vary. To solve the problem, by jointly optimizing the UAVs location while maximizing the data rate can be applied [81, 82].

In optimizing the position of the UAVs, the users can ensure that the video captured by the UAV's cameras is clear, stable, and free of obstructions. However, as the quality of streaming video may be varying, the bit rate for each UAVs traffic depends on the signal condition. This poses performance degradation for the nodes that can afford better bit rates. Real-time video streaming has higher requirements in terms of data rate, latency, and smoothness compared to other data types. As example, in a firefighting scenario, the network channel capacity fluctuates dramatically with the dynamic environment alongside the UAVs' movement, which can cause poor network performance and undesirable delays. This in turn makes it harder to learn the pattern variance of the channel capacity, thus resulting in failure to transmit with high capacity and high video quality. To capture the practical performance from testbed, authors in [184] used single UAV to conduct indoor experimental to measure the video streaming performance from BS. Ideally, the received of video data may be different from an individual transmission rate supported by each member. However, a higher performance can be achieved by transmitting at a rate affordable for all members of the UAVs rather than using the lowest transmission rate.

Furthermore, the user also might be interested to view from whole angle (360°) to give a full satisfaction to user. If the connection is poor or lost, UAVs cannot coordinate as a system and the mission objectives might be jeopardized. Therefore, the communication link between the UAVs that fly to get a real view while focused to ensure the quality of live-video transmission is important to be studied.

1.2.4 Emergency situation high-resolution video streaming

In an emergency situation, UAVs equipped with high-resolution cameras can play a crucial role in providing real-time video streaming and situational awareness in

a situation in mission of surveillance or an event coverage where the UAVs are deployed may require relaying data to the receiver. Consider a multiple large area of interest are to be captured using UAVs in the form of multimedia traffic to the targeted coverage area surveillance. The UAVs have limitation to capture the marked areas that may require data to maintain connectivity. This depends on how far the target areas away from the BS and the available number of UAVs can be used to provide the service. Requests can be received from users who can connect to the ground node or to the UAVs directly and receive their desired multimedia transmission. Streaming videos from UAVs can be used for SAR, surveillance, remote sensing, and post-disaster operations. Videos of different observed areas can be streamed simultaneously using multiple UAVs to be viewed by responders.

The challenge of this study is how to make multiple video reliable so that the end video is cover the full view of the targeted area without any missing information. However, the lost packets are retransmitted to the desired recipients should take into consideration, so that the same video at tth time are received to the receiver. The source cannot adapt the transmission rate when the receiver's link conditions are varied. Thus, the receiver node can suffer network congestion due to the bad link condition, or it can waste available network resources when it could afford higher bit rates. The objective is to achieve fairness through rate adaptation such that the source transmission rate is controlled based on the reception conditions of the members of the multi-video group. Adhering to strict delay bounds between packet transmission and reception, and QoS support for live video streaming is affected by fading, interference, and signal attenuation due to mobility. For example, delays higher than 250 ms are not acceptable for live video streaming but can be experienced when a receiver is three hops away from the source [95].

1.2.5 High dense urban with high number UAV

Due to increasing technology, the usage of UAV growing faster, especially in high dense urban area. There are mandatory requirements to achieve in order to provide coexistence and optimal performance for both UAVs and terrestrial users. However, today's cellular networks aren't built for aerial coverage, and deployments are primarily focused on providing excellent service to terrestrial users. Therefore, it caused UAV and terrestrial users face problems to maintain high performance when there are existing both type of users in each time slot.

Furthermore, in the current dynamic network environment with uncertain

user numbers, manual management becomes challenging in ensuring optimal service provision between terrestrial users and UAVs. When UAVs and terrestrial users shared the same resources, it may cause high interference.As such, the need to manage the interference and maintaining high QoS for UAV and allowing terrestrial users to operate even with minimal service provision and to main the high performance of both users. As a result, the combined with stringent regulatory requirements resulted in extensive research and standardization efforts to ensure that current cellular networks can reliably operate UAVs in a variety of deployment scenarios in excellent performance.

1.2.6 Overhead beam sweeping in mmWave

The mmWave frequencies beamforming capabilities can provide high capacities and LoS dominant connectivity to the UAV- terrestrial communications between BS and UAV. Fast mmWave beam alignment can enhance the data throughput for both UAV-UAV and BSUAV communications under 5G and beyond wireless systems. For example, the availability of UAV position information at lower frequencies may also provide reliable communication in addition to increasing throughput. Position information for fast beam alignment has been recently studied under vehicular context in mmWave systems [142]. This research focus on high mobility and autonomous operation of UAVs that require frequent beam realignment and faster, reliable beam alignment using UAV position information to enabling high data rate for mmWave UAVs for the high throughput. An effective beam alignment or tracking scheme is usually required to ensure the consistency of beam alignment in a high mobility environment.

1.3 Aim and Contributions

The overall aim of this research is to contribute to optimizing coverage, capacity, and performance in UAV-based cellular networks, mitigating interference, and enhancing beam alignment techniques. By achieving these aims, the study contributes to the advancement of wireless communication systems and meets the increasing demands of dynamic and challenging environments.

This study makes several contributions in the field of UAV-based cellular networks and their optimization as described below:

1. Dynamic Model for UAV-to-UAV and UAV-to-Ground-to-UAV Communication: The study proposes a dynamic model that optimizes the coverage

and capacity of UAV-based wireless networks. By considering the communication between UAVs and the ground station, the study develops an approach to position the UAV base station (UAV-BS) and UAV user equipment (UAV-UE) to maximize uplink data communication rates in real-time. This dynamic model addresses the challenges to capture large areas and providing seamless coverage, to improve the overall network performance.

2. Interference Mitigation and Resource Allocation Schemes: The study focus on mitigating inter-cell interference in scenarios where UAVs and terrestrial users (TUEs) share the same spectrum resources. The proposed dynamic cell muting interference and resource allocation scheduling schemes effectively manage interference and optimize resource allocation, leading to improved throughput and satisfactory levels for both UAVs and TUEs. These schemes mitigate interference and enhance the performance of the UAV-UE and TUE networks, enabling efficient coexistence in urban areas.

3. Enhance Beam Alignment Techniques: The study aim to develop advanced beam alignment techniques that could overcome the challenges posed by the high mobility and autonomous operation of UAVs. The study propose the deep reinforcement learning (DRL)-based framework using hierarchical Deep Q-Network (hDQN) for uplink UAV-BS beam alignment in the mmWave radio setting. The framework leverages location information to maximize beamforming gain during communication requests. Additionally, the convolution neural network radio mapping (C-RM) and DQN-based framework are introduced to enhance convergence time. This contribution improves the efficiency and reliability of UAV-BS communication and outperforms heuristic-based and exhaustive approaches.

1.4 Conclusion

This study distinguish itself from other researchers by offering a dynamic model for UAV-based wireless networks, focusing on UAV-to-UAV and UAV-to-ground-to-UAV communication scenarios. It addresses the limitations of UAVs in capturing large areas by deploying multiple UAVs and strategically positioning UAVs with the UAV-BS. One key gap in current research on UAV-based wireless networks for firefighting environments is the need to address real-time video streaming and QoE utility measurement to meet long-term requirements, which can enhance the real situation to give awareness and fast decision-making during fire-

fighting operations. The study also proposes interference mitigation and resource allocation schemes to manage inter-cell interference between UAVs and terrestrial users, enhancing throughput for both. Furthermore, it introduces the deep reinforcement learning-based beam alignment approach using hierarchical Deep Q-Network, combined with convolutional neural networks, to improve the efficiency and reliability of UAV to BS communication. These unique contributions contribute to the advancement of UAV-based wireless networks and provide novel solutions for optimizing coverage, mitigating interference, and enhancing beam alignment techniques.

In order to achieve the overall aims and goals, this thesis is divided into six chapters. Chapter 1 is dedicated to the introduction of this research followed by Chapter 2 which explains the concept of UAV, communication channel modelling, network traffic, interference and live-video streaming concept is explained. The dynamic model environment and the channel modelling are explained and used in Chapter 3, 4, and 5. Chapter 3 presents the performance of the proposed DQN algorithm in the co-design communication and control problem for the static environment, and dynamic and multiple environments for (i) moving UAV-BS and dynamic bit-rate selection and (ii) dynamic movement selection for both UAV-BS and UAV-UEs with dynamic bit-rate selection. The interference mitigation is presented to solve the issues of overwhelm number of UAVs and the scheduling control is considered as main point to mitigate the interference, and these will be described in Chapter 4. The context-information-beam-mapping to solve beam-pair alignment problem in uplink mmWave MIMO communication system is presented in Chapter 5. Chapter 6 concludes the study and discusses the future research direction.

Chapter 2

Literature Review

This chapter presents context information for the technical works discussed in the rest of the thesis. The basic of Unmanned Aerial Vehicle (UAV) types and functions is presented completely for a clear understanding of the whole thesis including basics of UAVs' communication, channel modelling, network traffic requirement, and their interference. To solve the problem in live-streaming from UAVs', the case study in surveillance areas with dynamic bitrate stream are discussed. Also, the issue in mitigate the interference has also been studied. Next, the mmWave beam alignment and interference mitigation will also discuss. The concept of Machine learning, especially Reinforcement Learning, Deep Q-Learning are then introduced for an essential understanding of technical works used in Chapter 3, 4 and 5.

2.1 Unmanned Aerial Vehicle

The unmanned aerial vehicle (UAV) or commercially known as a drone is an aircraft without a human pilot on board. UAV is typically a small type of aircraft and able to fly in any environment, i.e. complex and complicated area with many obstacles. Therefore, the UAV can be deployed to the surveillance area in high speed to oversee and provide visual information of the overall situation of surveillance area [131]. The UAV needs wireless connectivity for communication between UAV and controller (base station), and the controller needs a communication system to send the command to UAV. Wireless communication has experienced significant massive expansion since the early 1980s, and it is now the fastest-growing section of the communication industry and is one of the best means of addressing the accelerating demands for instant and extensive access to information [117]. Wireless communication has revolutionized the working en-

vironment and workforce mobility by eliminating the need for individuals to be tethered to a fixed location or formal work-based setting. In addition, UAVs can be deployed quickly whenever needed, which makes them promising candidates for providing cellular connectivity [58]. Furthermore, UAV also may work as temporary BS, i.e. serve the ground user as temporary telecommunication network [186].

UAV have emerged as powerful tools in various fields, and one area where they have proven to be indispensable is firefighting [132]. The integration of UAV technology in firefighting operations offers numerous benefits and addresses critical challenges faced by firefighters [132]. One of the primary reasons why we need UAVs for firefighting is their ability to provide enhanced situational awareness. Equipped with advanced cameras, sensors, and thermal imaging technology, UAVs offer real-time aerial surveillance of fire incidents. They capture high-resolution images and videos, providing firefighters with crucial information about the fire's behavior, size, and progression. This valuable data allows them to make informed decisions, develop effective strategies, and deploy resources in the most efficient manner.

2.1.1 Category of UAV

Based on studies, UAV can be categorized into three categories, based on generic and distinct application, namely as Internet delivery, Attack, and Payload [67, 86, 148].

Internet Delivery

For internet delivery, UAV helps in providing additional wireless infrastructure for broad coverage areas or disaster-struck areas. The UAV will hover in the area and be virtually stationary. The falling cost and increasing sophistication of consumer UAVs, combined with miniaturization of base station (BS) electronics, have made it technically feasible to deploy BSs on flying UAVs [58]. Since UAV BSs can be quickly deployed at optimum locations in 3D space, they can potentially provide much better performance in terms of coverage, load balancing, spectral efficiency, and user experience compared to existing ground based solutions. The deployment of UAV BSs, however, faces several practical issues. In particular, UAV placement and mobility optimization are challenging problems for UAV BSs, which have attracted significant attention from the research community. The optimization of UAV power consumption and the development

of practical recharging solutions for UAVs are also important challenges to overcome for sustaining the operation of UAV BSs. Finally, the optimization of the end-to-end link when moving UAV connects to terrestrial users (TUEs) of the network is an issue to be studied. When a large number of cells, TUEs, and UAVs exist in the network, with limited frequency bandwidth and spectrum resource reuse when BSs serve TUEs and UAVs, and causes severe interference to TUEs, especially when the network has a heavy loaded.

Military Attack

The UAVs are utilized in several civilian and military communication fields. While under the military attack category i.e., military target attack [67], the UAV will require to make forays into enemy territory, therefore the UAV must move fast and be able to plan autonomously. The UAV may operate with various degrees of autonomy, either control by a human operator or autonomously by a computer system. With the useful function and flexibility of UAV, the demand for it has increased dramatically over the last decade [67].

However, several UAVs cannot be deployed for military usage as they pose energy and security constraints. These issues are solved by introducing agents that can perform security and individual authentication, checking of integrity, image forgery detection, encryption, validation, and information collating and planning of paths [57].

Payload

UAV payload refers to the maximum weight that a UAV can carry, which measures its lifting capability. Payloads of UAV vary from tens of grams up to hundreds of kilograms. The larger the payload, the more equipment and accessories can be carried at the expense of a larger drone size, higher battery capacity, and shorter duration in the air. In addition, the UAV uplink also can cater for payload communication, especially video streaming [30, 104]. Typical payloads include video cameras and all sorts of sensors, which could be used for reconnaissance, surveillance, and commercial purposes. Normally, UAV equip with video camera requires much higher transmission rate than its control than non-payload communication (CNPC) in the downlink [104, 167]. UAV also can work as detection and collecting information, i.e., detection of forest fires or survey of crops. The UAV will to stream a video and send to ground BS [148].

2.1.2 Range and Altitude

The range of a UAV refers to the distance from which it can be remotely controlled. The range varies from tens of meters for small UAVs to hundreds of kilometers for large ones. Altitude here refers to the maximum height a UAV can reach regardless of the country-specific regulations [170]. The maximum flying altitude of a given UAV is a critical parameter for UAV-aided cellular communications, since a UAV BS needs to vary its altitude to maximize the ground coverage and satisfy different quality of service (QoS) requirements. Overall, UAV platforms can be classified into two types, depending on their altitude:
A. Low-altitude platforms (LAPs): LAPs are usually used to assist cellular communications, since they are more cost-effective and allow fast deployment. Moreover, LAPs usually provide short-range line-of-sight (LOS) links that can significantly enhance the communication performance [12].
B. High-altitude platforms (HAPs): HAPs or some call as balloons can also provide cellular connectivity. Compared to LAPs, HAPs have a wider coverage and can stay much longer in the air [12]. However, HAP deployment is more complex.

2.2 Communication

The wireless radio channel poses a severe challenge as a medium for reliable high-speed communication [64]. It is not only susceptible to noise, interference, and other channel impediments, but these impediments change over time in unpredictable ways due to user movement. In this section, we will characterize the variation in received signal power over distance due to path loss and shadowing. Path loss is caused by dissipation of the power radiated by the transmitter, as well as effects of the propagation channel. Path loss models generally assume that path loss is the same at a given transmit-receive distance. Shadowing is caused by obstacles between the transmitter and receiver that absorb power. When the obstacle absorbs all the power, the signal is blocked. Variation due to path loss occurs over very large distances (100-1000 meters), whereas variation due to shadowing occurs over distances proportional to the length of the obstructing object (10-100 meters in outdoor environments and less in indoor environments). Since variations due to path loss and shadowing occur over relatively large distances, this variation is sometimes referred to as large-scale propagation effects or local mean attenuation. Variation due to multipath occurs over very short distances, on the order of the signal wavelength, so these variations are sometimes referred to as small-scale propagation effects or multipath fading. All transmitted and

received signals we consider are real [64]. That is because modulators and demodulators are built using oscillators that generate real sinusoids (not complex exponentials).

2.2.1 Propagation Models

Communication is burdened with particular propagation complications, making reliable wireless communication more difficult than fixed communication between and carefully positioned antennas. The antenna height at a mobile terminal is usually very small, typically less than a few meters. Hence, the antenna is expected to have very little 'clearance', so obstacles and reflecting surfaces in the vicinity of the antenna have a substantial influence on the characteristics of the propagation path. Moreover, the propagation characteristics change from place to place and, if the terminal moves, from time to time. Usually the wireless channel is evaluated from 'statistical' propagation models: no specific terrain data is considered, and channel parameters are modelled as stochastic variables. Three mutually independent, multiplicative propagation phenomena can usually be distinguished: multipath fading, shadowing and 'large-scale' path loss.

Multipath propagation

Fading leads to rapid fluctuations of the phase and amplitude of the signal if the vehicle moves over a distance in the order of a wave length or more. Multipath fading thus has a 'small-scale' effect.

Shadowing

Shadowing is a 'medium-scale' effect: field strength variations occur if the antenna is displaced over distances larger than a few tens or hundreds of metres. Path loss: In general, path loss is a non-negative number since the channel does not contain active elements, and thus can only attenuate the signal. The path gain in dB is defined as the negative of the path loss: $Pt_t = P_L = 10log_{10}(P_r/P_t)$, which is generally a negative number. With shadowing, the received power will include the effects of path loss and an additional random component due to blockage from objects. The 3GPP model also captures the fact that the path loss exponent of ground- to-UAV links generally decreases as UAVs increase their height. Indeed, UAVs in LOS with their BSs experience a path loss similar to of free-space propagation ($\alpha = 2 : 2$).

Large-scale path loss.

The 'large-scale' effects cause the received power to vary gradually due to signal attenuation determined by the geometry of the path profile in its entirety. This is in contrast to the local propagation mechanisms, which are determined by terrain features in the immediate vicinity of the antennas. The large-scale effects determine a power level averaged over an area of tens or hundreds of metres and therefore called the 'area-mean' power. Shadowing introduces additional fluctuations, so the received local-mean power varies around the area-mean.

Rayleigh fading Rayleigh fading is caused by multipath reception. The mobile antenna receives a large number, N, reflected and scattered waves. Because of wave cancellation effects, the instantaneous received power seen by a moving antenna becomes a random variable, dependent on the location of the antenna. A sample of a Rayleigh fading signal. Signal amplitude (in dB) versus time for an antenna moving at constant velocity. Notice the deep fades that occur occasionally. Although fading is a random process, deep fades have a tendency to occur approximately every half a wavelength of motion.

Rician fading The model behind Rician fading is similar to that for Rayleigh fading, except that in Rician fading a strong dominant component is present. This dominant component can for instance be the line-of-sight wave. Refined Rician models also consider that the dominant wave can be a phasor sum of two or more dominant signals, e.g. the line-of-sight, plus ground reflection. This combined signal mostly treated as a deterministic (fully predictable) process, and that the dominant wave can also be subjected to shadow attenuation. This is a popular assumption in the modelling of satellite channels. Besides the dominant component, the mobile antenna receives a large number of reflected and scattered waves.

Rician factor The Rician K-factor is defined as the ratio of signal power in the dominant component over the (local-mean) scattered power. In the expression for the received signal, the power in the line-of-sight equals C2/2. In indoor channels with an unobstructed line-of-sight between transmitter and receiver antenna, the K-factor is between, say, 4 and 12 dB. Rayleigh fading is recovered for K = 0 (-infinity dB). However, in Rician fading the mean value of (at least) one component is non-zero due to a deterministic strong wave.

Nakagami fading The Rician and the Nakagami model behave approximately equivalently near their mean value. This observation has been used in many recent papers to advocate the Nakagami model as an approximation for situations where a Rician model would be more appropriate. While this may be accurate for the main body of the probability density, it becomes highly inaccurate for the tails. As bit errors or outages mainly occur during deep fades, these performance measures are mainly determined by the tail of the probability density function (for probability to receive a low power).

2.2.2 Network traffic requirement

In 3GPP highlight the UAV Traffic Requirements. The 3GPP identified the traffic types that cellular networks should cater for UAVs flying between ground level and 300 meters. These can be classified into three categories: 1) synchronization and radio control, 2) command control, and 3) application data.

1) Synchronization and radio control: The information contained within the synchronization and radio control messages is essential for a successful association and connectivity to the network. The transmission of these signals must be robust enough to guarantee that they can be decoded by flying UAVs. Examples of synchronization and radio control signalling include primary and secondary synchronization signals (PSS/SSS) and the physical downlink control channel (PDCCH), respectively.

2) Command control (CC) : The traffic enables beyond LoS UAV piloting and has strict quality of service requirements (QoS) in terms of latency and reliability. Cellular operators have identified an attractive business opportunity in the management of this traffic, since it can be offered as a complementary network service to organizations interested in reliably controlling their UAVs.

3) Application Data: The UAV application data transmissions are expected to be uplink-dominated. However, transferral of live video streaming data and photos captured by camera equipped UAVs contribute towards this traffic imbalance.

2.3 Video Streaming

UAV video streaming refers to the real-time transmission of video footage captured by an UAV to a remote location. With the advancements in UAV technology, live streaming capabilities have become increasingly popular and widely used in various applications such as aerial photography, surveillance, sports events coverage, and emergency response operations [72]. UAV live streaming offers several advantages in different industries. In the field of aerial photography and videography, it provides a unique perspective and allows professionals to capture stunning aerial shots and videos. In surveillance and security applications, live streaming from UAVs enables real-time monitoring of large areas, enhancing situational awareness and facilitating quick response to incidents [132]. During emergency response operations, such as firefighting or search and rescue missions, UAV live streaming provides valuable visual information to aid decision-making and coordination of rescue efforts.

Another challenge that needs to be taken into consideration is the satisfaction of the user during the live-streaming or in other words, good quality of experience (QoE) [70, 180, 30]. Live-streaming requires any media delivered and played back simultaneously from UAVs [180]. Then, humans will monitor the current situation and give feedback towards the action which they should take. The large dynamic environment poses a limitation to humans to learn the scenario and control action toward the system in order to maintain the maximum QoE.

2.3.1 Video Stream in Surveillance System

Wireless or mobile surveillance systems that integrate wireless cameras or ad hoc wireless video sensor networks with moving vehicles or mobile devices are necessity. The video streams captured by wireless or mobile camera stations (CSs) are uploaded via wireless channels to a control center where the acquired videos can be archived, analyzed, and/or distributed [155]. Surveillance systems of this kind have numerous applications, including real-time traffic monitoring, facility monitoring, combat/rescue operation monitoring, disaster relief, and damage assessment, etc. Depending on specific application scenarios, different quality of service (QoS) requirements may have to be imposed on the design of such systems. For example, video streams that report critical unfolding events will require high QoS levels compared with streams that contain no events. These existing works focused on QoS parameters in physical-layer performance metrics, such as packet loss rate, throughput, and jitter. The fluctuating performance of video stream-

ing by UAVs in firefighting scenarios necessitates high data rates, reliability, and smoothness. However, current methods of measuring QoE utility are impractical [132, 156]. The gap lies in the need for real-time video streaming and QoE utility measurement that can meet the long-term requirements of firefighting operations. To address this, research should focus on developing dynamic video streaming techniques and innovative QoE measurement approaches specific to firefighting environments. Closing this gap will enable more effective decision-making, enhance situational awareness, and improve firefighting outcomes. Therefore, the overall system needs to learn faster and respond quickly towards the action request to ensure the QoE is satisfied.

2.3.2 Dynamic bitrate streaming

Dynamic or adaptive bitrate streaming is a dynamic streaming technique that adjusts the quality of video playback based on the viewer's network conditions, ensuring a smooth and uninterrupted viewing experience. Adaptive bitrate streaming is a technique used in streaming multimedia over computer networks. While in the past most video or audio streaming technologies utilized streaming protocols such as RTP with RTSP, today's adaptive streaming technologies are almost exclusively based on HTTP [75] and designed to work efficiently over large distributed HTTP networks such as the Internet. It works by detecting a user's bandwidth and CPU capacity in real time and adjusting the quality of the media stream accordingly. Today's UAVs are struggling to deliver high-quality video in real time to ground receivers. The commercial UAVs adopt fixed-bitrate video streaming strategies which may result in severe rebuffering under poor Internet connection. The designs of wireless technologies that enable real-time streaming of high-definition video between UAVs and ground clients present a conundrum.

Real-time video streaming has higher requirements in terms of data rate, latency, and smoothness compared to other data types. In a firefighting scenario, the network channel capacity fluctuates dramatically with the dynamic environment alongside the UAVs' movement, which can cause poor network performance and undesirable delays. This in turn makes it harder to learn the pattern variance of the channel capacity, thus resulting in failure to transmit with high capacity and high video quality. To capture the practical performance from test bed, authors in [184] used single UAV to conduct indoor experimental to measure the video streaming performance from one LTE BS. Therefore, to overcome the limitation of fluctuate environment, the authors in [156] applied the Additive Variation Bitrate (ABR) method with Deep Reinforcement Learning (DRL) to select

proper video resolution based on previous communication rate and throughput. However, [156] only focused on a single video source ABR, which was guided by RL to make decisions based on the network observations and video playback states for selecting the optimal video resolution. While managing large firefighting areas, multiple UAVs are required, authors in [70] used multiple UAVs to stream a video and optimize the QoE to solved resource allocation using game theory technique, however, the QoE utility measurement used error statistic of PSNR and mean of sum (MOS) scale, which could lead to biased measurement.

Authors in [41] used UAV relay network and considered two factors, the bit rate of the video and the freezing time, to maintain the quality. However, the dynamic channel and different requests are not considered. Therefore, we improve the quality measurement by introducing three video quality factors, i.e., video resolution measurement, video smoothness, and latency penalty. This problem motivates us to exploit neural networks without relying on preconfigured information such as velocity and distance. However, in large search and rescue firefighting scenario, a nonordinary optical camera [65] should be considered to ensure the reception of a high quality video. To deal with a more complex environment and practical scenarios, such as search and rescue firefighting scenarios, the machine learning algorithm is a promising tool for solving the problem. The machine learning algorithms, especially deep reinforcement learning can be adapted to the fluctuated channel quality in networks and ensure the long-term QoE.

2.4 Interference Mitigation

To enhancing UAV communications and increase the communication effectiveness, the interference mitigation should be studied. A main challenge in providing connectivity to the low altitude UAVs through existing cellular network arises due to the increased interference in the network [118]. The increased altitude and favourable propagation condition cause UAVs to generate more interference to the neighbouring cells, and at the same time experience more interference from the downlink transmissions of the neighbouring BSs. The uplink interference problem may result in TUEs having degraded performance, whereas the downlink interference problem may make it challenging for a UAV to maintain connection with the network. The transmission performance of UAVs and TUEs is severely affected by interference among them, when BSs serving them in the same frequency simultaneously [23]. Using the model in [15], the study in [118] showed that highly loaded scenarios decreased UAV coverage due to high interference.

In addition, the authors in [88] gave theoretical interference analysis of cellular-connected UAV networks with TUEs based on radio characteristics, including UAVs' heights, ISD and signal-to-interference ratio level.

Consequently, authors in [8, 20, 23, 30, 160] considered interference mitigation schemes between TUEs and UAVs, by considering power control [8, 20, 30, 160], reducing UAV height [8], and antenna beam selection [23]. Although decreasing power allocation, reducing UAV height, and selecting proper antenna beams can mitigate interference and improve throughput, they can result in low coverage of UAVs and increase outage probability when BSs serving UAVs and TUEs simultaneously. To address this issue, authors in [103, 105] designed the cooperative beamforming technique to effectively suppress inter-cell interference (ICI) to the UAV, and authors in [119] designed a muting scheme to mute the cells with high interference to decrease interference between UAVs and TUEs.

The power control mechanism ensures that the transmit power of different uplink channels are controlled so as these channels are received at the BSs at appropriate power level [160]. The power control procedure aims to control the received power to be just enough to demodulate the channel (target received power), at the same time the transmit power at UEs are not unnecessarily high as it could create interference to the other uplink transmissions [160]. In many standards like LTE, the transmit power of the UE depends on the DL pathloss and target received power at the serving BS.

In addition, several downlink interference mitigation techniques to address in the 3GPP [160]. The full dimension MIMO (FD-MIMO) multi-antenna BSs defined in LTE Release 13 enhance the performance of UAV communications, which allow reducing the amount of interference generated towards the constrained spatial regions where UAVs lie, and their spatial multiplexing capabilities, which in turn enable a better utilization of the precious time/frequency resources.

UAVs with directional antennas and beam forming capabilities contribute to reduce the number of downlink interferers perceived by UAV devices. These interference mitigation gains can be further complemented with a boost of the useful signal power in UAVs beam steering towards their serving BS. Clearly, this solution entails a complexity increase in the design of hardware UAV transceivers.

2.4.1 Inter-cell Interference

Furthermore, when large number of moving TUEs and UAVs exist in 5G networks, there will be high inter-cell interference when BSs serving them. In [90, 91, 102], the authors considered cell muting and traditional optimization methods to

mitigate the ICI. Specifically, authors in [102] optimized UAV resource allocation based on their cell association to maximize throughput performances of TUEs and UAVs, and considered inter-cell interference coordination (ICIC) based on Release-10/11 to mitigate strong interference to TUEs. However, only a single UAV was considered in [102] and the UAV could only access to the resource block (RB) that had not been occupied by any TUEs, thus, the approach in [102] could not be adapted to the scenario with multiple UAVs. Authors in [90, 91] used the cell range expansion (CRE), enhanced inter-cell interference coordination (eICIC), and further-enhanced ICIC (feICIC) schemes to improve the overall spectral efficiency. However, the optimization methods in [90] and [91] aimed at optimal solutions in each time slot with high computation complexity and were not designed for long-term optimization problem.

With increasing number of devices in future terrestrial networks, the interference problem between UAVs and TUEs becomes more complicated. Therefore, efficient ICIC designs are required for enabling efficient spectrum sharing between UAVs and TUEs in future cellular-connected networks, in which, the resource allocation can be designed to mitigate interference and improve throughput of UAVs and TUEs. Based on the previous RB allocation and traffic patterns, authors in [140] proposed a deep Q-network (DQN) to select proper RBs for UAVs and TUEs to perform transmission with low interference.

Although 5G helps improve data rate performance, it has some drawbacks. Therefore, more 5G BSs are built to support 5G connectivity in multiple areas and brought BSs closer to users [16]. With the increase in the number of 5G BSs and TUEs, the interference among them increases, and with the increase in transmission opportunity, the situation becomes more complex to reduce interference while guaranteeing high quality of service (QoS) of UAVs and TUEs. To mitigate interference in complex scenarios, the DRL algorithm is considered.

To the best of our knowledge, none of these studies investigated multiple UAVs and TUEs' coordination in the cellular network and deployed dynamic RB scheduling to maximize long-term throughput performance. In practice, the control signal reception of UAVs is not only affected by the link quality of the communication channel, but also susceptible to interference. Thus, the control links between the BS and TUEs and UAVs are important, especially when the spectrum resources are constrained. To effectively solve the aforementioned problems, UAVs and TUEs require high-level coordination to ensure all users meet their minimum requirements and optimize their data-rate performance, especially in a highly dynamic environment.

To address this gap, we propose a dynamic scheduling, muting, and resource block management approach to effectively manage interference between UAVs and TUEs, ensuring a high Quality of Service (QoS) for both user groups. By implementing dynamic scheduling algorithms and muting techniques, we can adaptively allocate resources and manage interference in real-time, optimizing the overall network performance. This approach will enable efficient utilization of resources, mitigate interference, and ultimately enhance the QoS for UAVs and TUEs in complex and dynamic environments. The interference is decreased by muting the cells with the strongest interference and RBs are properly shared and scheduled to UAVs and TUEs, with the aim to satisfy high QoS requirements of UAVs and TUEs.

2.5 mmWave beam allignment

The mmWave beam alignment could enhance the reliability data for both UAV communications and BS-to-UAV communications under 5G and beyond wireless systems. The availability of user's position information, which could help for reliable communication and increase throughput. Position information can be leveraged for fast beam alignment in the upcoming sixth generation (6G) mmWave communications. The 5G radio wave direction can be obtained through mmWave frequencies and MIMO beamforming and to enable high speed data access and LoS dominant connectivity to UAVs. Position information for fast beam alignment has been recently studied under vehicular context in mmWave systems [142]. On the other hand, high mobility and autonomy UAV operation requires frequent realignment of the beam. To be faster and more reliable beam alignment, position information is highly needed in enabling high data rates for mmWave UAVs.

Beam alignment is the process of aligning the directional beams of both the transmitter and receiver antennas to establish a strong and stable communication link. An effective beam alignment or tracking scheme is usually required to ensure the consistency of beam alignment in a high mobility environment. Existing works [182] proposed beam tracking schemes using Kalman filters with high processing complexity. An alternate approach is to undergo beam training [173] and perform fast beam alignment after every significant change in UAV position along the BS coverage area. Existing works in vehicular environment proposed different training-based beam alignment approaches for terrestrial systems based on stochastic methods such as genetic and evolutionary algorithms [84, 133] and the use of contextual information [18, 46, 142, 144, 151]. Contextual information

generally involves data from the sensors such as position information, antenna configurations, channel state information and receiving signal power using low frequency carrier (e.g. during initial communication in 3rd generation partnership project (3GPP) beam access protocol [79, 43]) whenever needed, as this information is used abundantly to reduce the beam training overhead.

High mobility and autonomous operation of UAVs also requires frequent beam realignment and can be jointly optimized with reliable connectivity using reinforcement learning (RL)-based beam training [134, 47, 135, 137]. Authors in [47] jointly optimized UAV-BS trajectory and mmWave connectivity using deep reinforcement learning (DRL) techniques to obtain secure transmission and energy efficiency with eavesdroppers. In our previous work [135], we considered randomly moving mmWave UAVs and proposed a position-aided beam-pair alignment learning framework at terrestrial BS using deep Q-network (DQN). In this work, we have shown that a generic DQN-based framework at BS can enhance the mmWave beam-forming gains for any randomly moving UAV inside their coverage area in an online manner under different 3GPP conditions. However, the work assumed independent and fixed grid elements in the BS - UAV environment, increasing both action spaces and significant communication overhead for uniform planar array (UPA) antenna configurations under the learning framework. Hence, a better learning-based beamforming strategy is required for UPA antennas to enhance reliable connectivity for autonomous UAVs.

2.6 Optimization Problem

This section introduces several scenarios for optimizing joint UAV network to attain maximum network efficiency. Table 2.1 summarizes a selection of complex problems that involve optimizing multiple parameters. The study will explore numerous such intricate optimization problems and their corresponding solutions.

The authors in [125], briefly analysed to minimize the transmits power of the communication system by optimizing period, number of bits and packet error to ensure the control system's reliability and delays requirements needed to guarantee its stability. However, the overall packet loss is caused by decoding errors, transmission delay, and queuing delay violation [129], therefore, it is also important to optimize the uplink and downlink bandwidth configuration and delay, so it can minimize the total bandwidth. Another objective function commonly used to maximize the spectral efficiency [35] and/or uplink/downlink sum-rate [177]. In order to support an inadequate network, UAV has been initially proposed as

Table 2.1: Optimization problems to improve the network

Ref	Year	Optimization parameter	Objective
[125]	2014	sampling period, number of bits, packet error	minimize the power consumption
[129]	2018	uplink-downlink bandwidth configuration and delay	minimize the total bandwidth
[35]	2019	uplink resource allocation, communication, control	maximize the spectral efficiency
[177]	2019	subchannel allocation, UAV speed	maximize the uplink sum-rate
[171]	2019	trajectory, power allocation, transmission schedule, rate allocation	maximize the overall utility
[161]	2019	UAV location	maximize energy efficiency
[41]	2020	bandwidth, power allocation	maximize the total long-term QoE
[94]	2021	resource allocation, UAV trajectory	maximizing the total energy efficiency
[164]	2022	UAV's uplink cell associations, TX power allocations over multiple RB	maximize weighted UAV sum-rate of UAV, and TUEs

a relay to help other UAVs transmit to a nearby terrestrial BS with low signal-to-noise ratio (SNR) [177], therefore it is important to consider the subchannel allocation and UAV speed. Authors in [171] proposed a joint trajectory, power allocation, transmission schedule and rate allocation during the mission. The objective is to maximize the overall utility. The simulation validate the analytical findings against existing benchmark techniques.

The authors in [94, 161] focused on maximized energy by proposed to optimized UAV location [161] and resources allocation and UAV trajectory [94]. Another objective function commonly used is sum-rate UAV, authors in [164] find it by optimizing the cell association and power allocation. Authors in [41] used UAV relay network and considered two factors, the bit rate of the video and the freezing time to maintain the quality.

Most optimization problems that are related to UAVs placement and resource allocation can be found in the literature. We classify them into three categories: resource allocation for fixed UAV positions, 3D placement and UAV trajectory optimization, and UAV-BS 3D placement.

A. Resource Allocation for Fixed UAV Positions In [127], the authors present a distributed greedy approach to improve the user's sum-rate under backhaul capacity, bandwidth constraint, and maximum number of links limitation. The optimal power and spectrum allocation are investigated in [92] where the authors minimize the mean packet transmission delay for uplink communications. In [93], the authors' goal is to minimize the maximum energy needed to ensure a certain bit error rate target. Last but not least, the authors propose the global scheduling technique using standard optimization, and provide the light version of the algorithm to reach the suboptimal solution.

B. 3D Placement and UAV Trajectory Authors in [14] investigate the 3D placement of UAVs while maximizing the number of covered users. The UAV horizontal and vertical locations are optimized separately. The optimal altitude is found by solving a convex decoupled optimization problem, while the optimal 2D location is achieved by finding a solution to the smallest enclosing circle problem. In [51], authors optimize the UAV trajectory to accurately learn the environment propagation parameters. The authors introduced the map compression method and use dynamic programming to efficiently design the UAV trajectory. The optimal UAV position to maximize the end-to-end throughput is studied in [37] where information provided by the signal strength radio map is leveraged. In line with the previous cited work, authors in [36] provide an online algorithm, based on the theory of asynchronous stochastic approximation, for a fast deployment of flying relays, that minimizes the power consumption under constraints of outage probability and number of deployed drones.

C. Placement Optimization for UAV BSs The problem of optimum placement is more challenging for UAV BSs compared to the conventional terrestrial BSs because the UAV BS can be placed at many different heights in the sky [63]. However, the coverage as well as the different channels change with the altitude of the BS. Different researchers used different algorithms to solve the placement

optimization problem for UAV BSs. Some researchers considered the height of the UAV BS as a variable in their optimization formulation, thus treating it as a 3D placement problem, while others essentially solved 2D placement problems for constant heights. Optimizations also differed in whether the backhaul, interference from other BSs, and existence of terrestrial BSs in the same coverage area were considered in problem formulation.

