



**HAL**  
open science

# Machine Learning Techniques for UAV-assisted Networks

Arzhang Shahbazi

► **To cite this version:**

Arzhang Shahbazi. Machine Learning Techniques for UAV-assisted Networks. Machine Learning [cs.LG]. Université Paris-Saclay, 2022. English. NNT : 2022UPASG076 . tel-03889218

**HAL Id: tel-03889218**

**<https://theses.hal.science/tel-03889218>**

Submitted on 7 Dec 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Machine Learning Techniques for  
UAV-Assisted Networks  
*Techniques d'Apprentissage Automatique Pour Les  
Réseaux Assistés Par Drone*

**Thèse de doctorat de l'université Paris-Saclay**

École doctorale n 580, Sciences et technologies de l'information  
et de la communication (STIC)  
Spécialité de doctorat : Réseaux, Information et Communications  
Graduate School : Université Paris-Saclay  
Réfèrent : Faculté des sciences d'Orsay

Thèse préparée dans la unité de recherche **Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette, France**, sous la direction de **Marco DI RENZO**, Directeur de recherche CNRS, la co-supervision de **Petar POPOVSKI**, Professeur du Université d'Aalborg, la co-supervision de **Ioannis KRIKIDIS**, Professeur du Université de Chypre

**Thèse présentée et soutenue à Paris-Saclay, le 21 novembre 2022, par**

**Arzhang SHAHBAZI**

**Composition du jury**

Membres du jury avec voix délibérative

<b>Jalel Ben Othman</b> Professeur, University Paris 13, France	Président
<b>Lina Mroueh</b> Professeure, ISEP, France	Rapportrice
<b>Trung Duong</b> Professeur, Queens University Belfast, UK	Rapporteur
<b>Toktam Mahmoodi</b> Professeure, King's College London, UK	Examinatrice
<b>Alessio Zappone</b> Professeur, University of Cassino, Italy	Examineur



**Titre :** Techniques d'apprentissage automatique pour les réseaux assistés par drone

**Mots clés :** Réseaux cellulaires, Systèmes de communication sans fil, Véhicule aérien sans pilote, Apprentissage automatique, Apprentissage par renforcement, Communication radar à double fonction.

**Résumé :** L'objectif principal de cette thèse est la modélisation, l'évaluation des performances et l'optimisation au niveau du système des réseaux cellulaires de prochaine génération équipés de drones en utilisant l'intelligence artificielle. En outre, la technologie émergente de détection et de communication intégrées est étudiée pour être appliquée aux futurs réseaux sans fil des drones. En particulier, en s'appuyant sur la technique d'apprentissage par renforcement pour contrôler les actions des drones, cette thèse développe un ensemble de nouveaux cadres d'apprentissage automatique pour incorporer des mesures de performance importantes dans l'agent, telles que le débit du système de communication et l'erreur de localisation, qui peuvent être utilisées pour l'ana-

lyse et l'optimisation au niveau du système. Plus précisément, un nouvel algorithme basé sur l'apprentissage est proposé pour maximiser le débit du système en utilisant une connaissance préalable de la probabilité de présence des utilisateurs dans un réseau. Un cadre d'apprentissage fédéré a été introduit pour trouver une planification optimale de la trajectoire en formant un agent avec un algorithme d'apprentissage profond dans différents environnements afin d'obtenir une généralisation et une convergence plus rapide. Les performances d'un drone équipé d'un système de communication radar à double fonction sont étudiées et les avantages potentiels de ces systèmes sont démontrés en optimisant conjointement le débit du système de communication et l'erreur de localisation.

**Title :** Machine Learning Techniques for UAV-assisted Networks

**Keywords :** Cellular Networks, Wireless Communication, Unmanned aerial vehicle (UAV), Machine Learning (ML), Reinforcement learning (RL), Dual-Functional Radar Communication (DFRC).

**Abstract :** The main focus of this thesis is on modeling, performance evaluation and system-level optimization of next-generation cellular networks empowered by Unmanned Aerial Vehicles (UAVs) by using Machine Learning (ML). In addition, the emerging technology of Integrated Sensing and Communication is investigated for application to future UAV wireless networks. In particular, relying on Reinforcement Learning (RL) technique for controlling UAV actions, this thesis develops a set of new ML frameworks for incorporating important performance metrics in to the RL agent, such as the communication system throughput and localization error, which can be used for system-

level analysis and optimization. More specifically, a new learning-based algorithms proposed to maximize the system throughput by utilizing a prior knowledge of users likelihood of presence in a grid. A Federated Learning (FL) framework introduced to find an optimal path planning through training an agent with RL algorithm in different environment settings to achieve generalization and faster convergence. The performance of UAV equipped with Dual-Functional Radar Communication (DFRC) is investigated and the potential benefits of DFRC systems are shown by jointly optimizing communication system throughput and localization error.



# Table des matières

<b>1</b>	<b>Introduction</b>	<b>7</b>
1.1	Next Generation Strategies for UAV Communication Mobile Networks . . . . .	9
1.2	Machine Learning and Artificial Intelligence for UAV Networks Beyond 5G . . . . .	10
1.3	Thesis Overview and Major Contribution . . . . .	11
1.4	Publications . . . . .	12
<b>2</b>	<b>UAV for Next Generation of Cellular Communication - An Introduction</b>	<b>15</b>
2.1	UAV Aerial Base Station in 5G and Beyond . . . . .	17
2.1.1	Coverage and Capacity Enhancement for Wireless Cellular Networks . . . . .	17
2.1.2	UAVs Acting as Flying Base Stations for Disaster Scenarios . . . . .	18
2.1.3	UAV-Aided Terrestrial Networks for Information Transmission . . . . .	19
2.1.4	MIMO and Millimeter Wave Communications in 3D UAV Networks . . . . .	20
2.1.5	UAVs and IoT Communications . . . . .	21
2.1.6	Caching in UAV Base Stations . . . . .	22
2.1.7	Cellular-Connected Drones as User Equipments . . . . .	23
2.1.8	Flying Ad-Hoc Networks With UAVs . . . . .	23
2.1.9	UAV Air-to-Ground Channel Modeling . . . . .	24
2.2	Conclusion . . . . .	25
<b>3</b>	<b>Machine Learning for UAV-Enabled Wireless Networks</b>	<b>27</b>
3.1	Machine Learning for UAVs : An Introduction . . . . .	29
3.2	Supervised and Unsupervised Learning for UAVs . . . . .	30
3.2.1	Supervised Learning Overview . . . . .	30
3.2.2	Unsupervised Learning Overview . . . . .	31
3.2.3	Practical Issues of ML Implementation . . . . .	32
3.3	Reinforcement Learning for UAVs : An Overview . . . . .	33
3.3.1	Deep Reinforcement Learning (DRL) . . . . .	34
3.3.2	Q-Learning . . . . .	35
3.3.3	Q-Learning Overview . . . . .	35
3.3.4	Update Rule . . . . .	36
3.3.5	The Exploration/Exploitation Trade-Off . . . . .	36
3.3.6	Limitation of RL . . . . .	36
3.3.7	Federated Learning for UAVs . . . . .	37
3.3.8	Transfer Learning for UAVs . . . . .	39
3.4	Conclusion . . . . .	40

<b>4</b>	<b>Throughput Maximization with Learning Based Trajectory for Mobile Users</b>	<b>41</b>
4.1	Introduction . . . . .	43
4.2	System Model . . . . .	44
4.3	Learning Based Trajectory Design . . . . .	46
4.4	Numerical Results . . . . .	49
4.5	Conclusion . . . . .	51
<b>5</b>	<b>Federated Reinforcement Learning UAV Trajectory Design for Fast Localization of Ground Users</b>	<b>53</b>
5.1	Introduction . . . . .	55
5.2	System Model . . . . .	56
5.3	Proposed Method . . . . .	58
5.4	Federated learning . . . . .	59
5.5	Numerical Results . . . . .	61
5.6	Conclusion . . . . .	62
<b>6</b>	<b>Multi-Objective Trajectory Design for UAV-Assisted Dual-Functional Radar-Communication Network : A Reinforcement Learning Approach</b>	<b>63</b>
6.1	Introduction . . . . .	65
6.2	System Model . . . . .	68
6.2.1	Channel Model . . . . .	68
6.2.2	Power Consumption Model . . . . .	71
6.3	Problem Formulation . . . . .	72
6.4	Calculating localization error . . . . .	74
6.5	Preliminaries . . . . .	76
6.5.1	Proposed RL framework . . . . .	78
6.6	Numerical Results . . . . .	80
6.7	Conclusion . . . . .	84
<b>7</b>	<b>Conclusion</b>	<b>85</b>
7.1	Conclusion . . . . .	86
7.2	Future Work . . . . .	88
7.2.1	Machine Learning-aided Wireless Networks . . . . .	88
7.2.2	Federated Learning in Future Networks . . . . .	90
7.2.3	Machine Learning for Reconfigurable Intelligent Surfaces . . . . .	91
<b>8</b>	<b>Synthèse en français</b>	<b>101</b>

## 1 - Introduction



This chapter begins with section 1.1 which introduces overview of potential enablers in the next generation UAV communication networks alongside with the corresponding research challenges. Section 1.3 highlights the major contributions in this thesis work and the organization of the thesis. Section 1.4 provides lists of publications produced during my Ph.D. candidature.

## Sommaire

---

1.1	Next Generation Strategies for UAV Communication Mobile Networks . . . . .	9
1.2	Machine Learning and Artificial Intelligence for UAV Networks Beyond 5G . . . . .	10
1.3	Thesis Overview and Major Contribution . . . . .	11
1.4	Publications . . . . .	12

## 1.1 . Next Generation Strategies for UAV Communication Mobile Networks

In recent years, rapid progress have been done in the design and the improvement of Unmanned Aerial Vehicles (UAVs) of different sizes, shapes, and their communication capacities. Drones can move autonomously by attached microprocessors or can be operated from a far without taking any human personnel. Due to their adaptability, easy installation, low maintenance costs, versatility, and relatively small operating cost, the use of drones support new ways for commercial, military, civilian, agricultural, and environmental applications such as border surveillance, relay for ad hoc networks, managing wildfire, disaster monitoring, wind estimation, traffic monitoring, remote sensing, and search and destroy operations. Many of these applications needs a single UAV system and others like area monitoring for hazardous environments demand multi-UAVs systems. Although single drone systems are utilized for decades, by functioning and developing one large UAV, exploiting a set of small UAVs has many advantages. Each UAV acts as an isolated node in the single UAV systems, it can only communicate with the ground node. Consequently, the UAV communication system is established through only on UAV-to-infrastructure communication, and the communication between the UAVs can be based on the infrastructure. The capacity of a single UAV system is restricted compared to the multi UAV system which has many advantages. First and foremost, tasks are principally completed at a lower cost with multi UAV systems. Additionally, the collaborative work of UAVs can enhance the performance of the system. Moreover, if the UAV fails in a mission in a multi UAV system, the operation can continue to exist with the other UAVs, and tasks are generally finished more swiftly and efficiently with multi-UAV systems.

Multiple UAVs can be utilized for successful and efficient mission completion due to their capabilities, flight time, and limited payload. To enable cooperation, communication and networking are essential to organize multiple UAVs and achieve autonomous drones network. Also, Ad hoc networks from multiple UAVs can be an possible communication approach. In the ad hoc UAVs network, only some drones are connected to the ground base, but all of the drones structure an ad hoc network. In these systems, UAVs are able to communicate with other UAVs and the ground base. Ad hoc UAVs networks can be considered as a special structure of Mobile Ad-hoc Network (MANET), and Vehicular Ad-hoc Network (VANET). In fact, UAVs networks have some distinguished characteristics when compared to the existing Ad hoc networks. Nodes in the UAVs networks are characterized by their high mobility degree. VANET and MANET nodes are cars and walking men respectively and UAVs fly in the sky above them. The high mobility of UAVs impact the network topology that changes more frequently in comparison with the topology of the MANET or the VANET. Furthermore, MANET and VANET task is to create peer-to-peer connections. Drones network also require peer-to-peer connections to guarantee coordination and collaboration between UAVs. In

most cases, drones collect data and relay it to the ground station. Consequently, it is mandatory to make sure the UAV-to- UAV communication and the UAV-to-Infrastructure communication are functioning. Therefore, UAVs network should establish peer-to-peer communication and converge cast traffic at the same time. Moreover, distances between UAVs are much longer than nodes in the MANETs and VANETs. Thus, in an attempt to create stable communication links between UAVs, it is necessary to boost their communication range.

One of the most important design problems of multi-UAV systems is the communication, which is essential for coordination and collaboration between the UAVs. UAVs can be utilized in aerial sensor networks in which they are composed of multiple data sources assigned in different zones where UAVs nodes are used to gather information. It may contain different types of sensors, and each sensor may require different data delivery methods. If there is a need to use different sensors, they will be loaded on different UAVs, e.g., one UAV can be loaded with an infrared camera, while another UAV is equipped with a high-resolution camera. Furthermore, UAVs network have various challenging system parameters such as limited bandwidth, high mobility, irregular connectivity, restricted transmission range, and uncertain noisy channels. These challenges introduce different issues in the ad hoc multihop environment like collisions, and transmission delays. For example, it is very demanding to maintain the transmission range between two UAVs moving in opposite directions with very high velocity. Due to the aforementioned issues, more studies and deep investigation related to UAV communication systems are necessary. Subsequently, among the objectives of this thesis is to recognize the challenges and design characteristics and constraints of the UAVs networks. Furthermore, we investigate the fundamental needs and functions for communication in UAV-based systems, and we propose various solutions that can be utilized for UAV communication systems.

## **1.2 . Machine Learning and Artificial Intelligence for UAV Networks Beyond 5G**

In practice, cellular-connected drones will lead to various new application use cases, however, to obtain the benefits from these systems, different unique communication and security challenges for each of these applications need to be acknowledged. For this purpose, Machine Learning (ML) based solution techniques are recognized as a powerful tool for addressing the challenges of cellular-connected drones. It should be mentioned that such challenges can also be addressed at different levels such as the physical layer and 3D coverage enhancement. In this respect, ML-based methods can aid in meeting the technical challenges of cellular-connected UAVs while achieving new improvements in the design of the network. Even though many approaches exist for addressing the aforementioned challenges, we focus on machine learning solutions because of their built-in capability to predict

future network states, so allowing drones to adjust to the dynamics and randomness of the network in an online manner. Specifically, ML approaches permit drones to generalize their observations to hidden network states and can scale to large-sized networks, which thus makes them suitable for drone applications. Furthermore, for such UAV-based applications, energy efficiency and computation capacity are major design restrictions. As a result, the main scope of this thesis is to point out the advantages that AI brings for cellular-connected UAVs under various system configurations.

An important aspect of UAV systems is to maintain reliable cellular connectivity for the UAVs at each time instant along their corresponding trajectory while also minimizing the time required to carry out their objective. For instance, a delivery UAV must maintain a minimum signal-to-noise (SNR) ratio along its path to secure a reliable communication link for its control information. This generally depends on the UAV's location, cell association, transmit power, and the location of the serving ground users. For this fact, a key challenge for a UAV system is to optimize the UAVs' path planning so as to decrease their total delivery time while maintaining reliable wireless connectivity and consequently an instantaneous SNR threshold value. Even though a centralized approach can update the trajectory plan of each UAV, this would necessitate real-time tracking of the UAVs and control signals to be transmitted to the UAVs for all time instants. Furthermore, a centralized approach earns high round-trip latency and needs a central unit to obtain full knowledge of the current network state. For overcoming these challenges, one can implement online edge algorithms that must be individually run by each UAV to plan their corresponding future paths. In this respect, convolutional neural networks (CNNs) can be combined with a deep reinforcement learning (RL) algorithm based on a recurrent neural network (RNN) at the UAV level, resulting in a CNN-RNN techniques. The aforementioned algorithms exhibit dynamic temporal behavior and is characterized by its adaptive memory, which empowers it to collect necessary previous state information to estimate the future steps of each UAV. In the meantime, CNNs are mostly used for image recognition and consequently can be used for identifying the UAV's environment by extracting features from input images. For example, CNNs help drones in identifying the location of ground base stations, ground users, and other drones in the network. These extracted features are then fed to a deep RNN, which can be trained to learn an optimized sequence of the UAV's future steps that would minimize its mission time and maintain a reliable cellular coverage during mission time based on the input features.

### **1.3 . Thesis Overview and Major Contribution**

In the present thesis, motivated by the above stated research challenges for the upcoming 5G and beyond 5G, we investigate the performance evaluation of Unmanned Aerial Vehicle (UAV) communication networks by using Machine Learning

(ML) methods. In particular, we tackle the problem of UAV path planning while optimizing various system parameters. In particular we utilize Reinforcement Learning (RL) for finding the trajectory that can achieve the specific system objectives. The main contributions of this thesis are as follows :

- This thesis provides the detailed introduction on the use of UAVs in wireless networks. We investigate the main use cases of UAVs and explore the key challenges and applications. Moreover, this thesis explores in detail a novel research approach where ML methods applied to improve the performance of UAV networks. We provide an overview of RL and fundamentals of Federated Learning (FL).
- This thesis introduces a framework which is based on the likelihood of mobile users presence in a grid with respect to their probability distribution. We model a novel UAV-assisted communication system depending on the shortest flight path of the UAV while maximizing the amount of data transmitted to mobile devices. The approach we use is deep reinforcement learning technique for finding the trajectory to maximize the throughput for ground mobile users. Numerical results highlight how our method strike a balance between the throughput achieved, trajectory, and the complexity.
- This thesis propose an approach for localizing ground targets by using Received Signal Strength (RSS) and utilizing UAVs as aerial anchors. We introduce a new framework based on FL that includes multiple UAVs training in different environments settings for finding the optimal path which results in faster convergence of the RL model for minimum localization error.
- In this thesis, we explore the Dual-Functional Radar Communication (DFRC) in UAV networks where a single UAV serves a group of communication users and locate the ground targets simultaneously. To balance the communication and localization performance, we solve multi-objective optimization problem to jointly optimize communication system throughput and localization error over a particular mission duration that is limited by UAV's energy consumption and flying time. For this purpose, we introduce a new framework based on (RL) to allow the UAV to autonomously optimize its path which results in improving the localization accuracy and maximizing the number of transmitted bits.

#### 1.4 . Publications

- **Journal Papers** The following is a list of publications in refereed journals produced from the research outcomes of this thesis. This journal paper is used as the basis for this thesis.
  - (J) Arzhang Shahbazi, Christos Masouros and Marco Di Renzo. "Multi-Objective Optimization for UAV-Assisted Dual-Functional Radar-Communication

**Abstract :** *In this paper, we explore the optimal trajectory for maximizing communication throughput and minimizing localization error in a Dual-Functional Radar Communication (DFRC) in unmanned aerial vehicle (UAV) network where a single UAV serves a group of communication users and locate the ground targets simultaneously. To balance the communication and localization performance, we formulate a multi-objective optimization problem to jointly optimize two objectives : maximization of number of transmitted bits sent to users and minimization of localization error for ground targets over a particular mission period which is restricted by UAV's energy consumption or flying time. These two objectives are in conflict with each other partly and weight parameters are given to describe associated importance. Hence, in this context, we propose a novel framework based on reinforcement learning (RL) to enable the UAV to autonomously find its trajectory that results in improving the localization accuracy and maximizing the number of transmitted bits in shortest time with respect to UAV's energy consumption. We demonstrate that the proposed method improves the average transmitted bits significantly, as well as the localization error of the network.*

- **Conference Papers** The following is a list of publications in refereed conference proceedings that originated from the main findings of this thesis. The conference papers [1] contain material not presented in this thesis.
- **(C1)** Arzhang Shahbazi and Marco Di Renzo. "Analysis of Optimal Altitude for UAV Cellular Communication in Presence of Blockage." 2021 IEEE 4th 5G World Forum (5GWF). IEEE, 2021.