3D placement optimizations were studied in [112] but sought to minimize the total transmit power of the homogeneous network of UAV BSs. For such scenarios, the authors solved the 3D placement optimization problem by dividing the problem into two subproblems that are solved iteratively. Given the height of the UAV BSs, the first sub-problem obtains the optimal locations of the UAVs using the facility location framework. In the second sub-problem, the locations of UAV BS are assumed to be fixed, and the optimal heights are obtained using tools from optimal transport theory. Two types of mobilities based on the transport methods of UAV BSs:

1) UAVs are used only to transport the BS to a particular ground location where the BS autostarts to serve the users. If the BS needs to be relocated, it must shut down first before being transported to the new location. However, this type of UAV BSs cannot serve while it is in motion, but it can resume its service as soon as it reaches a target location.

2) UAVs continue to carry the BSs and the BSs can continuously serve the ground users while they are flying. Considering the first type of UAV BSs, Chou et. al., [42] studied the BS placement mechanism where the ground users are not served by the BS when it is moving to a new location. The loss of service time due to BS mobility therefore becomes a critical parameter for the optimization. The UAV BS in this case should consider both the user density of the target location as well as the moving time to the new location when deciding its target location. The authors of [42] have shown that this problem can be modelled as a facility location problem, where the transport cost represents the loss of service time due to the movement of the BS from the previous location to the new location.

The second type of UAV BS mobility opens up new opportunities to use UAV BSs due to their ability to serve ground users while in motion. In particular, under this scenario, the cost of BS mobility becomes negligible. It is then possible to design more advanced solutions where UAV BSs can continuously cruise the service area to maximize network performance under geospatial variance of demands. In both types, although a single UAV can perform plenty of tasks, multiple UAVs can form a cooperative group to achieve an objective more ef-

ficiently, and to increase the chance of successful task operation. Additionally, the robustness of the communications will increase by cooperative UAVs [181]. Maintaining the connectivity and controlling the distance between multiple UAVs is one of the main challenges in using cooperative UAVs [122]. Maximizing the coverage area [124], cooperative carrying task [40, 139], searching and localizing a target [45] are among the tasks that can be done by multiple UAVs.

Designing cruising UAV BSs requires autonomous mobility control algorithms that can continuously adjust the movement direction or heading of the BS in a way that maximizes system performance. These algorithms must also ensure that multiple UAV BSs cruising in an area can maintain a safe distance from each other to avoid collisions. Fotouhi et. al. [60] proposed distributed algorithms that take the interference signals, mobile users' locations and the received signal strengths at UEs into consideration to find the best direction for BS movements at any time. Controlling the mobility of a single serving UAV is also discussed in [58]. Game theoretic mobility control algorithms are proposed in [59, 60] for multiple UAV base stations cruising freely over a large service area without being subject to individual geofencing. The game theoretic mobility control not only increased packet throughput by 4 times compared to hovering BSs, it also helped avoid collisions as the BSs were implicitly motivated to move towards different directions to maximize coverage and throughput. The trajectory of a single UAV also can be optimized to improve the system performance. The UAV with a mission to fly between a source and destination point is studied in [178]. During this mission it has to maintain a reliable connection by associating with the ground BSs at each time. In [99] a UAV is used to offload data traffic from cell edge users and improve their performance. It is shown that by using one single UAV and optimizing its trajectory, the throughput improves significantly compared with the conventional cell-edge throughput enhancement scheme with multiple micro/small cells.

2.7 Machine Learning

In machine learning, there are three main types depending on training method, Supervised, Unsupervised and Reinforcement learning.

Supervised learning train the machine using data which are well "labeled." It means some data is already tagged with the correct answer. It can be compared to learning which takes place in the presence of a supervisor or a teacher. Furthermore, supervised learning algorithm learns from labeled training data, helps

to predict outcomes for unforeseen data.

Unsupervised learning is a machine learning technique that allows the model to work on its own to discover information without supervision. It mainly deals with the unlabelled data. Unsupervised learning algorithms allow to perform more complex processing tasks compared to supervised learning. Unsupervised machine learning finds all kind of unknown patterns in data. Unsupervised methods help to find features which can be useful for categorization.

Reinforcement learning is well-known and helpful to solve various problems in a highly dynamic environment [81]. Although there are many optimization methods, Reinforcement Learning (RL) seems to give an excellent solution where provides a great solution to a complicated and practical situation, and it is capable of interacting with the stochastic environment and giving feedback to the control. Various literature found RL helps to solve problem in control and also in communication [7, 76, 33, 96, 97, 156, 162]. RL also has shown good results in multiple disciplines such as medical, chemistry and many more. RL had effectively shown control of the path planning in [76, 33], and using the Echo State Network (ESN) solved the problem in planning and latency is thus optimized [33]. RL also helps in maximizing energy control whilst ensuring the fairness of communication connectivity [96]. All the research above say that RL methods guaranty the converged result. However, RL is a proving method to converge results and able to solve a dynamic environment effectively.

2.7.1 Applications of ML in UAV Communications

Nowadays, various applications of the ML solutions in UAV-enabled communications are depicted. Generally, the importance of distilling intelligence in wireless communication networks has been outlined in numerous works [133, 158, 175, 168, 56, 48, 29, 73, 25, 157, 47, 137, 159, 174, 66]. The authors have observed that the ever increasing heterogeneity and complexity of mobile networks has made monitoring and management of network elements intractable [172]. Moreover, ML allows systematic mining of valuable information from mobile data and automatically identifies correlations that are too complex to be derived by human experts. Likewise, the work in [85] has noted that, in wireless networks, ML enables the wireless devices to actively and intelligently monitor their environment, exploiting mobile data for training purposes in order to learn, predict, and adapt to the evolution of environmental features, including wireless channel dynamics, traffic and mobility patterns, as well as network composition, among others. In this way, they can proactively act towards maximizing the probability

of satisfying different performance metrics. It can be seen that ML takes as input data from different sources and through the application of various learning techniques, i.e., supervised/unsupervised, deep or reinforcement learning, allows the network to adapt to the wireless environment in a dynamic and autonomous manner. Thus, especially in networks consisting of a large number of nodes, such as those consisting of swarms of UAVs, centralized coordination and excessive overheads that must be acquired and exchanged among the network nodes are avoided, paving the way for distributed network optimization where intelligence plays a key role.

Motivations of Deep Reinforcement Learning (DRL) is to address the above issues in network optimizations, a wireless network should be intelligent enough to adjust itself in dynamic environments, explore unknown optimal policies, and transfer theoretical knowledge to practical scenarios. Motivated by these demands, machine learning technologies have been considered as viable solutions to solve communications issue [20]. There are numerous machine learning technologies that have been applied in communication systems [28]. Among them, DRL has shown great potential in beyond 5G wireless networks [172, 165]. Unlike optimization algorithms, DRL approaches can be model-free [98], and implemented in real-world communication systems. DRL leverages deep neural networks (DNN) and reinforcement learning to enable UAVs to learn from experience and make intelligent decisions in complex and dynamic environments. By combining perception, decision-making, and action, DRL algorithms allow UAVs to adapt and optimize their behavior over time.

In designing wireless networks for URLLC, the stringent QoS requirements should be satisfied. When using a DNN to approximate the optimal policy, the approximation should be accurate enough to guarantee the QoS constraints. However, when the environment changes, the pre-trained DNN can no longer guarantee the QoS constraints of URLLC. To handle this issue, the system needs to adjust the DNN in non-stationary environments with few or no training samples. However, during explorations, the DRL algorithm may try some bad actions while improve the policy in the unknown environment, which will deteriorate the QoS significantly and may lead to unexpected accidents in URLLC systems. Thus, the exploration safety will become a tradeoff for applying DRL in URLLC.

In recent years, researchers have explored various applications of DRL in the UAV domain, aiming to address challenges related to navigation, surveillance, communication, and more [159, 174]. Researchers have developed DRL-based navigation systems that enable UAVs to learn optimal trajectories and

avoid obstacles in real-time. These systems leverage neural networks to process sensor data and learn navigation policies through reinforcement learning algorithms [174]. Researchers have explored multiple aspects of UAV navigation using DRL, including trajectory optimization, fault tolerance, positioning, coverage path planning, and path optimization. By applying DRL algorithms, UAVs can autonomously navigate through complex environments, avoiding obstacles and achieving efficient and collision-free trajectories [183]. For example, study in [174] focused on capacity maximization in RIS-UAV networks, utilizing DRL to optimize trajectory and phase shift, leading to improved system capacity and resource allocation efficiency. In another research effort, a 3D positioning method for UAVs employed DRL to support configurable antennas and accurately measure parameters like angle of departure, angle of arrival, polarization status, and 3D positions [159]. DRL techniques have also been applied to address fault tolerance, ensuring stable and fixed-time tracking of UAV attitudes, even in the presence of nonlinear faults. Additionally, vision-based navigation techniques, when combined with DRL, provide UAVs with the ability to perceive their surroundings and make informed navigation decisions based on visual cues, which is particularly advantageous in GPS-blind environments [17].

Another significant application of DRL in UAVs has proven to be an effective approach for addressing challenges related to surveillance and its communication. Different with DRL-based navigation systems, DRL algorithm also able optimize trajectory control and power allocation for UAVs, leading to improved transmission latency [66]. Similarly, in wireless edge networks, DRL combined with communication transformer enables intelligent edge networks that provide quality-assured live streaming services while optimizing energy consumption. DRL-driven UAV-assisted edge computing enhances the quality of experience (QoE) for Internet of Things devices, while energy-efficient UAV movement control, guided by DRL, improves coverage, minimizes energy consumption, and ensures fairness [150]. Moreover, joint communication and action learning with DRL in multi-target tracking of UAV swarms improves performance, scalability, and robustness under communication failures. Cooperative multi-agent DRL facilitates reliable surveillance in smart city applications, surpassing other algorithms in terms of surveillance coverage, user support capability, and computational costs [175]. Finally, a deep reinforcement learning approach for joint trajectory design in multi-UAV networks explicitly considers the mutual influences among UAVs, optimizing the mission time [158]. However, the 3D positioning, network, and video resolution parameter also should take into consideration to enhance the

QoE in long-term situations, which can enhance situational awareness and improve mission efficiency.

Furthermore, DRL has shown promise in improving UAV communication and network management. By utilizing DRL algorithms, UAVs can optimize their communication protocols, spectrum allocation, and resource utilization [33]. DRL has shown promise in addressing interference challenges in UAV applications. By employing DRL, interference-aware, power control, and cooperative techniques can be developed to reduce interference, improve spectral efficiency, and enhance overall system performance [33, 106] . DRL-based approaches also optimize dynamic resource allocation [103], cancelling UAV-terrestrial interference [106], and enhancing connectivity with the cellular network. These studies highlight the potential of DRL in optimizing UAV operations, reducing interference, and improving network efficiency in UAV applications. This enables efficient and reliable communication between UAVs and ground stations, enhancing the overall performance of UAV networks.

Despite the significant advancements, several challenges remain in the field of DRL for UAVs. One key challenge is the high dimensionality and complexity of the UAV environment, which requires efficient training techniques and exploration strategies. Fast mmWave beam alignment could enhance the reliability and decrease the latency of 5G and beyond wireless systems for both UAV-UAV and BS-UAV communications [34]. Especially, the availability of UAV position information at lower frequencies (following the works [143, 145]) may also provide scope for reliable communication in addition to increasing throughput. Position information for fast beam alignment has been recently studied in mmWave systems [142, 146, 19]. The authors in [146, 19] proposed a learning-based beam training schemes using multi-armed bandit (MAB) approach, by building a database of finite beam-pairs useful for beam training based on vehicular position information. In these works, the key idea is that the machine learning (ML)-based approaches can effectively use the position information for fast mmWave beam alignment in an online manner.

High mobility and autonomous operation of UAVs also requires frequent beam realignment and can be jointly optimized with reliable connectivity using RL-based beam training [134, 47, 135, 137]. Authors in [47] jointly optimized UAV-BS trajectory and mmWave connectivity using DRL techniques to obtain secure transmission and energy efficiency with eavesdroppers. In [135], considered randomly moving mmWave UAVs and proposed a position-aided beam-pair alignment learning framework at terrestrial BS using DQN. In this work, we have

shown that a generic DQN-based framework at BS can enhance the mmWave beam-forming gains for any randomly moving UAV inside their coverage area in an online manner under different 3GPP conditions. However, the work assumed independent and fixed grid elements in the BS - UAV environment, increasing both action spaces and significant communication overhead for UPA antenna configurations under the learning framework. Hence, a better learning-based beamforming strategy is required for UPA antennas to enhance reliable connectivity for autonomous UAVs.

In conclusion, DRL offers exciting opportunities for advancing the capabilities of UAVs. Through its ability to learn from experience and optimize behavior, DRL can enable UAVs to navigate autonomously, perform surveillance tasks, and optimize communication. However, further research is required to address challenges related to practicality, training efficiency, safety, and scalability.

2.8 Conclusion

In this chapter, the study focused the fundamental concept of telecommunication and control in the context of UAV-based cellular networks. First, the study presents the UAV types and functions is presented completely for a clear understanding of the whole thesis. There is a need for UAV-to-UAV to stream real-time video and practical measure Quality of Experience (QoE) utility in ad-hoc scenarios. Secondly, efficient interference management and resource allocation schemes are required to address the interference problem between UAVs and terrestrial users in future cellular networks. Lastly, a better learning-based beamforming strategy is needed for UAVs equipped with Uniform Planar Array (UPA) antennas to optimize beam alignment and enhance reliable connectivity. Addressing these gaps will contribute to the development of practical solutions for real-time video streaming, interference management, and beam alignment in UAV applications, particularly in firefighting environments. The concept of Machine learning, especially Reinforcement Learning, Deep Q-Learning are then introduced for an essential understanding of technical works used in Chapter 3, 4 and 5.

Chapter 3

UAV-to-UAV Communications

3.1 Introduction

Live video become phenomena and can be accessed worldwide. The challenge in live-streaming especially in remote areas with lack of network bandwidth capacity and information is critical for the rescue team to do search and rescue (SAR). SAR network coverage, especially in forest fire areas, presents unique challenges due to the environmental conditions and the rapid spread of the fire. Forested regions often lack reliable network infrastructure, making it difficult to provide adequate coverage during fire emergencies [10].

In a firefighting scenario, the network channel capacity fluctuates dramatically with the dynamic environment alongside the UAVs' movement, which can cause poor network performance and undesirable delays [132]. This in turn makes it harder to learn the pattern variance of the channel capacity, thus resulting in failure to transmit with high capacity and high video quality. In this chapter, we consider UAV-to-UAV (U2U) communication to facilitate such scenario, where UAV at the high altitude acts as mobile base station (UAV-BS) to stream videos from other flying UAV-users (UAV-UEs) through the uplink. Over the years, numerous wildfires have caused challenges for firefighters to control and monitor fire in remote areas [114, 131]. Without new technology to monitor the incident area from the control station, the current practice of the fire station control lacks the technology to remotely visualize the dynamic fire situation in real-time for immediate action [131]. Therefore, it became challenge for the rescue teams in fighting against wildfire in remote areas when the information of the incidents did not receive in clear manner, such as the size and images of fire areas. As such, live-streaming from UAVs help to capture videos of dynamic fire areas, and is important for firefighter commanders in any location to monitor the fire

situation with quick response. The 5G network is promising wireless technology to support such scenarios. However, the position of the UAV will affect the performance of ad-hoc network. When a UAV is positioned imperfectly within an ad-hoc network, it can lead to several performance issues such as weak signal strength and interference [162, 38]. Due to the mobility of the UAV-BS and UAV-UEs, it is important to determine the optimal movements and transmission powers for UAV-BSs and UAV-UEs in real-time.

The increased altitude and favourable propagation condition cause UAVs to generate more interference to the neighbouring UAVs. The uplink interference problem may result in UAV-UE having degraded performance, and the interference problem may make it challenging for a UAV to maintain connection with the UAV-BS. Therefore, to maximize the data rate of video transmission with smoothness and low latency, while mitigating the interference according to the dynamics in fire areas and wireless channel conditions, the co-design of video resolution, the movement, and the power control of UAV-BS and UAV-UEs is proposed to maximize the Quality of Experience (QoE) of real-time video streaming. The algorithm will learn the dynamic fire areas and communication environment by using the Deep Q-Network (DQN) and Actor-Critic (AC) to maximize the QoE of video transmission from all UAV-UEs to a single UAV-BS. Simulation results shown the effectiveness of the proposed algorithm in terms of the QoE, delay and video smoothness as compared to the Greedy algorithm.

The contributions of this chapter are summarized as follows:

- To develop a framework for a dynamic UAV-to-UAV (U2U) communication model with a moving UAV-BS in multiple firefighting areas to capture a live-streaming panoramic view. The model of the dynamic fire arrival with different heights in every fire area is designed and UAVs' request arrival as Poisson process in each time slot, and design the UAV-UEs location spaces to capture a full panoramic view with multiple UAVs.

- To guarantee the smoothness and latency of the live video streaming among UAV-BS and UAV-UEs in this U2U network, the formulation of long-term Quality of Experience (QoE) is designed to maximization problem via optimizing the UAVs' positions, video resolution, and transmit power over each time slot.

- To solve the above problem, Deep Reinforcement Learning (DRL) approach based on the Actor-Critic (AC) and the Deep Q Network (DQN) are proposed. The results shown that the proposed AC and DQN approaches

outperform the Greedy algorithm in terms of QoE.

3.2 System Model and Problem Formulation

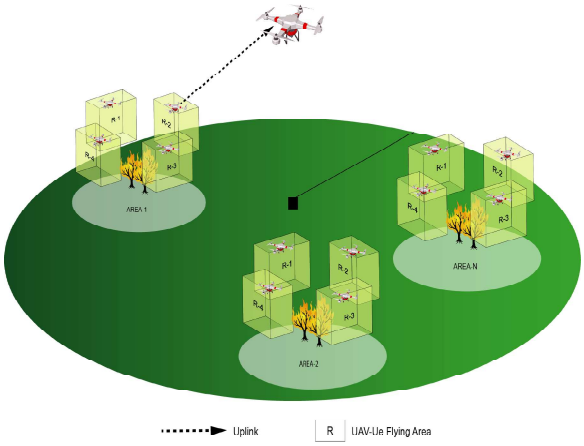

Figure 3.1: Illustration of System Model

As illustrated in Figure 3.1, this research will consider the single UAV-BS to provide the network coverage for multiple UAV-UEs to satisfy the network rate requirement of each UAV-UE to stream high quality video of multiple fire-fighting areas. The UAV-BS is located at the center of the environment, such as forest area, with the maximum coverage radius $r_{\max}$. The UAV-BS is connected through wireless network to the fixed or mobile control station. As the arriving distribution of the fire video streaming request is the same as that of the fire arrival distribution [121], which follows Poisson process distribution with density λ_a. The reason for this model is that the authors in [121] used the real data for 30 years of annual areas burned data to model the distributions, where the distributions of size and arrival time in real data are proved to follow the Poisson distribution. UAV-BS acts as agent to compute, communicate and control all UAVs' actions, including position and power allocation at each time slot. In our model, we assume using multirotor UAVs that have ability to hover in place, close-range inspection, mapping, and monitoring of fire-affected areas and equipped with cameras, sensors, and thermal imaging devices to capture detailed imagery and collect data [130]. The UAV-BS will receive the request when a fire event occurs, and the kth UAV-UE automatically flies to the center of kth flying region FR_k to serve the ith fire area $A_i(x_i, y_i)$.

We consider a video streaming task that lasts for T time slots with an equal duration t. The selection of the optimal location to stream the video plays an

important role in ensuring the UAV-UEs capture the full firefighting area of A_i. Therefore, the kth UAV-UE need to find the optimal position $U(x_k^*, y_k^*, h_k^*)$ to transmit the video to the UAV-BS. The size of the kth fire region FR_k for the kth UAV-UE depends on the number of UAV-UEs that perform the video streaming for the ith fire area A_i. To make sure that all UAV-UEs can jointly capture the panoramic video of A_i, K UAV-UEs are distributed evenly around A_i, as shown in Figure 3.1. Meanwhile, the UAV-BS also searches for the optimal location $P(x_{BS}^*, y_{BS}^*, h_{BS}^*)$ to satisfy the minimum data rate requirement for all UAV-UEs. In addition, the safety region of the A_i is considered to guarantee FR_k and A_i, and A_i and A_{i+1} are not overlapping to guarantee that the UAV-BS and UAV-UEs are safe from fire.

3.2.1 Request Arrival

The request contains the ith area A_i with its centre at (x_i, y_i) with radius r_i. We assume that K UAV-UEs serve each fire area and stream real-time videos simultaneously. We assume that the height of the fire h_i follows Log-normal distribution [147], thus, the minimum flying height of all UAVs is $h_{\min}$, which satisfies $h_{\min} = \max(h_i)$. All UAV-UEs in A_i will be operated at the same altitude. The environment is divided into W square grids, thus, the length, width and height of each grid are $\frac{X}{\sqrt[3]{W}}$, $\frac{Y}{\sqrt[3]{W}}$, $\frac{Z}{\sqrt[3]{W}}$, respectively. At the tth time slot, the flying position $\vec{U}(x_{i,k}, y_{i,k}, h_{i,k})$ of the kth UAV-UE can be calculated as

$$\vec{U}^{t+1}(x_{i,k}, y_{i,k}, h_{i,k}) = \vec{U}^{t}(x_{i,k}, y_{i,k}, h_{i,k}) + \vec{a}^{t}(x, y, z), \tag{3.1}$$

with

$$x_i - a \leq x_{i,k} \leq x_i + a, \tag{3.2}$$

$$y_i - a \leq y_{i,k} \leq y_i + b, \tag{3.3}$$

$$h_{\min} \leq h_{i,k} \leq h_{\max}, \tag{3.4}$$

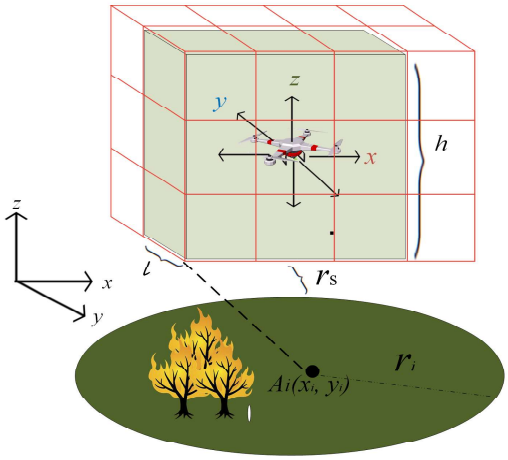

Figure 3.2: Flying boundry of the kth UAV-UE.

$$U^t_{(i,k=1)} = \{(x_1, y_1, h_1) | x_i - a \leq x_{i,1} \leq x_i + a, y_i + a \leq y_{i,1} \leq y_i + b, h_i \leq h_1 \leq h_{max}\}, \tag{3.5a}$$

$$U^t_{(i,k=2)} = \{(x_2, y_2, h_2) | x_i - b \leq x_{i,2} \leq x_i - a, y_i - a \leq y_{i,2} \leq y_i + a, h_i \leq h_2 \leq h_{max}\}, \tag{3.5b}$$

$$U^t_{(i,k=3)} = \{(x_3, y_3, h_3) | x_i - a \leq x_{i,3} \leq x_i + a, y_i - b \leq y_{i,3} \leq y_i - a, h_i \leq h_3 \leq h_{max}\}, \tag{3.5c}$$

$$U^t_{(i,k=4)} = \{(x_4, y_4, h_4) | x_i + a \leq x_{i,4} \leq x_i + b, y_i - a \leq y_{i,4} \leq y_i + a, h_i \leq h_4 \leq h_{max}\}. \tag{3.5d}$$

where $\vec{a}^t(x, y, z)$ is the action vector to determine the flying direction of the UAV-UE. The action vector $a = r_i + r_s$ limits the horizontal boundaries of flying UAV-UE, and $b = r_i + r_s + l$ is the vertical boundaries of the UAV-UE. r_s is the safe distance between A_i and FR_k to ensure the UAV cannot be affected by the fire and close enough to stream the fire area, l is the length of flying region, and h_{max} is the maximum height of UAV-UE regulated by the government (i.e. 120 m in UK [3]). The upper boundaries are introduced to ensure better uplink performance, capture a clear picture. This is because the picture frame can be clearer when the UAV-UEs are closer to the surveillance area. Furthermore, to capture full panoramic video, we propose the boundary flying area for UAV-UEs in each fire area, which can be written as Eq. (3.5).

3.2.2 Channel Model

In the wireless network, we assume that the channel model between the kth UAV-UE and the UAV-BS contains large-scale fading (path loss and channel gain) and small-scale fading [21]. We assume that the link between the UAVs are line-of-sight (LoS). Also, we consider that the wildfires have occurred in rural areas, and the height of the UAV should be higher than the fire to guarantee the UAV cannot be damaged by the fire. As all UAVs are flying in free space area, there are no blockages between the UAVs, and the UAVs can capture the videos following the Rural Macrocell Aerial Vehicular (RMa-AV) path loss model in 3GPP standard [4][Table B-2]. Also, to ensure the safety of UAVs, all UAVs are designed to fly above trees and fire, therefore no Non-LOS are considered in our model. The pathloss from the kth UAV-UE to the UAV-BS can be written as

$$PL^t_{\mathrm{LoS,k}} = 20\log\left(\frac{4\pi f_c d^t_k}{c}\right) + \eta_{\mathrm{LoS}}, \tag{3.6}$$

where f_c is the carrier frequency, c is the speed of light in vacuum, η_{Los} is the additional attenuation factors due to the LoS connection, and d^t_k is distance between the kth UAV-UE and the UAV-BS, as shown in Figure 3.3, which can be calculated as

$$d^t_k = \sqrt{\left(x^t_{BS} - x^t_k\right)^2 + \left(y^t_{BS} - y^t_k\right)^2 + \left(h^t_{BS} - h^t_k\right)\right)^2}. \tag{3.7}$$

In our model, we use the Rician distribution [49][63] to define small scale fading $p_\xi(d_k)$, which can be denoted as

$$p_\xi(d^t_k) = \frac{d^t_k}{\sigma_0^2}\exp\left(\frac{-{d^t_k}^2 - \rho^2}{2\sigma_0^2}\right) I_0\left(\frac{d^t_k}{\rho}\sigma_0^2\right), \tag{3.8}$$

with $d^t_k \geq 0$, and ρ and σ are the strength of the dominant and scattered (non-dominant) paths, respectively. The Rice factor κ can be defined as

$$\kappa = \frac{\rho^2}{2\sigma_0^2}. \tag{3.9}$$

It is possible that the selected position of each UAV-UE can generate more interference to the UAVs nearby, which can result in poor transmission performance and make it difficult for the UAV-UE to maintain the connection with the UAV-BS. Power control can be the solution to minimize the uplink interference

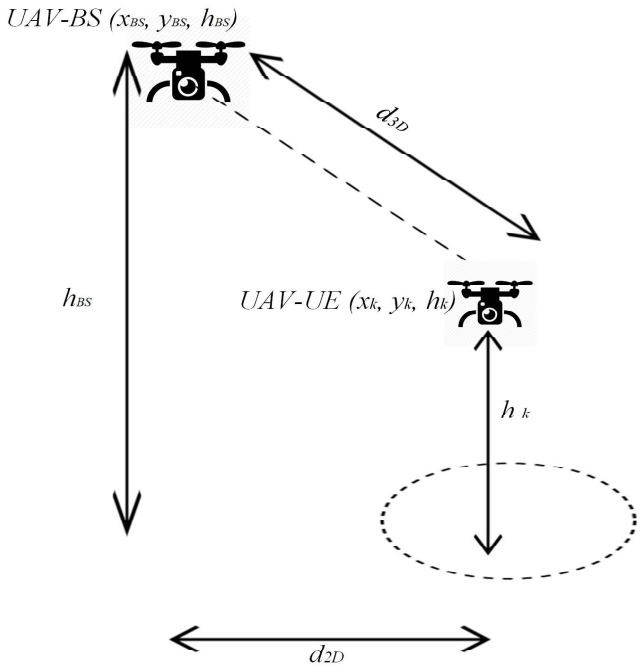

Figure 3.3: UAV-to-UAV communication.

among UAV-UEs at appropriate power level [160]. Through properly controlling the transmit power of each UAV-UE in the uplink transmission, the interference among UAV-UEs can be mitigated. According to the 3GPP guidelines [4], we consider fractional power control for all UAVs and the power transmitted by the kth UAV-UE while communicating with the UAV-BS can be given by

$$P_{U_k}^t = \min\left\{P_{U_k}^{\max}, \left(10\log_{10}(B)) + \rho_{\mathrm{u_k}} PL_{\mathrm{LoS,k}}^t \right)\right\}, \tag{3.10}$$

where $P_{U_k}^{\max}$ is the maximum transmit power of the UAV-UE, B is the channel bandwidth, and $\rho_{\mathrm{u_k}} = \{0, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$ is a fractional path loss compensation power control parameter [160].

In the proposed wireless UAV network, the received power from the kth UAV-UE to the UAV-BS at the tth time slot is presented as

$$P_k^t = P_{U_k}^t G\left(d_k^t\right)^{-\alpha} 10^{\frac{-p_\xi(d_k^t)}{10}}, \tag{3.11}$$

where $P_{U_k}^t$ is the transmit power of the kth UAV-UE, G is the channel power gains factor introduced by the amplifier and antenna [177], $(d_k^t)^{-\alpha}$ is the pathloss, α is the path loss exponent, and $p_\xi(d_k^t)$ is the Rician small scale fading. The interference from the mth UAV-UE to the UAV-BS at the tth time slot can be

Table 3.1: Type of Video Quality [1]

Video Quality	Resolution	FPS	Bitrate (average)	Data used/min	Data used/hr
144p	256x144	30	80-100 Kbps	0.5-1.5 MB	30-90 MB
240p	426x240	30	300-700 Kbps	3-4.5 MB	180-250 MB
360p	640x360	30	400-1,000 Kbps	5-7.5 MB	300-450 MB
480p	854x480	30	500-2,000 Kbps	8-11 MB	480-660 MB
720p (HD)	1280x720	30-60	1.5-6.0 Mbps	20-45 MB	1.2-2.7 GB
1080p (FHD)	1920x1080	30-60	3.0-9.0 Mbps	50-68 MB	2.5-4.1 GB

written as

$$I_{U2U}^{t} = \sum_{\mathbf{m}\in\mathbf{K}\setminus k} \psi_m^t P_m^t, \tag{3.12}$$

where $\psi_m^t = 1$ indicates that the transmission between the kth UAV-UE and the UAV-BS is active, otherwise, $\psi_m^t = 0$, and P_m^t is the transmit power of mth UAV-UE. The signal to interference plus noise ratio (SINR) of the UAV-BS is given by

$$\gamma_k^t = \frac{P_k^t}{N + \sum_{\mathbf{m}\in\mathbf{K}\setminus k} \psi_m^t P_m^t}, \tag{3.13}$$

where N is the noise power at the UAV-BS whose elements are average of independent random Gaussian variables with the variances σ_n^2. Then, the transmission uplink rate from the kth UAV-UE to the UAV-BS can be denoted as

$$R_k^t = B \log_2 \left(1 + \gamma_k^t\right). \tag{3.14}$$

3.2.3 Video Streaming Model

In this research, we consider the long-term video streaming that are modelled as consecutive video segments to maintain the live video streaming in the selected area. Each segment consists of multiple frames, and the frame is considered to be the smallest data unit. The resolution of each frame corresponds to its minimum data rate requirement. Table 3.1 presents the type of Video Quality [1]. For example, if the communication rate (bitrate) is between 300-700 kbps, the video type that we should consider to use is 240 p. Knowing that 144p corresponds to the smallest size of the video type, all UAV-UEs need to satisfy the minimum uplink bitrate, i.e., R_{min}=80 kbps.

Each UAV-UE is equipped with non-ordinary optical camera with the resolution of $r_{px} \times r_{py}$, and the video is consisted of multiple consecutive frames [65], which is used to monitor the fire area with three main goals: 1) detect the size of fire by continuous capturing the panoramic video; 2) verify and locate fires reported; and 3) closely monitor the known fire by streams using distribution relationship around the incident. The quality of the video frame depends on its resolution of the ith video frame at the tth time slot v_i^t. Furthermore, for each video frame, we assume that it has the same playback time T_l, i.e. 2ms to 4ms, which depends on 30 FPS or 60 FPS. In addition, the delay of the video streaming via UAVs is consisted of three elements, i.e. capture time, encoding time, and transmission time. As all UAVs capture the video using the same resolution, the capturing time and the encoding time are constant. Thus, we mainly focus on the uplink transmission time, which can be expressed as

$$T_{i,k}^t = \frac{D(v_i^t)}{R_k^t} = \frac{r_{px} \cdot r_{py} \cdot b}{B \log_2 (1 + \gamma_k^t)}, \tag{3.15}$$

where b is the number of bits per pixel, and $D(v_i^t)$ is the data size based on v_i^t. The video frames are processed in parallel in multi-core processors, and the time consumption at the tth time slot is $T^t = \max\{T_{i,k}^t\}$ [32]. To guarantee the smoothness and seamless of the video streaming, T^t must satisfy the delay constraint, namely, $T^t < T_l$.

3.2.4 Quality of Experience Model

The key parameters of video streaming are video quality, quality of variation, rebuffer time, and the startup delay [163]. Therefore, QoE is formulated by three factors, 1) the sum of video quality over K UAV users in ith area, 2) jitter between video frames (video smoothness penalty), and 3) video latency (delay penalty), where I is the maximum number of fire areas at the tth time slot. In practice, the video quality metric measured each video frame quality based on the selection of bitrate. However, the quality will decrease if the long-term video playback is not smooth, so we introduce two parameters, namely, video smoothness penalty and video latency. In long-term scenario, the drastic changes of video resolution can lead to uncomfortable to firefighters. Therefore, in our learning algorithm, we consider this element to ensure the smoothness of the playback. Finally, the latency is determined by streaming time and transmission time at the tth time slot, T^t, rebuffer time, and the startup delay [163]. According to [156], the rebuffering time and startup delay can be ignored. Thus, the video transmission

may be suffered from a delay, which can be calculated as $D^t = T^t - T_l$, where T_l is the delay constraint. The QoE is denoted as

$$QoE = \frac{\kappa_{i,k}^t}{IK}\left(\sum_{i=1}^{I}\sum_{k=1}^{K} q(R_{i,k}^t) - |q(R_{i,k}^t) - q(R_{i,k}^{t-1})|\right) - \omega^t D^t, \tag{3.16}$$

where $q(R_{i,k}^t)$ is video quality metrics [100], which can be written as

$$q(R_{i,k}^t) = \log\left(\frac{R_{i,k}^t}{R_{\min}(v_i^t)}\right), \tag{3.17}$$

where $\kappa_{i,k}^t$ and ω^t are the weights of video quality and delay, respectively. As our aim is to maximize the QoE, the condition of $\kappa_{i,k}^t > \omega^t$ must be guaranteed, and $R_{\min}(v_i^t)$ is the minimum rate that should be satisfied for the selected v_i^t.

3.2.5 Problem Formulation

The optimization problem in the context of UAVs positioning within an ad-hoc network involves finding the optimal locations for UAV placement to maximize overall performance. It is a complex problem that requires considering various factors, constraints, and objectives. Our aim is to maximize the QoE that jointly exploit the optimal positions of the UAV-BS and UAV-UEs, power control, and the optimal dynamic bitrate selection that result in maximize QoE and overall improved performance. The fluctuation of the transmission link will cause unstable network performance that leads to low QoE and high delay.

Thus, to minimize the delay and maintain the smoothness at each Transmission Time Interval (TTI) and maximize the quality of video streaming. We jointly consider the optimal UAV-BS location $BP = (x_{BS}^t, y_{BS}^t, h_{BS}^t)$, the position of the kth UAV-UE, $BU = (x_{i,k}^t, y_{i,k}^t, h_{i,k}^t)$, the maximum power control of UAV-UE P_{U_k}, the bitrate resolution $BV = \{144, 240, 360, 480, 720, \text{and } 1080\}$p, and UAV-UE's power $P_{U_k} = \{23, 25, \text{and } 30\}$ dBm [78], so that the adequate throughput can be achieved.

In this research, we aim to tackle the problem of optimizing the control factors defined as $A_t = \{BP, BU, BV, P_{U_k}\}$ in an online manner for every frame, where BP is UAV-BS's flying direction, BU is UAV-UEs' flying directions, BV is resolution of the ith UAV-UE, and P_{U_k} UAV-UE's power. At the tth time slot, the UAV-BS aims at maximizing the total long-term QoE in continuous time slots with respect to the policy π that maps the current state information s_t to the probabilities of selecting possible actions in A_t. Therefore, based on the QoE of

each UAV-UE, the optimization problem can be formulated as

$$\max_{\pi(A_t|S_t)} \sum_{i=t}^{\infty} \sum_{k=1}^{K} \gamma^{i-t} \, \mathrm{QoE}_k(i) \tag{3.18}$$

$$\text{s.t.} \max h_i > h_{BS}^t > h_{\max}, \tag{3.19}$$

$$R_{i,k}^t > R_{(\min)}^k(v_i^t), \tag{3.20}$$

$$v_i^t \in \{144\text{p}, 240\text{p}, 360\text{p}, 480\text{p}, 720\text{p}, 1080\text{p}\} \tag{3.21}$$

$$P_{(min)} > P_{U_k}^t > P_{(\max)}, \tag{3.22}$$

$$\sqrt{(x_{BS}^t - x_i)^2 + (y_{BS}^t - y_i)^2} > r_i + r_s, \tag{3.23}$$

$$\mathcal{U} \in \text{Eq.(3.1)}. \tag{3.24}$$

where the objective function in Eq. (3.18) captures the average QoE received at the UAV-BS and $\gamma \in [0, 1)$ is the discount factor to determine the weight accumulated in the future frames, and $\gamma = 0$ means that the agent concerns only the immediate reward. The UAV-BS's height must follow the condition in Eq. (3.19). The minimum requirement of data rate of UAV-UEs based on the dynamic bitrate selection guarantees R_k obtained from $\mathcal{U}_k$ as shown in Eq. (3.20) and follows minimum bitrate in Eq. (3.21) as shown in Table 3.1, while $P_{U_k}^t$ in Eq. (3.22) follows the maximum and the minimum power constraints. The maximum power constraint consideration is influenced by the available power capacity of the UAV's onboard battery, which also will affect the overall flight time and battery life. Then, Eq. (3.23) guarantees that the position of the UAV-BS will not intersect with the UAV-UE's flying region. $\mathcal{U}$ follows the requirement of the flying region FR_i presented in Eq. (3.1). In the experiment, the UAVs are hovering and flying at a constant speed. In our study, there are several trafe-offs in this problem: 1) throughput-bit rate trade-off, 2) throughput-power control trade-off, 3) throughput-distance trade-off, 4) power-distance trade-off, 5) throughput-video smoothness trade-off and 6) throughput-delay trade-off. Therefore, to achieve maximum QoE in long-term time slot, it is important to solve an optimal trade-off between data rate, bit-rate resolution selection, power control, and positions, which further motivates us to use the learning algorithms to jointly optimize the total long-term QoE of all UAV-UEs. All the factors mentioned above help to measure the QoE from the selected resolution to maintain the whole performance in long-term time slots. Also, the correlation between the video smoothness and the penalty delay is to ensure the overall video performance from the beginning

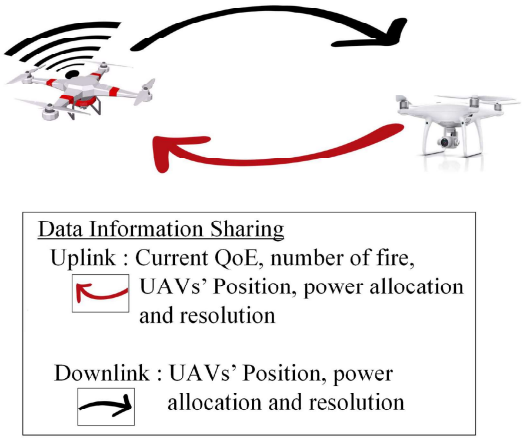

Figure 3.4: UAV-BS to UAV-UE communication information sharing.

to the end.

3.2.6 Channel State Information Sharing

Signal exchange happens in the uplink, the UAV-UEs have to send its locations, fire areas to the UAV-BS, and the QoE information of each UAV-UE will be readily available at the UAV-BS. After the learning is performed at the UAV-BS, the outputs are the actions, including the movement of the UAVs, selected video resolution, and power. After that, the selected actions will be sent through the downlink from the UAV-BS to each UAV-UE for its control. The whole process is illustrated in Figure 3.4.

3.3 Optimization Problem via Reinforcement Learning

In this section, we design several DRL algorithms to maximize the long-term QoE in UAV-to-UAV network to be compared with the existing traditional method - Greedy algorithm. Since the channel and the locations fire change over time, different numbers of UAVs are required at each time slot. Our problem cannot be solved by the traditional optimization method. It is because we have partially observed information, and our formulated optimization problem is a long-term problem, it cannot be solved by the traditional optimization method. In our problem, we consider long-term quality and smoothness of video streaming, which cannot be solved by the traditional optimization problem. Thus, we cannot compare with any state of the art traditional optimization problem. Thus, there is

no traditional optimization methods we can compare, and our problem can only be solved by deep reinforcement learning, because we focus on maximizing the long-term QoE. However, the traditional optimization method can only optimize for the current time slot and cannot optimize for long-term period. Specifically, we propose two DRL algorithms, which are Deep Q-Learning and Actor-Critic, to maximize the long-term QoE of live video streaming in U2U communication.

3.3.1 Reinforcement Learning

Our propose for RL-based method, the UAV-BS acts as centralized agent to collect video from UAV-UEs while maximizing QoE to solve problem which influenced by the delay, UAVs' positions, and bitrate selection during each TTI, and forms the partially observable Markov decision process (POMDP). At each TTI, the channel network condition, fire arrival, and network condition are different in each timeslot. Therefore, through learning algorithms, the UAV-BS (agent) is able to select the positions of the UAV-BS, positions of the UAV-UEs, the dynamic resolution and the maximum power allocation in order to maximize the individual QoE at each time slot and the long-term QoE objective.

State Representation

The current state s^t corresponds to a set of current observed information. The state of the UAV-BS can be denoted as $s = [\mathcal{P}, \mathcal{V}, \mathcal{U}, P_{U_k}, \text{QoE}]$, where $\mathcal{P}$= (x^t_{BS}, y^t_{BS}, h^t_{BS}) is the position of the UAV-BS, $\mathcal{V}$ is the bitrate selection, $\mathcal{U} = (x^t_k, y^t_k, h^t_k)$ is the positions of UAV-UEs, and P_{U_k} is k-UAV-UE's power.

Action Space

Q-agent will choose action $a = (BP, BU, BV, P)$ from set $\mathcal{A}$. The dimension of the action set can be calculated as $\mathcal{A} = BP \times BU^{i \times k} \times BV^i \times P$. The actions for UAVs include (i) UAV-BS's flying direction (BP), (ii) UAV-UEs' flying directions (BU), (iii) resolution of the ith UAV-UE (BV), and (iv) UAV-UE's power (P). The action space is presented as

- BP = (up, down, left, right, ascent, descent or hover)
- BU = (up, down, left, right, or hover)
- BV= (144, 240, 360, 480, 720, or 1080) p
- P = (23, 25, 30) dBm

To ensure the balance of exploration and exploitation actions of the UAV-BS, ϵ -greedy ($0 < \epsilon \leq 1$) exploration is deployed. At the tth TTI, the UAV-BS randomly generates a probability p_ϵ^t to compare with ϵ. If the probability $p_\epsilon^t < \epsilon$, the algorithm randomly selects an action from the feasible actions to improve the value of the non-greedy action. However, if $p_\epsilon^t \geq \epsilon$, the algorithm exploits the current knowledge of the Q-value table to choose the action that maximizes the expected reward.

Rewards

When the a^t is performed, the corresponding reward re^t is defined as

$$re^t = \frac{\psi_{i,k}^t}{IK}\left(\sum_{i=1}^{I}\sum_{k=1}^{K} q(R_{i,k}^t) - |q(R_{i,k}^t) - q(R_{i,k}^{t-1})|\right) - \omega^t D^t, \tag{3.25}$$

where $q(R_{i,k}^t)$ is video quality metrics [100], which can be written as

$$q(R_{i,k}^t) = \log\left(\frac{R_{i,k}^t}{R_{\min}(v_i^t)}\right), \tag{3.26}$$

$\psi_{i,k}^t$ and ω^t are the weights of video quality and delay, respectively. If $R_{i,k}^t$ is unable to satisfy the minimum transmission rate for $R_{\min}^k(v_i^t)$, namely, $R_{i,k}^t < R_{\min}^k(v_i^t)$, the system will receive negative reward, which means $\mathrm{re}^t < 0$.

3.3.2 Q-learning

The learning algorithm needs to use Q-table to store the state-action values according to different states and actions. Through the policy $\pi(s, a)$, a value function $Q(s, a)$ can be obtained through performing action based on the current state. At the tth time slot, according to the observed state s^t, an action a^t is selected following ϵ -greedy approach from all actions. By obtaining a reward re^t, the agent updates its policy π of action a^t. Meanwhile, Bellman Equation is used to update the state-action value function, which can be denoted as

$$\begin{aligned} Q(s^t, a^t) =& (1-\alpha)Q(s^t, a^t) \\ &+ \alpha\left\{\mathrm{re}^{t+1} + \gamma \max_{a^t \in \mathcal{A}} Q(s^{t+1}, a^t)\right\}, \end{aligned} \tag{3.27}$$

where α is the learning rate, $\gamma \in [0, 1)$ is the discount rate that determines how current reward affects the updating value function. Particularly, α is suggested

to be set to a small value (e.g., $\alpha = 0.01$) to guarantee the stable convergence of training.

3.3.3 Deep Q-learning

However, the dimension of both state space and action space can be very large if we use the traditional tabular Q-learning, which will cause high computation complexity. To solve this problem, deep learning is integrated with Q-learning, namely, Deep Q-Network (DQN), where a deep neural network (DNN) is used to approximate the state-action value function [168]. $Q(s, a)$ is parameterized by using a function $Q(s, a; \boldsymbol{\theta}_{\text{DQN}})$, where $\boldsymbol{\theta}_{\text{DQN}}$ is the weight matrix of DNN with multiple layers. s is the state observed by the UAV and acts as an input to Neural Networks (NNs). The output are selected actions in $\mathcal{A}$. Furthermore, the intermediate layer contains multiple hidden layers and is connected with Rectifier Linear Units (ReLu) via using $f(x) = \max(0, x)$ function. At the tth time slot, the weight vector is updated by using Stochastic Gradient Descent (SGD) and Adam Optimizer, which can be written as

$$\boldsymbol{\theta}_{\text{DQN}}^{(t+1)} = \boldsymbol{\theta}_{\text{DQN}}^{t} - \lambda_{\text{ADAM}} \cdot \nabla \mathcal{L}(\boldsymbol{\theta}_{\text{DQN}}^{t}), \tag{3.28}$$

where λ_{ADAM} is the Adam learning rate, and $\lambda_{\text{ADAM}} \cdot \nabla \mathcal{L}(\boldsymbol{\theta}_{\text{DQN}}^{t})$ is the gradient of the loss function $\mathcal{L}(\boldsymbol{\theta}_{\text{DQN}}^{t})$, which can be written as

$$\begin{aligned} \nabla \mathcal{L}(\boldsymbol{\theta}_{\text{DQN}}^{t}) = \mathbb{E}_{S^i, A^i, Re^{i+1}, S^{i+1}} \big[\big(Q_{\text{tar}} - Q(S^i, A^i; \\ \boldsymbol{\theta}_{\text{DQN}}^{t}) \cdot \nabla Q(S^i, A^i; \boldsymbol{\theta}_{\text{DQN}}^{t}) \big], \end{aligned} \tag{3.29}$$

where the expectation is calculated with respect to a so-called minibatch, which are randomly selected in previous samples $(S^i, A^i, Re^{i+1}, S^{i+1})$ for some $i \in \{t - M_r, t - M_r + 1, \ldots, t\}$, with M_r being the replay memory. The minibatch sampling is able to improve the convergence reliability of the updated value function [107]. In addition, the target Q-value Q_{tar} can be estimated by

$$Q_{\text{tar}} = re^{i+1} + \gamma \max_{a \in \mathcal{A}} Q(S^{i+1}, a; \bar{\boldsymbol{\theta}}_{\text{DQN}}^{t}), \tag{3.30}$$

where $\bar{\boldsymbol{\theta}}_{\text{DQN}}^{t}$ is the weight vector of the target Q-network to be used to estimate the future value of the Q-function in the update rule. This parameter is periodically copied from the current value $\boldsymbol{\theta}_{\text{DQN}}^{t}$ and kept fixed for a number of episodes. The DQN algorithm is a value-based algorithm, which can obtain an optimal strategy through using experience replay and target networks. It enables the

Algorithm 1: : Optimization by using DQN

Input: The set of UAV-BS position $\{x_{BS}, y_{BS}, h_{BS}\}$, bitrate selection V, the position of the kth UAV-UE $U_k = (x_k^t, y_k^t, h_k^t)$, $\sum QoE$ and operation iteration I.

Algorithm hyperparameters: Learning rate $\alpha \in (0, 1]$, $\epsilon \in (0, 1]$, target network update frequency K;

Initialization of replay memory M, the primary Q-network $\boldsymbol{\theta}$, and the target Q-network $\bar{\boldsymbol{\theta}}$;

for $e \leftarrow 1$ ***to*** I **do**

  Initialization of s^1 by executing a random action a^0;

  for $t \leftarrow 1$ ***to*** T **do**

    if $p_\epsilon < \epsilon$ **then**: Randomly select action a^t from $\mathcal{A}$;

    else select $a^t = \underset{a \in A}{argmax} Q\left(S^t, a, \theta\right)$;

    The UAV-BS performs a^t at the tth TTI ;

    The UAV-BS observes s^{t+1}, and calculate re^{t+1} using Eq. (3.25);

    Store transition $(s^t; a^t; re^{t+1}; s^{t+1})$ in replay memory M;

    Sample random minibatch of transitions $(S^i; A^i; Re^{i+1}; S^{i+1})$ from replay memory M;

    Perform a gradient descent for $Q(s; a; \boldsymbol{\theta})$ using (3.29) ;

    Every K steps update target Q-network $\bar{\boldsymbol{\theta}} = \boldsymbol{\theta}$.

  end

end

agent to sample from and train by the previously observed data online. This is due to the experience replay mechanism and randomly sampling in DQN, which use the training samples efficiently to smooth the training distribution over the previous behaviours. Not only does this massively reduce the amount of interactions needed with the environment, but also reduce the variance of learning updates. The DQN algorithm will create a sequence of policies whose corresponding value functions converge to the optimal value function, when both the sample size and the number of iteration go to infinity. The DQN algorithm is presented in Algorithm 1.

3.3.4 Actor-Critic

Different from the DQN algorithm, which obtains the optimal strategy indirectly by optimizing the state-action value function, while the AC algorithm directly determines the strategy that should be executed by observing the environment

state. The AC algorithm combines the advantages of value-based function method and policy-based function method. In the AC algorithm, the agent is consisted of two parts, i.e., actor network and critic network, and it solves the problem through using two neural networks. Meanwhile, the AC algorithm deploys a separate memory structure to explicitly represent the policy, which is independent of the value function. The policy structure is known as the actor network, which is used to select actions. Meanwhile, the estimated value function is known as the critic network, which is used to criticize the actions performed by the actor. The AC algorithm is an on-policy method and temporal difference (TD) error is deployed in the critic network. To sum up, the actor network aims to improve the current policies while the critic network evaluates the current policy to improve the actor network in the learning process.

The critic network uses value-based learning to learn a value function. The state-action value function $V(s^t, \boldsymbol{w}^t)$ in the critic network can be denoted as

$$V(s, \boldsymbol{w}^t) = \boldsymbol{w}^\top \boldsymbol{\Phi}(s^t), \tag{3.31}$$

where $\boldsymbol{\Phi}(s^t) = s^t$ is state features vector and $\boldsymbol{w}^t$ is critic parameters, which can be updated as

$$\boldsymbol{w}^{t+1} = \boldsymbol{w}^t + \alpha_c^t \delta^t \nabla_{\boldsymbol{w}} V\left(s^t, \boldsymbol{w}^t\right), \tag{3.32}$$

where α_c is the learning rate in the critic network. After performing the selected action, TD error δ^t is used to evaluate whether the selected action based on the current state performs well [180], which can be calculated as

$$\delta^t = \mathrm{re}^{t+1} + \gamma_{\boldsymbol{w}}(V\left(s^{t+1}, \boldsymbol{w}^t\right) - V\left(s^t, \boldsymbol{w}^t\right)). \tag{3.33}$$

Then, the actor network is used to search the best policy to maximize the expected reward under the given policy with parameters $\boldsymbol{\theta}_{\mathrm{AC}}$, which can be updated as

$$\boldsymbol{\theta}_{\mathrm{AC}}^{t+1} = \boldsymbol{\theta}_{\mathrm{AC}}^t + \alpha_a \nabla_{\boldsymbol{\theta}_{\mathrm{AC}}} J\left(\pi_{\boldsymbol{\theta}_{\mathrm{AC}}^t}\right), \tag{3.34}$$

where α_a is the learning rate in the actor network, which is positive and must be small enough to avoid causing oscillatory behaviour in the policy, and according to [180], $\nabla_{\boldsymbol{\theta}_{AC}} J\left(\pi_{\boldsymbol{\theta}_{AC}}\right)$ can be calculated as

$$\nabla_{\boldsymbol{\theta}_{\mathrm{AC}}} J\left(\pi_{\boldsymbol{\theta}_{\mathrm{AC}}^t}\right) = \delta^t \nabla_{\boldsymbol{\theta}_{\mathrm{AC}}} \ln\left(\pi\left(a^t | s^t, \boldsymbol{\theta}_{\mathrm{AC}}^t\right)\right). \tag{3.35}$$

The AC algorithm is presented in Algorithm 2.

Algorithm 2: : Actor-Critic Algorithm

Inputs: The set of UAV-BS position $\{x_{BS}, y_{BS}, h_{BS}\}$, bitrate selection V, the position of the kth UAV-UE $U_k = (x_k^t, y_k^t, h_k^t)$, $\sum QoE$ and operation iteration I.

Algorithm hyper-parameter: Learning rate $\alpha_c \in (0, 1]$, $\epsilon \in (0, 1]$, Target network update frequency K;

Initialization of policy parameter θ_{AC}, weight of the actor network $\mathbf{w}$, value of the critic network $\boldsymbol{V}$;

for $e \leftarrow 1$ ***to*** I **do**

  Initialization of s^0 by executing a random action;

  for $t \leftarrow 1$ ***to*** T **do**

    Select action a^t according to the current policy;

    The UAV-BS observes s^{t+1}, and calculate re^{t+1} using (3.25);

    Store transition $(s^t; a^t; re^{t+1}; s^{t+1})$;

    Update TD-error functions;

    Update the weights $\mathbf{w}$ of critic network by minimizing the loss;

    Update the policy parameter vector θ for actor network;

    Update the policy θ_{AC} and state-value function $V(s^t, \boldsymbol{w}^t)$.

  end

end

Finally, Figure 3.5 shows the network architecture design, where the current state is input to the neural network for both algorithms, DQN and Actor-Critic. DQN and Actor-Critic used an agent in sequential decision-making tasks, and they both belong to the family of policy-based methods. There are key differences in their architectures and learning approaches. DQN utilizes a single network and employs epsilon-greedy exploration, while Actor-Critic has separate networks for the actor and critic, allowing for more nuanced exploration and exploitation. DQN can suffer from training instability due to correlated observations, whereas Actor-Critic typically exhibits more stable training dynamics. The next key step is to determine the action to be sent to the environment and the reward to measure the QoE. Then, the new states are generated from observation for the next round of updates. Overall, considering DQN and Actor-Critic algorithms in our problem provides unique features and novelties contribute to their effectiveness towards adapting in different scenarios.

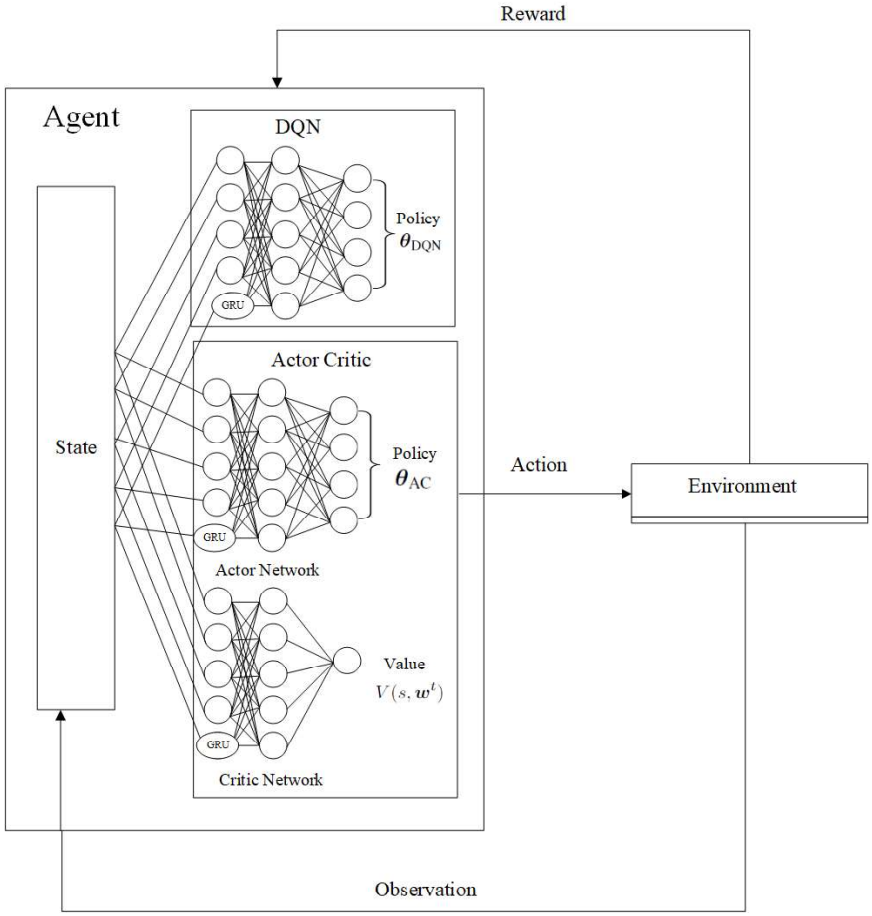

Figure 3.5: The network architecture designed.

3.3.5 Analysis Complexity of Reinforcement Learning Algorithms

The computational complexity of the DQN/AC algorithm, which includes DQN/AC learning architecture, the action selection of the UAVs, and the downlink transmission, are given by $O(mlogn + 2^A + N_i N_k)$, where m is the number of layers, n is the number of units per learning layer, A is number of action, N_i is number of fire area, and N_k is number of UAV for each fire area.

3.4 Simulation Results

In this section, we evaluate our proposed learning algorithms in our problem setup. The area of the region is 5000 m x 5000m x 100m. In the simulation, the maximum flying height $h_{\max}$ of the UAV-BS is 100m, which is satisfied with the maximum flying height 120m that is stipulated by the UK government. We assume that the available video bitrates of the dynamic video streaming for each

Table 3.2: Parameter

Parameter	Value
Number of UAV-UEs	12
Transmission power, PUe	23 dBm [177]
Bandwidth, B	3 MHz
Noise variance σ^2	-96 dBm [177]
Center frequency, f_c	2 GHz [128, pp. 3777]
Power gains factor, G	-31.5 dB [177]
Alpha, α	2
Channel parameter, LoS	0.1 [13, pp. 572]
Channel parameter, $NLoS$	21 [13]
Channel parameter, a	4.88 [128, pp. 3777],[11, pp. 7]
Channel parameter, b	0.43 [128, pp. 3777] ,[11, pp. 7]
Radius of target region	1250 m
Radius of Surveillance region, r_i	250 m
Learning Rate	0.1, 0.01
Initial, Final Exploration	1, 0.1
Discount Rate	0.8
Replay memory	1000

video frame are $(80, 300, 700, 1000, 2000, 3000)$kbps. The target area is captured by K UAV-UE(s), i.e., $K = 4$ in the ith fire area A_i $(i = 1, 2, \text{and}, 3)$. At the beginning, the UAV-BS will be deployed at the centre of the environment, i.e. (1250, 1250, $h_{\min}$), where $h_{\min}$ is the maximum height of the fire. When the fire occurs at the remote area, the UAV-UEs will immediately reach the fire location to stream and oversee the real-time situation. The height of the UAV-UEs in each fire area are fixed and follow the distribution of the fire height [121]. The network parameters for the system are shown in Table 3.2 and follow the existing approach and 3GPP specifications in [177], [4], and [13]. The performance of all results is obtained by averaging around 100 episodes, where each episode is consisted of 100 TTIs. The result is measured for the equal duration of the time slot at each time slot t and also called as TTI, where each TTI is equal to 0.5ms as follows in 3GPP [4]. Finally, the channel model parameters and grid environment parameters are set according to [177].

Figure 3.6 plots the average QoE value over different grid sizes via AC and

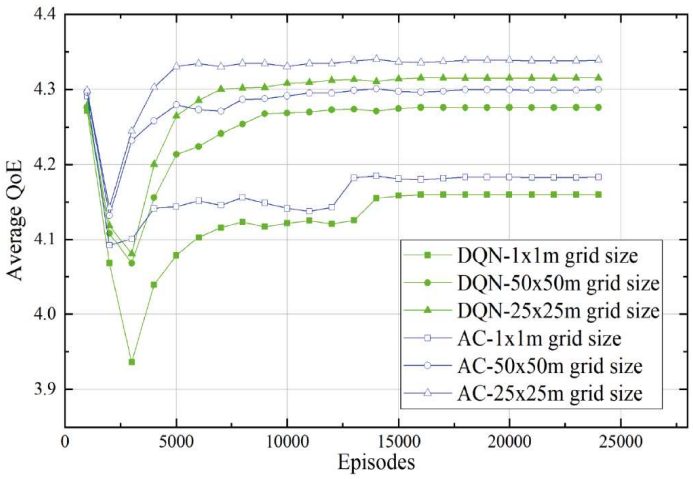

Figure 3.6: Average QoE of the UAV-BS with different schemes via different learning algorithms with different grid size of each episode.

DQN algorithms. From the result, it can be seen the 25m $\times$ 25m grid sizes produced the highest average QoE of the UAV-BS for both DQN and AC algorithms, therefore in the next simulation, we use 25 $\times$ 25m grid size. From the result, the number of grids will influence the movement of the UAV, the UAV will move more frequently in small grid size with more number of grids. In this case, the performance can be improved due to that the UAV can explore and exploit the environment more accurately. However, increasing the number of grids can lead to increased complexity in learning algorithms. It is because the action space will increase with more number of grids. In practice, the UAV operator has to decide what will be the best square size according to the movement step of each UAV. However, if we want to reduce the complexity by increasing the grid size or decreasing the number of grid, the result shows degraded performance of QoE, and it takes more time to obtain the convergence results due to difficulties in finding an optimal solution in long-term QoE analysis.

In each scenario, our proposed DQN and AC algorithms are compared with the Greedy algorithm. The Greedy algorithm selects the actions based on the immediate reward and local optimum strategy. The DQN is designed with 3 hidden layers, which each layer consists of 256, 128, 128 ReLU units, respectively. For the AC method, the critic DNN consists of an input layer with 19 neurons, a fully-connected neural network with two hidden layers, each with 128 neurons, and an output layer with 1 neuron. The UAV-BS is initially set at the centre of the environment with the height $h_{\min}$. In wildfires environment problem, the network

coverage with smooth streaming needs to overview the real-time situation. To guarantee high quality of video transmission from multiple UAVs in continuous time slots, the Recurrent Neural Network (RNN) is deployed. In temporal data, RNN based GRU network can approximate the value function or the policy of each DRL algorithm, where the stateless RNN does not need to re-initialize the memory at each training step, while the training progress is more resource-hungry and less stable [27]. The learning-based predictor uses a modern RNN model with parameters θ to predict the traffic statistic at each frame. The use of RNN is due to its ability to capture the time correlation of traffic statistics over multiple frames, which can aid in learning the time-varying traffic trend and improving prediction accuracy. Thus, RNN can capture the correlation among the state or action in over time, which can help DRL select more optimal action, and guarantee the high quality of video transmission.

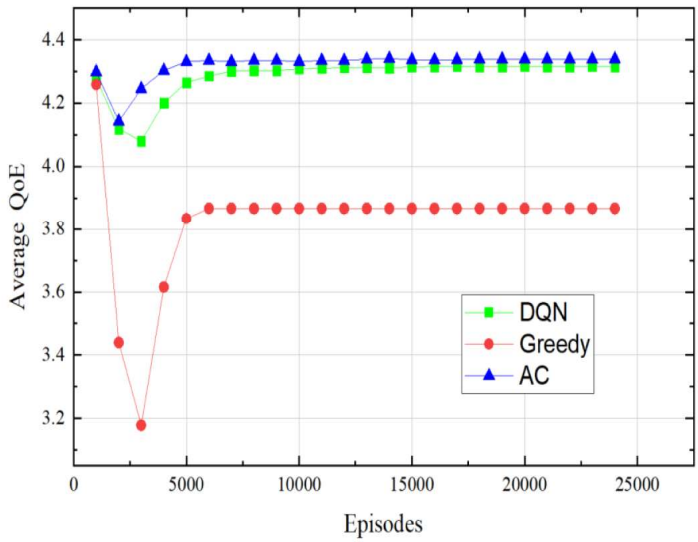

Figure 3.7: Average QoE value for each frame via AC, DQN and Greedy algorithms.

Figure 3.7 plots the average QoE value over all frames via AC, DQN and Greedy algorithms. It can be seen that DRL algorithms outperform the non-learning based algorithm, i.e., Greedy algorithm. The convergence of the reinforcement learning algorithms has been proved in [80], [77], an agent of the Q-learning algorithm is assured to converge to the optimal Q. Figure 3.7 plots the average QoE value over all frames in each episode via DQN/AC learning algorithms, which shows the convergence of the proposed two algorithms. It is observed that the total reward and the convergence speed of these two DRL learning algorithms follows: $AC > DQN$. This is due to the AC algorithm is updated in two steps, including the critic step and actor step. At each step, the critic

network judges the action selected by the actor network, which can select the actions more appropriately. Moreover, it can be seen that the DRL algorithms outperform the Greedy algorithm, where the convergence speed of the DRL algorithms is faster than the Greedy algorithm. Specifically, in the Greedy algorithm, the UAVs only consider exploiting the current reward, rather than exploring the long-term reward. Therefore, the UAVs are not able to achieve higher expected reward compared to the DRL algorithm.

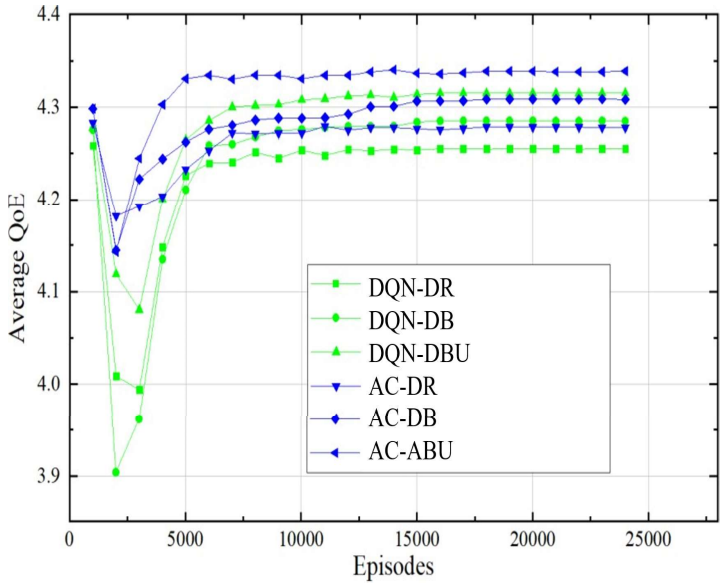

Figure 3.8: Average QoE of the UAV-BS with different schemes via different learning algorithms and with different optimization schemes of each episode.

Figure 3.8 plots the average QoE of the UAV-BS with different video transmission schemes via different learning algorithms in each episode. For simplicity, "DR" represents the scheme with dynamic resolution, "DB" is the scheme with dynamic resolution and dynamic UAV-BS, and "DBU" is the scheme with dynamic resolution, dynamic UAV-BS and UAV-UEs. It is observed that the average QoE of the AC algorithm outperforms all other algorithms, with it being able to achieve an optimal trade-off between data rate, bitrate resolution selection, power control, and positions. From the result, it is observed that with the dynamic environment and large size of the action, and the AC algorithm is able to select proper positions of UAVs and video resolution of video frames. This is mainly due to the experience replay mechanism, which efficiently utilizes the training samples, and the actor and critic functions are able to smooth the training distribution over the previous behaviours compared to DQN. In addition, we can observe that the strategies of selecting optimal positions for UAVs achieve higher performance compared to the UAVs with fixed locations. This result em-

phasizes the importance of the strategy with mobile UAVs. This is due to the fact that mobile UAVs can move through the network to reach the optimal positions that are able to adapt to dynamic fire scenarios.

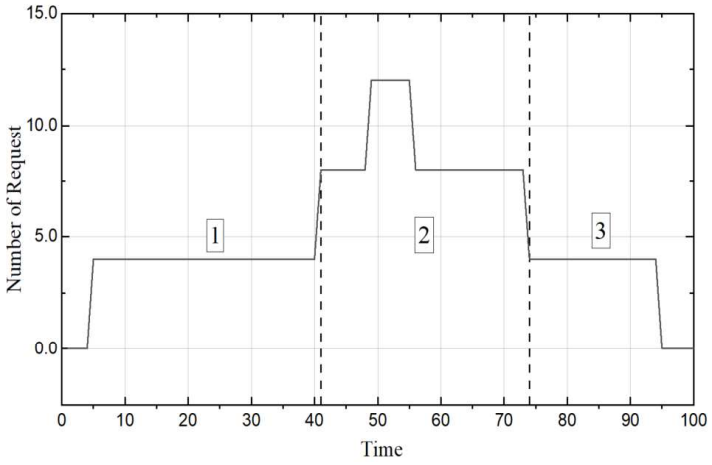

Figure 3.9: The request of the UAV-UEs in continuous time slots.

We provide more in-depth investigation of the relationship between the number of UAV request, dynamic video resolution, dynamic power control, and throughput with different learning algorithms in continuous 100 time slots. The results are also compared among the three algorithms, namely DQN, AC, and Greedy algorithms. The detailed results show how the optimization control helps UAVs to maximize the QoE at each time slot.

Figure 3.9 plots the UAV's requests follow the fire arrival distribution, which follow Poisson process distribution with density λ. In phase 1, there is a small number of fire arrival which leads to low request of UAV's number. But with time increases, the number of fire arrival is getting higher and leads high number of UAV's request is needed as shown in phase 2 and in phase 3, the request is dropped and less UAV's request is demanded. As the number of requests rapidly changes, we introduce power control to control the transmit power at UAV-UEs to mitigate the interference among UAV-UEs, thus maximizing the achievable rate of each UAV-UE.

Following the fire arrival requests depicted in Fig 3.9, Figure 3.10 shows the plots of the average power control over all UAV-UEs in continuous time slots with AC, DQN and Greedy algorithms. The power control helps mitigate the interference among UAV-UEs. As shown in phase 1 and phase 3 in Figure 3.9, there is a small number of fire requests with small number of UAVs to transmit

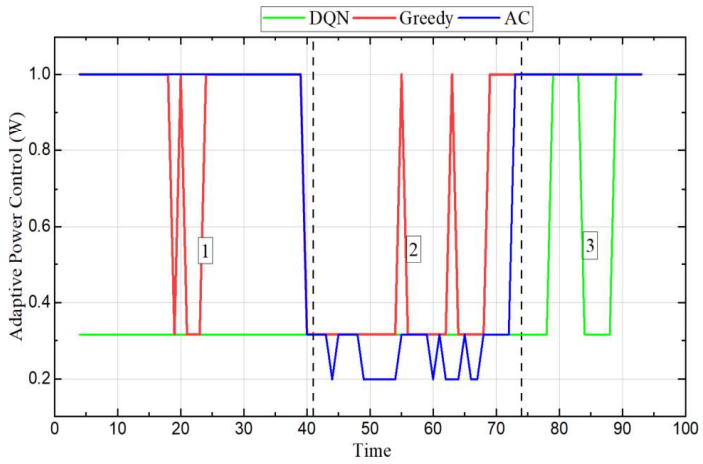

Figure 3.10: The power control of the UAV-UEs in continuous time slots with different learning algorithms.

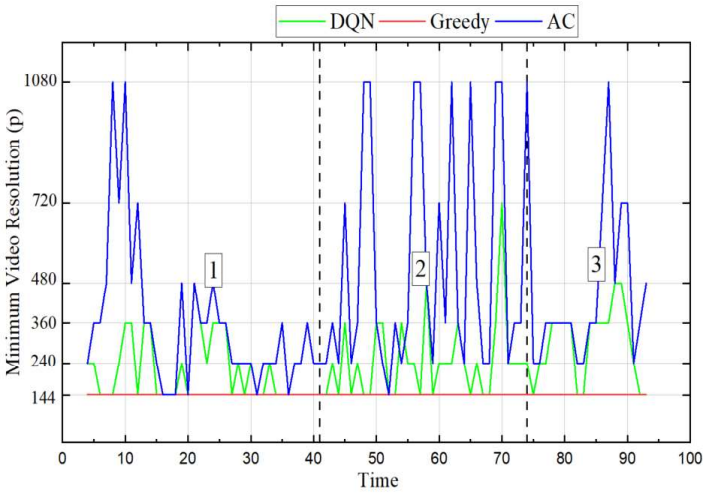

Figure 3.11: The average dynamic resolution of the UAV-UEs in continuous time slots with different learning algorithms.

the data. However, when the number of requests increases, a large number of UAVs are demanded, as shown in phase 2 of Figure 3.9. As can be seen from Phase 2 of Figure 3.10, the DRL algorithms learn the environment and effectively reduce the transmit power of each UAV-UE, to reduce the interference from UAV-UEs. We see that the Greedy algorithm maintains the higher power, even though high power can provide high received signal, it also causes high interference at the UAV-BS and failure in transmission.

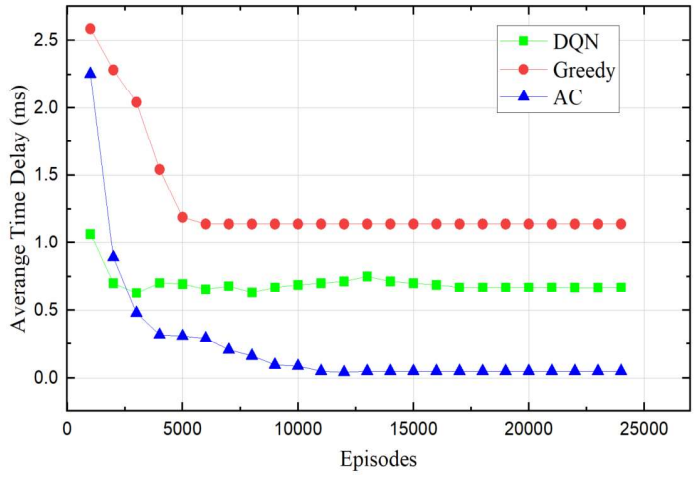

Figure 3.12: Average latency of video streaming with different learning algorithms.