**Abstract :** *In this paper, a novel framework for outage probability analysis consisting of unmanned aerial vehicles (UAV) base stations and ground users is proposed, which includes the blockage model for line-of-sight (LoS) and none-LoS (NLoS) probability and a tractable approach based on stochastic geometry. Specifically, a three dimensional (3D) LoS ball model is introduced to obtain the probabilistic propagation in UAV communication systems. By utilising this model, a tractable expression is derived for signal-to-noise ratio (SNR) outage probability. This approach leads to a closed-form expression for the optimal altitude of UAV which in turn helps to investigate the impacts of blockage height, density and length on the outage probability. Simulations are preformed to investigate the performance and accuracy of the proposed approach.*

- **(C2)** Arzhang Shahbazi and Marco Di Renzo. "Learning-based Localization of Mobile Users for Throughput Maximization in UAV Net-

works." 2021 IEEE 4th 5G World Forum (5GWF). IEEE, 2021.

**Abstract :** *In this paper, we design a new UAV-assisted communication system relying on the shortest flight path of the UAV while maximizing the amount of data transmitted to mobile devices. In the considered system, we assume that UAV does not have the knowledge of user's location except their initial position. We propose a framework which is based on the likelihood of mobile users presence in a grid with respect to their probability distribution. Then, a deep reinforcement learning technique is developed for finding the trajectory to maximize the throughput in a specific coverage area. Numerical results are presented to highlight how our technique strike a balance between the throughput achieved, trajectory, and the complexity.*

- (C3) Arzhang Shahbazi, Igor Donevski, Jimmy Jessen Nielsen and Marco Di Renzo. "Federated Reinforcement Learning UAV Trajectory Design for Fast Localization of Ground Users." EUSIPCO 2022, Special Session : AI/ML-based Disruptive Approaches for Next Generation Wireless Communication Systems

**Abstract :** *In this paper, we study the localization of ground users by utilizing unmanned aerial vehicles (UAVs) as aerial anchors. Specifically, we introduce a novel localization framework based on Federated Learning (FL) and Reinforcement Learning (RL). In contrast to the existing literature, our scenario includes multiple UAVs learning the trajectory in different environment settings which results in faster convergence of RL model for minimum localization error. Furthermore, to evaluate the learned trajectory from the aggregated model, we test the trained RL agent in a fourth environment which shows the improvement over the localization error and convergence speed. Simulation results show that our proposed framework outperforms a model trained with transfer learning by %30.*

## 2 - UAV for Next Generation of Cellular Communication - An Introduction



The use of flying platforms such as UAVs, popularly known as drones, is rapidly growing. In order to paint a clear picture on how UAVs can indeed be used as flying wireless base stations, in this chapter, we provide a comprehensive study on the use of UAVs in wireless networks. Specifically, with their inherent attributes such as mobility, flexibility, and adaptive altitude, UAVs grant several key potential applications in wireless systems. UAVs can be utilized as aerial base stations to enhance coverage, capacity, reliability, and energy efficiency of wireless networks. They also can operate as flying mobile terminals within a cellular network. Such cellular-connected UAVs can allow various applications ranging from real-time video streaming to item delivery. We study the main use cases of UAVs as aerial base stations and cellular-connected users. For each of the applications, we explore key challenges and fundamental problems.

## Sommaire

---

<b>2.1</b>	<b>UAV Aerial Base Station in 5G and Beyond</b>	<b>17</b>
2.1.1	Coverage and Capacity Enhancement for Wireless Cellular Networks . . . . .	17
2.1.2	UAVs Acting as Flying Base Stations for Disaster Scenarios . . . . .	18
2.1.3	UAV-Aided Terrestrial Networks for Information Transmission . . . . .	19
2.1.4	MIMO and Millimeter Wave Communications in 3D UAV Networks . . . . .	20
2.1.5	UAVs and IoT Communications . . . . .	21
2.1.6	Caching in UAV Base Stations . . . . .	22
2.1.7	Cellular-Connected Drones as User Equipments	23
2.1.8	Flying Ad-Hoc Networks With UAVs . . . . .	23
2.1.9	UAV Air-to-Ground Channel Modeling . . . . .	24
<b>2.2</b>	<b>Conclusion . . . . .</b>	<b>25</b>

---

## 2.1 . UAV Aerial Base Station in 5G and Beyond

In this section, we examine various aspects of UAV communication networks for 5G and beyond.

### 2.1.1 . Coverage and Capacity Enhancement for Wireless Cellular Networks

The hunger for high-speed wireless access has been constantly spreading, fueled by the rapid expansion of highly capable mobile devices such as smartphones, tablets, and more recently drone-UEs and IoT-style gadgets. By its very nature, the capacity and coverage of current wireless cellular networks have been broadly constrained, which led to the development of a plethora of wireless technologies that pursue to overcome this challenge [2]. These technologies, which consist of device-to-device (D2D) communications, ultra dense small cell networks, and millimeter wave (mmW) communications, are jointly realized as the center of next generation 5G cellular systems. Nonetheless, regardless of their invaluable benefits, those solutions have limitations of their own. For example, D2D communication will definitely demand better frequency planning and resource usage in cellular networks. Moreover, ultra dense small cell networks confront plenty of challenges in terms of back-haul, interference, and overall network modeling. Likewise, mmW communication is limited by blockage and high reliance on LoS communication to effectively deliver the possibility of high-speed, low latency communications. These challenges will be further intensified in UAV-UEs schemes [3].

The envision of a UAV carrying flying base stations as an necessary complement for such a heterogeneous 5G environment will allow overcoming some of the challenges of the existing technologies. Realizing LAP-UAVs can be a practical method for providing wireless coverage to geographical areas with limited cellular infrastructure. Furthermore, the utilizing UAV base stations is promising when deploying small cells for the particular purpose of servicing temporary events (e.g., sport events and festivals), is not economically viable, given the short period of time during which these events require wireless access. At the same time, HAP-UAVs can arrange a more long-term sustainable solution for coverage in such rural environments. On the other hand, mobile UAVs can provide on-demand connectivity, high data rate wireless service, and traffic offloading opportunity in hotspots and during temporary events such as football games or Presidential inaugurations [4]. For this purpose, *AT&T* and Verizon have recently declared several plans to use flying drones to provide temporarily boosted Internet coverage for college football national championship and Super Bowl. Evidently, flying base stations has the potential to become an important complement to ultra dense small cell networks.

Additionally, UAV-enabled mmW communications is a promising application of UAVs that can maintain LoS communication links to ground users. This solution is attractive for providing high capacity wireless transmission, while tackling the advantages of both UAVs and mmW links. Furthermore, incorporating UAVs with

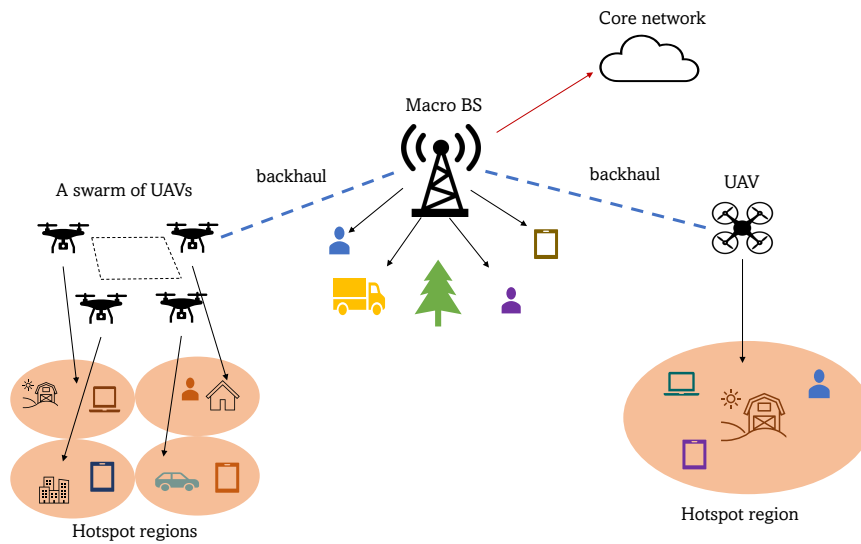


Fig. 2.1 – UAV communication networks over 5G.

mmW and potentially massive multiple input multiple output (MIMO) techniques can set up a whole new sort of dynamic, flying cellular network for providing high capacity wireless services, if well prepared and managed.

UAVs can also reinforce different terrestrial networks such as D2D and vehicular networks. For example, due to their mobility and LoS communications, drones can ease the rapid information dissemination amidst ground devices. Moreover, drones can possibly improve the reliability of wireless links in D2D and vehicle-to-vehicle (V2V) communications by exploiting transmit diversity. Specially, flying drones can aid in broadcasting common information to ground users, consequently decreasing the interference in ground networks by reducing the number of transmissions between users. Furthermore, UAV base stations can utilize air-to-air links to service other cellular-connected UAV-UEs, to mitigate the load on the terrestrial network. For the preceding cellular networking schemes, it is evident that the use of UAVs is quite logical because to their key features given in Tables III and IV such as agility, mobility, flexibility, and adaptive altitude. In fact, with benefiting from these unique features as well as establishing LoS communication links, UAVs can enhance the performance of existing ground wireless networks in terms of coverage, capacity, delay, and overall quality-of-service. These scenarios are certainly promising and one can see UAVs as being an integral part of beyond 5G cellular networks, as the technology blooms further, and new practical scenarios appear.

### 2.1.2 . UAVs Acting as Flying Base Stations for Disaster Scenarios

Natural disasters like floods, hurricanes, tornadoes, and severe snow storms usually bring devastating ramifications in many parts of the world. Amidst of wide-scale natural disasters and unexpected scenarios, the current terrestrial communication networks can be impaired or even completely broken, consequently becoming quite overloaded, as cleared out by the recent aftermath of floods in New York City subway stations. Specifically, cellular base stations and ground communications infrastructure can be often jeopardized during natural disasters. In such cases, there is a crucial need for public safety communications between first responders and victims for search and rescue operations. Thus, a robust, fast, and capable emergency communication system is necessary to facilitate effective communications

during public safety operations. In public safety scenarios, a reliable communication system will not only help to improve connectivity, but also it saves human lives. Correspondingly, FirstNet in the United States was set up to build a nationwide and high-speed broadband wireless network for public safety communications. The potential broadband wireless technologies for public safety cases involve 4G long term evolution (LTE), WiFi, satellite communications, and dedicated public safety systems such as TETRA and APCO25 [5]. Nonetheless, these technologies may not supply resilience, low-latency services, and swift adaptation to the environment during natural disasters. Thus, utilizing UAV-based aerial networks is a promising solution to facilitate fast, adaptive, and reliable wireless communications in public safety scenarios. Since UAVs do not demand highly constrained and expensive infrastructure (e.g., cables), they can effortlessly fly and adaptively change their positions to supply on-demand communications to ground users in emergency situations. Moreover, because of the unique features of UAVs such as mobility, flexible deployment, and rapid reconfiguration, they can establish on-demand public safety communication networks effectively. For example, UAVs can be expanded as mobile aerial base stations in order to deliver broadband connectivity to areas with damaged terrestrial wireless infrastructure. Furthermore, flying UAVs can repeatedly maneuver to bring full coverage to a given area within a minimum possible time. Thus, utilizing UAV-mounted base stations can be a suitable approach for supplying fast and ubiquitous connectivity in public safety scenarios.

### **2.1.3 . UAV-Aided Terrestrial Networks for Information Transmission**

Considering mobility and LoS opportunities of UAVs, they can support terrestrial networks for diffusing information and connectivity enhancement. For example, to aid a D2D network or a mobile ad-hoc network, UAVs can be utilized as flying base stations for information distribution among ground users. On the other hand, D2D networks can develop a successful approach for offloading cellular data traffic and enhancing network capacity and coverage, while their performance is restricted because of the short communication range of devices as well as increasing interference [6]. Generally, D2D networks can be considered as an effective approach for offloading cellular data traffic and improving network capacity and coverage, however their performance is restricted due to the short communication range of devices as well as potentially accumulating interference. Thus for these kind of scenarios, drones can facilitate rapid information circulation by intelligently broadcasting common files among ground devices. For instance, UAV-assisted D2D networks can help the or evacuation messages in public safety and rapid spread of emergency situations. Besides, flying UAVs can play a key role in vehicular networks (i.e., V2V communications) by disseminating safety information across the vehicles. Moreover, drones can increase the reliability and connectivity of D2D and V2V communication links. On the one hand, utilizing UAVs can reduce the interference by decreasing the number of required transmission links between ground

devices. On the other hand, mobile UAVs can bring opportunities for transmit diversity, consequently increasing reliability and connectivity in D2D, ad-hoc, and V2V networks. One of the practical approaches for keeping such UAV-assisted terrestrial networks is to leverage clustering of ground users. Thus, a drone can perform inside each of the clusters by directly communicate with the head of the clusters and the multi-hop communications. Here, by applying efficient clustering approaches and exploiting drones mobility, the connectivity of terrestrial networks can be substantially improved.

#### 2.1.4 . MIMO and Millimeter Wave Communications in 3D UAV Networks

Drones can be viewed as flying antenna systems that can taken advantage of for performing massive MIMO, 3D network MIMO, and mmW communications due to their aerial positions and their capability to deploy on demand at specific locations. For instance, in recent years, by exploiting both the vertical and horizontal dimensions in terrestrial cellular networks, a considerable interest in the use of 3D MIMO, also known as full dimension MIMO has risen. Specifically, 3D beamforming permits the creation of separate beams in the three-dimensional space at the same time, consequently decreasing inter-cell interference. While compared to the conventional two-dimensional MIMO, 3D MIMO approaches can provide higher overall system throughput and can aid a higher number of users [7]. In principle, for scenarios in which the number of users is high and they are spread in three dimensions with different elevation angles with regard to their coverage base station, 3D MIMO is more suitable. Because of the high altitude of Drone-carried flying base stations, ground users can be efficiently distinguishable at different heights and elevation angles measured with respect to the Drone. Moreover, LoS channel conditions in UAV-to-ground communications enable practical beamforming in 3D space for both azimuth and elevation angles. Consequently, Drone-carried flying base stations is a solid solution for employing 3D MIMO. Moreover, by utilizing drone-based wireless antenna array, a unique opportunity for airborne beamforming is provided. For effectively giving service to ground users in downlink and uplink scenarios, a UAV antenna array whose elements are single-antenna drones can become advantageous for MIMO and beamforming. In comparison to

In comparison with conventional antenna array systems, a UAV-based antenna array has the following benefits :

- The number of antenna elements is not limited by space constraints,
- Beamforming gains can be increased by dynamically adjusting the array element spacing
- The mobility and flexibility of UAVs permit efficient mechanical beamsteering in any 3D direction.

Furthermore, by utilizing a large number of small drones with an array formation, unique massive MIMO opportunists can be provided. Such drone-based massive antenna array can form any arbitrary shape and effectively perform beam-

forming. Drones can also be a key enabler for mmW communications. On the one hand, to establish LoS connections to ground users, the drones equipped with mmW capabilities can decrease propagation loss while operating at high frequencies. On the other hand, one can exploit advanced MIMO approaches such as massive MIMO in order to operate mmW communications by utilizing small-size antennas (at mmW frequencies) on drones. Meanwhile, to create reconfigurable antenna arrays in the sky, swarms of UAVs can be used.

### **2.1.5 . UAVs and IoT Communications**

Wireless networking technologies are developing exceptionally into a massive IoT environment that should incorporate a heterogeneous mix of devices varying from vehicles to conventional smartphones and tablets, sensors, wearables, and naturally, UAVs. For attaining the much needed applications of the IoT such as smart cities infrastructure management, healthcare, transportation, and energy management, it is necessary to have effective wireless connectivity among a massive number of IoT devices that should accurately convey their data, consistently at high data rates and ultra low latency. The extensive nature of the IoT needs a considerable rethinking to the course in which conventional wireless networks (e.g., cellular systems) function [8]. For example, energy efficiency, ultra low latency, reliability, and high-speed uplink communications become main challenges in an IoT environment that are not generally as critical in conventional cellular network use cases. Specifically, IoT devices are restricted by battery and are mostly unable to transmit over a long distance because of their energy limitations. For example, in areas which encounter regular poor coverage by terrestrial wireless networks, battery-restricted IoT devices may not be able to send their data to far-off base stations due to their power limitations. Moreover, due to the diverse applications of IoT devices, they may be installed in environments with no terrestrial wireless infrastructure such as mountains and desert areas. In this respect, the use of mobile drones is an encouraging approach to a number of challenges linked to IoT networks. Drones can be arranged as flying base stations to grant reliable and energy-efficient uplink IoT communication in IoT-centric scenarios. Due to the aerial nature of the drones and their high altitude, they can be efficiently installed to decrease the shadowing and blockage effects as the major cause of signal attenuation in wireless links. Consequently, with the effective placement of drones, the communication channel between IoT devices and drones can be substantially improved. Thus, battery-restricted IoT devices will need a substantially lower power to transmit their data to drones. Particularly, drones can be situated based on the position of IoT devices allowing those devices to successfully connect to the network using a minimum transmit power. Furthermore, drones can also provide service to massive IoT systems by adaptively updating their positions based on the activation pattern of IoT devices. This is in contrast to using ground small cell base stations which may need to be considerably increased to serve the predicated number of devices in the IoT. Thus, the connectivity and energy efficiency of IoT networks can be substantially

enhanced by taking advantage of unique features of drones.

### 2.1.6 . Caching in UAV Base Stations

It has been illustrated that Caching at small base stations (SBSs) is a promising approach to improve the communication system throughput and to reduce the transmission delay. However, it may not be effective to cache at traditional static ground base stations for covering mobile users in the case of recurrent handovers [9]. For this reason, when a user moves to a new cell, its corresponding demanded content may not be available at the new base station and, thus, the users may not achieve a proper coverage. To effectively give service to mobile users in these cases, each demanded content needs to be cached at different base stations which is not practical due to the signaling overheads and additional storage usages. Consequently, to increase the caching efficiency, it is mandatory to scatter flexible base stations that can track the users' mobility and effectively transmit the demanded contents. In consequence, one can foresee futuristic scenarios in which UAVs, operating as flying base stations, can dynamically cache the popular contents, track the mobility pattern of the matching users and, afterwards, effectively serve them. In fact, using cache-enabled drones for the case of traffic offloading in wireless networks is a promising method.

Cache-enabled UAVs can be optimally moved and positioned to deliver requested services to users by utilizing the user-centric information, like content request distribution and mobility patterns. Moreover, another advantage of distributing cache-enabled content is that the caching complexity can be reduced in comparison with a conventional static SBSs. For instance, when a mobile user moves to a new cell, its demanded content needs to be stored at the new base station. However, cache-enabled drones are capable of tracking the mobility pattern of users and, so, the content reserved at the drones will no longer require more caching at SBSs. In terms of practicality, a cache-enabled drone system and a central cloud processor can utilize varied user-centric information together with users' mobility patterns and their content demand dissemination to handle the drone installation. Indeed, these user-centric data can be learned by a cloud center by using any previous available users' information. Subsequently, to give service to ground users, the cloud center is able to effectively determine the position and mobility paths of cache-enabled drones. Consequently, a reduction in the overall overhead of updating the cache content is achieved. Also, content requests of a mobile user may need to be dynamically reserved at different SBSs, when caching with SBSs is performed. However, cache-enabled UAVs are capable of tracking the mobility pattern of users and bypass the regular updating of data demands of mobile users. This results to mobile cache-enabled UAVs that estimate the mobility patterns and content request information of users, thus, ground users can be efficiently receive communication service.

### 2.1.7 . Cellular-Connected Drones as User Equipments

In general, UAVs can operate as users of the wireless infrastructure. Specifically, drone-users can be surveillance, utilized for package delivery, remote sensing, and virtual reality applications. In fact, cellular-connected drones is envisioned to be a key enabler of the IoT. One of the recent applications for delivery-based drones is the Amazon's prime air drone delivery service, and autonomous delivery of emergency drugs. The major benefit of drone-users is their capability to quickly move and optimize their path to complete their objectives. To properly use UAVs as user equipments such as cellular connected drone-UEs, it is necessary to have reliable and low-latency communication between UAVs and ground BSs [10]. Indeed, to aid a large-scale deployment of UAVs, a reliable wireless communication infrastructure is necessary to efficiently control the drones' movement while supporting the traffic emerging from their application services. In addition to their need for ultra low latency and reliability, when used for surveillance purposes, drone-UEs will need a high-speed uplink connectivity from the terrestrial network and from other UAV-BSs. For this reason, modern cellular networks may not be able to fully incorporate drone-UEs as they were planned for ground users whose operations, mobility, and traffic characteristics are considerably varied from the drone-UEs. It should be noted that there are a numerous key differences between drone-UEs and terrestrial users. Firstly, drone-UEs usually encounter different channel conditions because of nearly LoS communications between ground BSs and flying UAVs. Thus, in this scenario, one of the major challenges for incorporating drone-UEs is significant LoS interference originated by ground BSs. Secondly, in contrast to terrestrial users, the on-board energy of drone-UEs is highly restricted. Thirdly, drone-UEs are in principle more dynamic than ground users as they are able to continuously fly in any angle. Consequently, supporting cellular-connected drone-UEs in wireless networks will establish novel technical challenges and design difficulties.

### 2.1.8 . Flying Ad-Hoc Networks With UAVs

Another important use cases of drones is in flying ad-hoc networks (FANETs), where multiple drones communicate in an ad-hoc manner. With respect to their mobility, no need for central control, and self-organizing nature, FANETs can establish the connectivity and communication range at geographical areas with restricted cellular infrastructure. Also, FANETs are crucial aspects in different applications such as traffic monitoring, remote sensing, border surveillance, disaster management, agricultural management, wildfire management, and relay networks. Specifically, a relaying network of drones provide communication links among a remote transmitter and receiver that are not able to have direct communication because of the obstacles or the long distance between them. In comparison with a single drone, a FANET with multiple small drones has the various advantages, such as :

- The coverage of FANETs can be easily increased by adding new drones and adopting optimal dynamic path planing method.



- The installation and maintenance cost of small drones is lower than the cost of a large drone with complex hardware and heavy payload.
- In FANETs, if one drone is out of service (due to weather conditions or any shortcomings in the drone system), FANET missions can still be carried on with the rest of flying drones. This kind of flexibility is not included in a single drone system.

### 2.1.9 . UAV Air-to-Ground Channel Modeling

Wireless signal propagation is impacted by the environment between the transmitter and the receiver. The air-to-ground (A2G) channel characteristics substantially vary from classical ground communication channels which, in turn, is capable of deciding the performance of UAV-based wireless communications in terms of coverage and capacity. Moreover, in comparison with air-to-air communication links that encounter dominant LoS, A2G channels are more prone to blockage. It is evident that the optimal design and deployment of drone-based communication systems demand utilizing a detailed A2G channel model. Although the ray-tracing method is a reasonable approach for channel modeling, it lacks satisfactory precision, specifically at low frequency operations. An accurate A2G channel modeling is crucial particularly when using UAVs in applications such as coverage improvement, cellular-connected UAVs, and IoT communications. The A2G channel characteristics is substantially different than ground communication channels. To be more specific, any movement or vibration by the UAVs can impact the channel characteristics. Also, the A2G channel is highly dependent on the operating altitude and type of the UAV, elevation angle, and type of the propagation environment. Consequently, finding a comprehensive channel model for UAV-to-ground communications requires exhaustive simulations and measurements in diverse environment settings. Furthermore, the effects of a UAV's altitude, antennas' movements, and shadowing caused by the UAV's body should be attained in channel modeling. Clearly, capturing such factors is challenging in A2G channel modeling.

One of the most widely preferred A2G path loss model for low altitude platforms is presented in [11] and, thus, we describe it in more detail. As shown in [11], the path loss between a UAV and a ground device depends on the position of the UAV and the ground device as well as the type of propagation environment (e.g., rural, suburban, urban, high-rise urban). In this case, based on the environment setting, A2G communication links can be either LoS or NLoS. It should be noted that without any further information about the exact positions, heights, and number of the obstacles, one should acknowledge the randomness associated with the LoS and NLoS links. Consequently, many of the existing literature on UAV communication adopted the probabilistic path loss model given in [11]. As discussed in this work, the LoS and non-LoS (NLoS) links can be treated separately with different probabilities of occurrence. The probability of occurrence is a function of the environment, density and height of buildings, and elevation angle between UAV and ground device. The general probabilistic LoS model is based on the common

geometrical statistics of various environments offered by the International Telecommunication Union (ITU-R). Specifically, for various types of environments, the ITU-R provides some environmental-dependent parameters to determine the density, number, and height of the buildings (or obstacles). For example, the buildings' heights can be modeled using a Rayleigh distribution as [3] :

$$f(h_B) = \frac{h_B}{\lambda^2} \exp \frac{-h_B}{2\lambda^2} \quad (2.1)$$

where  $h_B$  is the height of buildings in meters, and  $\lambda$  is an environmental-dependent parameter [2]. It is clear that because of the uncertainty associated with the height of buildings, one should consider a probabilistic LoS model while designing UAV-based communication systems. Thus, using the statistical parameters provided by ITU-R, other works such as [2] and [11] derived an expression for the LoS probability, which is given by :

$$P_{LoS} = \frac{1}{1 + C \exp -B[\theta - C]} \quad (2.2)$$

where  $C$  and  $B$  are constant values that depend on the environment (rural, urban, dense urban, or others) and  $\theta$  is the elevation angle in degrees. Clearly,  $\theta = \frac{180}{\pi} \times \sin^{-1}(\frac{h}{d})$ , with  $h$  being the UAV's altitude, and  $d$  is the distance between the UAV and a given ground user. For this scenario, the NLoS probability will be  $P_{NLoS} = 1 - P_{LoS}$ . We note that the probabilistic path loss model in (2.2) is an example of existing A2G channel models such as the one proposed by the 3GPP [74]. Equation (2.2) captures the fact that the probability of having LoS connection between the aerial base station and ground users is an increasing function of elevation angle. According to this equation, by increasing the elevation angle between the receiver and the transmitter, the blockage effect decreases and the communication link becomes more LoS. It is worth noting that the small-scale fading in A2G communications can be characterized by Rician fading channel model. The Rician K-factor that represents the strength of LoS component is a function of elevation angle and the UAV's altitude.

## 2.2 . Conclusion

In this chapter, we have provided a comprehensive study on the use of UAVs in wireless networks. We have investigated the main use cases of UAVs as aerial base stations and cellular-connected users. For each of the applications, we have explored key challenges and fundamental problems.



### 3 - Machine Learning for UAV-Enabled Wireless Networks

In this chapter, motivated by a wide set of new applications that can gain assistance from drone networks, such as smart cities and aerial base stations deployment, we cover in detail the new research directions when ML techniques are utilized to increase the performance of UAV networks. Recently, AI is growing rapidly and has been very successful, specifically due to the massive amount of the available data. As a result, a significant part of the research community has started to integrate intelligence at the core of UAVs networks by applying AI algorithms in solving several problems in relation to drones. In this chapter, we start by preparing an extensive overview of unsupervised and supervised ML techniques. Then we introduce RL in details, that have been broadly applied in UAV networks. Then, we discuss FL principles and advantages and where a FL approach can be used in the field of UAV networks.

## Sommaire

---

<b>3.1</b>	<b>Machine Learning for UAVs : An Introduction</b>	<b>29</b>
<b>3.2</b>	<b>Supervised and Unsupervised Learning for UAVs</b>	<b>30</b>
3.2.1	Supervised Learning Overview	30
3.2.2	Unsupervised Learning Overview	31
3.2.3	Practical Issues of ML Implementation	32
<b>3.3</b>	<b>Reinforcement Learning for UAVs : An Overview</b>	<b>33</b>
3.3.1	Deep Reinforcement Learning (DRL)	34
3.3.2	Q-Learning	35
3.3.3	Q-Learning Overview	35
3.3.4	Update Rule	36
3.3.5	The Exploration/Exploitation Trade-Off	36
3.3.6	Limitation of RL	36
3.3.7	Federated Learning for UAVs	37
3.3.8	Transfer Learning for UAVs	39
<b>3.4</b>	<b>Conclusion</b>	<b>40</b>

---

### 3.1 . Machine Learning for UAVs : An Introduction

The future of UAVs are envisioned as one of the promising technologies for the next-generation wireless communication networks. Their mobility and their capability to maintain LoS links with the ground users made them as a key solution for many potential applications. Similarly, artificial intelligence (AI) is expanding swiftly for the past decade and has been very successful, especially because of the massive amount of the available data. Therefore, an important aspect of the research community has been initiated to incorporate intelligence at the core of drone networks by applying AI algorithms in solving various problems in relation to UAVs.

In summary, AI is one of the trending sectors that brings intelligence to machines and makes them capable to complete objectives even better than a human can do. It is envisioned that bringing together the advantages of using AI within drone networks is a challenging and fascinating idea at the same time. Despite the fact that conventional approaches illustrated a major success in solving various problems in this sector, it is still interesting to study whether ML can contribute to more powerful and accurate solutions. It is worth opting for AI-assisted approaches given the unprecedented success realized by ML especially in decision-making problems, even when moving from classical methods to intelligent approaches needs sacrificing interpretability and tractability in some scenarios.

Nonetheless, the research community believes that intelligent approaches are not always guaranteed to outperform classical methods, instead, classical approaches might propose simple and powerful solutions in some cases. In fact, this duality is a proof that investigating the use of AI for the set of specific problems related to UAV networks is worth pursuing. In the past, UAVs were studied originally to be controlled fully manually by a person, however, with the recent evolution of AI, it became a trend to prepare smart drones in the markets. In light of this, AI can utilize the data accumulated by drone sensors to execute varied tasks. Also, AI is able to play a crucial role in resource management for drones to increase energy efficiency. The design of drones path planing and positioning are also subject to AI advancement by equipping the drone with the capability to dodge obstacles and design its path automatically. For example, in recent years, the drones that can follow users have seen a huge success in the markets. This kind of UAV provides high quality video footage by following and filming its owner while equipped with dynamic and intelligent obstacle avoidance and target tracking algorithms. Furthermore, comprehensive applications can be modernized in this context such as traffic management, surveillance, and landing site estimation. Imaging can also be enhanced for drones by applying the existing state of the art techniques related to computer vision for drones imaging.

In summary, it is the subgroup of AI that set up a computer to perform tasks accurately based on the experience gained by learning from some previous trials. Indeed, ML has been very advantageous over the last decade due to the large

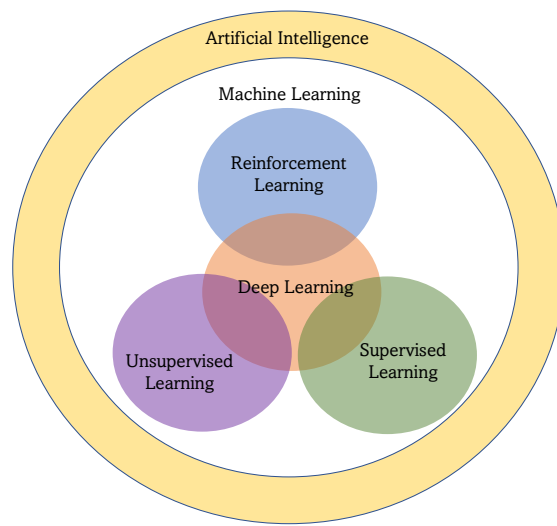


Fig. 3.1 – Machine learning overview.

available amount of data and powerful computers that were not accessible before. For this reason, research is now directed towards applying ML in drone-based problems. The field of ML is split into different categories of problems, for example, it can be divided to supervised learning problems, unsupervised learning problems, and RL-based problems. In the following, we separate between the supervised and the unsupervised learning and focus our attention to unsupervised learning and specifically RL.

### 3.2 . Supervised and Unsupervised Learning for UAVs

The areas of ML can be divided into various categories of problems, for example, it might be divided as shown in Fig. 3.1 to supervised learning problems, unsupervised learning problems, and RL-based problems. In the following, we distinguish between the supervised and the unsupervised learning and discuss advantages and limitations on each of them.

#### 3.2.1 . Supervised Learning Overview

In supervised learning, the provided data is labeled, in another manner, we provide for each data entry the ground-truth value, so that the algorithm utilizes these values to learn how to make a decision for a new unlabeled entry. For instance, one can predict a drone price from its characteristics. In this example, we need to grant the algorithm with a set of training data that includes each UAV characteristics and its corresponding label (the price). The dataset is often divided into a training set and a test set. The training set is utilized to learn the relationship between the input and the output and the test set is used to validate the model by measuring its precision. The supervised problems are usually divided into either regression problems or classification problems. Regression problems maintain continuous output values (e.g., predicting a price). On the other hand, classification problems provide discrete values specifying to which class the input belongs (e.g., classify benign or malignant cancer disease). In the following, we demonstrate the most well-known ML algorithms for supervised and unsupervised learning. We also focus on the algorithms that are used to solve the UAV-related problems in the literature.

Some Supervised Algorithms and NN Architectures :

- Combined Classification and Regression Algorithms : There are several supervised algorithms that can be utilized either for classification or regression. For example, Support Vector Machine (SVM) can do both the tasks, decision trees also can be formulated to solve regression or classification depending on the use case.
- Regression Algorithms : There exist algorithms that carry out pure regression objectives by predicting continuous value output. For example, we can point out two classical algorithms in ML which are linear regression and logistic regression.
- Classification Algorithms : It makes sense to talk about pure classifiers in ML. Although it is mentioned in some references that Naive Bayes classifier with “some modification” can be used for regression, we present it as a pure classifier example since it was derived initially for classification based on the probabilistic Bayes theorem.
- Multi Layer Perceptron (MLP) : To imitate the biological human neural networks, ANNs are mathematically formulated for ML. ANNs are built with several partially-connected nodes denoted by perceptrons and grouped into different layers. Each perceptron is responsible for processing information from its input and delivering an output. Also, MLP is the simplest form of an ANN that consists of one input layer, one or more hidden layers, and an output layer where a classification or regression task is accomplished.
- Convolutional Neural Networks (CNNs) : CNN is another type of ANN designed initially for computer vision tasks. A CNN usually takes an image as an input, assigns learnable weights and biases that are updated according to a specific algorithm. The CNN architecture is characterized by the convolutional layers which extract high-level features from the image that will be used later. In a typical CNN architecture a feature extraction is achieved in the first convolutional layers and classification is achieved via a fully connected layer.
- Recurrent Neural Networks (RNNs) : When the data is sequential in nature, RNNs will be used to solve the problem. For example, we can consider a text speech, a video, or a sound recording. RNNs are widely used in natural language processing (NLP), in speech recognition, and for generating image description automatically. The RNN architecture is similar to a regular neural network, only it contains a loop that allows the model to carry out results from previous neurons. RNN in its simplest form is composed of an output containing the prediction and a hidden state that represents the short-term memory of the system.

### 3.2.2 . Unsupervised Learning Overview

Dissimilar to supervised learning, the unsupervised learning does not utilize labeled data, instead, it searches for some underlying structure or hidden pattern in



the data and uncover it. For example, clustering the data, reducing data dimensionality, and data generation are considered typical tasks for unsupervised learning. In the following, we provide some classical unsupervised algorithms. Unsupervised Algorithms and NN Architectures :

- Clustering Algorithms : There are handful of popular clustering algorithm in ML. Here, we only mention K-means, Gaussian Mixture Modeling (GMM), DBSCAN, and agglomerate Clustering. Some of these algorithms are density-based algorithms such as DBSCAN, and others carry out hard association such as K-means. It should be noted that the GMM is a probabilistic model that uses soft association rule.
- Dimensionality Reduction Algorithms : Dimensionality reduction is a common method in ML consisting of transforming data from a high-dimensional space representation to a lower-dimensional space. In this context, we mention some spectral-based techniques such as autoencoders (AEs) which are a type of neural networks used to learn a representation of the data and encode it. Particularly, the architecture of an AE is remarkably simple. Also, we can mention another spectral-based algorithm which is principal component analysis (PCA) as a popular dimensionality reduction technique.
- Generative Adversarial Networks (GANs) : GANs are algorithmic architectures that use two neural networks in order to generate new, synthetic instances of data that can pass for real data. They are used widely in image generation, video generation, and voice generation.

### 3.2.3 . Practical Issues of ML Implementation

Due to the limited computing capacity onboard, the application of ML methods in UAV-based networks can be restricted. In fact, most commercially available UAVs are not equipped with the sophisticated processors that are essential to execute heavy ML algorithm. We must consider the drone's weight and power consumption even if plan to equip it with powerful CPU and GPU. Consequently, an identical issue will persist due to the power constraints of UAVs. One approach to solve this problem is to utilize the cloud to train models and make inferences at the UAV level.

Nonetheless, this approach will raise the communication costs, which in turn will bring us back to the energy constraint issue, due to the fact that the UAV have to communicate back and forth with the cloud. Thus, in different and encouraging approach, we can run the ML onboard, but this time adjusting the ML algorithms to the UAV's limited capacity. This method points us to a novel field usually referred to as on-device learning dedicated to constrained devices. Recently, many researchers have investigated device learning by addressing lightweight ML algorithms and examining the various ML and DL algorithms in terms of complexity and resource consumption. Also, one can propose a solution to address the execution of ML onboard based on FL. It contains executing ML in a decentralized way by sending model updates over networks instead of sharing raw data. We plan to

cover and discuss this technique in the following sections and chapters. In addition to the hardware and software restriction of drones mentioned above, the practical use of ML in UAV networks still faces other significant barriers related to existing rules and regulations. Although research is designed at partially or even fully autonomous UAV applications, most existing regulations do not allow such operations in practice. For instance, the U.S. Federal Aviation Administration (FAA), in its latest regulation did not lay out a single point concerning autonomous UAVs. They rather focused on regulations dedicated to the human operators who control a drone. However, it is important to mention that there is still a strong anticipation for autonomous UAVs to see the light of day. In fact, unlike the FAA, the European Aviation Safety Agency (EASA), in its latest regulation acknowledge the existence of autonomous drone operations by including them and classifying them in various classes according to the risk level of the application. Without doubt, this will propose new opportunities for innovative UAV solutions based on ML and AI in principle. In summary, it is vital to harmonize and unify the drone regulations around the world, as this will motivate future research in this area.

### 3.3 . Reinforcement Learning for UAVs : An Overview

RL is the area of ML dedicated to making decisions in a well-defined environment. Generally, a reinforcement problem always has 5 main elements as shown in the Fig. 4 :

- The Agent : An entity that can take an action denoted by  $A_t$  and receives a reward  $R_t$  accordingly.
- The Environment : A representation of the real-world in which the agent operates.
- The Policy : It is the mapping of each state  $S_t$  to an action  $A_t$ . We usually denote a policy by  $\pi$ .
- The Reward Signal : The feedback that the agent receives after performing an action. It is denoted on the Fig. 4 by  $R_t$ .
- The Value Function : It represents how good a state is, hence it is the total expected future rewards starting from a given state. A value function is usually denoted by  $V(s)$  where  $s$  is the state that we are interested in. Mathematically, it is formulated as follows :  $V(s) = E(G_t)$ , where  $G_t$  is the discounted sum of future rewards :  $G_t = \sum_t \lambda^{t-1} R_t, \lambda \in [0, 1]$ .