Following the fire arrival requests depicted in Figure 3.9, Figure 3.11 shows the plots of the minimum dynamic resolution over all UAV-UEs in continuous time slots with different learning algorithms. It is shown that the minimum video resolution of the AC algorithm is higher than the DQN and the Greedy algorithm in all scenarios. The AC algorithm is able to maintain an optimal video resolution at each time slot and guarantee high quality and smooth video playback with new request. However, the Greedy algorithm exploits with a minimum video resolution to maintain high rewards, and it only uses local optimal policy and causes poor performance. For phase 1 and 3, when the number of requests is low at the tth time slot, the power is high, and the throughput increases, thus, the resolution of video is high. However, when the number of request is increasing in phase 2, the AC algorithm is able to maintain a high resolution due to helps of dynamic power, which leads to better QoE for each UAV-UE. This will help to reduces the interference and improve the quality of the video resolution.

In Figure 3.12, we plot the average latency of video streaming of AC, DQN and Greedy algorithms. It can be seen that the latency performance of the AC algorithm outperforms that of the DQN algorithm. When multiple video streaming exist in the U2U communication, the interference among UAV-UEs occur and causes higher latency. Based on the observed state, the AC algorithm is able to select proper positions and transmission power of the UAV-UEs to mitigate the interference, which further decreases the latency. Thus, the AC algorithm is able to maximize the average QoE with the lowest average time latency. However, the Greedy algorithm is unable to exploit the violation of

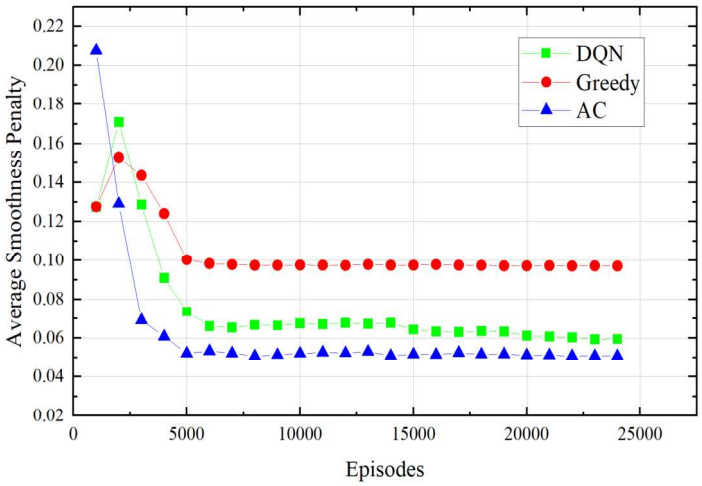

Figure 3.13: Average smoothness penalty with different learning algorithms.

latency constraints resulting in higher latency, which leads to lower QoE.

Figure 3.13 plots the average smoothness penalty of AC, DQN and Greedy algorithms. The smoothness penalty demonstrates the average video stability occupancy of UAV-UEs at each episode. When the learning algorithm is able to automatically choose the suitable resolution at the tth time slots and $(t-1)$th time slot, it will obtain lower smoothness penalty and higher QoE. Moreover, the AC algorithm is able to automatically choose the proper action based on actor and critic function, which leads to better smoothness of the AC algorithm compared to that of the DQN and Greedy algorithms. It has proved that the AC algorithm guarantees the smoothness of video transmission with high QoE. Meanwhile, the Greedy algorithm shows the worst performance, as it only makes local optimal selections.

Finally, the dynamic movement of UAVs as shown in Figure 3.14, and the duration time is 100s. In this simulation, we assume that all the UAVs moved at constant speed. At each time slot, the UAV-BS selects the direction from the action space, which contains 7 directions, while the action space of the UAV-UE contains 5 directions. Then, the dynamic movement maximizes the total long-term QoE of all UAVs. To reduce the complexity, we select only one UAV-UE for each fire area to illustrate the optimized trajectory of UAV-BS and UAV-UEs, which is shown in Figure 3.14.

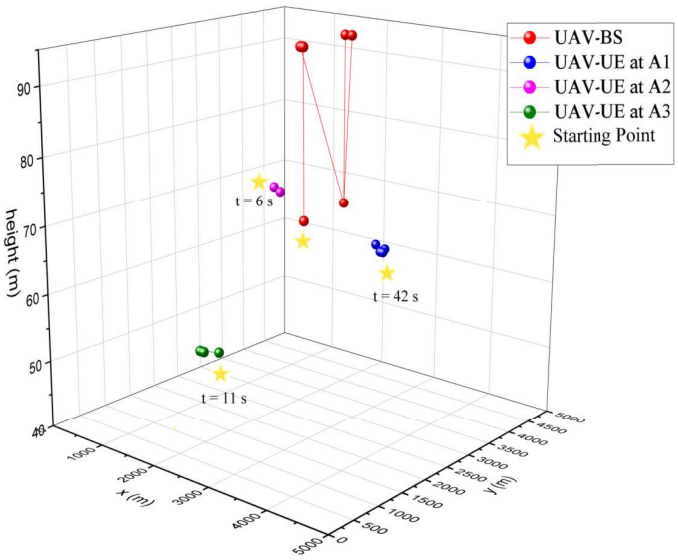

Figure 3.14: Dynamic trajectory of UAVs when dynamic fire arrival from t=0 to t=100s.

3.5 Conclusion

As conclusion, we developed a deep reinforcement learning approach for the mobile U2U communication to maximize the Quality of Experience (QoE) of UAV-UEs, through optimizing the locations for all UAVs, the additive video resolution, and the transmission power for UAV-UEs. The QoE function was designed to guarantee the smoothness of live video streaming among UAV-BS and UAV-UEs in this U2U network. The dynamic interference problem is resolved by utilizing the dynamic power control to achieve a higher achievable rate. Through our developed Deep Q Network and Actor-Critic methods, the optimal additive video resolution can be selected to stream real-time video frames, and optimal positions of the UAV-BS and UAV-UEs can be selected to satisfy the transmission rate requirement. Simulation results demonstrated the effectiveness of our proposed learning-based schemes compared to the Greedy algorithm in terms of higher QoE with low latency and high video smoothness. Therefore, AC achieved 12% higher achievable rate and QoE in the U2U communication scenario, because of integrating the advantages of the value-based and policy-based functions. However, since AC has two neural networks and needs more parameters to update, AC is more complex in terms of computation complexity compared to that of DQN. Thus, in future research, DQN can be more preferable to use if the scenario is more complex than our current scenario. DQN offers a novel and effective approach to address the challenges and complexities of optimizing dynamics U2U commu-

nication systems, which enhances the stability and convergence of the learning process, leading to improved overall performance and able to handle different scenarios, variations in network conditions and still achieve satisfactory results.

Chapter 4

Inter-cell Interference Mitigation for Cellular-connected UAVs

4.1 Introduction

UAV cellular network are widely deployed as a solution for providing large-scale radio connectivity towards current needs. However, today's cellular networks aren't built for aerial coverage, and deployments are primarily focused on providing excellent service to terrestrial users. These factors, combined with stringent regulatory requirements, have resulted in extensive research and standardization efforts to ensure the current cellular networks can reliably operate aerial vehicles in the variety of deployment scenario. Therefore, there is a need to investigate the performance of aerial radio connectivity in a typical urban area network deployment using extensive channel measurements and system simulations. The downlink radio interference play a key role, and yield relatively poor performance for the aerial traffic, when load is high in the network. However, due to different user type, i.e., flying and terrestrial user with high volume of them, the interference become the challenging for each user to receive strong connectivity network. Therefore, interference mitigation is a solution and will investigate in this chapter. Further, we introduce and evaluate the novel downlink inter-cell interference coordination mechanism applied to the aerial command and control traffic. Our proposed coordination mechanism is shown to provide the required aerial downlink performance increasing and degradation in the serving and interfering cells.

The contributions of this chapter are summarized as follows:

- We propose the dynamic muting scheme for moving UAVs and terrestrial users (TUEs) in the downlink scenario of the cellular network. The UAVs

and TUEs are uniformly distributed in the communication environment, and the dynamic requests from them follow Poisson process in each time slot.

- To guarantee excellent service among TUEs in the dynamic network, we formulate a long-term problem to mitigate the interference level of each UAV by muting cells, which can satisfy QoS requirements of TUEs and UAVs over time and maximize sum-rate of TUEs.

- To further increase the throughput of downlink transmission based on cell muting technique, we propose the dynamic muting and time-frequency scheduling algorithm. The muting scheme mutes proper number of interfering cells, and the time-frequency scheduling scheme allocates proper physical resource blocks (PRBs) to TUEs and UAVs.

- To solve the aforementioned problem, we deploy value function approximation solution (VFA), Tabular-Q, Deep Q Network (DQN), and MOSDS-DQN. Learning algorithms help the agent to select actions to maximize the long-term throughput of downlink transmission. The linear muting scheme from [119] is set as a benchmark as it using traditional optimization muting scheme to mitigate the inter-cell interference. Simulation results show that our proposed DQN approach outperforms the linear muting scheme in terms of higher throughput and lower interference. Furthermore, the proposed MOSDS-DQN guarantees the throughput performance of TUEs with increasing number of UAVs.

4.2 System Model and Problem

We assume that C base stations (BSs) with M antennas are deployed at the centre of C cells, using Orthogonal frequency-division multiplexing (OFDM) to serve the associated users, as shown in Figure 4.1. The OFDM has been used for over a decade and proved its robustness in multi-carrier technologies. It uses multiple smaller subcarriers to avoid the Inter-Channel Interference (ICI) and Inter-Symbol Interference (ISI) over wireless networks, and adds a Cyclic Prefix (CP) to demodulate the signal effectively on the receiver side [149].

According to [123, 4, 5], as show in Figure 4.2, the antenna element pattern

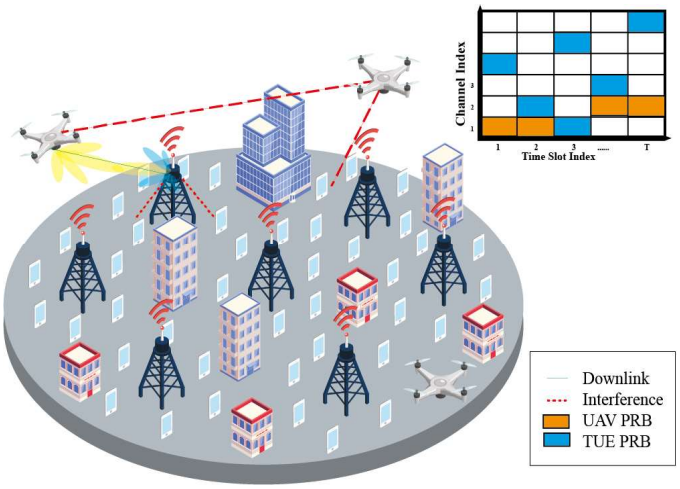

Figure 4.1: Illustration of UAV-cellular network model and resource block scheduling.

$A(\theta, \phi)$ for the mth antenna array is given by

$$A(\theta, \phi) = -\min\left\{-\Big[A_{E,V}(\theta) + A_{E,H}(\phi)\Big], A_m\right\}, \tag{4.1}$$

where $A_{E,V}(\theta)$ and $A_{E,H}(\phi)$ are vertical and horizontal radiation patterns of antenna elements, respectively, and $A_{E,V}(\theta)$ is denoted as

$$A_{E,V}(\theta) = -\min\left\{12\left(\frac{\theta - 90^o}{\theta_{3dB}}\right)^2, SLA_V\right\}. \tag{4.2}$$

In Eq. (4.1) and (4.2), θ is the vertical 3 dB beamwidth, SLA_V is the side-lobe level limit , and $A_{E,H}(\phi)$ is denoted as

$$A_{E,H}(\phi) = -\min\left\{12\left(\frac{\phi}{\phi_{3dB}}\right)^2, A_m\right\}. \tag{4.3}$$

In (4.3), ϕ is the horizontal 3 dB beamwidth, and A_m is the front-back ratio. Based on (4.2) and (4.3), the 3D antenna element gain for each pair of angles (θ, ϕ) is calculated as

$$A_G(\theta, \phi) = G_{max} - \min\left\{-\Big[A_{E,V}(\theta) + A_{E,H}(\phi)\Big], A_m\right\}, \tag{4.4}$$

where G_{max} is the maximum directional gain of the antenna element [5, 123, 24]. The above equations (4.1) - (4.4) provide the dB gain experienced by a ray with angle pair (θ, ϕ) based on the effect of the element radiation pattern. The cth

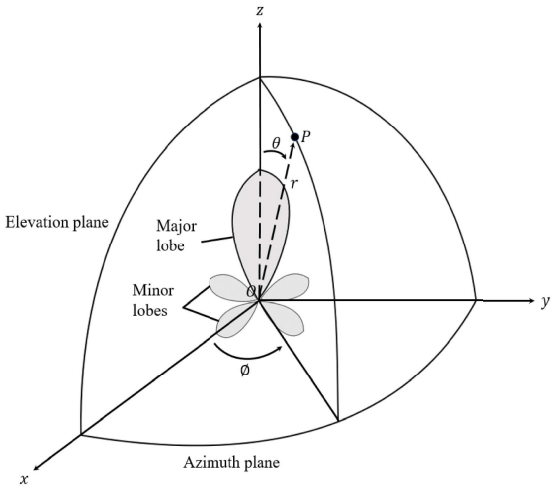

Figure 4.2: Illustration of antenna pattern.

BS ($c \in \mathcal{C}$) operates in a single-user mode serving either a terrestrial UE (TUE) with DL data or an UAV with Command and Control (C2) data. Both TUEs and UAVs are assumed to be equipped with a single antenna. Each cell consists of I uniformly distributed TUEs, while the total number of UAVs is J and they are uniformly distributed over the entire network with radius R_{NW}. The UAV in a cell is prioritized and assigned with PRBs, as it requires critical C2 data transmitted in a required data rate [4]. The distribution of TUEs is modelled as Poisson Process and the remaining available PRBs should be allocated to all TUEs.

4.2.1 Mobility Model

In the 3D environment, we assume that the UAV flies at the fixed height and speed, resulting in the mapping of the 3D environment into the 2D image with $W \times W$ grids. Each grid has the side length of a meters, and the UAV follows the center of each grid, creating the finite set of possible paths. Additionally, the UAV moves in four directions: right, left, forward, and backward. As the movement speed between grids is fixed, the latency experienced during the movement remains constant due to the consistent travel time between grids. The UAVs and UEs are distributed uniformly over the BS coverage area [62]. We consider the regular urban grid deployment of building blocks captured by a Point Poisson Process Process (PPP).

4.2.2 Channel Model

We adopt two different 3GPP standards to model the channels for TUEs and UAVs, respectively [4] [5]. Here, the UAV is assumed to be flying at a height where a line-of-sight (LoS) link is ensured. The small-scale channel gain of the UAV is modelled as Rician channel, while for TUEs, there are modelled as Rayleigh channels [5, 2, 34]. The pathloss from the uth UE to the BS is written as

$$PL^{t}_{\mathrm{LoS,u}} = \begin{cases} 15.3 + 37.6\log_{10}(d_{3D}), & 1.5m \leq h^{t}_{u} \leq 22.5m \\ 28.0 + 22\log_{10}\left(d_{3\mathrm{D}}\right) + 20\log_{10}\left(f_{c}\right), & \\ & 22.5m < h^{t}_{u} \leq 300m \end{cases} \tag{4.5}$$

where f_c is the carrier frequency, and $d^{u}_{3D}(t)$ is the distance between the uth user and the BS. We assume that each BS uses the same transmit power and each user has perfect knowledge of its channel state information (CSI), so that the signal to interference plus noise ratio (SINR) $\gamma_{c,\,j}$ between the cth cell and the jth UAV is written as

$$\gamma_{c,\,j} = \frac{P_{c}\|\boldsymbol{h}^{H}_{c,\,j} \cdot \mathbf{v}_{c,j}\|^{2}}{N_{c,\,j} + \sum_{k\in C\setminus c} P_{c}\|\boldsymbol{h}^{H}_{k,\,u} \cdot \mathbf{v}_{k,u}\|^{2}}, \tag{4.6}$$

where $P_{c} = \frac{P_{DL}}{10^{PL_{LOS,k}/10}} \times A_{G}(\theta,\phi)$, and P_{DL} is the downlink transmit power per PRB. In (4.6), $\boldsymbol{h}_{c,\,j} \in \mathbb{C}^{M\times 1}$ denotes the channel vector between the cth BS and the jth UAV, and $\boldsymbol{h}_{k,u}$ is the channel between the kth BS and the uth user in the cth cell. The channel model $\mathbf{h}$ includes both the small-scale fading and large-scale fading calculated by $h = g.\beta^{1/2}$, where g and β are small-scale fading and large-scale fading parameters, respectively. In (4.6), $\mathbf{v}_{k,u} = \left(\mathbf{g}_{k,u}\right)^{\mathrm{H}}\left(\mathbf{g}_{k,u}\left(\mathbf{g}_{k,u}\right)^{\mathrm{H}}\right)^{-1}$ represents the transmit zero-forcing precoding vector of the uth user in the cth cell [54], and $\boldsymbol{g}_{k,u} \in \mathbb{C}^{M\times 1}$ is the channel vector between the kth BS and the uth user in the kth cell. In addition, $N_{c,\,j}$ is the additive white Gaussian noise at the jth user.

4.2.3 User Association

According to [118], we consider the maximum Reference Signal Receive Power (RSRP) in the user association policy, which widely used cell association strategy. RSRP is defined as the linear average over the power contributions (in [W]) of the resource elements that carry cell-specific reference signals within the considered measurement frequency bandwidth [9]. However, if RSRP is calculated directly

through the receiving signals, there may be a lot of noise. Thus, the RSRP is denoted as

$$RSRP_{c,u} = P_c - PL_{LoS,k}. \tag{4.7}$$

In the maximum RSRP-based user association, the user is connected to the cell that provides the maximum RSRP. It is considered that the users can only be associated with one BS. The policy allows the jth user to be associated with the BS c that has the strongest RSRP.

$$u_j = \{u \mid \max RSRP_{c,u}, \forall j \in J\}. \tag{4.8}$$

Based on (4.6), the achievable rate of the jth UAV is calculated as

$$R_{c,j}^{UAV} = B_c \log_2(1 + \gamma_{c,j}), \tag{4.9}$$

where B_c is the bandwidth of the cth cell.

4.2.4 Latency Model

The overall packet loss is caused by downlink/uplink transmission delays, queueing delay and backhaul delay [129]. The queuing delay and transmission delay are two major bottlenecks against achieving the stringent latency constraints [53]. These delay components are hardly controlled in a dynamic system, and must control to be at very minimum [53]. The total downlink latency at the BS is the sum of the downlink UAV transmission $\tau_{c,j}^{trans}$ and queue latency $\tau_{c,j}^{sch}$ [129], which is written as

$$\tau_{c,j}^{total} = \tau_{c,j}^{trans} + \tau_{c,j}^{sch}. \tag{4.10}$$

Given the achievable date rate, the corresponding downlink transmission of jth UAV $\tau_{c,j}^{trans}$ is given by [69]

$$\tau_{c,j}^{trans} = \frac{L}{R_{c,j}}, \tag{4.11}$$

where L is the data size of the required of jth UAV, in B unit [119, 69].

4.2.5 Inter-Cell Interference Coordination (ICIC) for Macro-cell Muting

To improve received SINR of the UAV, the cells coordinate PRBs among TUEs and UAVs. The interfering BS $c \in C'$ leaves the PRBs blank/muted, allowing the UAV-serving BS to schedule its transmission within the same frame shared with TUEs. Therefore, Eq. (4.6) is rewritten as

$$\gamma_{c,\,j} = \frac{P_c \|\boldsymbol{h}_{c,\,j} \cdot \mathbf{v}_{c,j}\|^2}{N_{c,\,j} + \sum_{k \in C' \setminus c, k \notin C'} P_c \|\boldsymbol{h}_{k,\,u}^H \cdot \mathbf{v}_{k,u}\|^2}, \tag{4.12}$$

where $C' \subseteq \mathcal{C}$ is the set of cells being muted, and the second term in the denominator is the total interference from other cells. The SINR between the cth cell and TUE is given by

$$\gamma_{c,\,u} = \frac{P_c \|\boldsymbol{h}_{c,\,u}^H \cdot \mathbf{v}_{c,u}\|^2,}{N_{c,\,u} + \sum_{k \in C' \setminus c, k \notin C'} P_c \|\boldsymbol{h}_{k,\,u}^H \cdot \mathbf{v}_{k,u}\|^2}. \tag{4.13}$$

In (4.13), $\boldsymbol{h}_{c,\,u} \in \mathbb{C}^{M \times 1}$ denotes the channel vector between the cth BS and the uth user. Based on Eq. (4.12) and (4.13), the date rate of the uth TUE is defined as

$$R_{c,\,u}^{TUE} = B_c \log_2(1 + \gamma_{c,\,u}), \tag{4.14}$$

where B_c is the bandwidth of the cth cell and the data rate of the jth UAV is defined as

$$R_{c,\,j}^{UAV} = B_{c,j} \log_2(1 + \gamma_{c,\,j}). \tag{4.15}$$

In (4.15), $B_{c,j}$ is the fraction of bandwidth allocated to user j at cell c, with respect to the available bandwidth.

4.2.6 Antenna Beam Selection

When assuming the antenna elements are mounted on the UAV at the right spacing and angle/orientation, antenna selection with two or more directional antenna elements is equivalent to a simple beam selection [119]. For example, the UAV rotates its fuselage in the azimuth plane while keeping the right direction, then 1 or 2 antenna elements are sufficient to generate a 'beam' towards the serving cell. If degrees of freedom of the UAV are more restricted, at least 4 antenna elements need to be mounted to provide four beams in the azimuth plane.

Thus, we assume that antenna beam selection of the UAV is applied only in the azimuth plane, and an omnidirectional elevation radiation pattern is considered. An antenna beam radiation pattern is modelled as a $sinc()^2$ function, with -3 dB beam-widths of approximately 90°, or 50° in the azimuth plane with six beams [119]. The modelled beam patterns provide +6.6 dBi gain in the main direction and -3 dB gain in the front-to-side lobe attenuation according to [119], which can be used to compensate for the non-ideal orientation and shape of beams [119]. As shown in Figure 4.3, a simple setup with a grid of 2, 4 or 6 fixed beams is used (fixed relative to the UAV fuselage) to emulate a practical antenna selection mechanism.

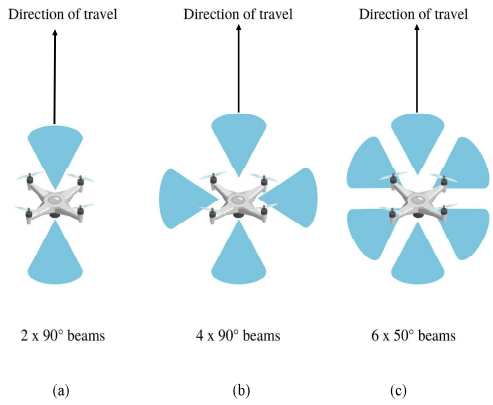

Figure 4.3: Modelled antenna beam configurations for the UAV.

4.2.7 Downlink Resource Block Scheduler

LTE transmission is segmented into frames, each one consists of 10 subframes, and each subframe is further divided into two slots. Each slot is 0.5 ms, so that the total time for each frame is 10 ms. Each time slot on the LTE downlink system consists of 7 OFDM symbols. The flexible spectrum allows the LTE system to use bandwidths ranging from 1.4 MHz to 20 MHz, where higher bandwidths are used for higher LTE data rates. The physical resources of the LTE downlink can be illustrated as a frequency-time resource grid, as shown in Figure 4.4. A Resource Block (RB) has a duration of 0.5 ms (one slot) and a bandwidth of 180 kHz (12 sub-carriers). Each RB has 84 resource elements in the case of a normal cyclic prefix and 72 resource elements for extended cyclic prefix.

In the RB scheduler technique, there are several types of scheduling algorithms, such as Round Robin (RR) [171] and proportional fairness (PF) [29]. RR scheduling is a non-aware scheduling scheme that allows users to take turns in

using the shared resources (time-RBs), without taking the instantaneous channel conditions into account. The radio resources in RR are assigned equally among all users, which compromises the throughput performance of the system. While PF is defined as the ratio of the average data rate to all users to maintain the equality of fairness [29]. To solve the problem, dynamic scheduling is introduced to schedule the available data for each Transmission Time Interval (TTI), which maximizes the scheduling gains. As shown in Figure 4.4, PRBs are allocated to sub-bands according to its channel and resource allocation models. However, the challenge is that it requires frequent coordination for exchanging control signals between cells, which increases the overhead among cells.

4.2.8 Problem Formulation

The objective is to maximize the throughput of TUE networks by selecting optimal actions in A^t subject to the UAV's QoS requirements (i.e., reliability). Thus, the optimization problem is formulated as

$$(P1): \quad \max_{\pi(A^t|S^t)} \sum_{i=t}^{\infty} \sum_{c=1}^{C} \sum_{u=1}^{U} \beta^{(i-t)} R_{c,u}^{TUE}(i,f) \tag{4.16}$$

$$\text{s.t.} \sum_{f=1}^{F} p_f \leq P_c, \quad p_f \geq 0, \tag{4.17}$$

$$R_{c,j}^{UAV}(i,\, f) \geq R_{Th}^{UAV}, \quad \forall j \in \mathcal{J} \tag{4.18}$$

$$R_{c,u}^{TUE}(i,\, f) \geq R_{Th}^{TUE}, \quad \forall u \in \mathcal{U} \tag{4.19}$$

$$\beta^i \in [0,1]. \tag{4.20}$$

where $\beta^i \in [0,1)$ is the discount factor determining the performance accumulated in the future reward. When $\beta^i = 0$, the agent only concerns about the immediate reward.

The optimization relies on the selection of actions in A_t according to the current and historical observations O_t with respect to the stochastic policy $\pi\left(A^t \mid S^t\right)$. The optimization problem aims at maximizing the total long-term reward in continuous time slots with respect to the policy π that maps the current state information s_t to the probabilities of selecting possible actions in A^t. The state S^t contains the set of instantaneous and cumulative data rates of both UAVs and TUEs, and the agent selects a specific action $A^t \in \mathcal{A}(S^t)$ that determines the index of cell(s) being muted. The throughput form of TUEs in (4.16) and

(4.19) should be the same, Eq. (4.17) in P1 means that the sum of power of each sub-frame should not larger than the maximum transmit power, p_f is the power of each sub-frame, and F is the total number of sub-frames. Eq. (4.17) guarantees the allocation of power [152] used by all total selected frames, F in maximum transmit power threshold at the BS, P_c , and Eq. (4.18) and Eq. (4.19) guarantee the transmission rate for UAVs and TUEs, R_{Th}^{UAV} and R_{Th}^{TUE}, respectively. The optimization problem aims at maximizing the total long-term reward in continuous time slots with respect to the policy π that maps the current state information s_t to the probabilities of selecting possible actions in A^t. The state S^t contains the set of instantaneous and cumulative data rates of both UAVs and TUEs, and the agent selects a specific action $A^t \in \mathcal{A}(S^t)$ that determines the index of cell(s) being muted.

To solve the problem, we consider muting schemes using reinforcement learning (RL), which are introduced in detail in the next section.

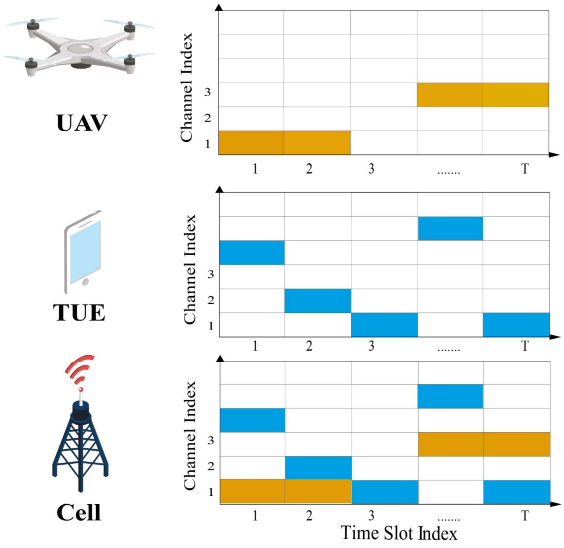

Figure 4.4: Dynamic PRB scheduling for UAVs and TUEs.

4.3 Muting Optimization Scheme using Reinforcement Learning

Since the channel and locations of UAVs and TUEs are changing over time also different muting and dynamic scheduling schemes are required in continuous time slots, the problem in P1 cannot be solved by a traditional optimization method, as the formulated optimization problem is a non-convex and long-term problem.

Therefore, in this section, we design several Reinforcement Learning (RL) algorithms to solve the problem in P1. The RL agent learns the optimal mapping from the input states to select the resource allocation action to maximize the long-term throughput.

4.3.1 Tabular Q-Learning

Consider the Q-agent deployed at the central unit to optimize the service provision for both UAVs and TUEs. To optimize the long-term reward, the agent first explores the environment. Let $s \in \mathcal{S}$, $a \in \mathcal{A}$, and $r \in \mathcal{R}$ denote the state, action, and reward, respectively.

State Representation

The current state S^t corresponds to a set of current observations. The state of the system is denoted as $\mathcal{S} = [\sum R_{TUE}, \sum R_{UAV}]$, where R_{TUE} is a set of data rate of TUEs and R_{UAV} is a set of instantaneous rate of UAVs.

Action Space

Q-agent selects action A from set $\mathcal{A}$. The action is denoted as $A_{t+1}^{P1} = \{N_m\}$, where N_m is the index of muting cells. To ensure the balance of exploration and exploitation actions of the agent, ϵ-greedy ($0 < \epsilon \leq 1$) exploration is deployed. At the tth TTI, the agent randomly generates a probability p_ϵ^t to compare with ϵ. If the probability $p_\epsilon^t < \epsilon$, the algorithm randomly selects an action from the feasible actions to improve the value of the non-greedy action. However, if $p_\epsilon^t \geq \epsilon$, the algorithm exploits the current knowledge of the Q-value table to choose the action that maximizes the expected reward.

Rewards

At the beginning of each TTI, the Q-agent observes the current state S^t and selects the specific action $A^t \in \mathcal{A}$. After performing the selected action A^t, the agent receives a reward R^{t+1} and observes a new state S^{t+1}. The optimization goal is to maximize the long-term throughput of TUEs while guaranteeing the quality of service (QoS) of UAVs, which is defined as:

$$Reward_t^{P1} = \sum_{u=1}^{U} R_u \cdot \mathbf{1}\left[R_{u,j} \geq R_{Th}\right], \tag{4.21}$$

where,

$$\mathbf{1}\left[R_{u,j} \geq R_{Th}\right] = \mathbf{1}\left[R_u \geq R_{Th}\right] \cap \mathbf{1}\left[R_j \geq R_{Th}\right]. \tag{4.22}$$

In Eq. (4.22), 1[x] is the indicator function, $\mathbf{1}[x] = 1$ when x is true, otherwise, $\mathbf{1}[x] = 0$, and $\cap$ is a logical and operation function. In the logical and operation function, $\mathbf{1}[x] \cap \mathbf{1}[y] = 1$ as x and y are true, otherwise, $\mathbf{1}[x] \cap \mathbf{1}[y] = 0$ [77].

In tabular Q, the state to action mapping is learned through value function $Q(s, a)$, which consists of a scalar value for all state and action spaces. The action that has the maximum value is selected from $\mathcal{A}$. To dynamically optimize the number of muted cells, the function learns the optimal policy π^* and optimizes the Q-table. The agent updates its Q-table using the immediate reward $\mathcal{R}^{t+1}$ and the next state-action value $Q(S^{t+1}, a)$, which is given by

$$\begin{aligned} &Q(S^t, A^t) = Q(S^t, A^t) + \\ &\alpha\Big[R^{t+1} + \gamma \max_{a \in A} Q(S^{t+1}, a) - Q(S^t, A^t)\Big]. \end{aligned} \tag{4.23}$$

In Eq. (4.23), $\alpha \in (0, 1)$ is the learning rate, and $\gamma \in [0, 1)$ is the discount rate that determines how much the current reward affects the future value. In each TTI, the agent selects the action with the highest probability with probability $p_e^t \geq \epsilon$, or vice versa. The learning rate α, most importantly, is set to be a small constant to guarantee stable convergence, as the reward can be biased due to unknown and unpredictable distribution of the observed states. The implementation of cell muting using tabular-Q method is shown in Algorithm 3.

4.3.2 Linear Value Function Approximation

However, the Tabular-Q requires large space to store state-action value, and needs to update each parameter to achieve convergence. To address these issues, we consider a linear value function approximation (VFA) method. VFA uses a 'Value Function' approximator to obtain a sub-optimal policy, but its efficiency depends on the deployed approximation function, such as Linear Approximator (LA), Deep Q-Learning (DQN), and decision trees.

LA approximates the value function $Q(S^t,\, A^t)$ by

$$Q(S^t,\, A^t) \approx \hat{Q}(S^t,\, A^t,\, \mathbf{w}^t), \tag{4.24}$$

Algorithm 3: : Tabular Q-Learning/Linear VFA to optimize cumulative terrestrial users' throughput

Algorithm hyperparameters: $\alpha \in (0, 1], \gamma \in [0, 1), \epsilon \in (0, 1]$

Tab-Q: Initialize Q-table $Q(s,\, a)$ **VFA:** Initialize $\mathbf{w}$

for *Iteration* $\leftarrow$ *1 to I* **do**

Initialize s^1 by executing a random action A^0;

UAVs identify the BS with the highest RSRP and associate with it.

for $t \leftarrow$ *1 to T* **do**

if $p_\epsilon < \epsilon$

Randomly select an action A^t from $\mathcal{A}$;

else

Tab-Q: select $A^t = \underset{A \in A}{argmax} Q\,(S^t, A)$;

VFA: select $A^t = \underset{A \in A}{argmax} Q\,(S^t, A, , \mathbf{w})$;

The agent performs A^t and mutes the selected cells.

The agent observes S^{t+1} and calculates R^{t+1} using Eq. (4.21).

Tab-Q: Update $Q(S,\, A)$ according to Eq. (4.23).

VFA: Update $\mathbf{w}$ according to Eq. (4.31).

end

Determine all active UEs (TUEs and UAVs) using

Bernoulli process.

Determine associate cell UEs and active UEs matrices.

Update the queue matrices.

Calculate SINR and transmission rate.

end

where $\mathbf{w}^t$ is the weight vector. The objective is to minimize the mean-squared error between these two values, given by

$$J(\mathrm{w}^t) = \mathbb{E}_\pi[(Q(S^t,\, A^t) \,-\, \hat{Q}(S^t,\, A^t,\, \mathbf{w}^t))^2]. \tag{4.25}$$

To obtain the optimal policy, $\mathbf{w}^t$ is updated by stochastic gradient descent (SGD), which is calculated as

$$\begin{aligned} -\frac{1}{2}\nabla_{\mathbf{w}} J(\mathbf{w}^t) =& [Q(S^t,\, A^t) \,-\, \hat{Q}(S^t,\, A^t,\, \mathbf{w}^t)] \\ & \nabla_{\mathbf{w}} \hat{Q}(S^t,\, A^t,\, \mathbf{w}^t), \end{aligned} \tag{4.26}$$

and

$$\nabla \mathbf{w}^t = \alpha[Q(S^t,\, A^t) \,-\, \hat{Q}(S^t,\, A^t,\, \mathbf{w}^t)]\nabla_{\mathbf{w}}\hat{Q}(S^t,\, A^t,\, \mathbf{w}^t). \tag{4.27}$$

In LA, $\hat{Q}(S^t,\, A^t,\, \mathbf{w}^t)$ is represented as a dot product of feature vector $\mathbf{x}(S^t, A^t)$ and weight vector $\mathbf{w}^t$, which is denoted as

$$\hat{Q}(S^t,\, A^t,\, \mathbf{w}^t) = \mathbf{x}^T(S^t,\, A^t)\mathbf{w}^t = \sum_{k=1}^{K} \mathrm{x}_k(S^t,\, A^t)\mathrm{w}_k^t, \tag{4.28}$$

and

$$\mathbf{x}(S^t,\, A^t) = \begin{bmatrix} \mathrm{x}_1(S^t,\, A^t) \\ \mathrm{x}_2(S^t,\, A^t) \\ \vdots \\ \mathrm{x}_K(S^t,\, A^t) \end{bmatrix}, \tag{4.29}$$

where $\mathbf{x}(S^t,\, A^t)$ corresponds to the entire state-action space. The current action is selected from the vector $\hat{Q}(S^t,\, A^t,\, \mathbf{w}^t)$ in Eq. (4.28), following the ϵ-greedy policy, which is the same as that of tabular Q-learning. The gradient descent in Eq. (4.27) is calculated as

$$\nabla_{\mathbf{w}}\hat{Q}(S^t,\, A^t,\, \mathbf{w}^t) = \nabla_{\mathbf{w}}[\mathbf{x}^T(S^t,\, A^t)\mathbf{w}^t] = \mathbf{x}(S^t,\, A^t), \tag{4.30}$$

and

$$\nabla \mathbf{w} = \alpha[Q(S^t,\, A^t) \,-\, \hat{Q}(S^t,\, A^t,\, \mathbf{w}^t)]\mathbf{x}(S^t,\, A^t). \tag{4.31}$$

The implementation of cell muting using VFA method is shown in Algorithm 3. However, the basic linear tabular- Q is not suitable, as the state-action space is so large and are increasing with the number of cell, and also function approximation technique is unable to train and get the optimal solution. Therefore, we consider DQN algorithm as a tool to solve large state-action space to find optimal cell muting solution to mitigate interference between users.

4.3.3 Deep Q-Network

When large number of cells, TUEs, and UAVs exist in the network, the state-action space increases exponentially. To address this issue, DQN is used to update the network's weights. The DQN algorithm for P1 is presented in Algorithm 4.

Algorithm 4: : DQN to optimize cumulative terrestrial users' throughput

Algorithm hyper-parameters: $\alpha \in (0, 1], \gamma \in [0, 1), \epsilon \in (0, 1]$

Initialize replay memory M, primary Q-network $\boldsymbol{\theta}$, and target Q-network $\overline{\theta}$

for $e \leftarrow$ *1 to I* **do**

Initialize s^1 by executing a random action a^0;

UAVs identify the BS with the highest RSRP and associate with it.

for $t \leftarrow$ *1 to T* **do**

If $p_\epsilon < \epsilon$: Randomly select action a^t from $\mathcal{A}$;

else select $a^t = \underset{a \in A}{argmax} Q\left(S^t, a, \theta\right)$;

Agent performs the selected A and mutes cells.