The aim is to choose correct actions (or policy) that maximizes a predefined reward function, which should be suitable to the type of RL problem. In addition to the 5 elements of RL mentioned above, another element can be expressed in some scenarios, which is the model. Conditional to its presence or not, RL problems can be branched into two main categories which are the model-based RL and the model-free RL. In the following, we differentiate between these two areas.

The model-based RL problem utilize a model as the sixth element to resemble

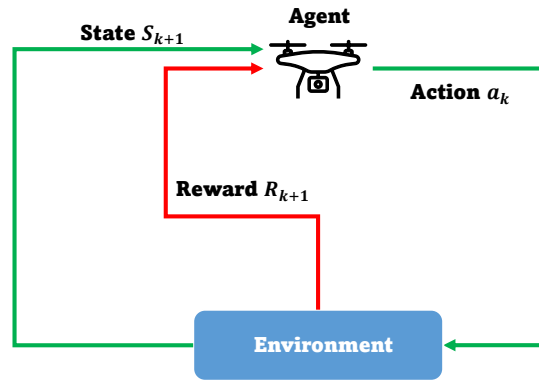


Fig. 3.2 – Reinforcement learning elements.

the behavior of the environment to the agent. Thus, the agent is capable to estimate the state and the action for time  $T + 1$  given the state and the action at time  $T$ . At this level, supervised learning could be a powerful tool to do the prediction work. Thus, dissimilar to the model-free RL, in model-based RL, the update of the value function is based on the model and not on any experience.

In model-free RL problems, the agent cannot predict the future and this is the main difference with the model-based RL framework explained previously. The actions are rather based on the trials and errors, where the agent, for example, can search over the policy space, calculate the different rewards, and decide finally an optimal reward. A common classic example for model-free RL is the Q-learning technique where it estimates the optimal Q-values of each action and reward and picks the action having the highest Q-value for the current state. In short, differentiating between model-based and model-free RL problems is a simple task. Just ask yourself the following question : Is the agent able to predict the next state and action, if the answer is yes then you are dealing with a model-based RL, alternatively, it is more likely a model-free RL problem.

### 3.3.1 . Deep Reinforcement Learning (DRL)

Even though RL had great accomplishment in solving different decision-making problems, it displayed a restricted performance in solving complex problems, specifically when using large action and state space [12]. Consequently, DRL began to gain a huge momentum in solving complex problems, especially after beating humans in many complex games, for example, chess and Go. The idea and novelty that lies behind this achievement is consist of approximating the states by the use of neural networks. This is the reason which makes the agent capable to deal effectively with unseen situations, in contrast to the classical RL method. Between the many algorithms introduced in the literature, in the following, we examine the most popular ones.

Deep Q Network (DQN) : DQN is the primary algorithm introduced in the context of DRL. For understanding the key concepts of DQN, a basic knowledge of Q-learning algorithm is recommended. It should be noted that DQN is introduced

as an improvement to Q-learning which utilizes a discrete state and action space in order to build the Q-table [13]. On the other hand, the Q-values of the DQN are approximated using ANN by stocking all the previous agent experience in a dataset and then feeding it to the ANN to generate the actions based on minimizing a predefined loss function derived from the Bellman equation. It should be also noted that the fact that the idea of DQN is inspired by Neural Fitted Q-learning (NFQ), that was suffering from overestimation problems and instabilities in the convergence [14]. There are many other improved variations of DQN such as double DQN, dueling DQN, and distributional DQN. Regardless of the phenomenal success of DQN, specifically when it was historically tested on ATTARI games, it has its own limitations such as the fact that it cannot deal with continuous space action and cannot utilize stochastic policies.

Deep Deterministic Policy Gradient (DDPG) : To overcome the limitation of discrete actions, Deterministic Policy Gradient (DPG) algorithm was primary introduced in Deepmind's publication in 2014 based on an Actor-Critic off policy method. For the sake of simplicity, lets say that Actor-Critic approaches are in principle composed mainly of two parts : a Critic that estimates either the action-value or the state-value and an Actor that updates the policy in the direction proposed by the Critic. Later on, in 2015, and based on the DPG algorithm, a new DRL algorithm called the Deep Deterministic Policy Gradient (DDPG) algorithm was proposed. DDPG is a model-free, off-policy technique that is based on Actor-Critic algorithm. In summary, DDPG is a DRL algorithm that aids the agent to find an optimal strategy by maximizing the reward return signal. The major advantage of this algorithm is that it functions well on high-dimensional/infinite continuous action space.

### 3.3.2 . Q-Learning

Motivated by its popularity among RL algorithms, we introduce Q-learning which is a classical free-model RL algorithm. Our intention in this section is to provide a comprehensive and practical explanation on how RL can be used in UAV path planning problems. We remain to a basic example where a drone is flying at a fixed altitude and learn how to reach a given target while achieving its designed objective.

### 3.3.3 . Q-Learning Overview

Q-learning algorithm is based on the Q-table utilized to choose actions for the agent at each step. The table is composed of the combination of every state with every possible action and thus its dimension is  $|States| \times |Actions|$ . The Q-table is used to store and update the maximum future reward indicated by  $Q(state_i, action_j)$  which is the  $(i_{th}, j_{th})$  entry of the Q-table. This Q-table has its important role in the Q-learning algorithm due to its application for determining which action should the agent picks such that the expected future reward is maximized.

### 3.3.4 . Update Rule

The update of the Q-table is done using a fundamental equation in RL which is the Bellman equation :

$$Q_{new}(s_t, a_t) = (1 - \alpha)Q_{old}(s_t, a_t) + \alpha(R_{t+1} + \gamma \max_a(Q(s_{t+1}, a))) \quad (3.1)$$

where  $s_t$ ,  $a_t$  are respectively the state and the action taken at time  $t$ ,  $\alpha$  is the learning rate, which allows the old value of the Q-table to influence current updates,  $\gamma$  is the discount factor, which is a measure of how future rewards will affect the system. After every picked action, the agent updates its Q-table values using (3.1), afterwards, at a given state, it selects the action having the highest Q-value.

### 3.3.5 . The Exploration/Exploitation Trade-Off

One of the basic concepts for RL, which is present also in Q-learning, is the exploration/exploitation trade-off. To explain this concept, let's uncover how the agent will succeed in reaching its objective. At the beginning, the agent makes a random step in the environment, then it starts updating the Q-table (initialized with zeros for instance) according to (3.1). Nonetheless, if the agent exclusively uses the Bellman equation, it is likely that it will be stuck in a good state forever, while better states exist on the environment. It is comparable to a case of an optimization process that is stuck in a local minimum or maximum while better solutions still exist by exploring the environment. To solve this problem, the exploitation/exploration trade-off is introduced. This concept propose randomness into the system so that the agent at each step can either exploit the environment by selecting actions that maximize the Q-values of the Q-table, or explore the environment by selecting some random actions. The parameter that usually corresponds to the probability threshold for exploration is described by  $\epsilon$ . In our implementation in the following chapters, we utilized a decay method that reduces the value for epsilon at each episode so that we encourage exploration at the beginning of the process, commonly known as early exploration, and then prioritize exploitation so that the agent can use the learned paths effectively.

### 3.3.6 . Limitation of RL

Dissimilar to supervised learning, RL is the area of ML that does not require the power of data to learn a new task. It rather uses the so-called "trial and error" methodology based on an agent's past experiences. In principle, this fact makes RL an extremely robust tool, specifically for done-based problems such as finding trajectory, resource management, and scheduling, where information is sometimes incomprehensible. Moreover, RL can echo supervised learning in one single point which is the objective of achieving full autonomy within a drone network by equipping drones with the capability to autonomously make decisions in a real-time

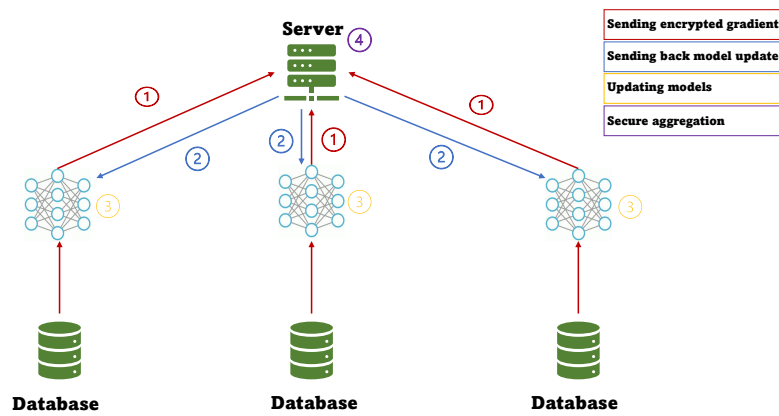


Fig. 3.3 – Federated learning architecture.

manner. In general, RL has shown its effectiveness by excelling in various problems and games, for instance, beating the world’s top chess grandmasters. Nonetheless, many scholars still hesitate about the applicability of RL in real-world tasks, specifically for autonomous flying or autonomous driving problems.

It should be noted that the difficulty of having a perfect perception of the environment from the agent’s perspective is combined with complexity of exploration/exploitation dilemma. To be more specific, let’s take the example of planning a path for a drone to reach a location in a fully autonomous control fashion. So, in order to apply RL, the drone (agent) needs to do exploration to discover its surroundings and learn how to interact. However, this is rather impossible for a high-dynamic and stochastic environments. In other words, the exploration task is fairly limited by the complexity of the environment and the cost of a drone crash. As for regulations, in many parts of the world, it is not allowed to use autonomous drones. For instance, utilizing UAVs for delivery has been completely excluded in the latest FAA regulations, as the new rules necessitate that the drone should always be in the operator’s field of sight, thus it is in contradiction with UAV delivery applications. For that reason, such rules can restrict the progress done so far in RL for different drone applications. On the other side, various large companies and research initiatives have been working on presenting alternatives to RL such as Evolution Strategies (ES) introduced by OpenAI. For conclusion, even if RL is not the ultimate solution for all drone-based tasks, it can be applied for some of them, which is illustrated by the numerical results of many research papers covered so far in this field.

### 3.3.7 . Federated Learning for UAVs

In previous section, we have covered a number techniques that could contribute to the development of intelligent UAV networks. Nonetheless, some of the algorithms discussed previously have certain drawbacks when incorporating with drone systems. Particularly, the limited computing capacity on-board of UAVs is one the major concerns. Thus, one can question the applicability of AI in UAV networks in a realistic scenario. In response, Google has recently implemented what is called FL, envisioning a practical way to implement ML algorithms in constrained networks. The FL concept is based on executing ML algorithms in a decentralized manner without the need to download the training set to a central node or a ser-

ver. It is not particularly aimed for a drone network, but for any type of network with central server (a base station in our scenario) and a number of clients (UAVs, mobile users).

Here, we present a comprehensive explanation of FL algorithm for a scenario in which a network of UAVs are served by a terrestrial base station. As a typical objective, we suppose that the UAVs are processing different ground images. We also assume that the optimization of the loss function is done through a simple stochastic gradient descent (SGD) algorithm. As illustrated in Fig. 9, the central server, which is the base station in our case, shares the current update of the global model, denoted by  $w_t$ , with a sub-set of the users. The subset size denoted by  $C$ , is randomly selected by the server. when the client UAV receives the current update of the global model, it utilizes its local training data to compute a local update of the global model related to each UAV. Those parameters are the mini-batch size denoted by  $B$  which indicates the amount of the local data used per each UAV, the index  $k$  of the UAV, and the number of training passes each client makes over its local dataset on each round, which is denoted by  $E$ . After updating process, the UAV only communicates the updated data, denoted by  $w_{t+1}^k$ , to the base station. For an SGD-based optimization, the update is calculated as follows :

$$w_{t+1}^k = w_t - \eta \nabla(w_t, B) \quad (3.2)$$

where  $\eta$  is the learning rate and  $l$  is the loss function. For instance, the UAV performs a full batch update and hence uses all its local data since  $B = \text{inf}$  . Then it repeats the (3.2) ten times since  $E = 10$  and delivers the output  $w_{t+1}^k$  to the base station. Once the local update  $w_{t+1}^k$  is received by the base station, it improves the global model and then removes these updates because they are no longer needed.

We have mentioned in the previous section that FL is a promising solution for constrained networks where extensive calculation could not be done onboard. It permits decoupling the model training and the access to the raw information because of the fact that it is not mandatory for drones to share any data with the server, instead, they only transmit their local update as explained already. Firstly, FL decrease the privacy and security issues by minimizing data traffic over the network. Consequently, it is considered an important approach for confidential systems where data does not need to be shared. For instance, one can consider a recommender system as an example of ML application where it is necessary that raw data will not be shared between the clients. In many scenarios the clients do not wish for others to know their preferences, thus FL preserves this privacy by keeping the local data of each user private and only share the model updates. Secondly, FL is well suited for applications where data is unbalanced. For instance, one client may be outside the region of interest and thus have a small amount of data in comparison with other clients. Let's take the example of detecting a car by utilizing a drone's camera, therefore even if one of the drones is displaced in a given location where cars do

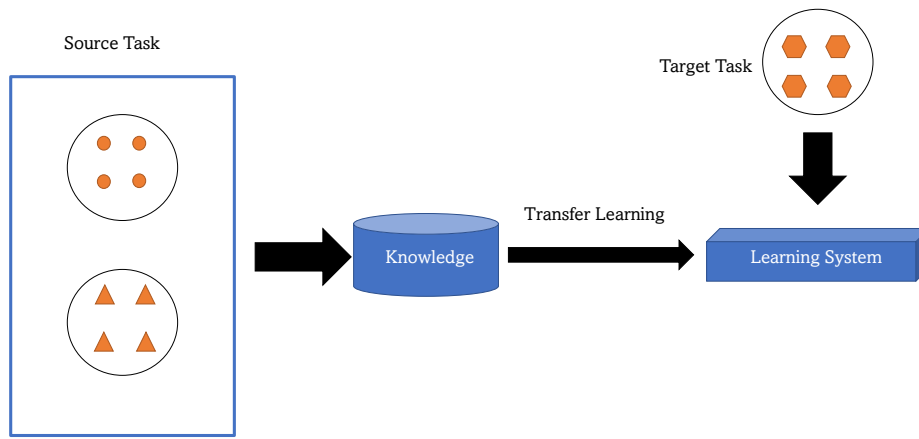


Fig. 3.4 – Schematic of transfer learning.

not often cross, that drone will efficiently detect a car when it is in the field of its camera. This is due to the fact that other drones communicating with the server have been involved in the training of the displaced UAV. Furthermore, the learning process in the FL framework can be active even if one of the nodes is in the idle state. For example, if one of the drones has to perform charging, an emergency landing or encounters a connectivity failure, the learning process continues and the drone can restore the updates when it reconnects to the network. Finally, FL execute well on non-independent and identically distributed data, for instance, the data partition realized by a single UAV cannot be representative of the overall information of the system simply because the drone can only conceive a part of a given process.

### 3.3.8 . Transfer Learning for UAVs

Transfer learning is a new machine learning method that transfers the learned model parameters to the new model to help the new model training [33]. Two basic concepts are referred to TL : (1) source domain, represents the object to be transferred ; (2) target domain, represents the target to be endowed with knowledge. As shown in Fig. 3, through transfer learning, the learned model parameters or knowledges from source domain task can be shared to the new model in target domain task, which can speed up the training process of new tasks and optimize the learning efficiency [34]. There are some common implementation methods of TL, such as instance-based approach, feature-based approach and parameter-based approach [35]. The instance-based TL method is to weight different data samples according to similarity and importance. However, this method needs to collect a large number of instance samples and calculate similarities between these instance samples and the new learning samples respectively, which consumes large amounts of memory resources and computational resources [36]. The feature-based TL approach needs to project the features of the source domain and the target domain into the same feature space, and utilize some machine learning methods to process the feature matrix [37], and this approach is mainly used for solving classification and recognition problems. The parameter-based TL approach applies the model trained in the source domain to the target domain, and completes the new similar task through a short retraining [38]. To build up UAVs tracking decision-making model in this research, parameter transfer is a simple and effective way, and



it can help UAVs learn similar strategies from a more reasonable initial network based on model parameters previously trained [39]. As a result, the tracking task is simplified into a set of simple sub-tasks. We can train the model to fulfill sub-tasks, and migrate the sub-tasks model to the final task through parameter-based transfer learning, which will be explained in detail in Section 3.

### **3.4 . Conclusion**

In this chapter, encouraged by a wide set of new applications that can gain assistance from drone networks, such as smart cities and aerial base stations deployment, we have covered in detail the new research directions when ML techniques are utilized to increase the performance of UAV networks. We begun by preparing an extensive overview of unsupervised and supervised ML techniques. Then we introduce RL in details, that have been broadly applied in UAV networks. Then, we discussed FL principles and advantages and where a FL approach can be used in the field of UAV networks.

## 4 - Throughput Maximization with Learning Based Trajectory for Mobile Users

In this chapter, we design a new UAV-assisted communication system relying on the shortest flight path of the UAV while maximizing the amount of data transmitted to mobile devices. In the considered system, we assume that UAV does not have the knowledge of user's location except their initial position. We propose a framework which is based on the likelihood of mobile users presence in a grid with respect to their probability distribution. Then, a deep reinforcement learning technique is developed for finding the trajectory to maximize the throughput in a specific coverage area. Numerical results are presented to highlight how our technique strike a balance between the throughput achieved, trajectory, and the complexity.

## Sommaire

---

<b>4.1</b>	<b>Introduction . . . . .</b>	<b>43</b>
<b>4.2</b>	<b>System Model . . . . .</b>	<b>44</b>
<b>4.3</b>	<b>Learning Based Trajectory Design . . . . .</b>	<b>46</b>
<b>4.4</b>	<b>Numerical Results . . . . .</b>	<b>49</b>
<b>4.5</b>	<b>Conclusion . . . . .</b>	<b>51</b>

## 4.1 . Introduction

Unmanned aerial vehicles (UAVs) have recently captivated interest as a rapid solution for providing communication services to ground users [15], [16]. In practice, it is not cost-effective or even practical to set up terrestrial base stations (BSs) in temporary hotspots or disaster areas. In contrast, due to the exceptional flexibility of deployment and maneuverability of UAVs, they can be employed in an efficient manner to serve as aerial BSs [17]. Moreover, the communication link between users and UAVs has typically high probabilities of line-of-sight (LoS) air-to-ground (A2G) channels, which can mitigate signal blockage and shadowing [5]. Wireless networks supported by UAVs constitute a promising technology for enhancing the network performance [18]. The applications of UAVs in wireless networks span across diverse research fields, such as wireless sensor networks (WSNs), caching, heterogeneous cellular networks, massive multiple-input multiple-output (MIMO), disaster communications and device-to-device communications (D2D). In all mentioned scenarios, a critical aspect for the system's ability to serve the highest possible number of users with the best achievable throughput is the user's location. Previous works have addressed the problem of path planning of UAV by neglecting the mobility of users in to the system model. Whereas fixed location of users may fulfill certain communication network scenarios, but in real life applications, one can not oversight the dynamic movement of users. In [19], the authors studied the joint 3D deployment and power allocation in a UAV-BS system that maximizes the system throughput. They proposed an algorithm which combined deep deterministic policy gradient with water-filling to allow the UAV to learn an optimal location in the continuous state and action spaces. In [20], the authors investigated the multi-UAV trajectory planning to provide a long-term energy-efficient content coverage. A multi-UAV trajectory planning problem was formulated as two related multi-agent cooperative stochastic games. For obtaining equilibriums of the games, the authors proposed a Q-learning based decentralized multi-UAV cooperative RL algorithm. The proposed algorithm enables UAVs to independently choose their policy and recharging scheduling. Also, in a decentralized manner, the UAVs share their learning results with each other over a timevarying communication network. In [21], authors proposed a 3D deployment based on the quality of experience and they considered the dynamic movement of ground users into their system model. They demonstrated that the proposed 3D deployment scheme based on Q-learning outperforms the K-means algorithm. However, the authors assumed the UAV has online knowledge of dynamic movement of ground users which is not always possible in real life applications.

In this chapter, we consider a system model relying on a single UAV to serve several mobile users. We propose a framework for finding the trajectory to maximize the achievable system throughput between all users. In our proposed model, the UAV is only aware of the initial position of users and needs to choose actions based on the stochastic model calculated from the mobility of users. For comparison, we

consider a scenario that UAV is connected through the GPS system and has the knowledge of user's location in each time instant.

The rest of this chapter is organized as follows : the system model and achievable system throughput are given in section 2. In Section 3, mobility model and stochastic model for localization of users are proposed. In Section 4, the deep reinforcement learning algorithm is utilized for obtaining the UAVs' dynamic movement when users are roaming. Numerical results are carried out in Section 5. Finally, the paper is concluded in Section 6.

## 4.2 . System Model

Consider a system consisting of a single UAV and  $U$  ground users with dynamic movement in the area and need to be covered. Let  $\mathbf{u}_u = [x_u, y_u]^T \in \mathbb{R}^{2 \times 1}$  represent the horizontal coordinate of  $u$ -th ground user where  $u \in \mathcal{U}$ . The 2D Cartesian coordinate of the UAV is presented as  $m = [x_m, y_m]^T$ . In practice, the ground users receive three different kinds of signals from UAVs including LoS, non-line-of-sight (NLoS), and multiple reflected signals. These signals occur with specific probabilities in different environments and the probability of multiple reflected signal which results multi-path fading is considerably lower than two other signals. Thus, their impact at the receiver side is typically ignored. Thus, we assume that the communication link between ground users and the UAV is overshadowed by the LoS signals. Based on this assumption, the channel power gain between  $u$ -th user and the UAV is only a function of their Euclidean distance as below

$$h_{u,m} = \rho_0 d_{u,m}^{-2} \quad (4.1)$$

where  $\rho_0$  is a constant shadowing power of the channel at the reference distance  $d_0 = 1\text{m}$  and  $d_u$  is the Euclidean distance between  $u$ -th user and UAV which can be written as

$$d_u = \sqrt{z_m^2 + \|\mathbf{u}_u - m\|} \quad (4.2)$$

Hence, we have

$$h_{u,m}(t) = \frac{\rho_0}{z_m^2 + \|\mathbf{u}_u - m\|} \quad (4.3)$$

The bit rate at time  $t$  for  $u$ -th user can be formulated as below

$$R_u(t) = \log_2(1 + \gamma_{u,m}(t)) \quad (4.4)$$

where  $\gamma_u(t)$  is the signal-to-noise ratio (SNR) corresponding to the  $u$ -th user at time  $t$ , which can be expressed as

$$\gamma_{u,m}(t) = \frac{P h_{u,m}(t)}{\sigma^2} \quad (4.5)$$

where  $P$  is the UAV transmit power and  $\sigma^2$  is the power of the additive white Gaussian noise (AWGN) at  $u$ -th user. Since users are mobile, for each user, there

are  $k$  possible locations with respect to time. So we have

$$Pr_u^{(x_k, y_k)}(t) = z \quad \forall u, \forall t, \forall u \quad (4.6)$$

Consequently, by utilising the above probability, the achievable system throughput can be expressed as

$$R_k^{(x_k, y_k)}(t) = \sum_{x_k, y_k} Pr_u^{(x_k, y_k)}(t) \times R_u^k(t) \quad (4.7)$$

Since the movement of users affect the system throughput, the UAV have to travel based on the real-time movement of users to maximize the throughput for ground users. Thus, to provide communication services for all ground users, we maximize the achievable system throughput subject to the location of each user based on their mobility model. So, we can write

$$\max_{x_m(t), y_m(t)} \left( \int_{t=0}^T \sum_{u=1}^U R_u^k(t) dt \right) \quad (4.8)$$

$$\text{s.t.} \quad x_1(0), \dots, x_u(0) = X_1(0), \dots, X_u(0), \forall u \quad (4.9)$$

$$y_1(0), \dots, y_u(0) = Y_1(0), \dots, Y_u(0), \forall u \quad (4.10)$$

$$x_u^k(t), y_u^k(t) = Pr_u^{(x_k, y_k)}(t), \forall k, \forall t, \forall u \quad (4.11)$$

$$z_m(t) = H_{uav} \quad (4.12)$$

$$P_{tx}(t) = P_m \quad (4.13)$$

$$V_c(t) = V_{uav} \quad (4.14)$$

where  $H_{uav}$  and  $V_{uav}$  are the altitude and velocity of UAV, while  $P_c$  is the value for transmit power from UAV to ground users. Furthermore, (9) and (10) denote that initial position of each user is known by the UAV; (11) indicates that the location of mobile users are estimated based on their probability distribution, (12),(13) and (14) set the constant values on altitude, transmit power and velocity of the UAV, respectively.

The memoryless mobility models such as Random Walk allow mobile nodes to move anywhere in the system with a stochastic random process for speed and direction. Consequently, the mobility patterns are very disordered and may not be able to reflect the real-time scenarios of mobile ad hoc networks. In reality, movements of mobile nodes are restricted by obstacles. Moreover, there is some correlation between the speed, direction, path, and destination of mobile nodes to meet their corresponding objectives. Since our objective is to let the UAV learn the trajectory based on the mobility of users, the choice of the mobility model has a major impact on the learned trajectory. If we consider a model that users change their direction or speed at each time step, the randomness in the environment is too chaotic in which, there is no meaningful trajectory to be learned. Also, border behavior of the environment and how users react when they reach the border

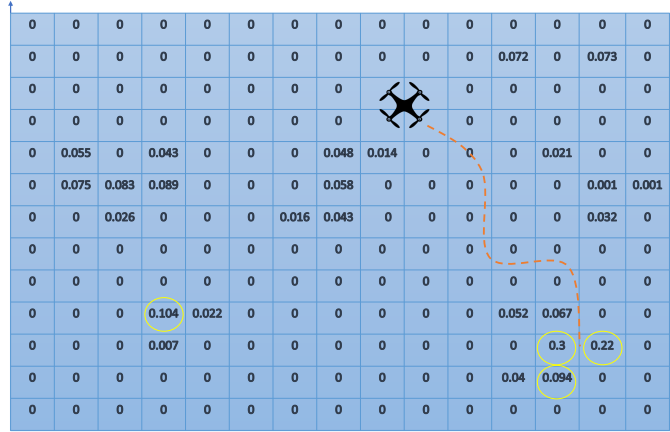


Fig. 4.1 – Probability distribution of a mobile user based on the grid model.

cannot be neglected. Therefore, we decide to choose a random mobility model for users that is realistic and practical. The Smooth Random Mobility describes how the correlation between the speed and the direction is used to provide the smooth movement patterns that are more realistic to be used in the real-life scenarios [22].

Now, with the given mobility model, as discussed in previous section we need to calculate the probability distribution in (4.8). There are different approaches for predicting the location or trajectory of an individual. The interested reader is referred to the following works, [23], [24] and [25]. Motivated by the work from [24], we partition the spatial area into a grid in which each cell has an area of  $25 \text{ m}^2$  and then counts the number of times a mobile user has visited each cell based on the simulation. With this information, we compute a probability distribution representing the likelihood of visiting each particular cell at the time instant  $t$ .

### 4.3 . Learning Based Trajectory Design

In this section, we describe the novel technique for localization of mobile users. In the considered scenario, we assume that the initial position of ground users are known to UAV. In our algorithm, with regard to probability distribution found by the grid model, the UAV makes the decision based on the most probable grids which have the highest probabilities. Here, because of the large action size, we limit the choices of UAV at each time instant to  $n_a = 4$  for each user. Also, since it is not necessary for the UAV to do the estimation at each time instant, we set a time period  $T_a$  in which the UAV will estimate the locations periodically. The localization algorithm is described in the following.

Given the location of mobile users, our goal is to obtain the optimal trajectory of the UAV to maximize the system throughput. Reinforcement Learning (RL) has a potential to deal with challenging and realistic models that include stochastic movements of nodes. In general, RL is a learning approach that is used for finding

the optimal way of executing a task by letting an entity, named agent, take actions that affect its state within the acting environment. The agent improves over time by incorporating the rewards it had received for its appropriate performance in all episodes [26]. In the Q-learning model, the UAV acts as agent, and the Q-learning model consists of four parts : states, actions, rewards, and Q-value. The aim of Q-learning is for attaining a policy that maximizes the observed rewards over the interaction time of the agent.

1. State Representation : Each state in the set is described as :  $(x_u, y_u)$ , where  $(x_u, y_u)$  is the horizontal position of UAV. As the UAV takes a trajectory in a specific episode, the state space can be defined as  $x_u : 0, 1, \dots, X_d$ ,  $y_u : 0, 1, \dots, Y_d$ , where  $X_d$  and  $Y_d$  are the maximum coordinate of this particular episode.
2. Action Space : The action space  $A$  is described by all possible movement directions, the action of remaining in the same place and 4 possible locations for each of the mobile users. By assuming that the UAV fly with simple coordinate turns, the actions related to movement of UAV is simplified to 7 directions. Combining the actions from the dynamic movement of UAV and estimation based on the grid model, the action size will be equal to 263.
3. State Transition Model : Considering a deterministic MDP, there is no randomness in the transitions that follow the agent's decisions. Thus, the next state is only affected by the action that the agent takes.
4. Rewards : The reward function is defined by the instantaneous throughput of users. If the action that the agent carries out at current time  $t$  can improve the throughput, then the agent receives a positive reward, otherwise, the agent receives a negative reward.

Due to the size of MDP, we create an RL agent as a feed-forward neural network (NN), with  $F$  input neurons,  $Y$  hidden states each with the same number of neurons  $Z$ , all using rectified linear (ReLU). When receiving the current state, described with  $F$  features as input, the NN agent outputs its evaluation for all seven actions that can be taken. However, the use of NNs in RL tasks may fail to converge especially in problems with stochastic environments, such as ours. Therefore, we rely on deep RL and using double Q-learning to solve our problem [27].

For the double-Q-learning RL algorithm, we need to keep two separate agents with the same properties but with different weight values  $w_P$  and  $w_T$ . As such they will output a different Q-action function when given the same state. One is used to choose the actions, called a primary model  $Q_P(s_t, a_t)$ , while the other model evaluates the action during the training, called a target model  $Q_T(s_t, a_t)$ . Therefore training occurs when taking a batch of experiences  $e_t$  from the buffer that is used to update the model as :

$$Q_P^{new} = (1 - \alpha)Q_P + \alpha [r_t + (1 - d_t)\gamma \max_a Q_T(s_{t+1}, a)] \quad (4.15)$$



Hyperparameter	value
optimizer for SGD	Adam
learning rate for optimizer	0.0001
discount factor $\gamma$	0.99
number of hidden layers	2
number of neurons	256
minibatch size	32
action space size	263
activation function	ReLU
replay buffer capacity	$10^6$

Table 4.1 – Training parameters.

where  $\max Q_T(s_{t+1}, a)$  is the action chosen as per the agent,  $\alpha$  is the learning rate which was an input to the Adam optimizer [28], and  $\gamma$  is a discount factor that reduces the impact of long term rewards. We implement this with soft updates where instead of waiting several episodes to replace the target model with the primary. The target model receives continuous updates discounted by value  $\tau$  as in  $w_T = w_T(1 - \tau) + w_P\tau$ .

Now, we examine how the agent makes the decision from the large action space at each time step and how invalid action masking and normalized probability distribution are realized to strict the agent for repeatedly taking invalid actions. It has been shown that invalid action masking scales better when the space of invalid actions is large and the agent solves the desired task while invalid action penalty struggles to explore even the very first reward.

First, let us see how a normalization carry out in to the discrete action space for when UAV has to decide the location of users after each  $t_c$  seconds. For illustration purposes, consider the 4 probabilities in Fig.6.1 which correspond to highest possible locations for one user at time  $t$ . Thus, let us acknowledge an MDP with the action set  $A = a_0, a_1, a_2, a_3$  and  $S = s, s'$  where the MDP reaches the state  $s'$  after an action is taken in the initial state  $s$ . Thus we have

$$\begin{aligned} P(s'|s, a) &= [p(a_0|s_0), p(a_1|s_0), p(a_2|s_0), p(a_3|s_0)] \\ &= [0.094, 0.3, 0.104, 0.22] \end{aligned} \quad (4.16)$$

Now, after normalization enforced, we can write

$$P(s'|s, a) = [0.13, 0.41, 0.14, 0.3] \quad (4.17)$$

Now for states that UAV actions are about the coordinates of UAV and come from the possible directions described in section ??, we have to mask

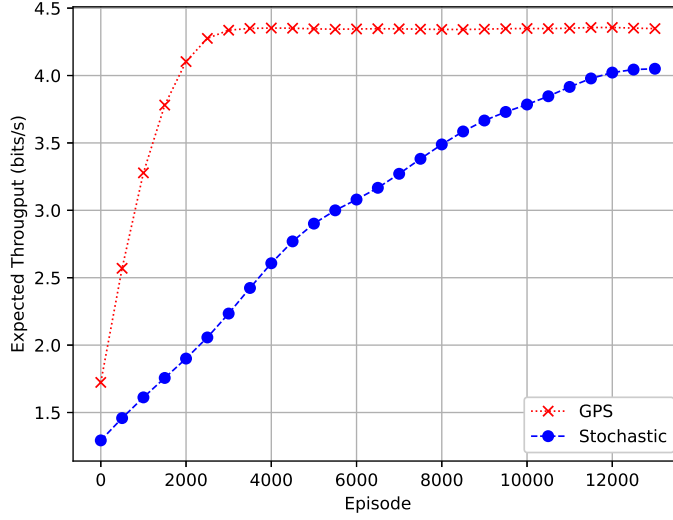


Fig. 4.2 – Convergence of the proposed algorithm vs. the number of training episodes.

the invalid actions which correspond to actions related to estimation of user’s location. Lets consider our actions space size which is equal to 263. We set the first 7 actions correspond to actions related to direction of UAV and other 256 actions related to user’s locations. Suppose that we have an action set  $A = a_0, \dots, a_6, \dots, a_{262}$  in which each action has same probability. Now let us assume that at time instants other than  $t_c$ , the actions  $[a_7, a_8, a_9, \dots, a_{262}]$  are invalid actions and only the first 7 actions are valid. Invalid action masking helps to avoid sampling invalid actions by “masking out” the probabilities corresponding to the invalid actions. This is usually achieved by replacing the probabilities of actions to be masked by zero. Let us use  $I_a$  which stands for this masking process and we can calculate the re-normalized probability distribution  $P(s'|s, a)$  as the following :

$$\begin{aligned}
 P(s'|s, a) &= I_A ([p_0, \dots, p_6, p_7, \dots, p_{262}]) \\
 &= [p'(a_0|s_0), \dots, p'(a_6|s_0), p'(a_7|s_0), \dots, p'(a_{262}|s_0)] \quad (4.18) \\
 &= [0.142, \dots, 0.142, 0, \dots, 0]
 \end{aligned}$$

#### 4.4 . Numerical Results

In this section, we present our numerical results characterising the optimization problem of UAV-assisted mobile networks. To highlight the efficiency of our proposed model, we compare it to a scenario when UAV is connected to GPS system and has the online knowledge of user’s location. We use Tensorflow 2.5.0 and the Adam optimizer for training the neural networks. The training parameters are provided in Table I. In deployment, a 2D area

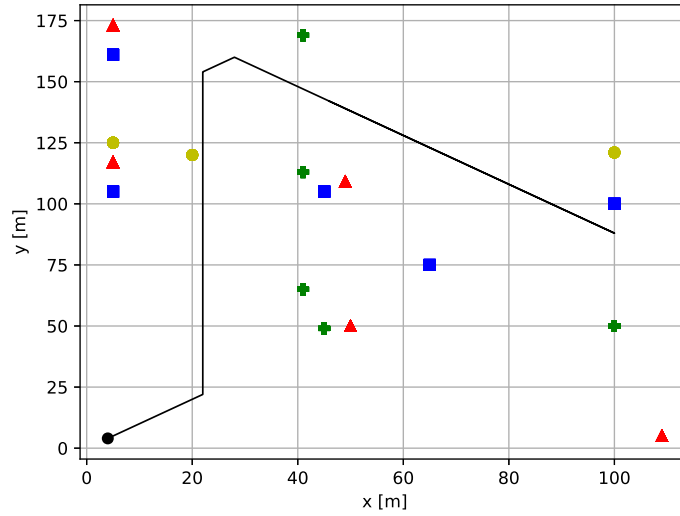


Fig. 4.3 – Trajectory obtained by UAV for the case that four ground users are roaming.

of  $1000^2$  m is considered. It is assumed that UAV flies at constant altitude and speed  $H_{uav} = 100\text{m}$  and  $V_{uav} = 20\text{m/s}$ , respectively. The UAV transmit power is set to  $P_c = 0.1\text{W}$  and the power of dense noise is assumed to be  $-174$  dB.

In Fig.4.2, we plot the expected throughput vs the number of training episodes. It can be observed that the UAV is capable of carrying out the actions in an iterative manner and learn from the mistakes for improving the system throughput. In this figure, we also compare our approach to a scenario when the UAV is connected through the GPS system and for the sake of comparison, we assume that the UAV is aware of the user's location at each time instant. As can be seen, the convergence rate of the proposed approach is much slower than the GPS approach. This is due to fact that of the large action space and the stochastic estimation of user's location, which results to necessity of more training episodes.

Fig.4.3 plots the trajectory of a UAV derived from the proposed approach when ground users move. In this figure, the trajectory of a UAV is shown for the mission duration time of 100 s. In this simulation, we assume that the UAV can move at a constant speed. At each time slot, the UAV choose a direction from the action space which contains 7 directions, then the trajectory will maximize the throughput of ground users. It should be noted that we can adjust the timespan to improve the accuracy of dynamic movement. This, in turn, increases the number of required iterations for convergence. Therefore, a trade-off exists between improving the throughput of ground users and the running complexity of the proposed algorithm.

## 4.5 . Conclusion

In this chapter, the DRL technique has been utilized to optimize the flight trajectory and throughput performance of UAV-assisted networks. The mobility of users is considered in to the system model and a novel approach for estimating the location of mobile users has been studied. A learning-based algorithm was proposed for solving the problem of maximizing the system throughput by utilising a prior knowledge of likelihood of presence in a grid. We designed a DRL based movement algorithm for obtaining the trajectory of UAV. It is demonstrated that the proposed approach performs well in comparison despite the fact of being simple to implement.



## 5 - Federated Reinforcement Learning UAV Trajectory Design for Fast Localization of Ground Users

In this chapter, we study the localization of ground users by utilizing unmanned aerial vehicles (UAVs) as aerial anchors. Specifically, we introduce a novel localization framework based on Federated Learning (FL) and Reinforcement Learning (RL). In contrast to the existing literature, our scenario includes multiple UAVs learning the trajectory in different environment settings which results in faster convergence of RL model for minimum localization error. Furthermore, to evaluate the learned trajectory from the aggregated model, we test the trained RL agent in a fourth environment which shows the improvement over the localization error and convergence speed. Simulation results show that our proposed framework outperforms a model trained with transfer learning by 30%.