Agent observes S^{t+1} and calculates R^{t+1}.

Store transitions (S^t, A^t, R^t, S^{t+1}) in replay memory, and sample random minibatch of transitions (S^t, A^t, R^t, S^{t+1}) from M.

Calculate $\hat{Q}(S^{t+1}, a, \boldsymbol{\theta})$ according to Eq. (**??**).

Calculate gradient descent using Eq. (4.31).

Update $\overline{\boldsymbol{\theta}}$ every K steps.

end

Determine all active UEs (TUEs and UAVs) using Bernoulli process.

Determine associate cell UEs and active UE matrices.

Update the queue matrices.

Calculate SINR and transmission rate.

end

4.3.4 Muting Optimization Scheme and Dynamic time-frequency PRB Scheduling (MOSDS)

In this section, solutions on solving interference among TUEs and UAVs while maximizing the TUEs' capacity is proposed. Dynamic requests from both TUEs and UAVs can cause higher interference, especially in a high dense urban area. To deal with this issue, we consider a MOSDS-DQN to maximize the total capacity of TUEs, mitigate the interference, and mute the cell causing high interference.

The effect of blank subframes is modelled by assuming that the downlink transmission from the corresponding cells is muted in the corresponding frequen-

cies. The main component is to suppress the blank sub-frames of the interfering cell, and the Almost Blank Sub-frames (ABS) scheme is applied according to ICIC Release 10 [90]. However, to reduce interferences, further-enhanced ICIC (feICIC) solutions are implemented in the system, which pre-allocate the packet in frequency and time domain for UAVs and TUEs as visualized in Figure 4.1 and Figure 4.4.

In this model, the available frequency bandwidth for the DL transmissions is divided into F sub-bands indexed by $f = 1, 2, \ldots, F$ and the time interval is slotted into transmission time intervals (TTIs) indexed by $i = 1, 2, \ldots, N$ as shown in Figure 4.4. The time-frequency resource grid consists of $F \times N$ RBs. Therefore, the date rate of the uth TUE is defined as

$$R_{c,\,u}^{TUE} = B_{i,f}^{c}\ \log_2(1 + \gamma_{c,\,u}), \tag{4.32}$$

where $B_{i,f}^{c}$ is the bandwidth of the RB (i, f) at the cth cell and the data rate of the jth UAV is defined as

$$R_{c,\,j}^{UAV} = B_{i,f}^{c}\ \log_2(1 + \gamma_{c,\,j}). \tag{4.33}$$

For dynamic resource scheduling, we mainly consider efficient dynamic scheduling, where different data sizes and requirements are considered in this scenario. As proposed in [4], the UAV data rate and latency requirements need to satisfy 60-100Kbps and 50ms. Specifically, for the resource allocation problems with different time and frequency requirements, quantized time-frequency resource block allocation scheme is considered, as shown in Figure 4.5. First, the controller classifies different services with the specific QoS requirements according to the service characteristics and the current network congestion. Second, according to the admission control policy, the resource block of each scheduled UAVs and TUEs are continuously mapped to the specific time and frequency domain. Finally, based on the current muting scheme, data sizes, and previous learning experience, dynamic resource scheduling for UAVs and TUEs are considered to reduce interference. Thus, to maximize TUE throughput and guarantee the reliability and latency of UAVs, optimizing both of the scheduling policies and cell muting selection are considered.

In addition, the omnidirectional antenna [119] is utilized in the algorithm to help mitigate the interference efficiently while maximizing the capacity of both TUEs and UAVs. It is assumed that each UAV transmits 1250B every 100ms [119]. Authors in [119] showed that TUEs could achieve the lowest capacity loss

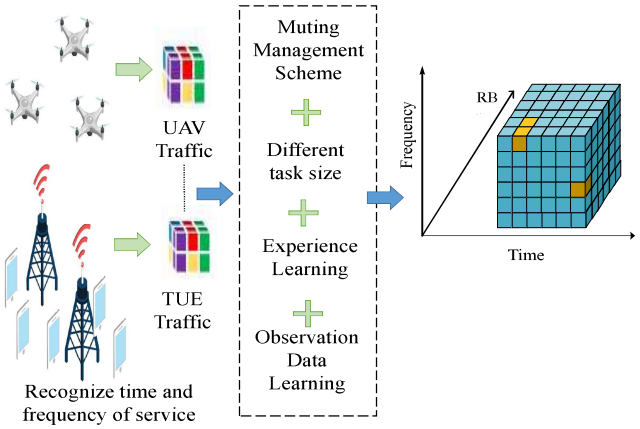

Figure 4.5: The dynamic scheduling design for MOSDS-DQN.

when the UAVs were scheduled to send information at every 10th and 50th TTI. However, the results are different when they have different number of UAVs in different scenarios, i.e., high load scenario. In real scenarios, it is difficult to predict the number of users. Thus, the main focus in this section is to jointly optimize the number of muting cells and UAVs' scheduling schemes, and the optimization problem is formulated as

$$(P2): \quad \max_{\pi(A_t|S_t)} \sum_{i=t}^{\infty} \sum_{c=1}^{C} \sum_{u=1}^{U} \beta^{(i-t)} R_{c,u}^{TUE}(i) \tag{4.34}$$

$$\text{s.t.} \mathcal{N}_{B_u} \cap \mathcal{N}_{B_j} = \varnothing, \quad \forall u \neq j, \tag{4.35}$$

$$\sum_{l=1}^{L_u} \mathcal{N}_{B_u} \leq Q, \quad \forall (u, l), \tag{4.36}$$

$$R_{c,j}^{UAV}(i,\ f) \geq R_{Th}^{UAV}, \quad \forall j \in \mathcal{J}, \tag{4.37}$$

$$R_{c,u}^{TUE}(i,\ f) \geq R_{Th}^{TUE}, \quad \forall u \in \mathcal{U}, \tag{4.38}$$

$$\beta^i \in [0, 1]. \tag{4.39}$$

where $\beta^i \in [0, 1)$ is the discount factor determining the performance accumulated in the future reward. If $\beta^i = 0$, it means that the agent only concerns the immediate reward. Eq. (4.35) shows a RB should always be allocated to one user. The scheduler length $\mathcal{N}_{B_u}$ in Eq. (4.36) should allocate no more than the maximum queue length Q. Next, Eq. (4.37) and Eq. (4.38) ensure a good service rate for UAVs and TUEs, respectively. As the arrival of UAVs cause a trade-off between available PRBs and interferences among all users, it is important to

consider an optimal trade-off among the RSRP, the group of UAV's RB, and muting scheme, which further motivates us to use the learning algorithms to jointly optimize long-term throughput of all users. The DRL agent then learns the optimal mapping from the input states to select the resource allocation action.

State Representation

The current state of the system includes commutative throughput of TUEs and UAVs, given by $S_2^t = \{\sum_{j=1}^{N_{TUE}} R_{j,t}^{TUE}, \sum R_t^{UAV}\}$, where R_{TUE} is a set of data rates of TUEs and R_{UAV} is the instantaneous rate of UAVs.

Action Space

Q-agent will choose an action a from set $\mathcal{A}$. The dimension of the action set is calculated as $\mathcal{A} = N_m \cdot t_s$. The action is denoted as $A_{t+1}^{P2} = \{N_m, t_s\}$, where N_m is the number of muting cells and t_s is the slice time allocation for UAVs to transmit data.

Rewards

After performing the selected actions, the accumulated reward function is given as

$$Reward_t^{P2} = \sum_{u=1}^{U} R_u \cdot \mathbf{1}\left[R_{u,j} \geq R_{Th}\right]. \tag{4.40}$$

The MOSDS-DQN algorithm is shown in Algorithm 5.

Figure 4.6 shows the proposed network architecture, where the current state is input into the neural network for the DQN algorithm. Next, an RNN-based GRU network is used to approximate the value function of the DRL algorithm. The GRU can capture the correlation between the state and action over time, and helps DRL to select the optimal actions.

4.3.5 Computational Complexity Analysis

In this section, we evaluate the computational complexity of one iteration of our proposed algorithm with respect to the size of the network, namely, the number of UEs and available resources. The computational complexity of the DQN algorithm, including DQN learning architecture, the action selection of the agent, and the downlink transmission, is given by $O(mlogn + 2^A + N_i N_k)$, where m is the number of layers, n is the number of units per learning layer, and A is the number of action.

Algorithm 5: : MOSDS

Initialization α, ϵ, M, $\boldsymbol{\theta}$, and $\bar{\theta}$.

UAV identifies the highest RSRP and associate with it.

Receive muting cell ID from Algorithm 4 and the packet scheduling.

Determine all active UEs (TUEs & UAVs) using Bernoulli process and time packet scheduling.

Determine associate cell UEs & active UEs matrices.

for $e \leftarrow$ *1 to I* **do**

  Initialize s^1 by executing a random action a^0;

  for $t \leftarrow$ *1 to T* **do**

    If $p_\epsilon < \epsilon$: Randomly select action a^t from $\mathcal{A}$;

    else select $a^t = \underset{a \in A}{argmax} Q\left(S^t, a, \theta\right)$;

    for $PRB_i \leftarrow$ *1 to I* **do**

      for $activecell_c \leftarrow$ *1 to C* **do**

        Update the queue matrices.

        Calculate pathloss.

        Calculate Antenna Gain.

        Calculate received power.

        Calculate the channel states.

        Calculate SINR and transmission rate.

      end

    end

  end

end

4.4 Numerical Results and Evaluation

In this section, we examine the effectiveness of our proposed muting optimization schemes using DQN algorithm. The network consists of 7 cells covering 1500m x 1500m area. In the simulation, the UAVs are distributed with a fixed flying height. The height of all TUEs is 1.5m, and the height of all UAVs is assumed to be of 120 m following UK regulations [3]. Both TUEs and UAVs are assumed to equipped with a single antenna. The TUEs and UAVs in each cell are uniformly distributed and the maximum number of UAVs in the entire network is 10. We assume that all users move within its corresponding cells. When a TUE reaches the boundary, it turns back and moves in a random direction. The network parameters for the system are shown in Table I, and follow the 3GPP specifications in [4], [13], and [177]. The proposed algorithms were implemented through the

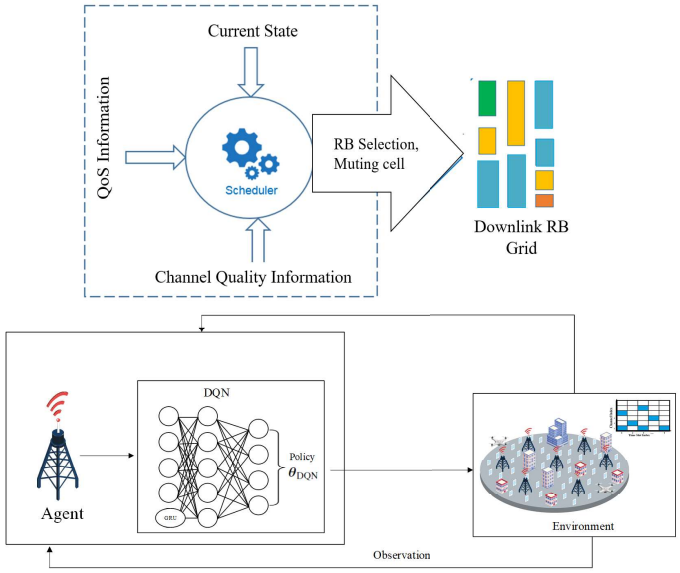

Figure 4.6: The learning network architecture for MOSDS-DQN

Python 3.7 programming tool, which is a powerful multi-purpose programming language with the TensorFlow framework in conjunction with multiple python packages and libraries, like keras and sklearn, because of their advanced and helper functions that make it much easier and faster to write DRL algorithms, due to the familiarity of the researchers with the tool. The results are obtained by averaging over 100 episodes, with each episode containing 100 TTIs.

In the downlink, the TUE traffic pattern is modeled as File Transfer Protocol (FTP) sessions [119], where both packet size and arrival time follow Poisson distribution. The downlink scheduler prioritizes the UAV transmission and C2 traffic over the FTP traffic, meaning that the BS schedules the UAV transmission first, and then the remaining TUEs and resources are divided equally among the connected TUEs that have FTP data to receive. If there is no downlink data to be transmitted, users are assumed to be in an idle mode. Otherwise, the user switches from the idle mode to a connected mode. Once the data buffer is clear, the user returns to the idle mode.

4.4.1 Muting Optimization Scheme using Deep Q-Learning

Figure 4.7 shows the reward of the dynamic muting scheme for different learning algorithms. From the simulation results, it is clearly shown that DQN outperforms both VFA and Tab-Q. The convergence of both VFA and Tab-Q are slightly faster than DQN, but unable to obtain the maximum reward as of DQN. Tab-Q fails to perform exploitation of each action in continuous time slots as it is fixed

Table 4.1: Simulation Parameter

Parameters	**Value**
Transmission power, P_c	30 dBm
Bandwidth, B	3 MHz
Noise power $N_{c,u}$	-142.39 dBm
Center frequency, f_c	2 GHz [61]
θ_{3dB}	65^o
ϕ_{3dB}	65^o
SLA_V	$30dB$
A_m	$30dB$
Antenna Gain, G_{max}	8dbi
UAV Threshold, R_{Th}^{UAV}	1Mbps [169]
UAV Threshold, R_{Th}^{TUE}	20 bps [94]
Learning Rate	0.1, 0.01
Discount Rate	0.8
Replay memory	1000

in a suboptimal strategy [81]. In addition, the VFA's target network might not fully works due to features and high number of state-action, which causes VFA cannot perform better exploration over time. DQN can explore and exploit actions, which enable it to obtain the maximum state-action value. Moreover, the convergence analysis of the reinforcement learning algorithms has been proven in [77],[80], so that the agent of the Q-learning algorithm can converge to the optimal Q value.

Figure 4.8 shows how dynamic muting actions affect the total throughput for all TUEs and UAVs via DQN muting scheme over time. In Figure 4.8, "Reward" represents the cumulative reward, "UAV" represents the total throughput for all UAVs, "TUE" represents the total throughput for all TUEs, and "Mute Cells" represents the number of muting cell in each time slot. At the early stage of learning, DQN learns to be adapted to the environment based on the observations, and the reward continues increasing. When $t = 155, 156, 189$, and 21617, the rewards drop to zero due to the difficulties in choosing the correct muting number that suits the current environment, which leads to the transmission rate of UAV not satisfying the threshold in Eq. (4.18). As time passes, DQN can predict and learn how to maximize the reward. However, the system can become unstable when the epsilon-greedy parameter is less than the threshold, namely, $p_e < \epsilon$, as

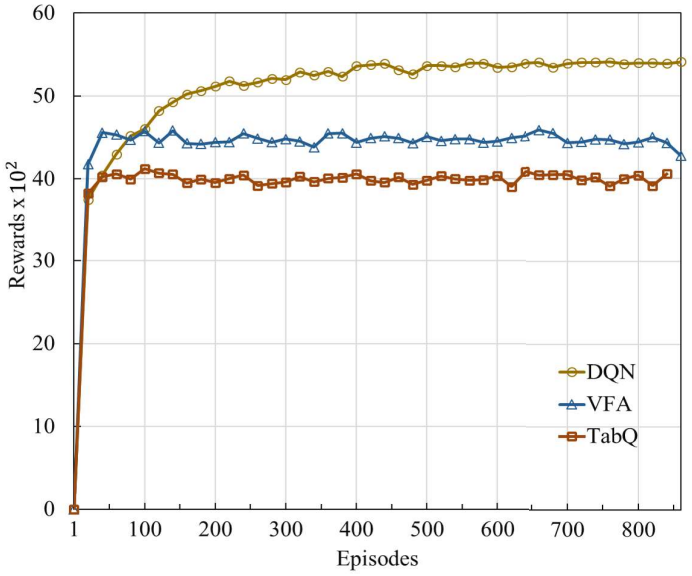

Figure 4.7: Rewards performance comparison between different learning algorithms.

it directly selects a random action and decreases the performances of TUE. When the algorithm converges, the performance of all TUEs and UAVs maintain at the maximum value with an optimized number of muting cell.

Figure 4.9 plots the convergence performance of DQN with different mitigation schemes [119]. For simplicity, "Highest benchmark" represents the linear muting scheme with 3 strong interfering neighboring cells are muted to allow UAVs to transmit the data without interference from TUEs, and "Lowest benchmark" shows the performance of the linear muting scheme when the system mutes a single neighboring cell with the highest interference, which can mitigate the interference between UAVs and TUEs. The "Highest benchmark" and "Lowest benchmark" use linear mitigation schemes in [119]. In [119], the "highest benchmark" has muted a maximum of 3 strongest RSRP interference signals to cancel the interference as following by the 3GPP Release-13 model [119, 71]. The DQN-based muting scheme shows 48% improvement compared to "Highest benchmark". It is proved that the DQN scheme can adequately select the correct number of muting cells to reduce interference, even though the proposed system changes dynamically. In addition, the DQN is able to perform in a dynamic scenario with a varying number of UAVs and TUEs, and select proper actions for the agent to maintain a higher data rate of TUEs. Compare to the lowest benchmark scheme, DQN improves nearly 80% of the overall data rate performance.

Figure 4.10 plots the interference comparison analysis for DQN-based muting and linear muting schemes in [119]. It can be seen that the proposed DQN muting scheme outperforms the Highest benchmark scheme. The result proves

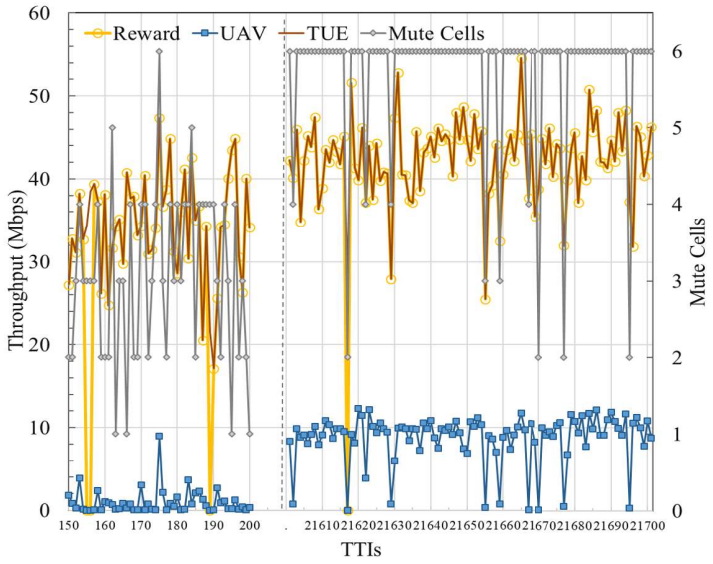

Figure 4.8: Dynamic action influence the rate for all TUEs and UAVs over time.

that the DQN muting scheme can accurately choose the cell muting index to reduce the interference in dynamic environments, and further maximize the average throughput.

Figure 4.11 and Figure 4.12 show the throughput performance of TUEs and UAVs in different situations, respectively. From Figure 4.11, we observed that when the number of TUEs increases, the interference increases, and more number of TUEs and UAVs cannot satisfy its minimum transmission requirements. From Figure 4.12, we can obtain that when the number of TUEs is small, UAVs achieve high throughput. However, when the number of TUEs increases up to 70, the average capacity of UAVs decreases because of high interference, and more UAVs cannot satisfy its minimum transmission requirements. It is because the muting schemes try to decrease the number of muting cells to let a high number of TUEs transmit data, which leads to the UAVs being unable to satisfy its minimum requirements of transmission rate. Also, high number of TUEs causes less bandwidth allocated to UAVs, which further leads to lower throughput of UAVs. In addition, the performance of the UAV with the Tab-Q algorithm decreases dramatically when the number of TUEs increases. This is because Tab-Q with high dimensional state space requires large memory, and has difficulty in selecting proper actions to achieve optimal results.

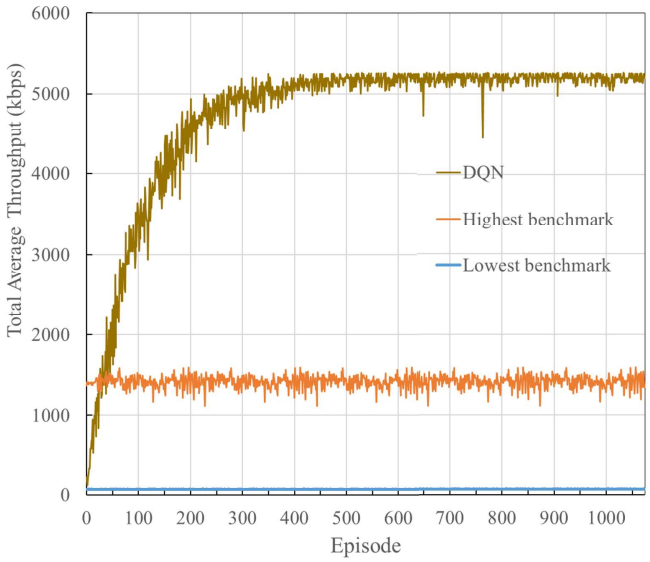

Figure 4.9: Average TUEs' throughput comparison between DQN-based muting scheme and linear muting.

4.4.2 Muting Optimization Scheme and Dynamic PRB Scheduling (MOSDS-DQN)

This section evaluates the proposed muting optimization scheme and dynamic PRB scheduling with MOSDS-DQN algorithm. Figure 4.13 shows the convergence performance of the MOSDS-DQN muting scheme. For instant, "MOSDS-DQN" represents DQN muting optimization scheme and dynamic PRB scheduling. It is observed that the MOSDS-DQN algorithm performs better than the DQN muting scheme. However, the MOSDS-DQN scheme shows the lowerst convergence speed. It is because MOSDS-DQN muting scheme has larger state and action space, and needs to select more proper actions to mute cells and allocate PRBs to UAVs and TUEs.

Figure 4.14 shows the throughput performance of dynamic actions over time. In Figure 4.14, "Reward" represents the cumulative reward, "UAV" represents the total throughput of all UAVs, and "Mute Cells" represents the number of muting cells each time slot. Figure 4.14a shows the muting cell action selected to maximize the overall throughput. At the beginning of the learning process, the reward continues increasing, it is because the algorithm is learning the environment based on previous experience. When the learning algorithms converge, proper number of muting cells is selected to decrease the interference and improve the throughput. However, in some time slots, the performance of the UAV severely decreases and cannot satisfy the minimum QoS because of some factors,

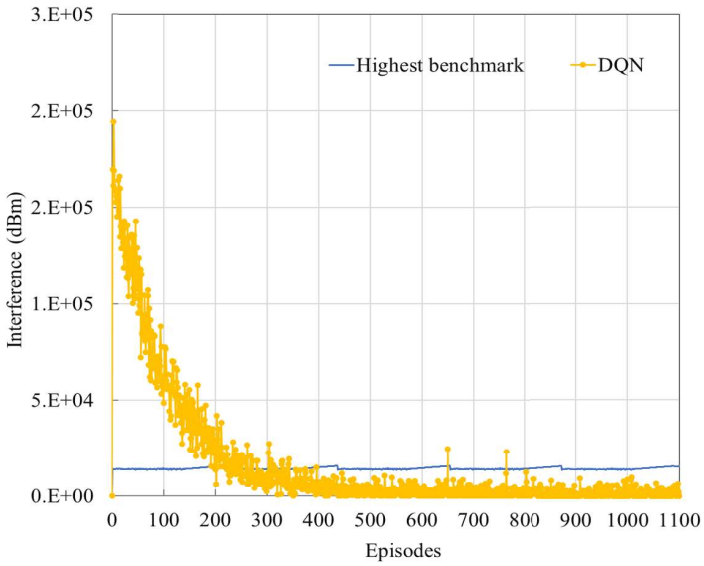

Figure 4.10: Comparison of interference analysis between DQN-based muting scheme and linear muting.

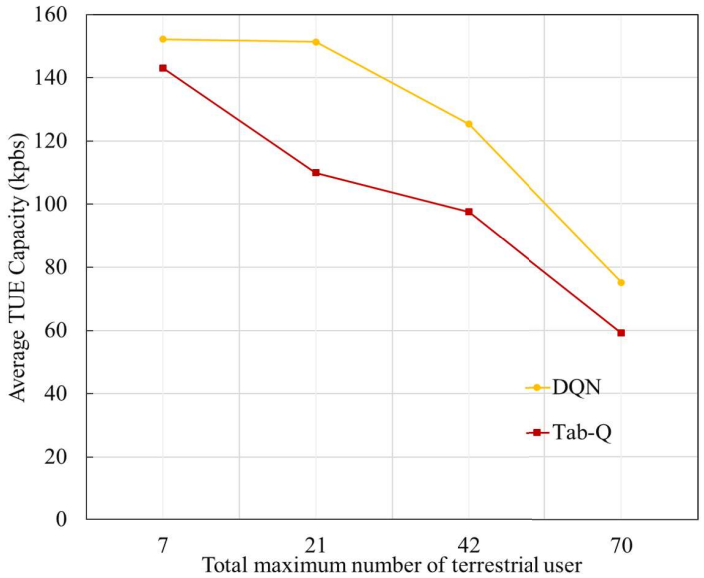

Figure 4.11: Average capacity rate for TUE based on different number of TUEs.

such as the data size of UAV, time allocation, and bandwidth allocation.

Figure 4.14b shows how the BS sends data to the UAV in 100ms. The network environment condition and the location of UAV play important roles in MOSDS-DQN to plan the number of data pack of UAV transmission. For example, if the UAV is far from the cell, the frequency of sending UAV's data pack should be reduced to decrease the transmission failure. Thus, less data pack of the UAV is transmitted, and less PRB is allocated to the UAV, which decreases the interference and improve the throughput performance of TUEs.

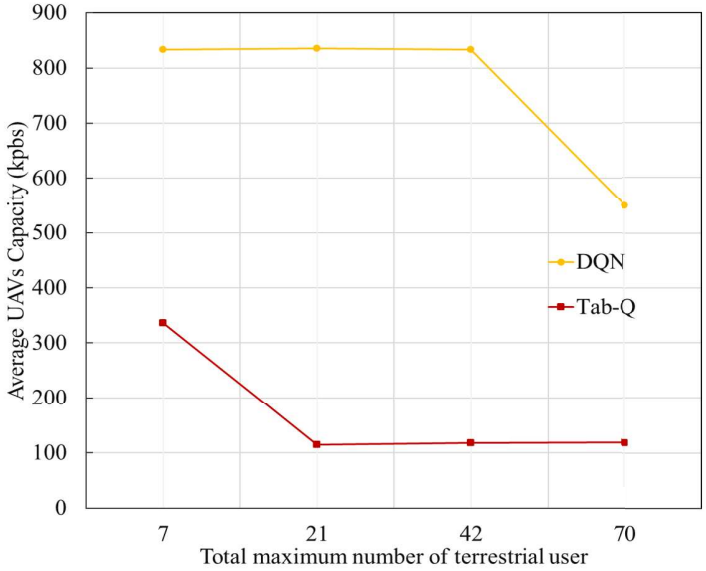

Figure 4.12: Average capacity rate for UAV based on different number of TUEs.

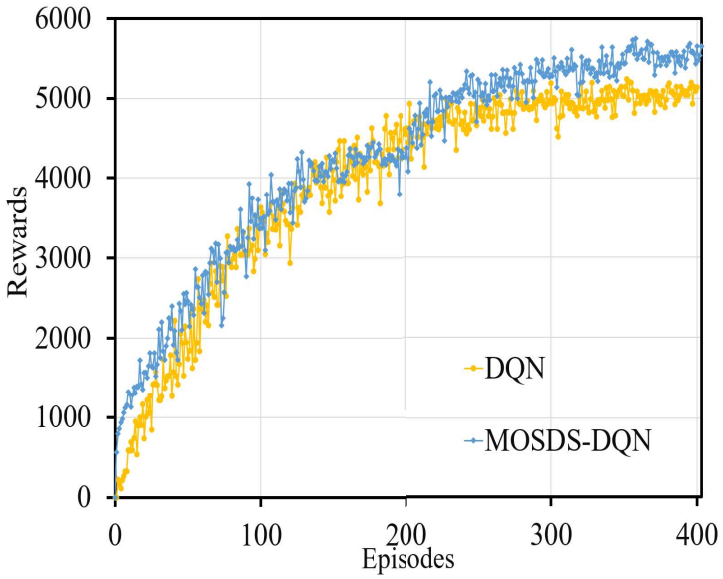

Figure 4.13: Rewards performance comparison between different schemes.

In the early phase of Figure 4.14c, the high-frequency range is used frequently, and it causes the performance of both UAVs and TUEs to decrease because of high interference. When the learning algorithms converge, proper frequency range for each PRB is selected to satisfy the QoS requirement. Therefore, the MOSDS-DQN algorithm is able to provide an effective way to select proper number of cells to mute and allocate proper PRBs to TUEs and UAVs, especially in different scenarios.

Figure 4.15 plots the interference comparison analysis for DQN-based muting and MOSDS-DQN schemes. It can be seen that the proposed MOSDS-DQN mut-

Figure 4.14: Dynamic action influence the reward for all UAVs in each episode.

ing scheme outperforms the DQN scheme. The result proves that the MOSDS-DQN muting scheme can accurately choose the cell muting index and allocate proper PRBs to TUEs to reduce interference in dynamic environments.

Figure 4.16 plots the average capacity rate for all TUEs with different number of UAVs based on muting schemes via DQN and MOSDS-DQN. It is observed

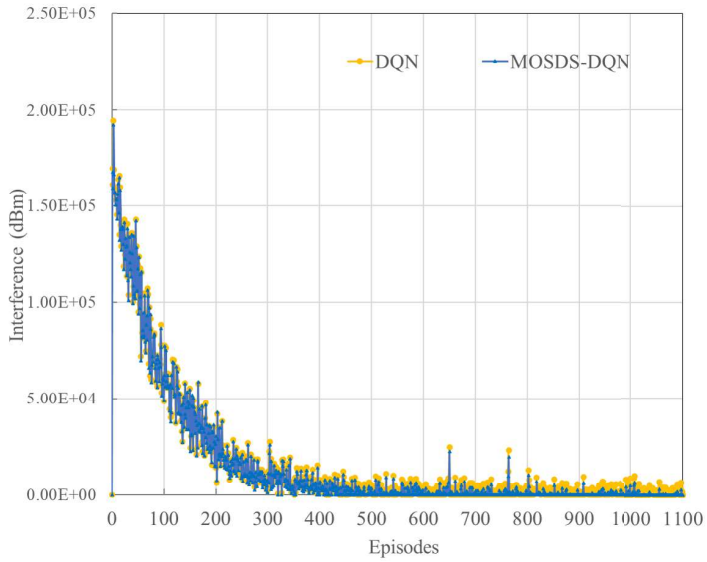

Figure 4.15: Comparison of interference analysis between DQN-based muting scheme and linear muting.

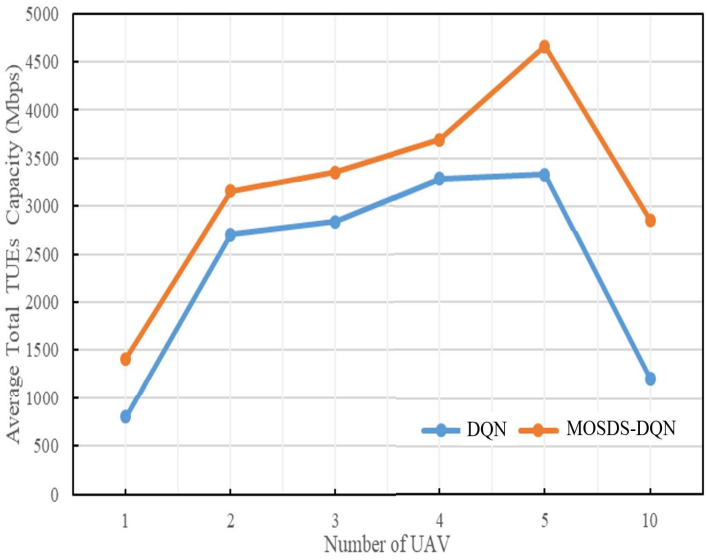

Figure 4.16: Comparison of average capacity rate for all TUEs based on different number of UAVs.

that when the number of UAVs is less than 4, both algorithms increase average capacity rate. However, when the number of UAVs increases from 4 to 5, the DQN algorithm is unable to increase average capacity. It is because it cannot select proper actions in a large action space. When the scenario becomes more complex, the algorithms need to balance the performance between UAVs and TUEs. In the simulation, UAVs have higher priority, thus, the performance of TUEs decreases tremendously when a high number of UAVs exist. Furthermore,

when the number of UAVs increases, the performance of MOSDS-DQN is higher than DQN.

4.5 Conclusion

As conclusion, the downlink inter-cell interference coordination mechanism is developed to mitigate the interference between BSs and TUEs while satisfying the rate requirements of UAVs. Dynamic muting optimization scheme and dynamic scheduling of PRBs were proposed to maximize the throughput of all users, and mitigate the interference by muting the cell(s) that caused high interference. The proper muting technique with proper number of muting interference cell and the time-frequency scheduling schemes to allocate the PRBs of TUEs and UAVs guarantee excellent service and satisfy QoS among TUEs and UAVs in long time slot. Simulation results showed that our proposed learning-based schemes achieved 80% and 48% performance improvement of throughput compared to the lowest and highest linear muting algorithms, respectively. Furthermore, the proposed MOSDS-DQN also showed 18% improvement compared to DQN algorithm.

Chapter 5

Radio Mapping-aided Beam Alignment for mmWave UAVs

5.1 Introduction

Beam forming alignment is vital component in millimeter wave (mmWave) wireless communication system. However, the usage of cellular-connected unmanned aerial vehicle (UAV) especially in dynamic scenario will cause challenges in conventional beam sweeping approach, especially when has a large overhead due to the high mobility and autonomous operation of UAVs.

In this chapter, we propose the deep reinforcement learning (DRL)-based framework for UAV-BS beam alignment using the hDQN in the 5G radio setting. Fast mmWave beam alignment could enhance the reliability and decrease the latency of 5G and beyond wireless systems for both UAV-UAV and BS-UAV communications [34]. Especially, the availability of UAV position information at lower frequencies (following the works [143, 145]) may also provide scope for reliable communication in addition to increasing throughput. Position information for fast beam alignment has been recently studied in mmWave systems [142, 146, 19]. The authors in [146, 19] proposed the learning-based beam training schemes using MAB approach, by building the database of finite beam-pairs useful for beam training based on vehicular position information. The key idea is that the ML-based approaches can effectively use the position information for fast mmWave beam alignment in an online manner. High mobility and autonomous operation of UAVs also requires frequent beam realignment and can be jointly optimized with reliable connectivity effectively using RL-based beam training.

The authors in [109] use radio map-based channel propagation environment information for efficient positioning of UAV. Radio map can be helpful in im-

proving learning and real-time update when there are sufficient changes in the radio source location (e.g., moving beyond the decorrelation distance power and others [26]). In such scenarios, radio maps that describe average outage channel probability, channel gains or signal-to-interference-plus-noise ratio (SINR) [168, 26, 179, 52, 109] help to extract such features needed to enhance the information needed to reduce training overhead and improve converges time. Radio map refers to the geographical signal power spectrum density, formed by the superposition of concurrent wireless transmissions, as a mathematical function of location, frequency and time. It contains rich and useful information regarding the spectral activities and propagation channels in wireless networks [26]. This shows that radio maps can be useful to better capture the spatial correlation information and decrease the overall learning convergence time. Recent works [153, 154] used channel knowledge map and the UAV location information, namely channel path map and beam index map, to derive network efficiency. However, the work assumes using the perfect channel which may be impractical for real-world environments. Therefore, we model the BS - UAV beam pair alignment problem using hDQN and convolution neural network radio mapping (CRM) with the aim to reduce the beam search complexity for UPA configurations. We consider the uplink environment where both BS and UAV beam direction pairs act as parameters of the learning problem, so that BS can be the receiver and the learning agent. The spatial position grid arrangement of the antenna elements that dependent on landmark configurations is designed by using vertical and horizontal angular, and configuring 360° of coverage area can be effectively exploited alongside the UAV position. The spatial information is considering mmWave beams with different beam width resolutions in a hierarchical manner during beam-training. Our simulations showed that the hDQN approach reduces the beam training overhead by 63% from our prior DQN method.

The contributions of this chapter are summarized as follows:

- We model the spatial position context-information-beam-mapping to solve beam-pair alignment problem in uplink mmWave MIMO communication system. The BS serves a UAV using 5G new radio (NR) in the communication protocol.

- We solve the BS - UAV beam pair alignment problem using hDQN. During uplink communication, the proposed method optimizes the BS - UAV beam-pair alignment generically across any UAV grid position inside the BS coverage area.

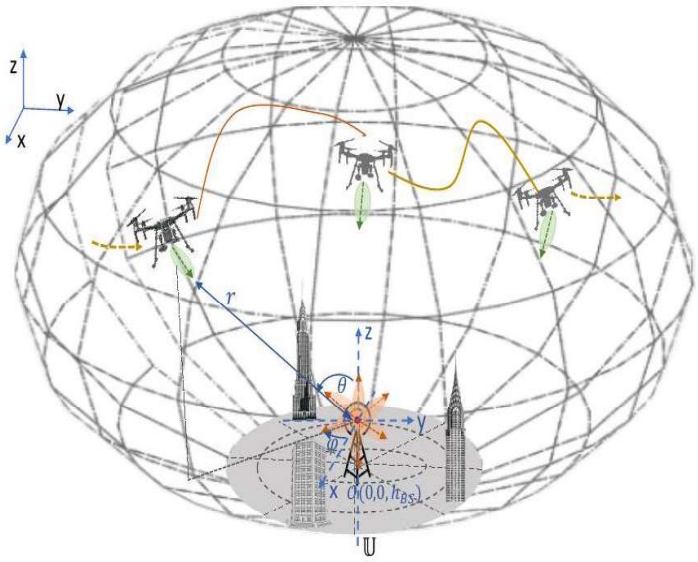

Figure 5.1: Illustration of System Model

- We develop the radio map as input to the 360° spatial-information-beam-mapping to give additional information to hDQN and help improve the convergence time.
- To maximize the performance of UAV, we maximize the beam-forming gain for every communication requests by using CRM-hDQN.
- We compare our CRM-hDQN-based proposed approach with vanilla method and hDQN beam alignment under ideal channel conditions. We analyse these approaches over different coverage areas and antenna configurations. Our results shown that CRM-hDQN-based approach achieves the optimal beam alignment with reduced number of training iterations.