## Sommaire

---

<b>5.1</b>	<b>Introduction</b>	<b>55</b>
<b>5.2</b>	<b>System Model</b>	<b>56</b>
<b>5.3</b>	<b>Proposed Method</b>	<b>58</b>
<b>5.4</b>	<b>Federated learning</b>	<b>59</b>
<b>5.5</b>	<b>Numerical Results</b>	<b>61</b>
<b>5.6</b>	<b>Conclusion</b>	<b>62</b>

## 5.1 . Introduction

In recent years, location-aware services have been recognized as a crucial component for broad applications in wireless communication. Generally, information regarding the location of objects can be exploited in different layers, from communication aided purposes to the application level where location information is desired to interpret the collected data [29]. For this purpose, the global positioning system (GPS) grants a suitable performance for outdoor applications. However, GPS is known of its expensive cost and vulnerability to jamming. Thus, alternative localization approaches have become more attractive for research focus over the past decade. In the literature, there are several ground anchor based localization techniques that have been broadly studied [30]. Specifically, the Received Signal Strength (RSS) technique is favorable because of its inherent simplicity and low complexity. This simplicity is due to the fact that RSS can be used without any modification to current systems, so it is the easiest way forward. Moreover, RSS based localization can achieve satisfactory performance in emergency situations [31]. Nonetheless, the variation around the mean signal power due to shadowing significantly impacts the reliability of this technique. This is especially important in urban and high urban environments where the shadowing effect is more severe and hence the localization accuracy drops significantly. To address this issue, unmanned aerial vehicles (UAVs) deployed as aerial anchors is an emerging solution in order to localize ground devices. The main benefits of UAV anchors are their higher probability of line-of-sight (LoS) with ground terminals and less shadowing effect at higher altitudes [32]. Thus, aerial anchors potentially are capable of resolving the main drawback of ground node localization when using RSS technique. In fact, UAV anchors can combine the benefits of satellites with a higher link probability of LoS and the advantages of ground anchors with a short link length and hence higher RSS resolution. Furthermore, UAVs are typically battery-limited which introduces an important challenge towards their deployment as aerial anchors. This fact restricts UAVs operational lifetime and hence reduces the number of measurements that can be collected during their mission, which can negatively affects the accuracy of localization. In fact, depending on the hovering duration, speed of the UAV, and length of the path, the energy consumption of the UAV varies.

The noteworthy success of Machine Learning (ML) is mainly associated to two key components – highly powerful computing and extremely efficient data analytic. However, such a impressive success in ML essentially relies on whether or not there are enough data to support ML algorithms so as to make them work convincingly, in which it becomes a crucial issue in many ML applications. Because of the proliferation of UAVs, collecting data through them becomes much practical and convenient such that a UAV anchor has gra-



dually been a vast live database abounding with real-time information, which can be utilized by ML to optimize network operations and organization. It has become an important issue to appropriately and effectively use ML techniques based on data distributed over a massive mobile network. Specifically, when transporting raw data from all UAVs to a server in a huge network due to the many issues, such as network congestion, energy consumption, privacy, security, etc. To avoid transporting a huge amount of distributed data to a server for conducting centralized ML and to preserve the privacy of users, a distributed learning methodology without raw data transportation, such as federated learning (FL) [33], becomes a viable solution.

In this paper, we introduce a novel framework for ground users (GUs) localization in urban environments using UAVs. Our proposed framework incorporate reinforcement learning with federated learning which enables us to explore the optimal trajectory of the UAVs for maximum localization accuracy for different types of propagation environments. First, by formulating the problem we investigate the paths that UAVs take for for minimum localization error for three environments with different parameters which impact the path loss and accuracy of localization. By utilizing federated learning technique we aggregate these models and finally we test the trained model in fourth environment. Our results show that the localization error achieved with same number of training episodes is %30 lower with trained FL model from three environment as compared to the model transferred sequentially from first environment to fourth environment.

The rest of this chapter is organized as follows. In Section II, we introduce the system model and the path loss model for localization based on RSS. Then, the machine learning framework for UAVs is introduced in Section III. In Section IV the simulation results are presented. Finally, the work is concluded in Section V.

## 5.2 . System Model

In this paper, we assume multiple UAVs flying over an urban area at a fixed altitude  $h$ , operating as an aerial anchors to localize multiple terrestrial users. These devices are equipped with a wireless communication device which periodically broadcast a probe request. We resort to utilizing the following log-normal shadowing pathloss model as it is capable of modeling wireless environments with acceptable precision [32]. We formulate the path loss as :

$$L = 20 \log(d) + 20 \log\left(\frac{4\pi f}{c}\right) + A_\tau(\theta) \quad (5.1)$$

where  $d$  is the distance between the UAV and ground user,  $f$  and  $c$  are respectively the system frequency and speed of light, and  $A_\tau(\theta)$  is a log-

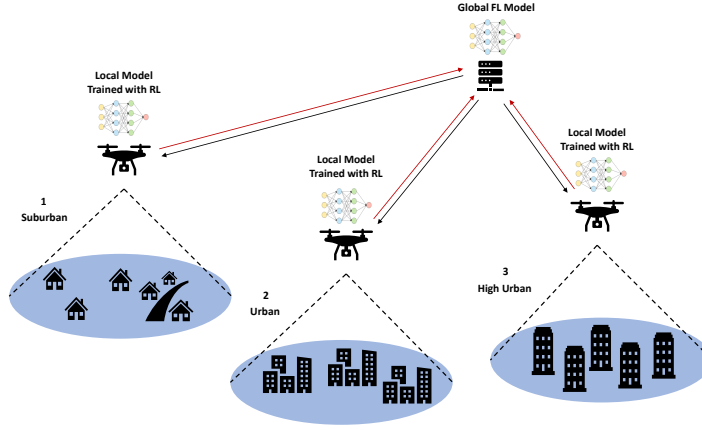


Fig. 5.1 – Federated learning architecture.

normal distributed random variable with mean  $\mu_\tau$  and variance  $\sigma_\tau^2(\theta)$ , i.e.,

$$A_\tau(\theta) \sim \mathcal{N}(\mu_\tau, \sigma_\tau^2(\theta)) \quad (5.2)$$

the variance can be defined as :

$$\sigma_\tau^2(\theta) = \mathbb{P}_{LoS}^2(\theta)\sigma_\tau^2(\theta) + [1 - \mathbb{P}_{LoS}^2(\theta)]\sigma_\tau^2(\theta) \quad (5.3)$$

where  $\sigma_\tau(\theta)$  corresponds to the shadowing effect of LoS and NLoS links between the UAV and the ground user, where  $\tau = \{0, 1\}$  is an indicator that can have value 1 for LoS link and 0 for NLoS link. Thus we have :

$$\sigma_\tau(\theta) = d_\tau \exp(-c_\tau \theta \frac{180}{\pi}) \quad (5.4)$$

and  $\mathbb{P}_{LoS}(\theta)$  is the probability of having LoS link, which is written as :

$$\mathbb{P}_{LoS}(\theta) = \frac{1}{1 + a \exp(-b(\theta \frac{180}{\pi} - a))} \quad (5.5)$$

where  $a, b, c_\tau, d_\tau$  and  $\mu_\tau$  are environment dependent parameters. Thus, the distance between the UAV and the device can be estimated as follows :

$$d = 10^\zeta \quad (5.6)$$

$$\zeta = \frac{P_t - P_r - 20 \log(\frac{4\pi f}{c}) - A_\tau(\theta)}{20} \quad (5.7)$$

where  $P_r$  and  $P_t$  denote the received and the transmitted power, respectively.

The position of a GU in 2D coordinates is described as  $(x_u, y_u)$ . Given the projection of UAV on the ground  $(x, y)$ , we can estimate the distances  $r_i = \sqrt{(x - x_u)^2 + (y - y_u)^2}$  based on (5.7). Moreover, the multilateration technique can be utilized to estimate the user's position. In multilateration

least squares are used to estimate the position of the user  $(\hat{x}, \hat{y})$  according to the estimated distances. In a two-dimensional space,  $n_i$  distance measurements from  $n_i$  dissimilar positions are calculated to generate  $n_i$  circles centered at the position where the measurements are taken with radii equal to the respective measurements. If the distance measurements are accurate, the  $n_i$  circles intersect in one point that establish the position of the user. Now, given  $(x_i, y_i)$  the ground position of the UAV at sample point  $i$ , and  $\hat{r}_i$  be the distance from sample point  $i$  to the middle of overlapping circles, then we can estimate the location  $(\hat{x}, \hat{y})$  using  $N$  number of samples from the following minimization formula :

$$(\hat{x}, \hat{y}) = \min_{\hat{x}, \hat{y}} \sum_{i=1}^N \sqrt{(x_{u_i} - \hat{x})^2 + (y_{u_i} - \hat{y})^2} - r_i \quad (5.8)$$

### 5.3 . Proposed Method

To solve the problem explained in the previous section, we resort to a reinforcement learning framework based on double Q-learning. Compared to the existing reinforcement learning algorithms such as Q-learning that may leads to a suboptimal trajectory, the double Q-learning algorithm permit the UAV to find the optimal flying trajectory to minimize the localization error of all users. Furthermore, in comparison with the traditional Q-learning algorithm that generally uses one Q-table to record and update the values coming from different states and actions [34], the double Q-learning algorithm uses two Q-tables to separately select and evaluate the actions. Consequently, the double Q-learning algorithm prevent the overestimation of Q values. Next, we introduce the components of the double Q-learning algorithm. We utilize a RL framework modeled as a Markov Decision Process (MDP) to solve the localization problem. Each UAV independently make decisions with respect to a tuple  $(P, A, R, S)$  in which :

1. **State Representation** : Each state considers the agent's location, represented by the UAV  $(x, y)$  coordinates in the trajectory taken, the localization error and estimated distances calculated by RSS signals explained in Section I.
2. **Action Space** : The action space is defined by all possible movement directions on the sides of the hexagon plus the action of remaining in the same place formatted into a 7-tuple.
3. **State Transition Model** : Considering a deterministic MDP, there is no randomness in the transitions that follow the agent's decisions. Thus, the next state is only affected by the action that the agent takes.

---

**Algorithm 1** : Federated averaging with DDQN.

---

```
1 : Execution on Server :
2 : Initialize  $w_0$ 
3 : for  $j = 1$  to  $\text{Max}_{\text{rounds}}$  do
4 :    $M = \text{set of UAVs}$ 
5 :   for Each UAV in parallel do
6 :      $w_{t+1}^k = \text{ClientUpdate}(k, w_t)$ 
7 :   end for
8 :    $w_{t+1} = \sum \frac{1}{M} w_{t+1}^k$ 
9 : end for
10 : Return  $w_{t+1}$  to UAVs.
11 :
12 : Execution on UAV :
13 : Construct reward function  $R$ 
14 : Init : UAV position,  $s, Q^i \ i \in [A, B]$ 
15 : Repeat
16 :   if  $\text{Local}_{\text{rounds}} < \max(\text{local}_{\text{rounds}})$ 
17 :     Choose action :
18 :      $a = \text{argmax}_a Q^i(s, a)$  from  $Q^i (i \in [A, B])$ 
19 :     Receive immediate reward
20 :     Update table  $Q^i$ 
```

---

4. **Rewards** : The reward function is defined by the average localization error from the ground users at each step,

$$r[n] = \frac{L_s}{e[n]} \quad (5.9)$$

where  $L_s$  is desired localization error which is set to  $10m$  and  $e[n]$  is the evaluated localization error at time instant  $n$ .

#### 5.4 . Federated learning

In the UAV network proposed in Section II, our aim is to investigate the performance of FL over the UAV network that localize ground users via RSS reading, which lead to continuous FL between the edge server and the UAVs. Thus, we propose a FL model over the network in Fig. 1 as follows. Suppose there are 3 UAVs distributed in the network and their task is to jointly learn a global model with the edge server in  $T$  training rounds. To characterize the impact of different environment parameters on localization error, we assume each UAV is operating in a different environment setting i.e from sub urban to high urban.

	$a$	$b$	$\mu_1$	$\mu_0$	$d_1$	$d_0$	$c_1$	$c_0$
$env_1$	4.88	0.43	0.1	21.0	11.25	32.17	0.06	0.03
$env_2$	9.61	0.16	1.0	20.0	10.39	29.6	0.05	0.03
$env_3$	12.08	0.11	1.6	23.0	8.96	35.97	0.04	0.04
$env_4$	14.32	0.08	2.3	34	7.37	37.08	0.03	0.03

Table 5.1 – The path loss parameters for : Suburban (1), Urban (2), Dense Urban (3) and Highrise Urban (4) environments [11].

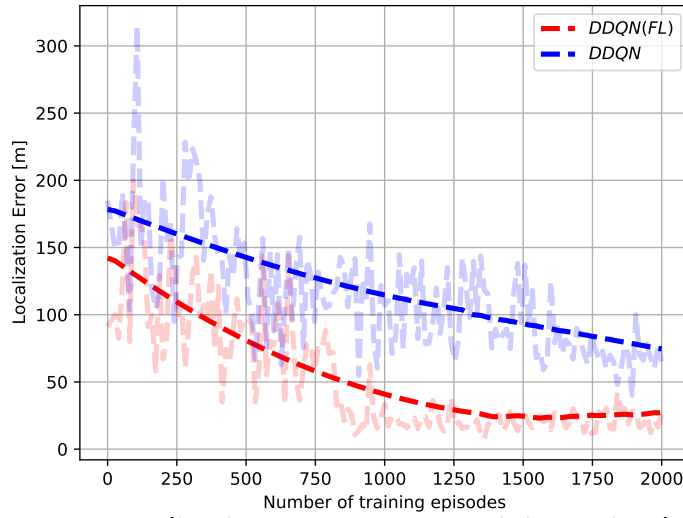


Fig. 5.2 – Localization error versus training episodes in  $env_1$ . Comparison between FL model and baseline DDQN.

FedAvg orchestrates training with a central server which hosts the shared global model  $w_t$ , where  $t$  is the communication round. The algorithm initialize by randomly setting the global model  $w_0$ . One communication round of FedAvg can be described in the following : At the beginning, the server distributes the current global model  $w_t$  to all UAVs. After updating their local models  $w_t^k$  to the shared model,  $w_t^k \leftarrow w_t$ , each UAV partitions its local data into batches and performs epochs of Stochastic Gradient Decent (SGD). Finally, UAVs upload their trained local models  $w_{t+1}^k$  to the server, which then generates the new global model  $w_{t+1}$  by computing a weighted sum of all received local models. Our approach for utilizing FedAvg reinforcement learning for localization is represented in Algorithm 1.

## 5.5 . Numerical Results

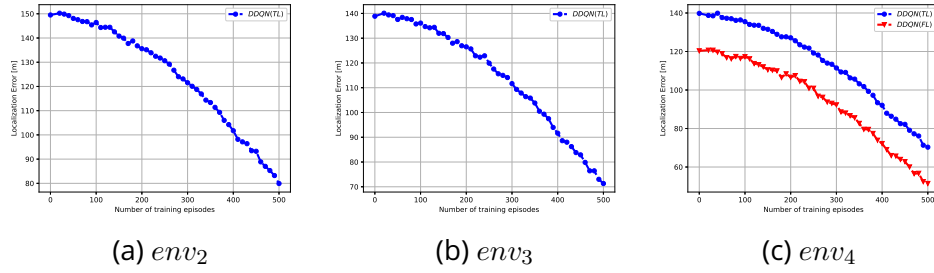


Fig. 5.3 – Localization error versus training episodes (a) pre-trained model from  $env_1$  transferred and retrained in  $env_2$ ; (b) pre-trained model from  $env_1$  and  $env_2$  transferred and retrained in  $env_3$ ; (c) test and comparison of model transferred from previous environments and FL architecture in  $env_4$ .

We assume  $N$  GUs uniformly distributed in a circular area with a radius of 750m, centered at  $(x, y) = (0, 0)$ . The values for the path loss model considered in this paper are chosen as recommended in [11] for urban environments and are summarized in Table 6.2. We assume all UAVs are flying at a fixed altitude and they can measure the RSSI from all users in their communication range. We resort to Python as a programming language to simulate the operation of our proposed method, and the numerical results are averaged over ten runs.

Fig. 6.5 shows the convergence of the proposed FL method. From Fig. 6.5 we observe that the FL algorithm required approximately 1300 episodes to reach convergence, which is much less than the number of episodes required for convergence of the DDQN. Fig. 6.5 also shows that the FL algorithm achieves the localization error of  $25m$  after only 1200 episodes, which is about %75 lower than the one reached by the DDQN baseline. This stems from the fact that the FL algorithm has already trained a set of weights from training in 3 environments and starts the training process with a pre-trained model.

In Fig. 6.6, we test the performance of the FL trained model from 3 urban environment on the 4th environment with a scenario when Transfer Learning (TL) is applied to transfer the model RL agent trained in one environment to next environment. Transfer learning aims at improving the process of learning new tasks using the experience gained by solving predecessor problems which are somewhat similar. Fig. 6.6 shows the results obtained in the scenario when considering different training options for the DDQN algorithm : in Fig. 6.6 (a) a training of  $N_e = 500$  is done in the environment 2 on the basis of pre-trained model in environment 1 ; followed by a training of  $N_e = 500$  in the environment 3 based on the transferred model from environment 2 Fig. 6.6 (b), and finally the agent is trained with  $N_e = 500$  episodes in environ-

ment 4 based on the pre-trained model from previous environments and also  $N_e = 500$  episodes is training with FL pre-trained model from environment 1 – 3, Fig. 6.6 (c). As we can see the localization error achieved with 500 episodes of training in the 4th environment with the pre-trained model from transfer learning is approximately equal to  $70m$ , while with 500 episode, the FL pre-trained model reaches the localization error of  $50m$ . This result shows that our proposed framework is efficient in reducing convergence speed by %30 and achieving better generalization performance in comparison with transfer learning approach.

## 5.6 . Conclusion

The enhancement in localization accuracy of ground users when using UAV as base station and relying on RSS techniques has been studied. Specially, we utilized a FL framework to find an optimal trajectory through training an agent with RL algorithm which reached convergence faster. This paper validated the effectiveness of placing anchors at different position with respect to different environment setting in terms of both localization error and the required number of episodes for training an RL agent. Finally, the reported results motivate inspecting other localization methods, such as angle-of-arrival, and possibly integrate them with the proposed FL-based framework for further improvements.

6 - Multi-Objective Trajectory Design for UAV-Assisted Dual-Functional Radar-Communication Network : A Reinforcement Learning Approach



In this chapter, we explore the optimal trajectory for maximizing communication throughput and minimizing localization error in a Dual-Functional Radar Communication (DFRC) in unmanned aerial vehicle (UAV) network where a single UAV serves a group of communication users and locate the ground targets simultaneously. To balance the communication and localization performance, we formulate a multi-objective optimization problem to jointly optimize two objectives : maximization of number of transmitted bits sent to users and minimization of localization error for ground targets over a particular mission period which is restricted by UAV's energy consumption or flying time. These two objectives are in conflict with each other partly and weight parameters are given to describe associated importance. Hence, in this context, we propose a novel framework based on reinforcement learning (RL) to enable the UAV to autonomously find its trajectory that results in improving the localization accuracy and maximizing the number of transmitted bits in shortest time with respect to UAV's energy consumption. We demonstrate that the proposed method improves the average transmitted bits significantly, as well as the localization error of the network.

## Sommaire

---

<b>6.1</b>	<b>Introduction</b>	<b>65</b>
<b>6.2</b>	<b>System Model</b>	<b>68</b>
6.2.1	Channel Model	68
6.2.2	Power Consumption Model	71
<b>6.3</b>	<b>Problem Formulation</b>	<b>72</b>
<b>6.4</b>	<b>Calculating localization error</b>	<b>74</b>
<b>6.5</b>	<b>Preliminaries</b>	<b>76</b>
6.5.1	Proposed RL framework	78
<b>6.6</b>	<b>Numerical Results</b>	<b>80</b>
<b>6.7</b>	<b>Conclusion</b>	<b>84</b>

---

## 6.1 . Introduction

Unmanned aerial vehicle (UAV) or drone is marked as a critical component for future mobile networks that can arrange both ubiquitous communication and radar sensing functions due to its flexible on-demand deployment and ability in trajectory design [35]. Specially in practical scenarios in emergency situations, such as natural or man-made disasters, UAVs can not only maintain communication link with users, but also localize targets for successful environment sensing to avoid obstacle and potential attacks [36].

Because of the constraints on UAVs, such as weight and power, it is very demanding to install both communication system and radar system. Meanwhile, deploying a large number of UAVs, in which some provide communication services while the others perform radar sensing, will not only introduce co-channel interference between communication systems and radar systems, but also increase the resource consumption. Joint communication and radar sensing (JCAS) [37], also known as dual-functional radar-communication (DFRC) [38], is a promising solution to aforementioned problems. In DFRC, a single transmitted signal is used, and a majority of hardware and signal processing are shared between communication and radar. Thus, the payload and resource usage can be minimized.

In [38], the authors introduced a dual-function system with joint radar and communication platforms, where sidelobe control of the transmit beamforming was used to enable communication links. In [39], the authors developed a single transmitter with multiple antennas to communicate with downlink cellular users and detect radar targets simultaneously. In [40], the authors proposed the performance trade-off between radar and communication, and utilized a DFRC MIMO system to minimize the downlink multiuser interference under both a constant modulus constraint and a similarity constraint with respect to referenced radar. In [41], the authors studied a framework in which a beampattern was used to enhance the radar sensing performance while guaranteeing the performance of the downlink communications for the DFRC system. In [42], the authors studied a joint UAV location, user association, and UAV transmission power control in a DFRC multi-UAV network, where multiple UAVs are employed to simultaneously serve a group of ground users for communications and cooperatively sense the targets. In [43], a beamforming design for joint radar sensing and multi-user communications was proposed in which they formulated a optimization problems to minimize the CRB of target estimation by imposing SINR constraints for multiple communication users. In [44], an OFDM system for simultaneous radar and communication operations was considered, and the characteristics of OFDM signals were utilized in radar processing to reduce the typical drawbacks of correlation based processing. In [45], the authors studied a new multibeam framework that allows seamless integration of communication and sensing.

In [46], a closed-form solution for optimizing the coefficients in the analog antenna arrays to generate a multibeam for joint communication and radio sensing was introduced. Moreover, the authors in [47] proposed a novel technique for embedding communication information into MIMO radar waveform via sparse antenna array. In [48], the authors investigated the power minimization issue in DFRC system via joint subcarrier assignment and power allocation.

Although the advantages of alternative localization techniques such as, AOA (angle of arrival), TOA (time of arrival), or TDOA (time difference of arrival) have been demonstrated in enhancing the performance of wireless networks, the radio received signal strength (RSS) is more attractive due to its simplicity and cheap functionality (does not require extra antennas or time synchronization) [49]. Despite having low complexity, its localization accuracy is fairly affected by the randomness of the received signal and shadowing, notably in urban areas. However, a UAV may be used to localize ground targets as an enhancement. The UAV has the capacity to measure the RSS of multiple targets from different positions with higher probability of line-of-sight (LoS), and thus better localization accuracy [11]. Furthermore, besides accurate positioning, timely localization is also crucial for many operations like in search and rescue missions. For instance, finding locations of trapped people after a disaster or a patient who needs rescue in a serious life threat [50]. Consequently, finding the correct flight path (trajectory) is essential for both timely and accuracy of the targets' localization. Additionally, a UAV has limited energy which reduce its operational lifetime. Thus, different factors such as UAV's velocity, hovering time, and path length affect the energy consumption of the UAV, and as a result impact the localization accuracy due to fewer collected RSSI measurements. Another challenge is that the UAV, before its mission, does not know the number and locations of the objects, therefore, none of the existing pre-path planing algorithms from the literature are efficient for the fast localization operation. To this end, the necessity in creating an autonomous UAV so as to observe the environment while localizing becomes crucial [51].

In the literature, there are many works that studied the localization problem. In [49], the authors investigated the main factors that impact the accuracy of the RSS measurements and proposed an approach to mitigate the negative impacts of these factors. In [52], the authors introduced a distributed based localization technique to attain high accuracy without dense deployment. In [53], new schemes (cooperative and noncooperative) based on convex optimization are designed to enhance the localization accuracy. In [54], the authors analyzed the accuracy achieved through changing the height and distance of the anchors to terrestrial targets.

Furthermore, [55] proposed three different pre-determined trajectories for

a mobile anchor to travel the whole area, and demonstrated that any deterministic trajectory display significant benefits compared to a random movement. In [55], the authors proposed a location verification using a random anchor movement. In [56], a novel trajectory is proposed, where in this approach, all deployed nodes are localized with high precision and short required time. In [57], the authors introduced a trajectory named LMAT. The authors in [58] presented a novel localization algorithm, where in their technique, one mobile anchor combine least square method to estimate the location of terrestrial nodes. In [59], multiple location-aware mobile anchors localize the unknown nodes. To implement this, the authors introduced two algorithms in which one is to control the trajectory of the mobile anchor, and another is to extract the direction and distance of unknown nodes.

Moreover, localizing ground targets by utilizing UAV is studied thoroughly in the literature. In [60], the authors studied the advantages of using drone anchor. In [61], a multiple path planing algorithm based on traveling salesman problem is proposed for a UAV to localize all targets positions. Also, in [62] a technique using triangulation that guarantees the localization precision is introduced. In [63], the authors improved the localization accuracy by equipping a UAV with directional antennas. [64] extended the approach even further by using omnidirectional antenna. In [31], the authors proposed a framework using RL to let a UAV traverse a trajectory that results in finding the position of multiple ground targets with minimum average localization error under fixed amount of UAV energy consumption, trajectory length, number of waypoints, or flying time. In [65], the authors proposed a method to localize users in disaster scenes having regions with varying importance that may be set according to the damage and population level. In [66], the authors studied 3-D localization via autonomous UAV that works independently of the GPS or other detectable mobile signals transmitted by the UAV. For this purpose, they utilized the existing cellular infrastructure to enable the UAV to determine its location using the locations of four surrounding base stations of the cellular network. In [67], a novel localization and path planning approach based on UAVs is proposed in which the UAVs can extract one-hop neighbor information from the devices that may have run out of power by using directed wireless power transfer.

To the best of our knowledge, no work has considered using a smart UAV to autonomously observe the environment and find the trajectory that results in faster multipleobject localization with minimum errors, by only relying on RSS information, and taking into account the variation of shadowing with UAV elevation angle in urban areas. By leveraging the advantages of DFRC systems, the performance of communication and localization can be improved with reduced power consumption. However, a number of important issues need to be addressed, such as the path planing and speed of the UAV.

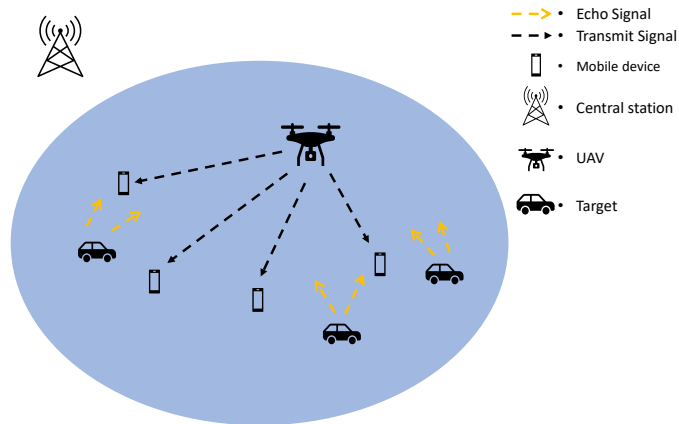


Fig. 6.1 – UAV-Assisted DFRC System Architecture

In this paper, we study a UAV enabled DFRC system, where a single UAV is employed to simultaneously serve a group of communication users and cooperatively localize the targets in the area. We introduce a framework using reinforcement learning (RL) to optimize the operation of the UAV in urban areas. Based on the UAV limitations, such as UAV energy, operational time, UAV speed, a Markov decision process (MDP) model is formulated. Then, the introduced RL algorithm (known as double-Q-learning algorithm) allow the UAV the necessary artificial intelligence to autonomously find the path to optimize the communication system throughput and achieve a localization precision with considered capacity factor. The novelty of our work focused on the fact that a smart UAV autonomously discover the environment and identify the path that will result to providing the maximum communication service in terms of average throughput and the fastest multi-object localization with desired error, by just counting on RSS information, and considering the variation of shadowing with UAV elevation angle in urban areas.

The rest of this chapter is organized as follows. In Section II, we introduce the system model, the path loss model for localization based on RSS and the power consumption model for rotary UAV. Then in section III, we describe the multi objective optimization problem. The machine learning framework for UAVs is introduced in Section IV. In Section V the simulation results are presented. Finally, the work is concluded in Section VI

## 6.2 . System Model

### 6.2.1 . Channel Model

We study a downlink UAV dual-functional radar-communication system where a single UAV is localizing and communicating with  $K$  ground users. The UAV is able to fly in the target area with the fixed altitude,  $h$ , for safety

considerations. The  $x - y$  location of the UAV is denoted by  $x_u, y_u$ . The location of  $k$ -th ground user can be given by  $x_k, y_k$ . We resort to utilising the following log-normal shadowing pathloss model as it is capable of modeling wireless environments with acceptable precision. We formulate the path loss as [32] :

$$L = 20 \log(d) + 20 \log\left(\frac{4\pi f}{c}\right) + A(\theta) \quad (6.1)$$

where  $d$  is the distance between the receiver and transmitter,  $f$  and  $c$  are respectively the system frequency and speed of light, and  $A(\theta)$  is a log-normal distributed random variable with mean  $\mu$  and variance  $\sigma^2(\theta)$ , i.e.,

$$A(\theta) \sim \mathcal{N}(\mu, \sigma^2(\theta)) \quad (6.2)$$

given that  $\mu = 0$ , and  $\sigma^2(\theta)$  can be defined as :

$$\sigma^2(\theta) = \mathbb{P}_{LoS}^2(\theta)\sigma_{LoS}^2(\theta) + [1 - \mathbb{P}_{LoS}^2(\theta)]\sigma_{NLoS}^2(\theta) \quad (6.3)$$

where  $\sigma_{LoS}(\theta)$  and  $\sigma_{NLoS}(\theta)$  correspond respectively to the shadowing effect of LoS and NLoS links between the UAV and object, and they are given by :

$$\sigma_{LoS}(\theta) = a_{LoS} \exp(-b_{LoS}\theta) \quad (6.4)$$

$$\sigma_{NLoS}(\theta) = a_{NLoS} \exp(-b_{NLoS}\theta) \quad (6.5)$$

and  $\mathbb{P}_{LoS}(\theta)$  is the probability of having LoS link, which is written as :

$$\mathbb{P}_{LoS}(\theta) = \frac{1}{1 + a_0 \exp(-b_0\theta)} \quad (6.6)$$

where  $a_0, b_0, a_{LoS}, b_{LoS}, a_{NLoS}$  and  $b_{NLoS}$  are environment dependent parameters. Thus, the distance between the UAV and the device can be estimated as follows :

As described in [68], many localization techniques can be used in wireless networks like trilateration, multilateration, triangulation and others. The aforementioned techniques are based on GPS, RSSI, AOA (angle of arrival), TOA (time of arrival), or TDOA (time difference of arrival) measurements to perform localization of devices with unknown positions. RSSI-based techniques have been shown to provide an effective trade-off between accuracy, feasibility and complexity and, thus, are suitable for our proposed solution approach. Once an RSSI reading is captured, it needs to be converted to distance using an appropriate channel model. Thus, by considering the pathloss model from eq.6.1, we can write :

$$P_{ref}(dB) = P_t(dB) - L \quad (6.7)$$

where  $P_{ref}$  and  $P_t$  denote the reflected signal and the transmitted signal power, respectively. The received signal at the UAV coming from the reflection at the target can be defined as :

$$P_r(dB) = \delta P_{ref}(dB) - L \quad (6.8)$$

where  $\delta$  is reflection coefficient and is defined as standard normal distribution. Consequently, the distance between the UAV and the target to be localized can then be calculated as follows :

$$d = 10^{(P_t - P_r - (40 \log(d) + 40 \log(\frac{4\pi f}{c})) - \zeta)} \quad (6.9)$$

After mapping the received RSSI reading to its corresponding distance, well-known trilateration-based localization techniques can be used. In a two-dimensional space, three distance measurements from three distinct positions are recorded to generate three circles centered at the position where the measurements are taken with radii equal to the respective measurements. Should the distance measurements be accurate, the three circles intersect in one point that constitute the position of the object to be localized. Unfortunately, converting RSSI values to distances does not yield accurate measurements due to the statistical variations in wireless channels. As a result, the circles do not end up intersecting in one point but rather have an intersection area as demonstrated in Fig. 6.3, and the device's position is then estimated by minimizing the least square error. Due to variations in different environments, it is not possible in practice to estimate a fixed value for the shadowing component to be factored in the distance calculation in (4) . As a result, we address this problem by bounding the shadowing component between two designated values  $\zeta_{min}$  and  $\zeta_{max}$  and calculating the corresponding bounding distance values  $d_{min}$  and  $d_{max}$ , respectively, to form the radii of two concentric circles centered at the position of the UAV when the corresponding measurement is taken. The user is then expected to reside in the circular ring formed by the area enclosed by the two concentric circles. The UAV then moves and collects measurements from at least two other positions to satisfy the requirement of trilateration. Two concentric circles are generated from each measurement as depicted in the right side of Fig. 6.3 and the user location is then bounded to the area of intersection of all circular rings. The user's location is estimated to be the center of the resulting formed area.

For giving communication service to ground users, the effective rate of user  $k$  associated with the UAV is obtained by :

$$R_k = \log_2(1 + \gamma_k) \quad (6.10)$$

where  $\gamma_u$  is the signal-to-noise ratio (SNR) corresponding to the  $u$ -th user

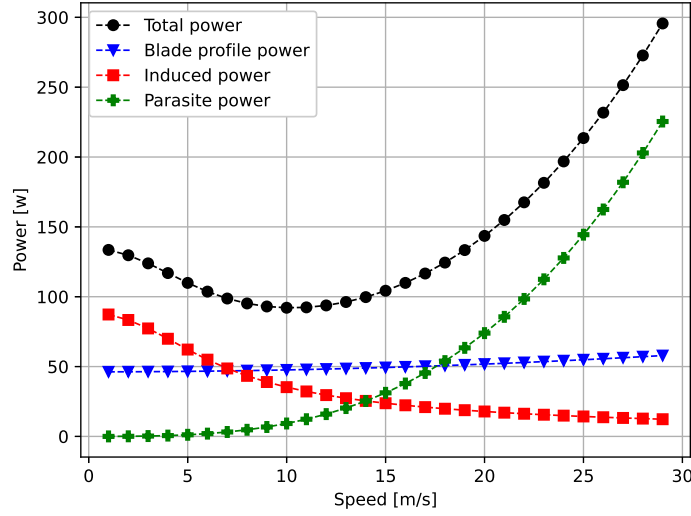


Fig. 6.2 – Power consumption versus UAV speed.

at time slot  $n$ , which can be expressed as

$$\gamma_k = \frac{P_t}{N10^{L_k/10}} \quad (6.11)$$

where  $P_t$  is the UAV transmit power,  $N$  is the power of the additive white Gaussian noise (AWGN) at  $u$ -th user and  $L_k$  represent signal attenuation as given in Eq. (6.1). We also assume orthogonal frequency-division multiplexing (OFDM) data transmission which enables the UAV to be less susceptible to interference and enables more efficient data bandwidth.

### 6.2.2 . Power Consumption Model

In this subsection, we present a suitable simple power consumption model for a UAV following the work presented in [69]. From the fact that the energy consumption of data communication is negligible compared to the energy required to keep the UAV aloft and fly, we compound the model into three main power consumption sources. The total power consumption of UAV when it is on the move can be written as follows :

$$P_{total} = P_{blade} + P_{parasite} + P_{induced} \quad (6.12)$$

where  $P_{blade}$  is the power required to turn the rotors' blade, and it is given by :

$$P_{blade} = K \left( 1 + 3 \frac{v^2}{v_b^2} \right) \quad (6.13)$$

where  $v$  is the UAV velocity,  $v_b$  is the blade's rotor speed, and  $K$  represents a constant which depends on the dimensions of the blade.



Parasite Power is the power used to overcome the drag force resulted from moving through the air.

$$P_{parasite} = \frac{1}{2}\rho v^3 F \quad (6.14)$$

$\rho$  is the air density, and  $F$  represents a constant that depends on the UAV drag coefficient and reference area. Note that this power is proportional to the UAV velocity  $v$ ; it is zero when hovering and gradually increases by the speed of the UAV.

This power is required to lift the UAV and overcome the drag caused by the gravity. Whenever a UAV is moving, the airflow coming at it redirects the UAV and helps to lift it. Hence, the induced power has inverse proportion to the airspeed. When hovering, all the airflow needed to lift the UAV has to be created by the blade rotors, which results in more power consumption. The induced power can be written as follows :

$$P_{induced} = mgv_i \quad (6.15)$$

where  $m$  and  $g$  respectively denote the mass of the UAV and the standard gravity, whereas,  $v_i$  represents the mean propellers' induced velocity in the forward flight, and it is given by :

$$v_i = \sqrt{\frac{-v^2 + \sqrt{v^4 + \left(\frac{mg}{A}\right)^2}}{2}} \quad (6.16)$$

with  $A$  being the area of the UAV. In the case of hovering, (i.e., when  $v = 0$ ), the total power consumption is limited to hovering power and is calculated accordingly :

$$P_{total} = P_{hover} = K + \sqrt{\frac{(mg)^3}{2\rho A}} \quad (6.17)$$

In Fig. 6.2, we show the trend of the three power consumption factors as well as the total power versus the UAV speed. As it is shown in the figure, we can conclude that at optimal speed ( $10[m/s]$ ), the UAV consumes less power compared to hovering time. Thus, in order to minimize the localization error with the knowledge of limited UAV battery, it is not always desirable to increase the number of RSS samples.

### 6.3 . Problem Formulation

In this work, we aim to maximize the number of transmitted bits and minimize the localization error of ground users at the same time while taking

in to account the constraint on UAV energy consumption. The UAV is required for perception of the urban environment and to implement real-time path planning. The decision of UAV flight trajectory and the choose of hovering position should consider the quality of communication, the precision of localization for ground users and energy consumption of UAV. As for the number of transmitted bits, its maximization depends on the amount of data that are sent over the UAV mission period. It can be easily concluded that to maximize  $R_{sum}$ , on one hand, the UAV should fly at a lower speed so that it can have a higher flight time, which means more transmitted bits. On the other hand, the hovering location should be close to the target users so as to improve the data rate. From this aspect, hovering over the intersection of all users is the best choice. As for the minimization of localization error, besides the maximization of RSS samples, we hope that more samples are taken from different positions on the UAV. It may conflict with the UAV's hovering directly over the intersection point of ground users to get the maximum data rate. As for the constraint of UAV's energy consumption and flight time, it is clear that slower speed can achieve minimum energy consumption and higher flight time. However, it may be not fast enough to collect more RSS samples and reduce the localization error.