5.2 System Model

Figure 5.1 considers a cellular mmWave MIMO uplink communications between the UAV and the BS. The BS serves multiple UAVs in the time domain multiple access (TDMA) manner under its spherical coverage area. The BS is fixed at centre, $\mathcal{O}\left(0,0,h_{\mathrm{BS}}\right) \in \mathbb{R}^3$ and communicates with the moving UAV with a multi-path mmWave beam-forming. We assume single UAV and single BS for the urban macro-cellular (UMa) environments. The multi-antenna UAV hovers randomly and communicates with the multi-antenna BS in the urban environment following 5G NR standard protocol [31]. The environment is 3D spherical coverage area composed of multiple grids, the set enclosing them is denoted as $\mathbb{U}$. Following

the 3D spherical coordinate system, let r_h, θ_h, ϕ_h represent the radial distance, elevation and azimuthal angles of grid index $h \in \{0, 1, \ldots |\mathbb{U}|\}$ with respect to BS.

We consider an analog beamforming equipped with one radio frequency (RF) chain and UPA structures of N_t and N_r antennas for both BS and UAV, respectively. The UAV transmit (TX) while the BS receives received (RX) a radio signal in multiple beam directions following $\mathcal{B}_{\text{TX}}$ and $\mathcal{B}_{\text{RX}}$ codebook, respectively with angles defined as

$$b_i = (i-1)\frac{\pi}{N}, i \in \{1, 2, \ldots, N\}, \tag{5.1}$$

where b_i represents a RF radio beam direction with a fixed narrow beam width $\left(\frac{\pi}{N}\right), N$ represents N_t, N_r antennas for $\mathcal{B}_{\text{TX}}$ and $\mathcal{B}_{\text{RX}}$ codebook, respectively. The codebook values are defined using the beamforming vectors w_{TX} and w_{RX} forUAVand BS, respectively, given by

$$\text{w}\left(b_i\right)]_{n=0}^{N-1} = \frac{1}{\sqrt{N}} \exp\left(j\frac{2\pi n d}{\lambda}\sin\left(b_i\right)\right), b_i \in \mathcal{B}, \tag{5.2}$$

where $N = N_t, \mathcal{B} = \mathcal{B}_{\text{TX}}$ and $N = N_r, \mathcal{B} = \mathcal{B}_{\text{RX}}$ for $\text{w} = \text{w}_{\text{TX}}$ and $\text{w} = \text{w}_{\text{RX}}$, respectively. Here, d is the antenna spacing assumed to be $\frac{\lambda}{2}$ in this work. λ is the wavelength and b_i is the i^{th} codebook direction Eq. (5.1).

5.2.1 User Mobility

The UAV moves randomly along the 3D coverage area which is divided into multiple (x, y, z) grids of equal size. $\text{UAV}_h(t)$ in represents the user equipment (UE) position on grid index $h \in \{0, 1, ...|\mathbb{U}|\}$ in the BS coverage area $\mathbb{U}$ at any time instant t is given by

$$\text{UAV}_h(t) = \left(r_h \sin\theta_h \cos\phi_h, r_h \sin\theta_h \sin\phi_h, r_h \cos\theta_h\right), \tag{5.3}$$

where $Ux_t, y_t, z_t \in \mathbb{U}$ and r_h is the UE radical distance from BS and grid elevation angle and azimuth angle. The UE acts as TX, BS as RX equipped with single RF chains and uniform linear array (ULA) structure of N_t and N_r antennas, respectively. Here, $\mathbb{U}$ and $\{z_t\} \in \mathbb{Z}$ are the coverage areas of the serving Node BS and altitude ranges of UE, respectively. The BS receives (RX) radio signal through one of its multiple beam directions each time, following the same codebook set $\mathcal{B}$. We assume UE_h (with respect to BS locations) is known during each P_1 procedure of 3GPP beam access protocol. Here we assume UAV hovers on every hop with random mobility and enters the grid position defined in $\mathbb{U}$.

The communication begins with the TX request from UAV, while the BS RX radio unit at BS, and starts with the random beam-pair at time $t = 0$ and learns to choose the beam direction $(b_p, b_q)\,, b_p \in \mathcal{B}_{\mathrm{TX}}, b_q \in \mathcal{B}_{\mathrm{RX}}$ over time for each TX grid position with index $h \in \{0, 1, \ldots |\mathbb{U}|\}$. We assume TX and RX beam directions to be the same for all UAVs movements within each grid position.

The BS receives the initial radio beam b_q at broader angular-resolution level from $\mathcal{F}$ and then switch to narrow radio-beams over time, to reduce the beam search space and still achieve efficient beamforming gains for UPA antenna configurations. Here, we assume the moving UAV transmit radio signals in the same narrow beam directions within each grid position. Thus, the BS selects the sequence of beam-pair directions for TX and RX, with every change in grid position as the substantial change in TX location induces a variance in the radio measurements, following 3GPP fifth 5G NR beam alignment protocol [43]. The 3GPP 5G NR beam alignment protocol for physical layer consists of initial communication (used as $(P1)$ procedure), beam selection (used as $(P2)$ procedure) and an optional beam refinement (used as $(P3)$ procedure) [43]. We consider BS and UAV following $(P1)$ and $(P2)$ procedures at every grid position, along the coverage area set $\mathbb{U}$. During $(P1)$ procedure, the UAV is assumed to send the communication request with respect to its position, while the learning framework at BS responds with a hierarchical sequence of radio beam-pairs to be considered for next phase of uplink based beam access protocol. $(P2)$ generally implies the radio beam selection procedure at mmWave frequencies later used for the data transmission [43]. Similar to the works in [144, 133, 135], the BS and UAV in $(P2)$ are assumed to undergo the beam-training procedure following the sequence of beam-pairs configured by the BS-side learning framework from initial communication procedure.

The received signal measurement can be observed at the BS for different TX-RX beam pairs during these procedures and its' timing information can be estimated using 5G protocol frame structure [43]. We define travel time unit (TTU) as the orthogonalorthogonal frequency division multiple access (OFDM) symbol time during every beam transmission or reception from the 5G frame structure. In this work, we use this definition to measure the communication overhead for the learning-based beam sweeping procedure in TTU units.

5.2.2 Communication Model

We consider the multi-path link (LoS or non-line-of-sight (nLoS)) radio channel between UAV at time t and BS location $\mathcal{O} \in \mathbb{R}^3$. The BS and UE are equipped

with single radio frequency (RF) chains of $\left(N_x^{\text{rx}}, N_y^{\text{rx}}\right)$ receive and $\left(N_x^{\text{tx}}, N_y^{\text{tx}}\right)$ transmit antennas respectively. As the BS serves multiple UEs in a TDMA manner, we model the communication between single UAV and single BS with UPA for the UMa environments [43]. We assume each UPA beam at both BS and UAV projected with azimuthal ϕ and elevation θ main lobe broadside direction. Let M denote the number of multi-paths or reflection points in the environment, the channel matrix corresponding to the m^{th} path is given by

$$\mathbf{H}_m \triangleq \beta_m \mathbf{a}_R\left(\theta_{\text{m}}^{\text{rx}}, \phi_{\text{m}}^{\text{rx}}\right) \mathbf{a}_T^H\left(\theta_{\text{m}}^{\text{tx}}, \phi_{\text{m}}^{\text{tx}}\right) \tag{5.4}$$

where, β_{m} is the antenna channel gain, $\theta_{\text{m}}^{\text{tx}}, \theta_{\text{m}}^{\text{rx}}$ are the azimuthal angle of departure (AoD) and angle of arrival (AoA), $\phi_{\text{m}}^{\text{tx}}, \phi_{\text{m}}^{\text{rx}}$ are the elevation AoD and AoA of m^{th} communication link between BS and UE. a $_R\left(\theta_{\text{m}}^{\text{rx}}, \phi_{\text{m}}^{\text{rx}}\right) \in \mathbb{C}^{N_x^{\text{rx}} N_y^{\text{rx}}}, \mathbf{a}_T\left(\theta_{\text{m}}^{\text{tx}}, \phi_{\text{m}}^{\text{tx}}\right) \in \mathbb{C}^{N_x^{\text{tx}}} N_y^{\text{tx}}$ are the antenna array steering vectors for $\left(\theta_{\text{m}}^{\text{rx}}, \phi_{\text{m}}^{\text{rx}}\right)$ and $\left(\theta_{\text{m}}^{\text{tx}}, \phi_{\text{m}}^{\text{tx}}\right)$, respectively. Let $\omega_x = \frac{2\pi}{\lambda} d_x \sin\theta\cos\phi, \omega_y = \frac{2\pi}{\lambda} d_y \sin\theta\sin\phi, \lambda$ is the wavelength, $\otimes$ denote the Kronecker product, N_x and N_y are the antenna elements along x and y-axis, d_x and d_y are the antenna element spacing in x and y-direction, respectively. Then, the array steering vector is given by

$$a(\theta, \phi) = \frac{1}{\sqrt{N_x N_y}} \begin{bmatrix} 1 \\ e^{j\omega_y} \\ \vdots \\ e^{j(N_y-1)\omega_y} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ e^{j\omega_x} \\ \vdots \\ e^{j(N_x-1)\omega_x} \end{bmatrix} \tag{5.5}$$

where $(\theta, \phi) = \left(\theta_{\text{m}}^{\text{rx}}, \phi_{\text{m}}^{\text{rx}}\right), (N_x, N_y) = \left(N_x^{\text{rx}}, N_y^{\text{rx}}\right)$ and $(\theta, \phi) = \left(\theta_{\text{m}}^{\text{tx}}, \phi_{\text{m}}^{\text{tx}}\right), (N_x, N_y) = \left(N_x^{\text{tx}}, N_y^{\text{tx}}\right)$ for $\mathbf{a}_R\left(\theta_{\text{m}}^{\text{rx}}, \phi_{\text{m}}^{\text{rx}}\right)$ and $\mathbf{a}_T\left(\theta_{\text{m}}^{\text{tx}}, \phi_{\text{m}}^{\text{tx}}\right)$, respectively. For a unit-norm transmit and receive beamforming vectors namely, $\mathbf{w}_k \in \mathbb{C}^{N_x^{\text{tx}}\ N_y^{\text{tx}}}$ and $\mathbf{f}_k \in \mathbb{C}^{N_x^{\text{rx}}}\ N_y^{\text{rx}}$, baseband equivalent of the received signal at discrete symbol time k is given by

$$y_k = \underbrace{\sum_{m=0}^{M} \sqrt{P_{tx}} \mathbf{f}_k^H \mathbf{H}_m \mathbf{w}_k x_k}_{r_k} + \nu_k, \tag{5.6}$$

where P_{tx} is transmission power, $\nu_k \sim \mathcal{CN}(0, WN_0)$ is the effective noise with zero mean and two-sided power spectral density $\frac{N_0}{2}$, x_k represents one OFDM symbol of the time-domain transmitted signal with bandwidth W and TTU time period with $\frac{1}{K}\sum_{k=0}^{\text{K}} \|x_k\|^2 = 1$. We assume $\mathbf{H}_m$ to follow 3GPP UMa conditions [44] and

$k = 0, 1, ...K$ denotes the number of samples spanned over TTU time. $\mathbf{w}_k$ and $\mathbf{f}_k$ for UPA beams are measured using (5.5) for selected codebook directional pairs (θ_k, ϕ_k) from $\mathcal{W}$ and $\mathcal{F}$, respectively.

5.2.3 Antenna Configuration Model

Follow to our previous work on linear codebook direction sets [136] and following (5.5), the UPA radio beam directions are determined by linear array angular resolutions namely, $\frac{2}{N_x}$ and $\frac{2}{N_y}$ with their physical angles $(-\frac{\pi}{2}, \frac{\pi}{2})$ along x and y direction, respectively. For example, the maximum number of ULA antenna elements considered either at TX or RX side is 8. Similarly under UPA, the maximum number of antenna elements at TX side and RX side are set as $2 \times 2 = 4$ and $8 \times 8 = 64$, respectively. Therefore, if a configuration of 8 ULA antenna elements at both TX and RX, thus, the total possible beam pair is 64. We consider a mmWave radio signal with 30 GHz carrier frequency and perform the simulations on both aerial LoS and nLoS 3GPP UMa channel conditions. The path loss models for LoS and nLoS are denoted by UMa-avLoS and UMa-avnLoS, respectively, following five parameter alpha-beta-gamma model (from [6]) as shown in Table 5.1.

We assume that the UE transmits radio signals with a narrowest angular resolutions in $\mathcal{W}$ codebook directions while the BS receives the signal through one of its hierarchical multi-angular resolution codebook directions from $\mathcal{F}$. The hierarchical directional set $\mathcal{F}$ consists of L (for example, $0 \leq l \leq L, L = \log_2(N_x^{\text{rx}})$ along x-axis) multiple angular resolution levels along x and y directions separately, with the l^{th} level codebook directional subset $\mathcal{F}_l = \{f_1^{(l)}, f_2^{(l)}, ..., f_{\mathbf{card}(\mathcal{F}_l)}^{(l)}\}$ designed to uniformly cover all the spatial frequency range $(-1, 1)$ (physical angles $(-\frac{\pi}{2}, \frac{\pi}{2})$) along x and y-directions separately and satisfy the relation $\mathbf{card}(\mathcal{F}_1) < ... < \mathbf{card}(\mathcal{F}_L)$ as shown in Figure 5.2. Here, $\mathbf{card}(\mathcal{F})$ denotes the cardinality of $\mathcal{F}$. For this work, we consider a two level angular-resolution subsets at RX, namely, $\mathcal{F}_\mathcal{B}$ and $\mathcal{F}_\mathcal{N}$ ($\mathbf{card}(\mathcal{F}_\mathcal{B}) < \mathbf{card}(\mathcal{F}_\mathcal{N})$) with their beam widths $\psi = \frac{2}{\mathbf{card}(\mathcal{F})}$, where $\psi = \psi^\mathcal{B}$ and $\psi = \psi^\mathcal{N}$ ($\psi^\mathcal{B} > \psi^\mathcal{N}$) for $\mathcal{F}_\mathcal{B}$ and $\mathcal{F}_\mathcal{N}$, respectively. We select $(\psi_x^\mathcal{N}, \psi_y^\mathcal{N}) = (\frac{2}{N_x^{\text{rx}}}, \frac{2}{N_y^{\text{rx}}})$ and $(\psi_x^\mathcal{B}, \psi_y^\mathcal{B}) = (\frac{2^{(L-l+1)}}{N_x^{\text{rx}}}, \frac{2^{(L-l+1)}}{N_y^{\text{rx}}}), l \in [0, L)$. We assume an elliptical surface for every rectangular grid element in $\mathbb{U}$ and is proportional to $\psi^\mathcal{B}$ given by $(\psi_x^\mathcal{B}, \psi_y^\mathcal{B}) = (\eta_\theta cos\theta_0, \eta_\phi)$ where η_θ and η_ϕ are the elevation and azimuthal angular resolution in $\mathbb{U}$, respectively[50]. θ_0 is the elevation angle of a UE grid element $g, g \in \mathbb{U}$ from the BS. Thus, with $\psi^\mathcal{B}$ selection, η_θ and η_ϕ can be chosen to favour single broad beam projection from BS, for every grid element in $\mathbb{U}$. We define $r_k = \sum_{m=0}^{M} \sqrt{P_{tx}} \mathbf{f}_k^H \mathbf{H}_m \mathbf{w}_k x_k$. Then, the signal-to-noise ratio (SNR)

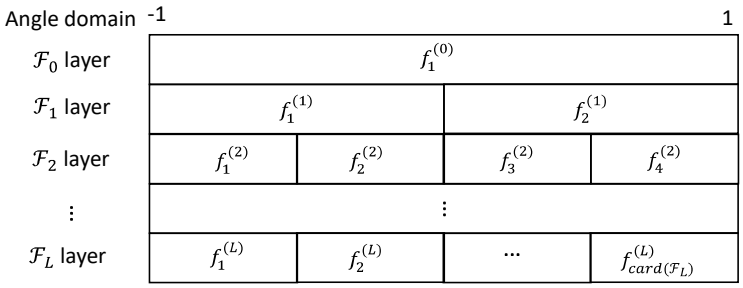

Figure 5.2: Beam coverage of a hierarchical beam structure codebook.

is given as SNR $= \frac{\frac{1}{K}\sum_{k=0}^{K} {r_k}^2}{N_0 W}$ and overall rate measurement R in bits per channel use is log(1+SNR). Thus, the optimal beam-pair for UE-BS during P_1 procedure is selected based on the data rate measurements.

5.2.4 Problem Formulation

We consider an uplink communication between BS and UAV following 3GPP beam access protocol [44]. We formulate the learning-based beam-pair alignment as the partially observable Markov decision process (POMDP) during P_1 and P_2 procedures, and maximize the beamforming gain for any UAV position around the BS coverage area $\mathbb{U}$. We consider received signal strength (RSS) of radio signal and radio beam pair directions (both TX and RX) as the known and unknown parameters of this multi-location environment, respectively.

In this work, we model the interactive RL-based beam-pair alignment problem as the partially observable Markov decision process (POMDP). At any time instant t, we define the parameters $s_t = \{(\text{UE}_h, b_r, b_s, b_u),$
$\text{UE}_h \in \mathbb{U}, b_r \in \mathcal{W}, b_s \in \mathcal{F}_\mathcal{B}, b_u \in \mathcal{F}_\mathcal{N}\}$, $a_t^\mathcal{B} = \{(b_p, b_q), b_p \in \mathcal{W}, b_q \in \mathcal{F}_\mathcal{B}\}$, $a_t^\mathcal{N} = \{(b_m, b_n), b_m \in \mathcal{W}, b_n \in \mathcal{F}_\mathcal{N}\}$ where s_t, $a_t^\mathcal{B}$, $a_t^\mathcal{N}$, r_t are the state, broad beam-pair action, narrow beam-pair action and reward at time instant t. Data rate measurements computed for each applied action are considered as the rewards for the problem. As shown in Figure 5.2, every broad beam codebook direction $f_c^{(l)}$ in $\mathcal{F}_\mathcal{B}$ comprise a finite subset of narrow beam codebook directions $\{f_c^{(l)} - \frac{\psi^\mathcal{B}}{2} \leq f_d^{(L)} \leq f_c^{(l)} + \frac{\psi^\mathcal{B}}{2}, f_d^{(L)} \in \mathcal{F}_\mathcal{N}\}$ with cardinality defined as V. Let π_1 and π_2 denote the broad beam action and narrow beam action policies for state transitions $(s_t, a_t^\mathcal{B}, r_{t+V}, s_{t+V})$ and $(s_t, a_t^\mathcal{N}, r_t, s_{t+1})$, respectively.

After UE's P_1 procedure, BS starts with the random receiving beam direction and then proceeds towards the maximum beamforming gain by applying actions and undergoing state transitions, accordingly. The current applied action becomes part of the next state, undergoing state transition. We define an episode

e_{π_1,π_2} as the consecutive set of such broad beam and narrow beam actions until the terminal state following policies π_1 and π_2, respectively. The objective of this problem for broad beam and narrow beam actions can be formulated as

$$\begin{aligned}
&(P1): \max_{\{\pi_2(a_1^{\mathcal{N}})\}} \sum_{t\le i<\infty} \gamma^{i-t}\mathbb{E}_n\left[r\left(a_{iV}^{\mathcal{N}}\right)\right], \\
&(P2): \max_{\{\pi_2(a_1^{\mathcal{N}})\}} \sum_{t\le i<\infty} \gamma^{i-t}\mathbb{E}_n\left[r\left(a_{i}^{\mathcal{N}}\right)\right], \\
&\text{s.t.} \\
&r\left(a_t^{\mathcal{N}}\right) = \begin{cases} 1 & \text{if } R\left(a_t^{\mathcal{N}}\right) \ge R_{\max}\left(s_t\right) \\ -1 & \text{otherwise} \end{cases}, \\
&\gamma \in (0,1],
\end{aligned} \tag{5.7}$$

where $R_{max}(s_t)$ is the optimal data rate measurement observed among the information history o_t until its previous episode e_{π_1,π_2}, $r(a^{\mathcal{N}_t})$ and $R(a^{\mathcal{N}_t})$ are the rewards and data rate measurement observed on applying action beam-pair $a^{\mathcal{N}_t}$, respectively. We maximize the objective formulation by learning the hierarchical sequence of beam-pair actions starting with broad beam level selection from $\mathcal{F}_{\mathcal{B}}$ and switch to narrow-beam level selection from $\mathcal{F}_{\mathcal{N}}$ following the same reward function (5.9). We consider the hDQN approach to solve this objective problem in (P1).

5.3 Learning Methods Formulation

In this research, we tackle the beam alignment problem at every grid location as a learning problem. Moreover, the performance of RL approach also comparable to that of traditional exhaustive search method. Once RL learning-based methods converged, can significantly reduce the communication overhead during initial access procedure and maximize beamforming in $\mathcal{O}(1)$ time. On the other hand, the traditional approach always results in exhaustive search over entire action space $\mathcal{A}$ each time. The focus of this work is to design an online learning framework that is generic across both location and time, suitable to the considered environment.

5.3.1 The Exhaustive method

The method mainly involves exhaustive search among the set of actions $\mathcal{A}$, to find the best beam pair with maximum possible beamforming between UE and BS. Since we consider the multi-location environment, exhaustive beam scanning is required for every change in grid element unit of UE inside $\mathbb{U}$. This frequent scanning results in significant communication overhead, especially with higher antenna elements. However, this method can determine the best possible beam alignment between BS and UE. If $s_t \in \mathcal{S}$ is the UE state information available at time instant t, then this method can be formulated as

$$\begin{aligned} (P1) : & \max_{(a_t|s_t)} R(s_t, a_t), \\ & \text{s.t. } a_t \in \mathcal{A} \end{aligned} \tag{5.8}$$

where $R(s_t, a_t)$ is the measured data rate on applying a_t to state s_t between BS and UE.

5.3.2 Reinforcement Learning

RL is an interactive learning problem consisting of set of states $\mathcal{S}$, actions $\mathcal{A}$ and rewards, following the markov decision process (MDP) or POMDP process [138]. The state transition is involved on applying each action until the terminal state is reached. The objective of the problem is to learn an optimal policy of state transitions with actions over time and reach the terminal state through reward accumulation [138].

$$\begin{aligned} (P3) : & \max_{\{\pi(a_t|o_t)\}} \sum_{i=t}^{\infty} \gamma^{i-t} \mathbb{E}_{\pi}[r_{a_i}(i)], \\ & s.t. \\ & r_{a_t}(t) = \begin{cases} 1 & \text{if } R(a_t) \geq R_{max}(s_t) \\ -1 & \text{otherwise} \end{cases}, \\ & \gamma \in (0, 1]. \end{aligned} \tag{5.9}$$

In this work, the RL based beam alignment problem is modelled as a POMDP problem. At any time instant t, we define the parameters $s_t = \{(s', a')\ s' \in \mathcal{S}, a' \in \mathcal{A}\}$, $a_t \in \mathcal{A}$ and $r_t \in \mathcal{R}$ where s_t, a_t, r_t are the state, action and reward at time instant t. Here, $\mathcal{S}$ and $\mathcal{A}$ correspond to state and action spaces for scenario. a' corresponds to the set of previous actions applied for state transitions until the

time instant t.

$$\begin{aligned} \mathcal{S} &= \{(\mathrm{UE}_l, b_r, b_s), \mathrm{UE}_l \in \mathbb{U}, b_r \in \mathcal{B}_{\mathrm{TX}}, b_s \in \mathcal{B}_{\mathrm{RX}}\} \\ \mathcal{A} &= \{(b_p, b_q), b_p \in \mathcal{B}_{\mathrm{TX}}, b_q \in \mathcal{B}_{\mathrm{RX}}\}, \end{aligned} \tag{5.10}$$

where UE_l is the location of UE within coverage area $\mathbb{U}$ while $\mathcal{B}_{\mathrm{TX}}$, $\mathcal{B}_{\mathrm{RX}}$ are the beam codebook sets at UE and BS side respectively. $\mathcal{E}_2$ are the state-action space learning environments for RL methods. As shown in (5.10), b_r, b_s are the beam codebook directions corresponding to UE_l previous time instant, following $\mathcal{B}_{\mathrm{TX}}$ and $\mathcal{B}_{\mathrm{RX}}$, respectively. This information is helpful to instantiate the state-transition model at UE_l, required for RL POMDP formulation [138]. Data rate measurements computed on applying each action are considered as the rewards for the problem. We denote $o_t = \{a_{t-1}, s_{t-1}, a_{t-2}, s_{t-2}, ..., a_1, s_1\}$ as the observed history of all such state information and past actions. After the 3GPP initial communication procedure with UE, BS starts with a random receiving beam direction and then proceeds towards the maximum beamforming gain by applying actions and undergoing state transitions, accordingly. The current applied action becomes part of the next state, undergoing state transition. We define an episode e_π as the consecutive set of such actions until the terminal state following a policy π. The objective of this problem can be formulated as mentioned in (5.9), where $R_{max}(s_t)$ is the optimal data rate measurement observed among the information history o_t until its previous episode e_π, γ^{i-t} is the discount factor applied on the rewards received from future actions a_i in the episode, $r_{a_t}(t)$ is the reward and $R(a_t)$ is the data rate measurement observed on applying action beam-pair a_t, respectively. We follow DQN approach to solve this RL objective problem.

5.3.3 Deep Q-Network

Complete steps followed by DQN for RL based beam alignment problem are shown in Algorithm 6.

We define episode as the consecutive set of actions applied on the starting state until it reaches the terminal state with maximum beam alignment for that location. In order to prevent episodes with infinite set of actions during training, we confine maximum episode length to exhaustive set of beam pairs possible under the chosen antenna configuration. For example, the configuration of 8 ULA antenna elements at both TX and RX can result in maximum episode length of 64.

As the RL learning objective formulation involves both current data rate R_t

and best observed data rate $R_{\max}(s_t)$ measurements (shown in (5.9)), we consider the overall online training procedure of DQN framework under two phases namely, Warmup phase and Training phase. During the Warmup phase, the exploration is set to maximum, in order to observe the best possible data rate for the given UE location by applying maximum episode length of actions. During the Training phase, the algorithm continues to reduce its exploration and move towards exploitation following ϵ-greedy policy. The episode starts with initial random action and applies next actions to reach the terminal state as quickly as possible. The Warmup phase results in extra training time at the start but this is later helpful in quick learning of DQN framework during the training phase. This procedure also favours quick convergence of beam alignment process for the current location based on its neighbourhood beam alignment convergence through experience replay memory buffer, thereby leading to overall faster training of DQN framework for multi-location environment.

5.3.4 Hierarchical DQN-based beam allignment

DQN is a value-based approach, learning the optimal approximated policy of states mapping to actions $\pi(s) = a$ by parameterizing and estimating state-action value function $Q(s, a; \theta)$ where θ denotes the weight matrix of the primary deep neural networks (DNN) [108]. The hDQN framework integrates hierarchical action-value functions operating at different temporal scales using DQN approach and learns optimal approximated policies $\pi_1(s) = a, a \in \mathcal{A}_{\mathcal{B}}$ and $\pi_2(s) = a, a \in \mathcal{A}_{\mathcal{N}}$, respectively [89, 74]. Under our hDQN framework, we consider the broad beam (BB) and narrow beam (NB) DQN agents over the same state space $\mathcal{S}$ but different action spaces $\mathcal{A}_{\mathcal{B}}$ and $\mathcal{A}_{\mathcal{N}}$, respectively as shown in Figure 5.3.

For the BB agent, we denote the primary DNN network weight matrix and target DNN network weight matrix as θ_1 and $\overline{\theta_1}$, respectively [108]. We consider the fully connected DNN for both the networks where $\overline{\theta_1}$ is updated with primary network parameters θ_1, after every K_1 iterations. The input of DNN is given by the variables in s_t. The intermediate layers are fully connected linear units with rectifier linear units (ReLU) and the output layer is composed of linear units in a correspondence with $\mathcal{A}_{\mathcal{B}}$. We consider both DNNs with zeros initialization bias and Kaiming normalization weights. The memory buffer of experiences $D_1 = \{e_1, e_2, e_3, \ldots, e_t\}$, $e_i = \left(s_i, a_i^{\mathcal{B}}, r_{i+V}, s_{i+V}\right)$ are collected, where the mini batch of them $U(D_1)$ are randomly sampled and sent into BBDQN [108]. For the NB agent, we follow the same procedure with network weight parameters as $\theta_2, \overline{\theta_2}$, target network updated every K_2 iterations, the output layer mapped to $\mathcal{A}_{\mathcal{N}}$ with

Algorithm 6: RL approach using DQN

$M \rightarrow$ Training Episodes;
Algorithm hyper-parameters: learning rate $\xi \in (0, 1]$, discount rate $\gamma \in [0, 1)$, ϵ-greedy rate $\epsilon \in (0, 1]$, update steps K;
Initialization of replay memory M to capacity C, the primary Q-network with parameters θ_1, the target Q-network with parameters θ_2
$\mathcal{S}, \mathcal{A}$: State and Action space of DQN agent
for *episode* $\leftarrow 1$ *to* M `// for each episode`
do
  Any random UAV transmits the communication request from the (x,y,z) location.
  $N \rightarrow$Episode limit
  BS responds with sequence of N action beam-pairs over the channel with policy π
  Initialization of s_1 by executing a random action a_0 and (x,y,z) location information
  n=0,
  while $True$ **do**
    `// Episode with` ϵ`-greedy policy` π
    if $p_\epsilon < \epsilon$ **then**
      select a random action $a_t \in A$
    else
      select $a_t = argmax_{a \in \mathcal{A}} Q(s_t, a, \theta)$
    end
    BS applies a_t beam-pair over the channel, receive signal for $(t+1)^{th}$ iteration during uplink communication
    UAVobserves s_{t+1}, compute data rate and calculate the reward following (5.9)
    Store transition $e = (s_t, a_t, r_{t+1}, s_{t+1})$ in replay memory D
    Sample random minibatch of transitions $U(D)$
    Compute Loss and Perform gradient descent for $Q(s, a; \theta)$
    Update the target network parameters $\theta_2 = \theta_1$ after every K steps
    $n = n + 1$ `// Increment episode time`
    if *done or* $(n = N)$ **then**
      Update the optimal data rate measurement $R_{\max}(s_t)$
      break `// End episode`
  end
  DQN updates the sequence of action beam-pairs for (x,y,z) location
  `// BS uses the updated sequence on next TX request from (x,y,z) location`
end

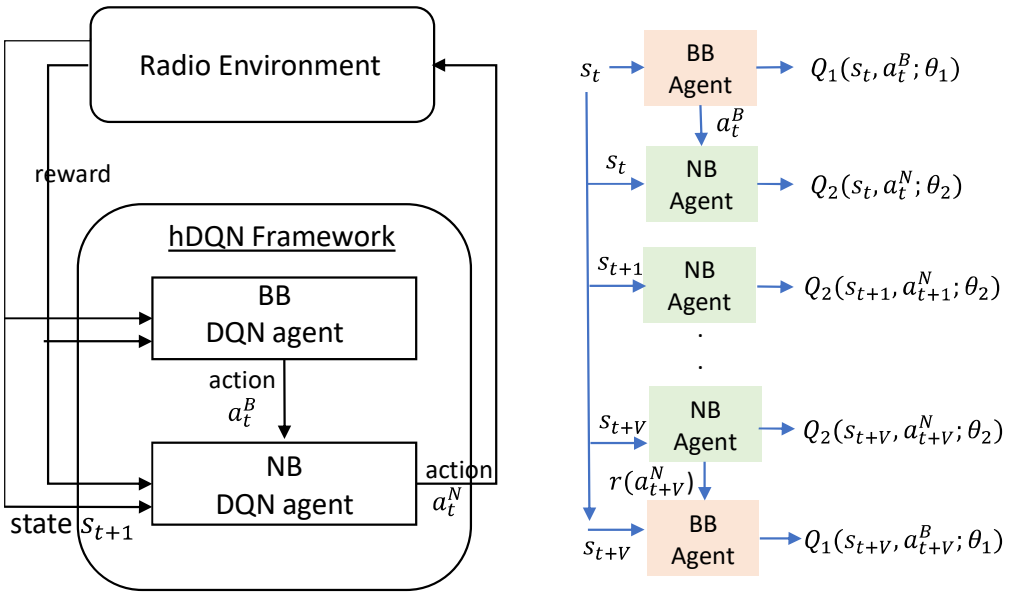

Figure 5.3: hDQN framework with BB and NB DQN agents

disjoint (from D_1) memory buffer of experiences as $D_2 = \{e_1', e_2', e_3', \ldots, e_t'\}$ and collected transitions as $e_i' = \left(s_i, a_i^{\mathcal{N}}, r_{i+1}, s_{i+1}\right)$ respectively.

Let $Q_1(s, a; \theta_1)$ and $Q_2(s, a; \theta_2)$ denote the state-action value functions of BB and NB agents, respectively as shown in Figure 5.3. For both the DQN agents, mean squared error (MSE) loss function is computed between primary, target networks during back propagation, and θ is updated using stocastic gradient descent (SGD) and Adam Optimizer as

$$\theta_{t+1} = \theta_t - \zeta_{\text{Adam}} \nabla \mathcal{L}(\theta_t),$$

where ζ_{Adam} is the learning rate, $\nabla \mathcal{L}(\theta_t)$ is the gradient of the DQN loss function. Here, $(\theta, \nabla\mathcal{L}) = (\theta_1, \nabla\mathcal{L}_1)$ and $(\theta, \nabla\mathcal{L}) = (\theta_2, \nabla\mathcal{L}_2)$ for BB and NB agents, respectively. Thus, we note that hDQN is practical in applying only narrow beams over the channel by using BB agent as the meta-controller. Here, we define episode as the consecutive set of hierarchical actions applied on the starting state until it reaches the terminal state with maximum beamforming gain for that location. In order to prevent episodes with infinite set of actions during training, we confine maximum episode length to exhaustive set of beam pairs under the chosen UPA configuration. Hence, in the proposed hDQN, the maximum episode lengths are card $(\mathcal{A}_{\mathcal{B}})$ (say $K_{\mathcal{B}}$) and V for the BB and NB agents, respectively.

Algorithm 7: Hierarchical DRL using DQN

$M \rightarrow$ Training Episodes; Algorithm hyper-parameters: BB learning rate $\zeta_1 \in (0,1]$, BB ϵ-greedy rate $\epsilon_1 \in (0,1]$, BB episode limit $K_{\mathcal{B}}$, NB learning rate $\zeta_2 \in (0,1]$, NB ϵ-greedy rate $\epsilon_2 \in (0,1]$, NB episode limit V;
Initialization of replay memory D_1 to capacity C_1, D_2 to capacity C_2, BB network parameters $\theta_1, \overline{\theta_1}$ and NB network parameters $\theta_2, \overline{\theta_2}$;
$\mathcal{S}$: State space of BB, NB agent;
$\mathcal{A}_{\mathcal{B}}, \mathcal{A}_{\mathcal{N}}$: Action space of BB and NB agent, respectively;
for *episode* $\leftarrow 1$ *to* M `// for each episode`
do
    Any random UAV transmits the communication request from the (x,y,z) location;
    BS responds with a sequence of $VK_{\mathcal{B}}$ action beam-pairs over the channel with π_1, π_2 policies;
    Initialization of s_0 by executing a random action $a_0^{\mathcal{B}}$, $a_0^{\mathcal{N}}$ and (x,y,z) location information;
    $k = 0$
    while $True$ **do**
        if *done or* $(k = K_{\mathcal{B}})$ **then**
            `// End Training episode`
            Reset Env and obtain new s_0
        select $a_t^{\mathcal{B}}$ from BB network following ϵ_1
        BS selects the NB action subset $\mathcal{P}$ ($|\mathcal{P}| \leq V$) corresponds to $a_t^{\mathcal{B}}$;
        $p = 0$;
        for $p \leq V$ **do**
            if *done or* $(p = V)$ **then**
                `// End NB episode`
                Update $R_{\max}^{\mathcal{N}}(s_{t+p})$;
            if *warmup* **then**
                Randomly select $a_t^{\mathcal{N}} \in \mathcal{A}_{\mathcal{N}}$
            else
                select $a_t^{\mathcal{N}}$ from NB network following ϵ_2
            BS applies $a_t^{\mathcal{N}}$ over the channel, receive signal for $(t+1)^{th}$ episode during uplink communication;
            UE observes s_{t+1} and calculate the reward $r(a_t^{\mathcal{N}})$;
            Store the experience $(s_t, a_t^{\mathcal{N}}, r_t, s_{t+1})$ to D_2;
            Train and update NB parameters θ_2;
        Store the experience $(s_t, a_t^{\mathcal{B}}, r(a_{t+p}^{\mathcal{N}}), s_{t+p})$ to D_1;
        Train and update BB parameters θ_1;
    $k = k + 1$ `// Increment episode time`
    hDQN updates the sequence of action beam-pairs for (x,y,z) location;

We consider the overall hDQN training procedure into Warmup and Training phases, similar to our prior work in [136]. As the reward formulation in (5.9) involve computing $R_{\max}(s_t)$ measurements over $\mathcal{A}_{\mathcal{N}}$, we consider the warmup phase only for NB agent of hDQN. During Training phase, the hDQN perform exploration and exploitation using ϵ-greedy policies ϵ_1 and ϵ_2 for BB and NB agents, respectively. The Warmup phase results in extra training time at the start but favours quick convergence of hDQN during training phase resulting in faster beam-alignment training for the multi-location environment.

5.3.5 Convolution neural network radio mapping (CRM)

To solve the convergence and optimize problem in (5.9), we propose the learning architecture based on convolution neural network (CNN), and radio mapping namely as CRM, as shown in Figure 5.4. As beam alignment is the key problem to maintaining the quality of backhaul link between UAV and BS under mmWave communications. For example, a major challenge for achieving high 3D beamforming gain in UAV communication is to effectively track the channel variation arising from UAV's high mobility and thereby obtain accurate channel state information (CSI) at BS. CNNs offer the ability to learn complex patterns and relationships from data, which can be leveraged to improve the spatial correlation of the channel state information (CSI) between the transmitter and receiver, such as beamforming alignment in wireless communication systems. Beamforming alignment often requires accurate estimation CNNs can be utilized to estimate the channel characteristics by learning the complex relationships between transmitted signals and received signals. By training the CNN with a large dataset of known/unknown channel conditions, it can learn to extract features from the received signals and estimate the channel parameters necessary for beamforming alignment. The authors in [109] use radio map-based channel propagation environment information for efficient positioning of UAV. Radio map can be helpful in improving learning and real-time update when there are sufficient changes in the radio source location (e.g., moving beyond the decorrelation distance power and others [26]).

In such scenarios, radio maps that describe average outage channel probability, channel gains or SINR [168, 26, 179, 52, 109] help to extract such features needed to enhance the information needed to reduce training overhead and improve converges time. The image of radio mapping is captured and feed to CNN algorithm. The radio mapping process will elaborate in detail in Section 5.4. The CRM will gather the data features to feed the system with the current and previ-

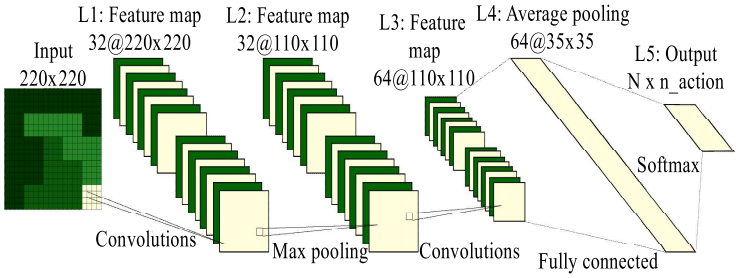

Figure 5.4: Proposed CNN to classify the channel status and strength

ous features of the environment. Once the features are captured, the hDQN will learn faster and fasten the convergence process. The different locations' channel strength are labelled with different colours, and the features are mapped into the 3D image in each timeslot. The CNN is a multi-layer network evolved from the traditional neural network. The CNN mainly includes an input layer, several convolution layers, several pooling layers, a fully-connected layer, and an output layer. It is used for feature extraction and mapping through fast training and possesses high classification and prediction accuracy. We assume that the proposed CNN model consists of one data input layer, N_c convolution layers, N_p pooling layers, N_f fully-connected layers, and N output layer. It is assumed that the size of the input image is 220×220. The detailed description of each layer is introduced as follows:

(a) Data Input Layer: The BS location, the UAV users with different locations and radius, and signal stregth denoted by different colors in the 3D image with the groundtruth label of the best beam-pair allignment. The preprocessed images are used as input data of the convolution network. All images are projected into a characteristic subspace of $N_0 \times N_0 \times 3$, where 3 presents the color of the image is in RGB model.

(b) Convolution Layer: The characteristics of the input image are extracted by the randomly initialized filter. It is possible that the input image has various characteristics, thus, multiple filters are used to extract all features of the original image. Zero padding is also used for each convolution layer to keep the size of features extracted from the input image.

(c) Pooling Layer: It plays an important role in sub-sampling via using features extracted from the convolution layers. The time complexity is decreased in the next convolution layer or fully-connected layer by reducing the number of operations in sub-sampling. The Max-pooling method is usually deployed to extract the largest value in the sliding window for the subsampling among all methods used in the pooling layer.

(d) Fully-connected Layer: The features extracted by the convolution and pooling layers are inserted into the neural network. The softmax layer that is often used for the classification of multiple classes is employed at the end of the fully-connected layer. The classification result corresponds to a probability that the overall probabilities of all classes is equal to 1, and the class with the highest probability is the estimated label for the corresponding input image.

5.4 Radio map in radio networks

In this section, the spatial features are used to show the coverage area of BS to maximize the transmission reliability. Radio mapping could solve the entire coverage over the large and continuous geographical area [26]. To maximize in long-term spectral activity, the radio map is useful to improve the learning and update when there are sufficient changes in the radio source location (e.g., moving beyond the decorrelation distance power and others).

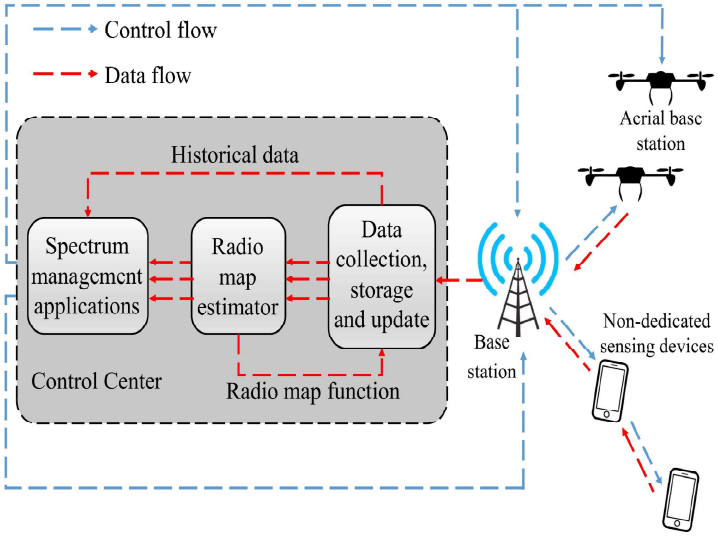

Figure 5.5: Radio map network

As shown in Fig. 5.5, the use of the radio map in wireless networks consists of three major steps: (i) measurement collection and processing; (ii) radio map construction and update; and (iii) radio map-assisted BS / UAV management. Specifically, the system operator first collects and filters the distributed measurements, which are then used by the estimator to compute the radio map. The constructed radio map is then used to derive useful knowledge about the spectrum usage pattern and essential parameters in the network, for example, wireless device location, interference level, and channel models. The radio map for each

cth BS refers to the spatial distribution of its large-scale channel gain over the 3D space, i.e., height's with UAV at locations $u \in \mathbb{R}^3$. As the space is infinite and continuous, it is not feasible to store the entire data $c_b \in \mathbb{R}^3$ for all locations of u, due to the finite storage in practice [179]. Therefore, we propose to discretize the space into the 3D grid follow in Figure 5.1 and 5.6 where ΔD is chosen to be sufficiently describe the channel gain as being approximately constant within each grid cell.

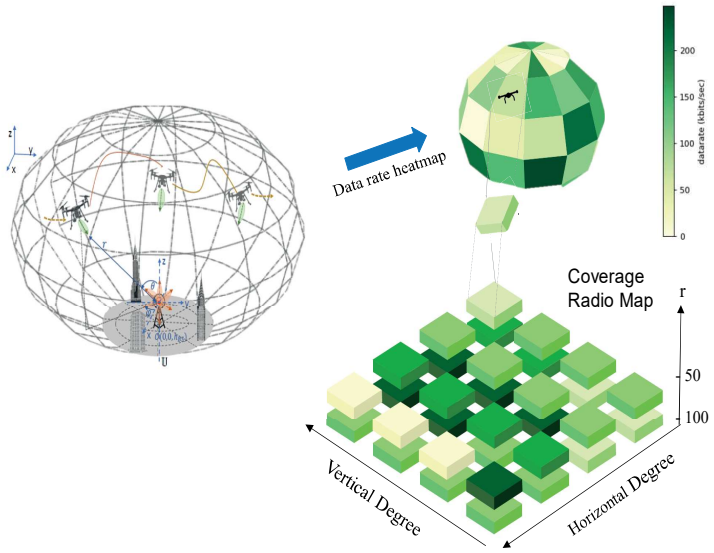

Figure 5.6: Illustration of scenario and 3D radio map projection

5.4.1 3D and 2D map projection

As shown in Figure 5.6a, we consider the large spherical environment area with the BS at the center. The spatial information can be captured from the geographical signal power spectrum density, which is formed by the superposition of concurrent wireless transmissions in different locations, frequencies, and times. The spatial features are different in each time, therefore we decided to use the rate maps heatmap using ranges of data rates in each vertical and horizontal degree to illustrate the information. The measurements are computed from the DQN/ hDQN simulations at the end of each episode. The data are collected via offline or online manner and will create 3D data rate radio map and update the received data of coverage area as in Figure5.6b. The 3D illustration is important to ensure all the information is well captured to maximize the usage of spatial features. So we can get 360° of the current network once UAVs fly to target allocation. However, in the real situation, the receive data is like a small hexag-

onal, and cell size cannot fix in the big sphere or hexagonal which will cause in accuracy during the data measurement [110], therefore to improve the accuracy by model the projection of angular angle and illustrate as 3D grid radio map projection. This helps radio mapping in capturing spatial information in a real scenario and online manner. Then, as shown in Figure 5.6b, once we received the updated radio mapping following the yellow-green heat colour map. To ease the overall view, we create 2D rectangular projections as shown in Figure 5.6c, from which we borrow the concept in 3D video tiles projection in [73] to maximize the beam-forming gain. Then, we arrange based on the horizontal degree and the vertical degree, which are the angle and elevation angle, respectively. When the UAV is moving to the different radius, the radio map will capture and fill the tile with the current information and update the layer in z-axis as shown in Figure 5.6c. 3D coverage area will automatically update by the end of the episode.

5.4.2 Offline

A high-level overview of offline and online CRM for UAV radio network is illustrated in Figure 5.7. The key difference between CRM and a standard RL setup is the utilization of a radio map in CRM, which provides precomputed information about the environment's wireless characteristics. This helps the hDQN in estimating initial beam-pair features for all locations during early episodes, facilitating more informed decision-making and efficient learning in the context of cellular-connected UAVs and beamforming. In the offline phase, the CNN model uses the radio map ground truth data set that we received from the offline dataset based on the Vanilla hDQN dataset, which initially feeds into the database to train, learn and predict the initial radio beam pair alignment, a_t^B. Then it will be used to predict the initial pair in every initial episode. The flow is shown in Figure 5.7.

5.4.3 Online

Given the limitations in practical implementation, obtaining the updated Vanilla hDQN dataset in an offline manner is not feasible, as it can only be collected after the training process has been completed. Therefore, we propose an online configuration hDQN. In each episode, hDQN will revise an online update and generate the radio map image to feed into the CNN database. The CNN will train the image dataset in each episode from current hDQN training, and select the

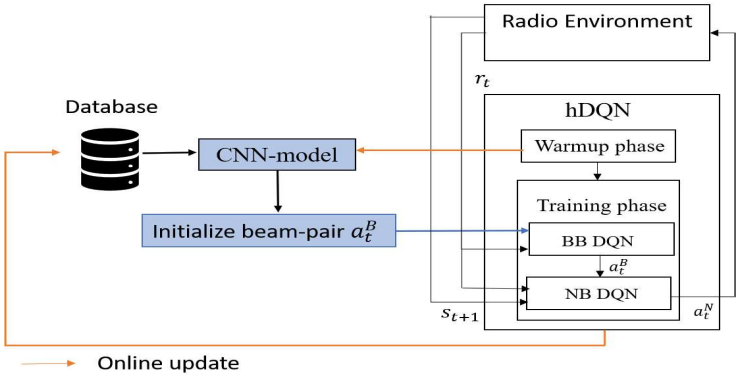

Figure 5.7: Flow methodology of Radio map in hDQN network

best model and predict the best acurate initial beam-pair (action, a_t^B). Initially, the image datasets are only collect from the warm-up phase, as the algorithm still exploring in order to exploit the best possible data rate for the given UE location by applying maximum episode length of actions. When the hDQN explore the environment, the dataset with the current new data is collected, and the radio map is generated in each episode to feed in real-time into CNN framework. While in training the CNN model, we train the CNN with the previous and current dataset, and this we called an online training manner. The benefit of online model is to reduce the probability of misselect an action, a^B and able to find the correct beam-pair, and also reduce the time for NB DQN to predict a^N in maximizing the overall beamforming gain. However, with additional training in every episode, the online CNN in hDQN will cause the additional time to converge compared to the offline manner.

However, since the wireless channel conditions between the UAV and the base station can change rapidly due to the UAV's movement, environmental factors, and interference. Fast beam tracking and real-time adjustment are required to compensate for UAV motion and maintain a stable and uninterrupted backhaul link. Continuous and accurate beam alignment is necessary to adapt in an online (real-time) algorithm in adapting to these changing channel conditions and maintain a stable and high-quality backhaul link. Another benefit of the online-hDRM approach is able to monitor the UAV-beam pair alignment in real time because it capture real scenario and able to maintain the network connectivity with enough training time. Since the training, i.e., command and control is conducted at the terrestrial-based station, it will be the drawback to the offline-hDRM approach, due to the algorithm needs to have the complete data set before the training will be started, which is not practical in control and command for dynamic scenarios.

Table 5.1: Simulation Parameter

Parameters	Value
mmWave freq	30GHz
Bandwidth W	120MHz
antenna element spacing d	0.5
Radius r_h	$50, 100, 150m$
Transmit power P_{tx}	0dB
Transmit antenna elements N_{tx}	$\{(2 \times 2), (4 \times 4)\}$
Receiving antenna elements N_{rx}	$\{(4 \times 4), (8 \times 8)\}$
Noise Level N_0	-174 dBm/Hz
BS location (in m)	$[0, 0, 25]$
UMa-nLoS pathloss coefficients	$\alpha : 4.6 - 0.7 \log_{10}(\mathbb{U}_{\text{zloc}})$, $\beta : -17.5, \gamma : 2.0, \sigma : 6.0$ $\kappa : 20 \log 10\left(\frac{40\pi}{3}\right)$ [11]

5.5 Simulation Results

As described in Section 5.2-C and Section 5.3, we implement the hDQN-based beam-pair alignment, following P1 and Algorithm 6. Similarly, we implement the state of-the-art DQN-based approach [136] over UPA configuration and compare our results. We note here that both the RL-based methods, once converged, can significantly reduce the communication overhead during P_2 procedure [44] and maximize the beam-forming gain in $\mathcal{O}(1)$ time. In this section, we investigate the training performance of our proposed hDQN-based training procedures over different UPA configurations and UMa -nLoS channel conditions. We select 13 random reflection points within BS coverage and fix them throughout the hDQN and DQN nLoS simulations. For simplicity, we consider UAV hovers in $\mathbb{U}$ with fixed radial distance from BS $r = 20$ m. The simulation conditions for all the numerical results are listed in Table 5.1.

5.5.1 hDQN vs DQN Training Performance

As shown in Figure. 5.8, the red plot shows the DQN overall reward performance under the UMa-LoS channel while the green and blue plots depict the rewards (following (5.7)) over hDQN overall training time under UMa-LoS and UMa-nLoS conditions, respectively. We note that all the simulations are carried out

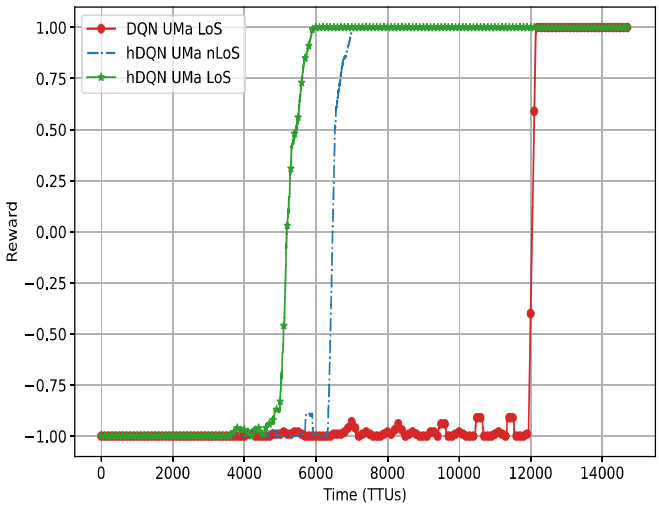

Figure 5.8: hDQN, DQN training convergence for $(N_{\mathrm{TX}}, BN_{\mathrm{RX}}, N_{\mathrm{RX}}) = (2 \times 2, 4 \times 4, 8 \times 8)$ UPA configuration.

with $(N_{\mathrm{TX}}, N_{\mathrm{RX}}) = (2 \times 2, 8 \times 8)$ and $(l_x, l_y) = (2, 2)$ i.e. $BN_{\mathrm{RX}} = 4 \times 4$UPA is selected under UMa channel with thermal noise but no shadow fading and channel variation conditions. DQN simulations are performed over $A_{\mathcal{N}}$ action space while the hDQN method apply broad $(\mathcal{A}_{\mathcal{B}})$ and narrow beams $(\mathcal{A}_{\mathcal{N}})$ using BB and NB networks, respectively. We observe that under both UMa -LoS and UMa -nLoS conditions, the hDQN training procedure attains the maximum reward with significantly less training time compared to the DQN method, resulting in faster training convergence.

5.5.2 hDQN For Different UPA Configurations

In this subsection, we plot the training times (TTUs) and maximum achievable data rates (5.7) of hDQN and DQN-based approaches under different UPA antenna configurations with UMa-nLoS conditions. As shown in Figure 5.9, blue and red bars shown the training times of hDQN and DQN respectively, while the black plot depict maximum learnt data rates obtained in both methods. We note that the same DNN architecture and hyperparameters values are used for all the hDQN simulations. We observe that hDQN converges faster than DQN and achieves the maximum data rate with average reduction in training overhead of 43% among all UPA simulations. Under the same $(N_{\mathrm{TX}}, N_{\mathrm{RX}})$ configuration, we observe that selection of higher BN_{RX} increases the reliability of providing maximum achievable rate across narrow grid element area in $\mathbb{U}$. This also impacts the

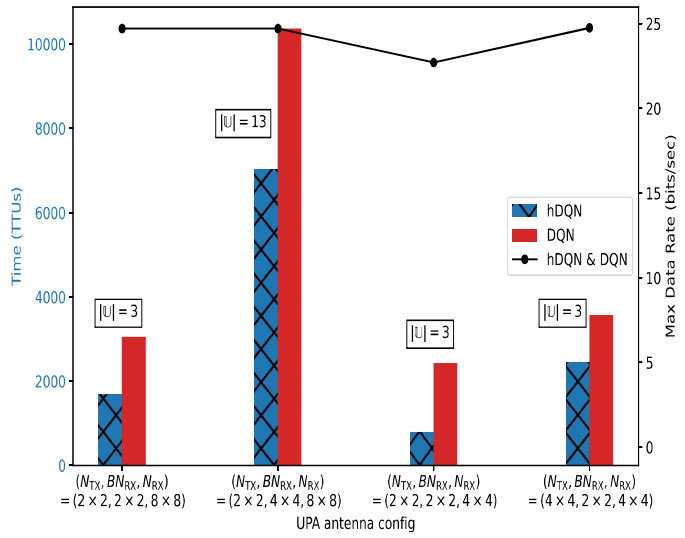

Figure 5.9: hDQN overall training performance under UMa nLoS conditions.

training time due to the increase in state space S for DQN, both S and $\mathcal{A}_{\mathcal{B}}$ in hDQN. However, we notice that hDQN converges faster as the selection of broad beam actions depends on both ϵ_1 policy and the convergence of NB network. Now, increasing the BN_{RX}, decreases the cardinality (V) of narrow beam subset for each broad beam pair in $\mathcal{A}_{\mathcal{B}}$, resulting in faster overall convergence. Thus, the observed results show that broad beam level selection is crucial and has more impact on both training and rate performance under the hDQN approach. This can be useful to trade-off rate and training performances over broad beam level selections for different cellular UAV applications.

5.5.3 hDQN with increasing coverage area

In this subsection, we study the training performance of hDQN with the increase in coverage area requirement under UMa channel condition. Figure 5.10 shows the reward performance of hDQN over time across with difference BS radial distance as in Eq. (5.3). The radius is compared to measure the performance of UAV of it converging time when the distance between BS and UAV differs. We note that the accumulated reward plots are shown against the number of TTUs for better analysis of its convergence. With the increase in radial of the BS, more grid elements are needed to hover around and support its radio link. It is observed that the hDQN-based approach converges well with different coverage area requirements. The learning is observed to be relatively quicker in convergence with increase in the coverage area of BS under $r = 20m$. However, when

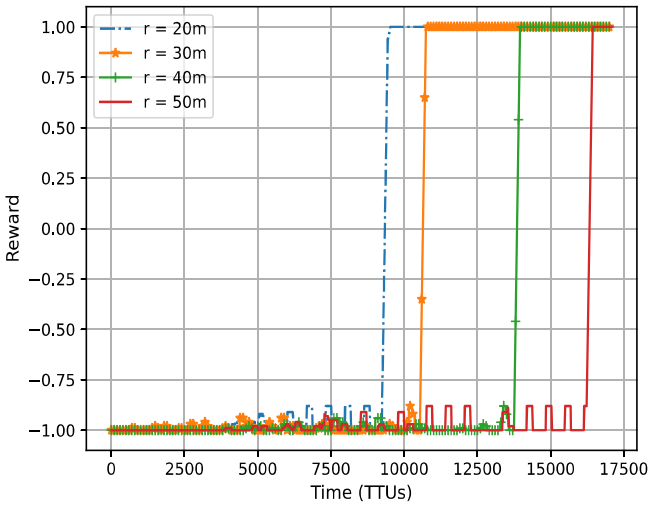

Figure 5.10: hDQN different radius under UMa nLoS conditions.

$r = 30m, 40m, 50m$, it required more exploration, therefore, the chosen $r = 20m$ is suitable with the grid element that we set at the initial state with help to fasten exploit and converge process.

5.5.4 DRM performance

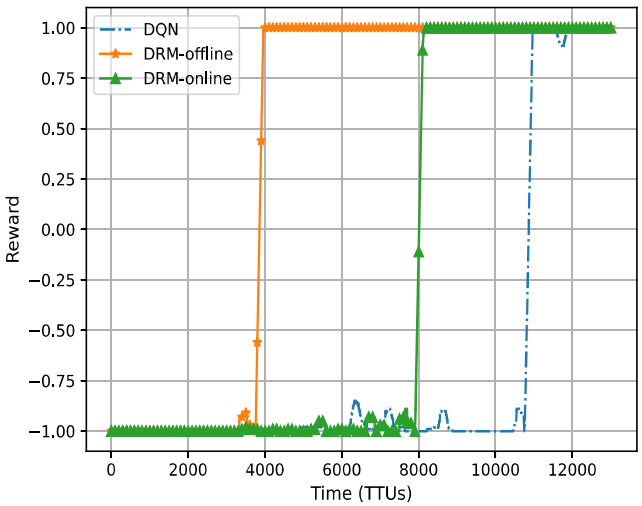

Figure 5.11: Comparison between CRM-DQN and Vanilla DQN.

In this subsection, firstly, we measure the performance of CNN in accuracy and model loss before we deploy the model into CRM and DQN-based framework,

namely as DRM. The accuracy performance and loss function is evaluated to ensure the selected model is accurate to help DQN to predict the initial action. This process is important in CRM algorithm as it will future direct the direction of exploration and exploitation of the algorithm. Figure 5.11 plot the average reward of CRM-DQN and DQN-based approaches under UMa-nLoS conditions. For instant, blue line is Vanilla DQN namely as 'DQN', orange line is DQN-CRM offline namely as 'DRM-offline' and green line is DQN-CRM online namely as 'DRM-online'. We note that the same DNN architecture and hyper-parameters values are used for all the DQN and CRM simulations. We observe that DRM converges 60% and 20% faster than DQN in UPA simulations for offline and online, respectively. The DRM-offline shows better performance compared to DRM-online, this is due to the DRM-offline algorithm using the previous Vanillla-DQN dataset to choose the initial action for the training phase, this can help shorten the process of exploration of DQN, as previous training is already converged. However, in the real situation, DRM-online is more practical because the DRM-online is using the real-time data to generate radio-map to feed to CNN, so it can be able to get the fresh dataset for initial beam-pair prediction and use in DQN algorithm.

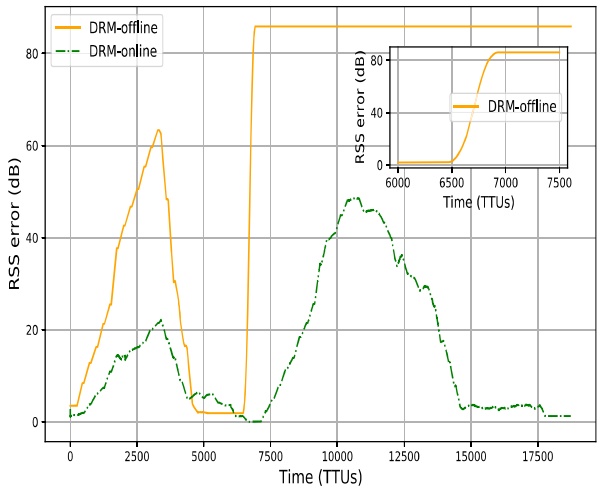

Figure 5.12: RSS Error plot for offline and online DRM.

Figure 5.12 plot the RSS errors of the agent with respect to the exhaustive approach, respectively. Average RSS error is defined as the mean difference in RSS values of a proposed DRM approach with respect to an exhaustive approach (measured in dB scale) over UE locations. This metric helps us estimate the accuracy of learning a beam-pair in the proposed DRM approach with respect to the traditional exhaustive approach, at every time instant during the training procedure.

In this environment, we are plotting different channel condition UMA-nLoS to UMA-LOS. DRM training performance in real-time conditions in an online manner by considering change in channel conditions, thermal noise, slow fading and slow channel variation as shown Fig. 5.12. We observed that the DRM-offline received higher RSS errors compared to DRM-online. However, DRM-offline converges faster when the environment changes, this is because DRM-offline used the previous model and dataset, and also convolutional neural network (CNN) is able to supervise learning to choose the initial action, which leads to minimum error compared to online-based. Even though, DRM-online takes a longer time to adapt to new environments, it learns faster and received lower errors, as it continues to learn during the training process. Therefore, it strengthens our hypothesis, the DRM-online is more practical and able to adapt to different environments because the DRM-online is using real-time data to generate radio-map to feed to CNN, and able to get a fresh dataset for initial beam-pair prediction and use in DQN algorithm.

5.5.5 hDRM performance

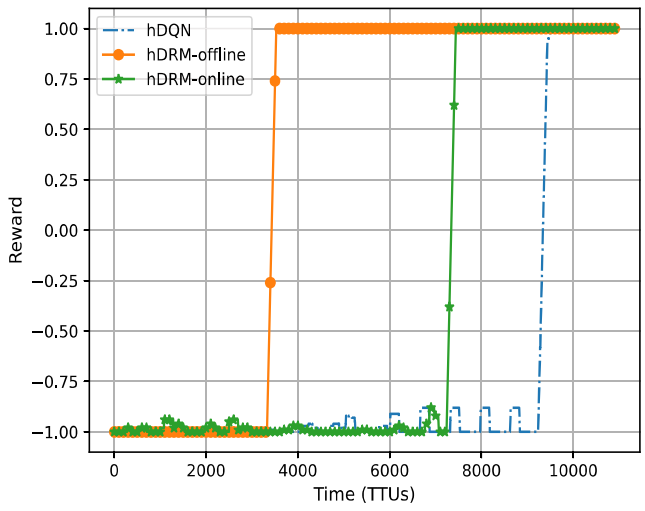

Figure 5.13: Comparison between CRM-hDQN and Vanilla hDQN.

Figure 5.13 plots the TTU and average reward of vanilla hDQN-based approaches and CRM and hDQN-based framework, namely as hDRM. The simulation is simulated under the same UPA antenna configurations with UMa-nLoS conditions. For instant, the blue line is Vanilla hDQN namely as 'hDQN', and green line and purple line are integration between hDQN and radio mapping us-

ing CRM in an offline and online manner, namely as 'hDRM-offline' and 'hDRM-online', respectively. The random location is generated and based on the location, we generate the radio map and predict the initial beam pair using CRM and feed it to vanilla hDQN. Based on the CRM database, we replace the initial random selection with a^B which generate from CNN and CRM algorithm, which are able to reduce the exploration of hDQN and reduce the convergence time. It clearly shows hDRM outperform and successfully reduce 63% of convergence time compared to vanilla hDQN. The radio map of channel strength and its features helps the hDRM to reduce the convergence in the training phase, the predicted initial beam-pair was fed to BB-DQN as shown on in Figure 5.7. It helps the process of NB-DQN to have the correct group to predict action in NB DQN. As it also can reduce the hDQN complexity and helps NB-DQN to exploit the beam-pair action faster.

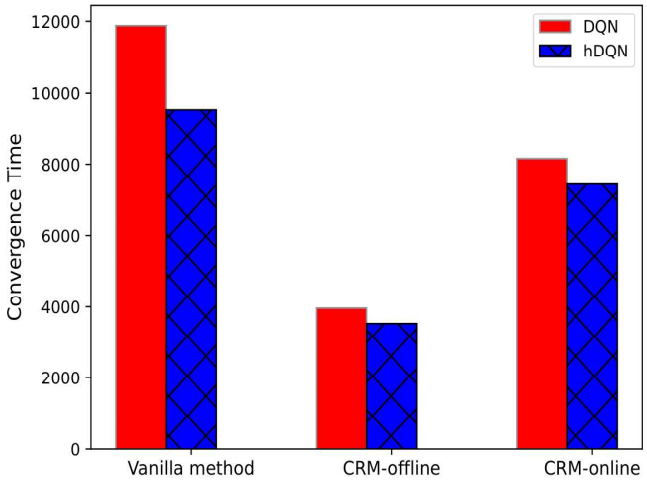

Figure 5.14: Comparison between DQN, hDQN, CRM-DQN, CRM-hDQN.

Finally, the performance of standard vanilla DQN, Vanilla hDQN, and DRM, hDRM for both offline and online manner are present in Figure 5.14. The 'standard' represent Vanilla DQN and hDQN with red bar and blue bars, respectively. While DRM is represented by red bars and hDRM with blue bars for both CRM-offline and CRM-online. From the bar-plot, it showed, hDRM-offline have the best performance compared to other algorithms. However, there is a problem for the hDRM-offline that used data from vanilla hDQN, and consider not that practical because the prior data collected after vanilla hDQN finished train, whereas even the hDRM-online is the second-best algorithm, but it wins in practicality. Even we can be observed hDRM-offline faster compared to online-based approaches.

Therefore, it became the tradeoff between practicality and convergent time.

5.6 Conclusion

In this research, we proposed the hDQN-based position-aided beam alignment framework for cellular-connected mmWave UAVs and maximize the beam-forming gain within the BS coverage area in an online manner. We also analyzed the hDQN approach over state-of-the-art DQN-based method under different UPA antenna configurations and diverse channel conditions. Our results showed that the proposed hDQN converges faster than the DQN method with an average overall training reduction of 43% for UPA configurations. 360° spatial-information-beam-mapping are proposed to give additional information to hDQN and help improved the convergence time. The CRM and DQN-based framework showed 60% improve over the convergence time compared to the proposed framework against vanilla-method approaches.

Chapter 6

Summary and Concluding Remarks

In this chapter, the main contributions and the future research directions of this thesis are summarized and presented as below.

6.1 Conclusion

In conclusion, this study has highlighted the significant contributions and advancements made in the field of UAV-based cellular networks for urban and suburban (firefighting) scenarios. By optimizing coverage and capacity, together with mitigating interference and improving beam alignment techniques, the study has demonstrated the potential of UAVs to enhance communication and data transmission in dynamic environments.

The study's outcomes showcase improved coverage and capacity through the strategic deployment of multiple UAVs in coordination with base stations. This optimization enables better communication rates and ensures smooth video streaming, ultimately enhanced the Quality of Experience (QoE) for UAV users.

Furthermore, the study addresses the challenge of interference between UAVs and terrestrial users, offering interference coordination mechanisms and resource allocation schemes. These innovations effectively mitigate inter-cell interference, resulting in significant throughput improvements for both UAVs and terrestrial users.

Additionally, the study introduces deep reinforcement learning-based beam alignment techniques, combining hierarchical Deep Q-Networks and convolutional neural networks. These techniques enhanced the efficiency and reliability of UAV to BS communication, reducing convergence time and improving overall perfor-

mance in mmWave radio settings.

Overall, this study findings has contributed to the advancement of UAV-based wireless networks, offering novel solutions for optimizing coverage, mitigating interference, and improving beam alignment techniques.

6.2 Future Research

Based on the findings and advancements made in this study, there are several promising areas for future research in UAV-based cellular networks.

One important aspect to consider is the optimization of battery life usage in UAVs especially in firefighting operations. UAVs typically have limited battery capacity, which can impact their operational time and effectiveness in extended firefighting missions. Future research can focus on developing energy-efficient algorithms and techniques to optimize UAV battery usage. This can include intelligent path planning, dynamic power management, and energy harvesting methods to prolong the UAV's operational time and ensure sustained communication and surveillance capabilities during firefighting operations.

For the development 6G networks in future, which are envisioned to provide ultra-high data rates, ultra-low latency, and massive connectivity, the backhaul plays a crucial role in enabling reliable and high-capacity communication between UAVs and base stations, and interference management remains a critical challenge in UAV-based cellular networks. Future research can explore advanced interference management techniques and scheduling algorithms to mitigate interference and improve overall network performance. This can involve dynamic spectrum allocation, interference coordination mechanisms, and intelligent scheduling algorithms that consider the dynamic nature of UAV movements. These advancements would enhance the coexistence of UAVs and terrestrial users while maximizing the efficiency of wireless communication in ad-hoc and high-dense scenarios.

Future research could focus on addressing the challenges associated with backhaul in beam pair alignment for 6G. This can involve the development of efficient backhaul architectures, protocols, and optimization algorithms to establish reliable and high-capacity connections between UAVs and base stations. The scalable backhaul solutions need to be designed to handle the increased traffic and connectivity demands of UAVs in a 6G environment.

By exploring these future research directions, researchers can further advance the field of UAV-based cellular networks for urban, suburban, and ad-hoc sce-

narios such as firefighting applications. The outcomes of such research would contribute to the development of more efficient and effective communication systems, extending the capabilities of UAVs and ultimately enhancing the safety and effectiveness of firefighting teams.

References

[1] Recommended upload encoding settings - youtube help.

[2] Technical specification group (TSG) RAN WG4; RF system scenarios. TR 25.942, 3GPP, Jun. 2001. V15.0.0.

[3] Drones: how to fly them safely and legally, Sep 2017.

[4] Study on enhanced lte support for aerial vehicles. TR 36.777, 3GPP, Dec. 2017. V15.0.0.

[5] Study on channel model for frequencies from 0.5 to 100 GHz. TR 38.901, 3GPP, Jun. 2018. V15.0.0.

[6] Study on Enhanced LTE support for Aerial Vehicles. Technical Report 36.777, version 15.0.0, Jan, 2018.

[7] Mohamed A Abd-Elmagid, Aidin Ferdowsi, Harpreet S Dhillon, and Walid Saad. Deep reinforcement learning for minimizing age-of-information in uav-assisted networks. *arXiv preprint arXiv:1905.02993*, 2019.

[8] Aly Sabri Abdalla, Keith Powell, Vuk Marojevic, and Giovanni Geraci. UAV-assisted attack prevention, detection, and recovery of 5G networks. *IEEE Wirel. Commun.*, 27(4):40–47, 2020.

[9] Evolved Universal Terrestrial Radio Access. Evolved universal terrestrial radio access (e-utra); physical layer; measurements. *Measurements*, 2018.

[10] Fahad Taha AL-Dhief, Naseer Sabri, S. Fouad, N.M. Abdul Latiff, and Musatafa Abbas Abbood Albader. A review of forest fire surveillance technologies: Mobile ad-hoc network routing protocols perspective. *Journal of King Saud University - Computer and Information Sciences*, 31(2):135–146, 2019.

[11] Akram Al-Hourani, Sithamparanathan Kandeepan, and Abbas Jamalipour. Modeling air-to-ground path loss for low altitude platforms in urban environments. In *2014 IEEE global communications conference*, pages 2898–2904. IEEE, 2014.

[12] Akram Al-Hourani, Sithamparanathan Kandeepan, and Simon Lardner. Optimal lap altitude for maximum coverage. *IEEE Wireless Communications Letters*, 3(6):569–572, 2014.

[13] Akram Al-Hourani, Sithamparanathan Kandeepan, and Simon Lardner. Optimal LAP altitude for maximum coverage. *IEEE Commun. Lett.*, 3(6):569–572, Dec. 2014.

[14] Mohamed Alzenad, Amr El-Keyi, Faraj Lagum, and Halim Yanikomeroglu. 3-d placement of an unmanned aerial vehicle base station (uav-bs) for energy-efficient maximal coverage. *IEEE Wireless Communications Letters*, 6(4):434–437, 2017.

[15] Rafhael Amorim et al. Radio channel modeling for UAV communication over cellular networks. *IEEE Wirel. Commun. Lett.*, 6(4):514–517, 2017.

[16] Jianping An, Kai Yang, Jinsong Wu, Neng Ye, Song Guo, and Zhifang Liao. Achieving sustainable ultra-dense heterogeneous networks for 5g. *IEEE Communications Magazine*, 55(12):84–90, 2017.

[17] Muhammad Yeasir Arafat, Muhammad Morshed Alam, and Sangman Moh. Vision-Based Navigation Techniques for Unmanned Aerial Vehicles: Review and Challenges. *Drones*, 7(2):89, jan 27 2023.

[18] Juan C Aviles and Ammar Kouki. Position-aided mm-wave beam training under nlos conditions. *IEEE Access*, 4:8703–8714, 2016.

[19] Irmak Aykin, Berk Akgun, Mingjie Feng, and Marwan Krunz. Mamba: A multi-armed bandit framework for beam tracking in millimeter-wave systems. In *Proc. of the IEEE INFOCOM 2020 Conference, Toronto, Canada*, July, 2020.

[20] Amin Azari, Mustafa Ozger, and Cicek Cavdar. Risk-aware resource allocation for URLLC: Challenges and strategies with machine learning. *IEEE Commun. Mag.*, 57(3):42–48, 2019.

[21] M Mahdi Azari, Giovanni Geraci, Adrian Garcia-Rodriguez, and Sofie Pollin. Cellular UAV-to-UAV communications. In *Proc. IEEE 30th Annu. Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, pages 120–127, Sep. 2019.

[22] M Mahdi Azari, Giovanni Geraci, Adrian Garcia-Rodriguez, and Sofie Pollin. UAV-to-UAV communications in cellular networks. *IEEE Trans. on Wireless Commun.*, 19(9):6130–6144, Jun. 2020.

[23] Mohammad Mahdi Azari, Fernando Rosas, Alessandro Chiumento, and Sofie Pollin. Coexistence of terrestrial and aerial users in cellular networks. In *2017 IEEE Globecom Workshops (GC Wkshps)*, pages 1–6. IEEE, 2017.

[24] Constantine A Balanis. *Antenna theory: analysis and design*. John wiley & sons, 2015.

[25] Shuvabrata Bandopadhaya, Soumya Ranjan Samal, and Vladimir Poulkov. Machine learning enabled performance prediction model for massive-mimo hetnet system. *Sensors*, 21(3):800, 2021.

[26] Suzhi Bi, Jiangbin Lyu, Zhi Ding, and Rui Zhang. Engineering radio maps for wireless resource management. *IEEE Wireless Communications*, 26(2):133–141, 2019.

[27] Petros S Bithas, Emmanouel T Michailidis, Nikolaos Nomikos, Demosthenes Vouyioukas, and Athanasios G Kanatas. A survey on machine-learning techniques for uav-based communications. *Sensors*, 19(23):5170, 2019.

[28] Mario Bkassiny, Yang Li, and Sudharman K. Jayaweera. A survey on machine-learning techniques in cognitive radios. *IEEE Communications Surveys Tutorials*, 15(3):1136–1159, 2013.

[29] Ishan Budhiraja, Neeraj Kumar, and Sudhanshu Tyagi. Deep-reinforcement-learning-based proportional fair scheduling control scheme for underlay d2d communication. *IEEE Internet of Things J.*, 8(5):3143–3156, 2021.

[30] Liyana Adilla binti Burhanuddin et al. QoE optimization for live video streaming in uav-to-uav communications via deep reinforcement learning. *IEEE Trans. Veh. Technol.*, pages 1–14, 2022.

[31] Javier Campos. Understanding the 5g nr physical layer. *Keysight Technologies release*, 2017.

[32] Pablo Carballeira, Julián Cabrera, Antonio Ortega, Fernando Jaureguizar, and Narciso García. A framework for the analysis and optimization of encoding latency for multiview video. *IEEE J. Sel. Topics Signal Process.*, 6(5):583–596, Sept. 2012.

[33] Ursula Challita, Walid Saad, and Christian Bettstetter. Deep reinforcement learning for interference-aware path planning of cellular-connected uavs. In *Proc. 2018 IEEE Int. Commun. Conf. (ICC)*, pages 1–7. IEEE, 2018.

[34] Ursula Challita, Walid Saad, and Christian Bettstetter. Interference management for cellular-connected UAVs: A deep reinforcement learning approach. *IEEE Trans. Wireless Commun.*, 18(4):2125–2140, 2019.

[35] Bo Chang, Lei Zhang, Liying Li, Guodong Zhao, and Zhi Chen. Optimizing resource allocation in urllc for real-time wireless control systems. *IEEE Transactions on Vehicular Technology*, 68(9):8916–8927, 2019.

[36] Arpan Chattopadhyay, Avishek Ghosh, and Anurag Kumar. Asynchronous stochastic approximation based learning algorithms for as-you-go deployment of wireless relay networks along a line. *IEEE Transactions on Mobile Computing*, 17(5):1004–1018, 2017.

[37] Junting Chen and David Gesbert. Optimal positioning of flying relays for wireless networks: A los map approach. In *2017 IEEE international conference on communications (ICC)*, pages 1–6. IEEE, 2017.

[38] Qinbo Chen. Joint position and resource optimization for multi-uav-aided relaying systems. *IEEE Access*, 8:10403–10415, 2020.

[39] Shuying Chen, Xi Li, Changqing Luo, Hong Ji, and Heli Zhang. Energy-efficient power, position and time control in UAV-assisted wireless networks. In *2019 IEEE Globecom Workshops (GC Wkshps)*, pages 1–6. IEEE, 2019.

[40] Tao Chen, Qi Gao, and Mingyu Guo. An improved multiple uavs cooperative flight algorithm based on leader follower strategy. In *2018 Chinese Control And Decision Conference (CCDC)*, pages 165–169. IEEE, 2018.

[41] Yan Chen, Hangjing Zhang, and Yang Hu. Optimal power and bandwidth allocation for multiuser video streaming in UAV relay networks. *IEEE Trans. on Veh. Tech.*, 69(6):6644–6655, 2020.

[42] Shih-Fan Chou, Te-Chuan Chiu, Ya-Ju Yu, and Ai-Chun Pang. Mobile small cell deployment for next generation cellular networks. In *2014 IEEE Global Communications Conference*, pages 4852–4857. IEEE, 2014.

[43] Erik Dahlman, Stefan Parkvall, and Johan Skold. *5G NR: The next generation wireless access technology.* Academic Press, 2020.

[44] Erik Dahlman, Stefan Parkvall, and Johan Skold. *5G NR: The next generation wireless access technology.* Academic Press, 2020.

[45] Patricia de Sousa Paula, Miguel Franklin de Castro, Gabriel A Louis Paillard, and Wellington WF Sarmento. A swarm solution for a cooperative and self-organized team of uavs to search targets. In *2016 8th Euro American Conference on Telematics and Information Systems (EATIS)*, pages 1–8. IEEE, 2016.

[46] Francesco Devoti, Ilario Filippini, and Antonio Capone. Facing the millimeter-wave cell discovery challenge in 5g networks with context-awareness. *IEEE Access*, 4:8019–8034, 2016.

[47] Runze Dong, Buhong Wang, Jiwei Tian, Tianhao Cheng, and Danyu Diao. Deep Reinforcement Learning Based UAV for Securing mmWave Communications. *IEEE Transactions on Vehicular Technology*, 72(4):5429–5434, 4 2023.

[48] Baojia Du et al. Mapping wetland plant communities using unmanned aerial vehicle hyperspectral imagery by comparing object/pixel-based classifications combining multiple machine-learning algorithms. *IEEE J. Sel. Top Appl. Earth Obs Remote Sens.*, pages 1–1, 2021.

[49] Gregory David Durgin. *Space-time wireless channels.* Prentice Hall Professional, 2003.

[50] Robert S Elliott. Antenna theory and design, a john wiley & sons. *INC., Publication*, 2003.

[51] Omid Esrafilian, Rajeev Gangula, and David Gesbert. Learning to communicate in uav-aided wireless networks: Map-based approaches. *IEEE Internet of Things Journal*, 6(2):1791–1802, 2018.

[52] Omid Esrafilian, Rajeev Gangula, and David Gesbert. Map reconstruction in uav networks via fusion of radio and depth measurements. In *ICC*

2021-IEEE International Conference on Communications, pages 1–6. IEEE, 2021.

[53] Ali A Esswie and Klaus I Pedersen. Capacity optimization of spatial pre-emptive scheduling for joint URLLC-eMBB traffic in 5G new radio. In *2018 IEEE Globecom Workshops (GC Wkshps)*, pages 1–6. IEEE, 2018.

[54] Ali A Esswie and Klaus I Pedersen. Null space based preemptive scheduling for joint URLLC and eMBB traffic in 5G networks. In *2018 IEEE Globecom Workshops (GC Wkshps)*, pages 1–6. IEEE, 2018.

[55] Dian Fan et al. Channel estimation and self-positioning for UAV swarm. *IEEE Trans. on Commun.*, 67(11):7994–8007, 2019.

[56] Hassan Fawaz, Melhem El Helou, Samer Lahoud, and Kinda Khawam. A reinforcement learning approach to queue-aware scheduling in full-duplex wireless networks. *Computer Networks*, 189:107893, 2021.

[57] Alem H Fitwi, Deeraj Nagothu, Yu Chen, and Erik Blasch. A distributed agent-based framework for a constellation of drones in a military operation. In *2019 Winter Simulation Conference (WSC)*, pages 2548–2559. IEEE, 2019.

[58] Azade Fotouhi, Ming Ding, and Mahbub Hassan. Dynamic base station repositioning to improve spectral efficiency of drone small cells. In *2017 IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, pages 1–9. IEEE, 2017.

[59] Azade Fotouhi, Ming Ding, and Mahbub Hassan. Service on demand: Drone base stations cruising in the cellular network. In *2017 IEEE Globecom Workshops (GC Wkshps)*, pages 1–6. IEEE, 2017.

[60] Azade Fotouhi, Ming Ding, and Mahbub Hassan. Flying drone base stations for macro hotspots. *IEEE Access*, 6:19530–19539, 2018.

[61] Yuan Gao, Weigui Zhou, Hong Ao, Jian Chu, Quan Zhou, Bo Zhou, Kang Wang, Yi Li, and Peng Xue. A novel optimal joint resource allocation method in cooperative multicarrier networks: Theory and practice. *Sensors*, 16(4):522, 2016.

[62] Margarita Gapeyenko, Dmitri Moltchanov, Sergey Andreev, and Robert W. Heath. Line-of-sight probability for mmwave-based uav communications in

3d urban grid deployments. *IEEE Transactions on Wireless Communications*, 20(10):6566–6579, 2021.

[63] Niklas Goddemeier and Christian Wietfeld. Investigation of air-to-air channel characteristics and a UAV specific extension to the rice model. In *2015 IEEE Globecom Workshops (GC Wkshps)*, pages 1–5, Dec. 2015.

[64] Andrea Goldsmith. Path Loss and Shadowing. *Wireless Communications*, pages 27–63, 2013.

[65] Kinshuk Govil, Morgan L Welch, J Timothy Ball, and Carlton R Pennypacker. Preliminary results from a wildfire detection system using deep learning on remote camera images. *Remote Sensing*, 12(1):166, 2020.

[66] Shaoai Guo and Xiaohui Zhao. Multi-Agent Deep Reinforcement Learning Based Transmission Latency Minimization for Delay-Sensitive Cognitive Satellite-UAV Networks. *IEEE Transactions on Communications*, 71(1):131–144, 1 2023.

[67] Lav Gupta, Raj Jain, and Gabor Vaszkun. Survey of important issues in uav communication networks. *IEEE Communications Surveys & Tutorials*, 18(2):1123–1152, 2015.

[68] Sang Ik Han. Survey on uav deployment and trajectory in wireless communication networks: Applications and challenges. *Information*, 13(8):389, 2022.

[69] Ayaka Hanyu, Yuichi Kawamoto, and Nei Kato. Adaptive channel selection and transmission timing control for simultaneous receiving and sending in relay-based uav network. *IEEE Transactions on Network Science and Engineering*, 7(4):2840–2849, 2020.

[70] Chao He, Zhidong Xie, and Chang Tian. A QoE-Oriented uplink allocation for Multi-UAV video streaming. *Sensors*, 19(15):3394, 2019.

[71] Harri Holma, Antti Toskala, and Jussi Reunanen. *LTE small cell optimization: 3GPP evolution to Release 13*. John Wiley & Sons, 2016.

[72] Fenghe Hu, Yansha Deng, and A Hamid Aghvami. Correlation-aware cooperative multigroup broadcast 360 ˆ {deg} video delivery network: A hierarchical deep reinforcement learning approach. *arXiv preprint arXiv:2010.11347*, 2020.

[73] Fenghe Hu, Yansha Deng, and A Hamid Aghvami. Cooperative multi-group broadcast 360° video delivery network: A hierarchical federated deep reinforcement learning approach. *IEEE Transactions on Wireless Communications*, 2021.

[74] Fenghe Hu, Yansha Deng, and A Hamid Aghvami. Cooperative multi-group broadcast 360° video delivery network: A hierarchical federated deep reinforcement learning approach. *IEEE Transactions on Wireless Communications*, 2021.

[75] Han Hu, Cheng Zhan, Jianping An, and Yonggang Wen. Optimization for HTTP adaptive video streaming in UAV-Enabled relaying system. In *Proc. 2019 IEEE Int. Conf. on Commun. (ICC)*, pages 1–6. IEEE, 2019.

[76] Jingzhi Hu, Hongliang Zhang, and Lingyang Song. Reinforcement learning for decentralized trajectory design in cellular uav networks with sense-and-send protocol. *IEEE Internet of Things J.*, 2018.

[77] Jingzhi Hu, Hongliang Zhang, and Lingyang Song. Reinforcement learning for decentralized trajectory design in cellular UAV networks with sense-and-send protocol. *IEEE Internet Things J.*, 6(4):6177–6189, 2018.

[78] Yijia Huang and Yong Zeng. Simultaneous environment sensing and channel knowledge mapping for cellular-connected uav. In *2021 IEEE Globecom Workshops (GC Wkshps)*, pages 1–6. IEEE, 2021.

[79] E. T. S. Institute. Study on 5g new radio(nr) access technology, technical report 38.912. Tr, European Telecommunications Standards Institute,, May. 2017. V15.0.0.

[80] Tommi Jaakkola, Michael I Jordan, and Satinder P Singh. On the convergence of stochastic iterative dynamic programming algorithms. *Neural computation*, 6(6):1185–1201, 1994.

[81] Nan Jiang, Yansha Deng, Arumugam Nallanathan, and Jonathon A Chambers. Reinforcement learning for real-time optimization in NB-IoT networks. *IEEE J. Sel. Areas Commun.*, 37(6):1424–1440, Jun. 2019.

[82] Nan Jiang, Yansha Deng, Arumugam Nallanathan, and Jinhong Yuan. A decoupled learning strategy for massive access optimization in cellular IoT networks. *IEEE J. on Sel. Areas in Commun.*, 39(3):668–685, 2021.

[83] Abhishek Joshi, Sarang Dhongdi, Shubham Kumar, and KR Anupama. Simulation of multi-UAV Ad-Hoc network for disaster monitoring applications. In *2020 Int. Conf. on Inf. Network. (ICOIN)*, pages 690–695, Jan. 2020.

[84] Tobias Kadur, Hsiao-Lan Chiang, and Gerhard Fettweis. Effective beam alignment algorithm for low cost millimeter wave communication. In *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, pages 1–5. IEEE, 2016.

[85] Nei Kato, Zubair Md. Fadlullah, Fengxiao Tang, Bomin Mao, Shigenori Tani, Atsushi Okamura, and Jiajia Liu. Optimizing space-air-ground integrated networks by artificial intelligence. *IEEE Wireless Communications*, 26(4):140–147, 2019.

[86] Aziz Altaf Khuwaja, Yunfei Chen, Nan Zhao, Mohamed-Slim Alouini, and Paul Dobbins. A survey of channel modeling for uav communications. *IEEE Communications Surveys & Tutorials*, 20(4):2804–2821, 2018.

[87] Paulo V Klaine, João PB Nadas, Richard D Souza, and Muhammad A Imran. Distributed drone base station positioning for emergency cellular networks using reinforcement learning. *Cognitive computation*, 10(5):790–804, 2018.

[88] Istvan Kovacs et al. Interference analysis for UAV connectivity over LTE using aerial radio measurements. In *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, pages 1–6. IEEE, 2017.

[89] Tejas D Kulkarni, Karthik Narasimhan, Ardavan Saeedi, and Josh Tenenbaum. Hierarchical deep reinforcement learning: Integrating temporal abstraction and intrinsic motivation. *Advances in neural information processing systems*, 29, 2016.

[90] A. Kumbhar, İ. Guvenç, S. Singh, and A. Tuncer. Exploiting LTE-advanced HetNets and FeICIC for UAV-assisted public safety communications. *IEEE Access*, 6:783–796, 2018.

[91] Abhaykumar Kumbhar, Hamidullah Binol, Ismail Guvenc, and Kemal Akkaya. Interference coordination for aerial and terrestrial nodes in three-tier LTE-advanced HetNet. In *Proc IEEE Radio Wirel Symp*, pages 1–4. IEEE, 2019.

[92] Jin Li and Youngnam Han. Optimal resource allocation for packet delay minimization in multi-layer uav networks. *IEEE Communications Letters*, 21(3):580–583, 2016.

[93] Kai Li, Wei Ni, Xin Wang, Ren Ping Liu, Salil S Kanhere, and Sanjay Jha. Energy-efficient cooperative relaying for unmanned aerial vehicles. *IEEE Transactions on Mobile Computing*, 15(6):1377–1386, 2015.

[94] Yabo Li, Haijun Zhang, Keping Long, Chunxiao Jiang, and Mohsen Guizani. Joint resource allocation and trajectory optimization with QoS in UAV-based NOMA wireless networks. *IEEE Trans. on Wireless Commun.*, 20(10):6343–6355, 2021.

[95] Morten Lindeberg, Stein Kristiansen, Thomas Plagemann, and Vera Goebel. Challenges and techniques for video streaming over mobile ad hoc networks. *Multimedia Systems*, 17(1):51–82, 2011.

[96] Chi Harold Liu, Zheyu Chen, Jian Tang, Jie Xu, and Chengzhe Piao. Energy-efficient uav control for effective and fair communication coverage: A deep reinforcement learning approach. *IEEE J. Sel. Areas Commun.*, 36(9):2059–2070, 2018.

[97] Xiao Liu, Yuanwei Liu, and Yue Chen. Reinforcement learning in multiple-uav networks: Deployment and movement design. *arXiv preprint arXiv:1904.05242*, 2019.

[98] Nguyen Cong Luong, Dinh Thai Hoang, Shimin Gong, Dusit Niyato, Ping Wang, Ying-Chang Liang, and Dong In Kim. Applications of deep reinforcement learning in communications and networking: A survey. *IEEE Communications Surveys Tutorials*, 21(4):3133–3174, 2019.

[99] Jiangbin Lyu, Yong Zeng, and Rui Zhang. Uav-aided offloading for cellular hotspot. *IEEE Transactions on Wireless Communications*, 17(6):3988–4001, 2018.

[100] Hongzi Mao, Ravi Netravali, and Mohammad Alizadeh. Neural adaptive video streaming with pensieve. In *Proc. Conf. of the ACM Special Interest Group on Data Communication*, pages 197–210, Aug. 2017.

[101] Antonino Masaracchia et al. The concept of time sharing NOMA into UAV-Enabled communications: An energy-efficient approach. In *2020 4th*

Int. Conf. on Recent Advances in Signal Processing, Telecommunications & Comput. (SigTelCom), pages 61–65, Aug. 2020.

[102] Weidong Mei, Qingqing Wu, and Rui Zhang. Cellular-connected UAV: Uplink association, power control and interference coordination. *IEEE Trans. Wirel. Commun.*, 18(11):5380–5393, 2019.

[103] Weidong Mei and Rui Zhang. Cooperative downlink interference transmission and cancellation for cellular-connected UAV: A divide-and-conquer approach. *IEEE Trans. on Commun.*, 68(2):1297–1311, 2019.

[104] Weidong Mei and Rui Zhang. Uav-sensing-assisted cellular interference coordination: A cognitive radio approach. *IEEE Wireless Communications Letters*, 9(6):799–803, 2020.

[105] Weidong Mei and Rui Zhang. Aerial-ground interference mitigation for cellular-connected UAV. *IEEE Wirel. Commun.*, 28(1):167–173, 2021.

[106] Weidong Mei and Rui Zhang. Aerial-ground interference mitigation for cellular-connected UAV. *IEEE Wireless Commun.*, 28(1):167–173, 2021.

[107] Volodymyr Mnih et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529, Feb. 2015.

[108] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529, Feb, 2015.

[109] Xiaopeng Mo, Yuwei Huang, and Jie Xu. Radio-map-based robust positioning optimization for uav-enabled wireless power transfer. *IEEE Wireless Communications Letters*, 9(2):179–183, 2019.

[110] Mohammad Mozaffari, Ali Taleb Zadeh Kasgari, Walid Saad, Mehdi Bennis, and Mérouane Debbah. Beyond 5g with uavs: Foundations of a 3d wireless cellular network. *IEEE Transactions on Wireless Communications*, 18(1):357–372, 2018.

[111] Mohammad Mozaffari, Xingqin Lin, and Stephen Hayes. Toward 6G with connected sky: UAVs and beyond. *IEEE Commun. Mag.*, 59(12):74–80, 2021.

[112] Mohammad Mozaffari, Walid Saad, Mehdi Bennis, and Merouane Debbah. Optimal transport theory for power-efficient deployment of unmanned aerial vehicles. In *2016 IEEE international conference on communications (ICC)*, pages 1–6. IEEE, 2016.

[113] Mohammad Mozaffari, Walid Saad, Mehdi Bennis, Young-Han Nam, and Mérouane Debbah. A tutorial on uavs for wireless networks: Applications, challenges, and open problems. *IEEE communications surveys & tutorials*, 21(3):2334–2360, 2019.

[114] MM Müller, L Vilà-Vilardell, and H Vacik. Forest fires in the alps–state of knowledge, future challenges and options for an integrated fire management. *EUSALP Action Group*, 8, 2020.

[115] Raheeb Muzaffar, Evşen Yanmaz, Christian Raffelsberger, Christian Bettstetter, and Andrea Cavallaro. Live multicast video streaming from drones: an experimental study. *Autonomous Robots*, pages 1–17, 2019.

[116] Saif Najmeddin. *UAV-Enabled Wireless Powered Communication Networks*. PhD thesis, Concordia University, 2021.

[117] Kamesh Namuduri, Serge Chaumette, Jae H Kim, and James PG Sterbenz. *UAV networks and communications*. Cambridge University Press, 2017.

[118] Huan Cong Nguyen et al. Using LTE networks for UAV command and control link: A rural-area coverage analysis. In *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, pages 1–6. IEEE, 2017.

[119] Huan Cong Nguyen et al. How to ensure reliable connectivity for aerial vehicles over cellular networks. *IEEE Access*, 6:12304–12317, 2018.

[120] Giancarlo Eder Guerra Padilla, Kun-Jung Kim, Seok-Hwan Park, and Kee-Ho Yu. Flight path planning of solar-powered UAV for sustainable communication relay. *IEEE Robot. Automat. Lett.*, 5(4):6772–6779, Aug. 2020.

[121] Justin J Podur, David L Martell, and David Stanford. A compound poisson model for the annual area burned by forest fires in the province of ontario. *Environmetrics*, 21(5):457–469, 2010.

[122] Yaohong Qu, Jizhi Wu, Bing Xiao, and Dongli Yuan. A fault-tolerant cooperative positioning approach for multiple uavs. *IEEE Access*, 5:15630–15640, 2017.

[123] Mattia Rebato, Laura Resteghini, Christian Mazzucco, and Michele Zorzi. Study of realistic antenna patterns in 5G mmWave cellular scenarios. In *Proc. 2018 IEEE Int. Commun. Conf. (ICC)*, pages 1–7. IEEE, Jul. 2018.

[124] Lang Ruan, Jinlong Wang, Jin Chen, Yitao Xu, Yang Yang, Han Jiang, Yuli Zhang, and Yuhua Xu. Energy-efficient multi-uav coverage deployment in uav networks: A game-theoretic framework. *China Communications*, 15(10):194–209, 2018.

[125] Yalcin Sadi, Sinem Coleri Ergen, and Pangun Park. Minimum energy data transmission for wireless networked control systems. *IEEE Trans. on Wireless Commun.*, 13(4):2163–2175, Feb. 2014.

[126] Mahmoud M Selim et al. On the outage probability and power control of D2D underlaying NOMA UAV-assisted networks. *IEEE Access*, 7:16525–16536, Jan. 2019.

[127] Syed Awais W Shah, Tamer Khattab, Muhammad Zeeshan Shakir, and Mazen O Hasna. A distributed approach for networked flying platform association with small cells in 5g+ networks. In *GLOBECOM 2017-2017 IEEE Global Communications Conference*, pages 1–7. IEEE, 2017.

[128] C. She et al. Ultra-reliable and low-latency communications in unmanned aerial vehicle communication systems. *IEEE Trans. Commun.*, 67(5):3768–3781, May 2019.

[129] Changyang She, Chenyang Yang, and Tony QS Quek. Joint uplink and downlink resource configuration for ultra-reliable and low-latency communications. *IEEE Trans. Commun.*, 66(5):2266–2280, 2018.

[130] Ali Magdi Sayed Soliman, Suleyman Cinar Cagan, and Berat Baris Buldum. The design of a rotary-wing unmanned aerial vehicles–payload drop mechanism for fire-fighting services using fire-extinguishing balls. *SN Applied Sciences*, 1:1–10, 2019.

[131] Ki Won Sung, Edward Mutafungwa, Riku Jäntti, Minseok Choi, Joohyung Jeon, Dohyun Kim, Joongheon Kim, Jose Costa-Requena, Anders Nordlöw, Sachin Sharma, et al. Primo-5g: making firefighting smarter with immersive videos through 5g. In *Proc. 2019 IEEE 2nd 5G World Forum (5GWF)*, pages 280–285. IEEE.

[132] Ki Won Sung, Edward Mutafungwa, Riku Jäntti, Minseok Choi, Joohyung Jeon, Dohyun Kim, Joongheon Kim, Jose Costa-Requena, Anders Nordlöw, Sachin Sharma, et al. Primo-5g: making firefighting smarter with immersive videos through 5g. In *Proc. 2019 IEEE 2nd 5G World Forum (5GWF)*, pages 280–285. IEEE.

[133] Praneeth Susarla, Yansha Deng, Giuseppe Destino, Jani Saloranta, Toktam Mahmoodi, Markku Juntti, and Olli Sílven. Learning-based trajectory optimization for 5g mmwave uplink uavs. In *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*, pages 1–7. IEEE, 2020.

[134] Praneeth Susarla, Yansha Deng, Giuseppe Destino, Jani Saloranta, Toktam Mahmoodi, Markku Juntti, and Olli Sílven. Learning-based trajectory optimization for 5g mmwave uplink uavs. In *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*, pages 1–7. IEEE, June, 2020.

[135] Praneeth Susarla, Bikshapathi Gouda, Yansha Deng, Markku Juntti, Olli Sílven, and Antti Tölli. DQN-based beamforming for uplink mmWave cellular-connected uavs. In *2021 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6. IEEE, 2021.

[136] Praneeth Susarla, Bikshapathi Gouda, Yansha Deng, Markku Juntti, Olli Sílven, and Antti Tölli. Dqn-based beamforming for uplink mmwave cellular-connected uavs. In *2021 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6. IEEE, 2021.

[137] Praneeth Susarla, Bikshapathi Gouda, Yansha Deng, Markku Juntti, Olli Silvén, and Antti Tölli. Learning-Based Beam Alignment for Uplink mmWave UAVs. *IEEE Transactions on Wireless Communications*, 22(3):1779–1793, 3 2023.

[138] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, Nov, 2018.

[139] Yu Herng Tan, Shupeng Lai, Kangli Wang, and Ben M Chen. Cooperative heavy lifting using unmanned multi-agent systems. In *2018 IEEE 14th International Conference on Control and Automation (ICCA)*, pages 1119–1126. IEEE, 2018.

[140] Fengxiao Tang, Yibo Zhou, and Nei Kato. Deep reinforcement learning for dynamic uplink/downlink resource allocation in high mobility 5G hetnet. *IEEE Journal on Selected Areas in Commun.*, 38(12):2773–2782, 2020.

[141] Shams ur Rahman, Geon-Hwan Kim, You-Ze Cho, and Ajmal Khan. Positioning of uavs for throughput maximization in software-defined disaster area uav communication networks. *Journal of Communications and Networks*, 20(5):452–463, 2018.

[142] Vutha Va, Junil Choi, Takayuki Shimizu, Gaurav Bansal, and Robert W Heath. Inverse multipath fingerprinting for millimeter wave v2i beam alignment. *IEEE Transactions on Vehicular Technology*, 67(5):4042–4058, 2017.

[143] Vutha Va, Junil Choi, Takayuki Shimizu, Gaurav Bansal, and Robert W Heath. Inverse multipath fingerprinting for millimeter wave v2i beam alignment. *IEEE Transactions on Vehicular Technology*, 67(5):4042–4058, Dec 2017.

[144] Vutha Va, Takayuki Shimizu, Gaurav Bansal, and Robert W Heath. Online learning for position-aided millimeter wave beam training. *IEEE Access*, 7:30507–30526, 2019.

[145] Vutha Va, Takayuki Shimizu, Gaurav Bansal, and Robert W Heath. Online Learning for Position-Aided Millimeter Wave Beam Training. *IEEE Access*, 7:30507–30526, Mar 2019.

[146] Vutha Va, Takayuki Shimizu, Gaurav Bansal, and Robert W Heath. Online Learning for Position-Aided Millimeter Wave Beam Training. *IEEE Access*, 7:30507–30526, Mar 2019.

[147] Maria Val Martin, Ralph Kahn, and Mika Tosca. A global analysis of wildfire smoke injection heights derived from space-based multi-angle imaging. *Remote Sensing*, 10(10):1609, Oct. 2018.

[148] Evgenii Vinogradov, Hazem Sallouha, Sibren De Bast, Mohammad Mahdi Azari, and Sofie Pollin. Tutorial on uav: A blue sky view on wireless communication. *arXiv preprint arXiv:1901.02306*, 2019.

[149] Ankur Vora and Kyoung-Don Kang. Effective 5G wireless downlink scheduling and resource allocation in cyber-physical systems. *Technologies*, 6(4):105, 2018.

[150] Shuoyao Wang, Suzhi Bi, and Ying-Jun Angela Zhang. Deep Reinforcement Learning With Communication Transformer for Adaptive Live Streaming in Wireless Edge Networks. *IEEE Journal on Selected Areas in Communications*, 40(1):308–322, 1 2022.

[151] Yuyang Wang, Nitin Jonathan Myers, Nuria González-Prelcic, and Robert W Heath. Site-specific online compressive beam codebook learning in mmwave vehicular communication. *IEEE Transactions on Wireless Communications*, 20(5):3122–3136, 2021.

[152] Chen-Yu Wei and Wanjiun Liao. Multi-cell cooperative scheduling for network utility maximization with user equipment side interference cancellation. *IEEE Transactions on Wireless Communications*, 17(1):619–635, 2017.

[153] Di Wu, Yong Zeng, Shi Jin, and Rui Zhang. Environment-aware and training-free beam alignment for mmwave massive mimo via channel knowledge map. In *2021 IEEE International Conference on Communications Workshops (ICC Workshops)*, pages 1–7. IEEE, 2021.

[154] Di Wu, Yong Zeng, Shi Jin, and Rui Zhang. Environment-aware hybrid beamforming by leveraging channel knowledge map. *arXiv preprint arXiv:2206.08707*, 2022.

[155] Po Han Wu, Chih Wei Huang, Jenq Neng Hwang, Jae Young Pyun, and Juan Zhang. Video-Quality-Driven Resource Allocation for Real-Time Surveillance Video Uplinking Over OFDMA-Based Wireless Networks. *IEEE Trans. on Vehicular Technology*, 64(7):3233–3246, 2015.

[156] Xuedou Xiao et al. Sensor-augmented neural adaptive bitrate video streaming on UAVs. *IEEE Trans. on Multimedia*, pages 1–12, Oct. 2019.

[157] Hao Xie, Dingcheng Yang, Lin Xiao, and Jiangbin Lyu. Connectivity-aware 3d uav path design with deep reinforcement learning. *IEEE Transactions on Vehicular Technology*, 70(12):13022–13034, 2021.

[158] Shu Xu, Xiangyu Zhang, Chunguo Li, Dongming Wang, and Luxi Yang. Deep Reinforcement Learning Approach for Joint Trajectory Design in Multi-UAV IoT Networks. *IEEE Transactions on Vehicular Technology*, 71(3):3389–3394, 3 2022.

[159] Yuntao Xue and Weisheng Chen. A UAV Navigation Approach Based on Deep Reinforcement Learning in Large Cluttered 3d Environments. *IEEE Transactions on Vehicular Technology*, 72(3):3001–3014, 3 2023.

[160] Vijaya Yajnanarayana et al. Interference mitigation methods for unmanned aerial vehicles served by cellular networks. In *2018 IEEE 5G World Forum (5GWF)*, pages 118–122, Jul. 2018.

[161] Shengzhi Yang et al. Energy efficiency optimization for UAV-assisted backscatter communications. *IEEE Commun. Lett.*, 23(11):2041–2045, 2019.

[162] Yirga Yayeh Munaye, Hsin-Piao Lin, Abebe Belay Adege, and Getaneh Berie Tarekegn. Uav positioning for throughput maximization using deep learning approaches. *Sensors*, 19(12):2775, 2019.

[163] Xiaoqi Yin, Abhishek Jindal, Vyas Sekar, and Bruno Sinopoli. A control-theoretic approach for dynamic adaptive video streaming over HTTP. In *Proc. 2015 ACM Conf. on Special Interest Group on Data Communication*, page 325–338, Aug. 2015.

[164] Ayat Zaki-Hindi, Rafhael Amorim, István Z Kovács, and Jeroen Wigard. Uplink coexistence for high throughput uavs in cellular networks. In *GLOBECOM 2022-2022 IEEE Global Communications Conference*, pages 2957–2962. IEEE, 2022.

[165] Alessio Zappone, Marco Di Renzo, and Mérouane Debbah. Wireless networks design in the era of deep learning: Model-based, ai-based, or both? *IEEE Transactions on Communications*, 67(10):7331–7376, 2019.

[166] Linzhou Zeng, Xiang Cheng, Cheng-Xiang Wang, and Xuefeng Yin. Second order statistics of non-isotropic uav ricean fading channels. In *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, pages 1–5. IEEE, 2017.

[167] Yong Zeng, Jiangbin Lyu, and Rui Zhang. Cellular-connected UAV: Potential, challenges, and promising technologies. *IEEE Wireless Commun.*, 26(1):120–127, 2018.

[168] Yong Zeng, Xiaoli Xu, Shi Jin, and Rui Zhang. Simultaneous navigation and radio mapping for cellular-connected UAV with deep reinforcement learning. *IEEE Trans. on Wireless Commun.*, 2021.

[169] Yong Zeng, Xiaoli Xu, and Rui Zhang. Trajectory design for completion time minimization in UAV-enabled multicasting. *IEEE Trans. on Wireless Commun.*, 17(4):2233–2246, 2018.

[170] Yong Zeng, Rui Zhang, and Teng Joon Lim. Wireless communications with unmanned aerial vehicles: Opportunities and challenges. *IEEE Communications magazine*, 54(5):36–42, 2016.

[171] Cheng Zhan et al. Unmanned aircraft system aided adaptive video streaming: A joint optimization approach. *IEEE Trans. Multimedia*, 22(3):795–807, 2020.

[172] Chaoyun Zhang, Paul Patras, and Hamed Haddadi. Deep learning in mobile and wireless networking: A survey. *IEEE Communications Surveys Tutorials*, 21(3):2224–2287, 2019.

[173] Chiya Zhang, Weizheng Zhang, Wei Wang, Lu Yang, and Wei Zhang. Research challenges and opportunities of uav millimeter-wave communications. *IEEE Wireless Communications*, 26(1):58–62, 2019.

[174] Haijun Zhang, Miaolin Huang, Huan Zhou, Xianmei Wang, Ning Wang, and Keping Long. Capacity Maximization in RIS-UAV Networks: A DDQN-Based Trajectory and Phase Shift Optimization Approach. *IEEE Transactions on Wireless Communications*, 22(4):2583–2591, 4 2023.

[175] Liang Zhang, Bijan Jabbari, and Nirwan Ansari. Deep Reinforcement Learning Driven UAV-Assisted Edge Computing. *IEEE Internet of Things Journal*, 9(24):25449–25459, dec 15 2022.

[176] Shuhang Zhang et al. Joint trajectory and power optimization for UAV relay networks. *IEEE Commun. Lett.*, 22(1):161–164, Oct. 2017.

[177] Shuhang Zhang, Hongliang Zhang, Boya Di, and Lingyang Song. Cellular UAV-to-X communications: Design and optimization for multi-UAV networks. *IEEE Trans. Wireless Commun.*, 18(2):1346–1359, Feb. 2019.

[178] Shuowen Zhang, Yong Zeng, and Rui Zhang. Cellular-enabled uav communication: A connectivity-constrained trajectory optimization perspective. *IEEE Transactions on Communications*, 67(3):2580–2604, 2018.

[179] Shuowen Zhang and Rui Zhang. Radio map-based 3d path planning for cellular-connected uav. *IEEE Transactions on Wireless Communications*, 20(3):1975–1989, 2021.

[180] Zhicai Zhang et al. QoE aware transcoding for live streaming in SDN-Based Cloud-Aided HetNets: An actor-critic approach. In *Proc. 2019 IEEE Int. Commun. Conf. Workshops (ICC Workshops)*, pages 1–6, May 2019.

[181] Jianwei Zhao, Feifei Gao, Guoru Ding, Tao Zhang, Weimin Jia, and Arumugam Nallanathan. Integrating communications and control for uav systems: Opportunities and challenges. *IEEE Access*, 6:67519–67527, 2018.

[182] Jianwei Zhao, Feifei Gao, Weimin Jia, Shun Zhang, Shi Jin, and Hai Lin. Angle domain hybrid precoding and channel tracking for millimeter wave massive mimo systems. *IEEE Transactions on Wireless Communications*, 16(10):6868–6880, 2017.

[183] Xiaohong Zheng, Hongyi Li, Choon Ki Ahn, and Deyin Yao. Nn-based fixed-time attitude tracking control for multiple unmanned aerial vehicles with nonlinear faults. *IEEE Transactions on Aerospace and Electronic Systems*, 59(2):1738–1748, 2022.

[184] Hui Zhou et al. Real-time video streaming and control of cellular-connected UAV system: Prototype and performance evaluation. *IEEE Wireless Communications Letters*, pages 1–1, 2021.

[185] Maowu Zhou, Hongbin Chen, Lei Shu, and Ye Liu. Uav assisted sleep scheduling algorithm for energy-efficient data collection in agricultural internet of things. *IEEE Internet of Things Journal*, 2021.

[186] Christopher Zygowski and Arunita Jaekel. Optimal formulation for maximizing area coverage in wireless sensor networks with mobile nodes. In *2018 IEEE 10th Latin-American Conference on Communications (LATINCOM)*, pages 1–6. IEEE, 2018.