It is evident that these two objectives are in conflict with each other partially. Due to the random distribution of devices and their dynamic numbers, it is considerably complicated and may impose significant computational cost to identify an optimal trajectory and hovering location decision. Moreover, the environment is partly observed, traditional model-based methods like dynamic programming method are incapable to solve this problem. Recently, DRL has accomplished an excellent ability to solve complex problems and is considered as one of the core technologies of machine learning. With integration of deep learning and RL, it owns the strong understanding ability and decision-making ability and thus can realize end-to-end learning. It has shown great potential in solving sophisticated network optimizations. DDQN, which is one of the DRL algorithms, has been proved that can learn effective policies in problems with complex optimal policy in great state space. It is suitable for our proposed UAV's flight decision problem where UAV is operating in a stochastic environment. Since the reward of original DDQN algorithm is scalar, we extend it to weighted sum reward for the multi objective optimization problem. The problem can be formulated as :

$$\begin{aligned}
& \text{rCI} * \max_{\mathbf{x}, \mathbf{y}, \mathbf{v}} \left( W_1 \mathbb{E} \sum_{k=1}^K R_n[k] + W_2 \text{MSE}(\hat{x}_k, \hat{y}_k), (x_k, y_k) \right) \forall k \\
& \text{s.t. } E_{total}[n] \leq \lambda B_u \\
& l_{min} \leq x[n] \leq l_{max}, \quad \forall n \\
& l_{min} \leq y[n] \leq l_{max}, \quad \forall n \\
& v_{min} \leq v[n] \leq v_{max}, \quad \forall n \\
& z[n] = H_u, \quad \forall n
\end{aligned}$$

$$P_t[n] = P_u, \forall n$$

In summary, we aim to find a control policy that can 1) maximize the system throughput; 2) minimize the localization error, and 3) ensure that the energy consumption of UAV does not exceed the battery capacity and the UAV is capable to return safely to recharging station. It is quite challenging to achieve all of these objectives because on one hand, to provide effective communication, it is preferred for the UAV to hover at a optimal position, on the other hand to minimize the localization error, it is preferred for the UAV to move around to different locations; and to minimize the energy consumption, it is preferred to reduce UAV movements (for energy savings). Hence, a good solution to this problem is supposed to well address this trade-off. Furthermore, (6.18b) ensures the UAV energy consumption to not exceed  $\lambda$  percentage of UAV on-board battery, (6.18c), (6.18d) and (6.18e) indicates the boundary of horizontal movements and speed of UAV in the environment, respectively. Also, (6.18f) and (6.18g) set the constraints for UAV's altitude and transmit power, respectively.

#### 6.4 . Calculating localization error

In this section, we clarify how the UAV estimates the location of ground users with received RSSI, and utilizing multilateration repeatedly to minimize the average position errors. To be more specific, we will describe how to calculate  $e[n]$  from reward function described in (6.25) and estimate future Q-value function  $Q(s_{t+1}, a_{t+1})$  for RL agent. Here, we depicts the localization process for single user and then it can as well be applied to other users. Finally, the average localization errors from all users will be the measured metric for the RL reward and Q-value at each state. In Fig. 6.3, we show the localization error reduction of a user by utilizing multilateration technique. The user to be located is highlighted by red dot, the UAV path by blue dashed lines, and user estimated location area by shaded green color. In the initial stage, by receiving RSSI measurement at on time stamp, following the channel model and (6.9) in section ??, the location of of the user is estimated in the shaded green zone between the inner ( $I_1$ ) and outer ( $O_1$ ) circles. The radius of these circles is dependent on the shadowing parameter and path loss exponent. In the next stage, when the UAV traverse to the next position and measure another RSSI measurement, the localization zone shrinks. Whenever the number of measurements becomes three, the position of the user can be estimated using trilateration, and consequently, the calculation of the localization error. As the number of samples and RSSI measurements increase, the localization error correspondingly reduces.

In Fig. 6.3, we illustrate how the error for one user using three samples can be calculated. The intersection point between three lines and connect

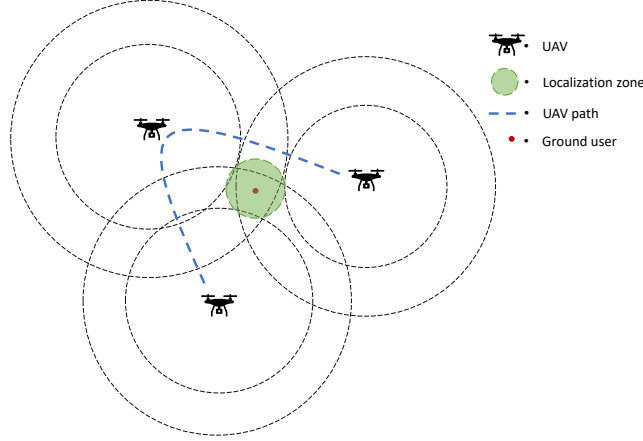


Fig. 6.3 – Tilateration for the case of one node in which shadowing component is bounded between two values.

inner and outer circles presents the estimated position of the user. Thus, the localization error can be obtained by finding the farthest border point to the estimated point as shown in the black line in the figure. Here, we consider the Cartesian coordinate for the estimated location of the user is  $(\hat{x}, \hat{y})$ . Let  $(x_{s_i}, y_{s_i})$  be the known ground position of the UAV at sample point  $i$ , and  $\bar{r}_i = \frac{O_i + I_i}{2}$  be the distance from sample point  $i$  to the middle of the two circles, then the estimated position  $(\hat{x}, \hat{y})$  using  $M$  number of samples can be calculated from the following optimization model :

$$(\hat{x}, \hat{y}) = \underset{\hat{x}, \hat{y}}{\operatorname{argmin}} \left\{ \sum_{i=1}^M \left( \sqrt{(x_{s_i} - \hat{x})^2 + (y_{s_i} - \hat{y})^2} - \bar{r}_i \right)^2 \right\} \quad (6.18)$$

The border points of the estimated zone of the user are generated each by the intersection of two RSSI circles. Fig. 6.4 shows how a border point is found. As shown in the figure,  $r_1$  and  $r_2$  are respectively the radius of sample points  $s_1$  and  $s_2$ , and  $k$  is the distance between the two sample points.  $P_1$  and  $P_2$  are the required intersection points between two circles, and  $P_0$  is the intersection point of the perpendicular line connecting  $P_1$  and  $P_2$  with line  $k$ . Respectively,  $q_1$  and  $q_2$  denote the distances from  $s_1$  to  $P_0$ , and from  $P_0$  to  $w_2$ , respectively. Now, if we let  $(x_{s_1}, y_{s_1})$ ,  $(x_{s_2}, y_{s_2})$ ,  $(x_{P_0}, y_{P_0})$ ,  $(x_{P_1}, y_{P_1})$ , and  $(x_{P_2}, y_{P_2})$  define respectively the Cartesian coordinates for points  $w_1, w_2, P_0, P_1$ , and  $P_2$ , then the border points are calculated through the following equations :

$$x_{P_1, P_2} = x_{P_0} \pm \frac{(y_{s_2} - y_{s_1})}{k} \quad (6.19)$$

$$y_{P_1, P_2} = y_{P_0} \mp \frac{(x_{s_2} - x_{s_1})}{k} \quad (6.20)$$

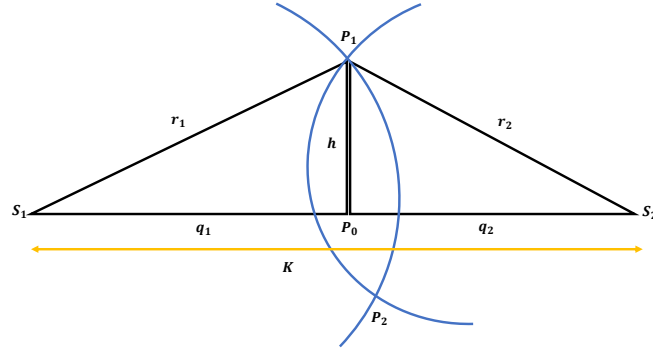


Fig. 6.4 – Illustration of attaining the intersection point between two RSSI arcs.

where  $(x_{P_0}, y_{P_0}) = (x_{s1} + \frac{(x_{s2} - x_{s1})q_1}{k}, (y_{s1} + \frac{(y_{s2} - y_{s1})q_1}{k}))$ ,  $q_1 = \frac{r_1^2 - r_2^2 + k^2}{2k}$  and  $h = \sqrt{r_1^2 - q_1^2}$ . After a new RSSI sample received by the UAV, the accuracy of the estimated user localization zone is updated by first removing the the previous border points and then add new intersection points (described above) and finally find distances from all obtained zone points to the estimated user point, and the one with farthest distance is the user's localization error. After obtaining the localization error for all ground users in the current state  $s_t$ , we average over all error values. Then, we evaluate the reward function corresponding to localization in eq. (6.25) by dividing the localization error calculated in the current state by  $e_{min}$  which is the set to arbitrary value for minimum possible localization error, i.e  $10[m]$ . Similarly, we estimate the future average localization errors for all available neighbor sample points and actions, and we update the approximated Q-value function for all actions and store them into the table. Subsequently, for the next iteration, we choose the action that results in higher reward by looking at the stored Q-value functions.

## 6.5 . Preliminaries

As shown in Fig. 1, utilizing the multilateration technique to find the position of targets with lower localization errors, the UAV needs to travel to more waypoints. However, with limited flight time due to the UAV battery capacity and the path length, a certain UAV trajectory results in optimal localization precision. Therefore, to find the best trajectory, we let the UAV interact and observe the environment by using RL and learn to autonomously find the optimal trajectory that can achieves the minimum localization errors. In this section, we review the RL framework, a machine learning approach

which is suitable for controlling an autonomous machine such as UAV.

RL is a learning approach that is used for finding the optimal way of executing a task by letting an entity, named agent, take actions that affect its state within the acting environment. In RL, the environment is typically formulated as an MDP, which is described by four tuples  $(S, A, R, P)$ , set of possible state  $S$ , set of available actions  $A$ , and reward function  $R : S \times A$  and transition probability  $P(\hat{s}|s, a) \rightarrow [0, 1]$ . The agent interacts with an unknown environment through the repeated observations, actions, and rewards to construct the optimal strategy. When interacting with the environment, after choosing an action  $a_t \in A$ , the agent receives a reward  $r(s_t, a_t)$  and moves to the next state  $s_{t+1}$ . The goal of RL is to learn from the transition tuple, and find an optimal policy  $\pi^*$  that will maximize the cumulative sum of all future rewards. Note that the policy  $\pi = (a_1, a_1, \dots, a_T)$  defines which action  $a_t$  should be applied at state  $s_t$ . If we let  $r(s_t, \pi(a_t))$  denote the reward obtained by choosing policy  $\pi$ , the cumulative discount sum of all future rewards using policy  $\pi$  is given by :

$$R_\pi = \sum \gamma^{t-1} r(s_t, \pi(a_t)) \quad (6.21)$$

where  $\gamma \in [0, 1)$  is a discount factor, which measures the weight given to the future rewards (i.e., when  $\gamma = 0$ , the agent considers only the current received rewards, whereas, when the factor approaches one, the agent strives for future higher reward). Now, let  $\Lambda$  denote the set of all admissible policies. Then, the optimal policy is given by :

$$\pi^* = \operatorname{argmax}_{\pi \in \Lambda} R_\pi \quad (6.22)$$

Note that RL is modeled as a Markov Decision Process (MDP), where the tuple  $(s_t, a_t, r(s_t, a_t), s_{t+1})$  is conditionally independent of all previous states and actions. Therefore, the agent does not need to memorize or save all the state-action tuples, just the last one, and subsequently updates it at each cycle or iteration.

In this work, we rely on double-Q-learning algorithm to solve our problem which allows us to keep two separate agents with the same properties but with different weight values  $w_P$  and  $w_T$ . As such they will output a different Q-action function when given the same state. One is used to choose the actions, called a primary model  $Q_P(s_t, a_t)$ , while the other model evaluates the action during the training, called a target model  $Q_T(s_t, a_t)$ . Therefore training occurs when taking a batch of experiences  $e_t$  from the buffer that is used to update the model as :

$$Q_P^{new} = (1 - \alpha)Q_P + \alpha [r_t + (1 - d_t)\gamma \max Q_T(s_{t+1}, a)] \quad (6.23)$$

UAV's altitude ( $h$ )	100[m]
Rotor solidity ( $s$ )	0.05
Profile drag coefficient ( $\delta$ )	0.012
UAV weight ( $N$ )	20[N]
Air density ( $\rho$ )	1.225[kg/m <sup>3</sup> ]
Rotor disc area ( $A$ )	0.503[m <sup>2</sup> ]
Rotor blade tip speed ( $U_{tip}$ )	120[m/s]
Induced power correction factor ( $k$ )	0.1
Environment constant for $P_{LoS}$ ( $a_0$ )	45
Environment constant for $P_{LoS}$ ( $b_0$ )	10
Shadowing constant for $P_{LoS}$ ( $a_{LoS}$ )	10
Shadowing constant for $P_{LoS}$ ( $b_{LoS}$ )	2
Shadowing constant for $P_{NLoS}$ ( $a_{NLoS}$ )	30
Shadowing constant for $P_{NLoS}$ ( $b_{NLoS}$ )	1.7

Table 6.1 – Simulation Parameters

where  $\max Q_T(s_{t+1}, a)$  is the action chosen as per the agent,  $\alpha$  is the learning rate which was an input to the Adam optimizer [28], and  $\gamma$  is a discount factor that reduces the impact of long term rewards. We implement this with soft updates where instead of waiting several episodes to replace the target model with the primary. The target model receives continuous updates discounted by value  $\tau$  as in  $w_T = w_T(1 - \tau) + w_P\tau$ .

### 6.5.1 . Proposed RL framework

It is of great importance to cast the optimization problem into the MDP in a proper way. The agent rely on the interaction with the environment to adapt its behavior and learn optimal policies. In the following, we present a detailed description of state space, action space and reward in our model.

1) State Space : Collecting the real-time parameters in the environment depends on frequent information exchange between the UAV and ground devices. This will cause delay, considerably reducing the efficiency of the system and occupy a large amount of wireless resources. To be more realistic, we consider that the UAV can only observe its own state and partial network

Table 6.2 – Network Configuration

Parameters	Values
Number of training episodes	16000
Learning rate	0.0001
Discount factor	0.99
Replay memory size	8000
Batch size	32
Number of neurons	256
SGD optimizer	Adam
Activation function	ReLU

information. Specifically, UAV can record its own location, the estimated distances and localization error of ground targets, communication rate of users and UAV energy consumption. Thus the state space is defined symbolically as :

$$S = \{s_t\} = \{x_u(t), y_u(t), d_j(t), e_j(t), R_k(t), E_u(t)\} \quad (6.24)$$

where  $d_j(t)$  is the distance between the target device and the UAV under the Cartesian coordinates,  $e_j(t)$  is the calculated localization error described in previous section,  $R_k(t)$  is the communication rate of the corresponding user and  $E_u(t)$  is the energy consumption of the UAV . In practical scenarios, most of the information is not necessary for decision-making. In our setup, we extract a small amount of essential information to represent the state of the environment. These elements of state space will enable the UAV to have a reasonable general perception of the environment. Moreover, it overcomes the lack of network information which is common problem that exists in DFRC systems.

2) Action Space : The action space  $A$  is described by all possible movement directions and the action of remaining in the same place. By assuming that the UAV fly with simple coordinate turns, the actions related to movement of UAV is simplified to 7 directions.

3) Reward : The reward function incorporates instantaneous throughput of users and localization error of targets. It can be written as follows :

$$r[n] = w_R \frac{R[n]}{R_{max}} + w_L \frac{e_{min}}{e[n]} \quad (6.25)$$

where  $R_{max}$  is the maximum achievable rate in the environment,  $e_{min}$  is minimum desired localization error,  $w_R$  and  $w_L$  are corresponding weights for each reward function.



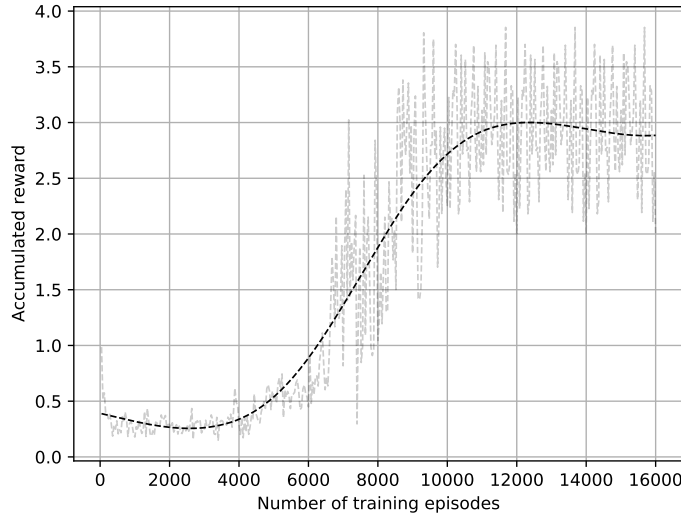


Fig. 6.5 – Accumulated reward vs number of training episodes.

## 6.6 . Numerical Results

In this section, we evaluate the performance of our RL approach in localizing terrestrial targets and giving service to ground users numerically. We generate randomly the locations of the targets and ground users, in which we want the UAV to localize and giving service communication, respectively. Based on the environment parameters and the probability of LoS, we evaluate the range between the inner-circle and outer-circle where the target is located from the ground reflection of UAV's position. Thus, the zone attained from the intersection of multiple inner and outer circles is considered as the location zone of the target. Consequently, the localization error or accuracy can be evaluated by calculating the distance from the farthest border point to the center of the zone. Moreover, by adding more samples from different positions, the location error is reduced.

For the numerical study, we assume 10 terrestrial targets and 20 ground users which are randomly distributed in a region of  $750 \times 750 \text{ m}^2$ . We also assume the UAV is flying at a fixed altitude  $H_u$ . The parameters used in this section and their corresponding values (taken and recommended by [5], [28], [30], [38] for urban environments) are listed in Table 1. In summary, we first illustrate the convergence of the proposed DDQN in the considered multi objective optimization problem. Then we study the performance of our system for communication and sensing by varying UAV's speed. Later on, we study the performance of our RL method by varying the corresponding weights for communication rate and localization error in our RL reward function. Finally, we show the performance trade-off between localization error and number of

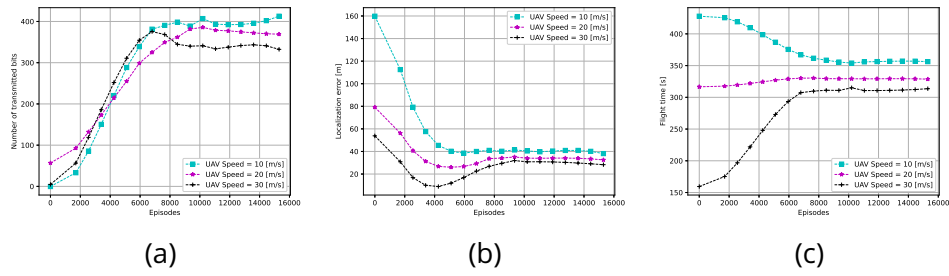


Fig. 6.6 – Training curves tracking optimization objectives : (a) Number of transmitted bits; (b) Localization error; (c) UAV flight time.

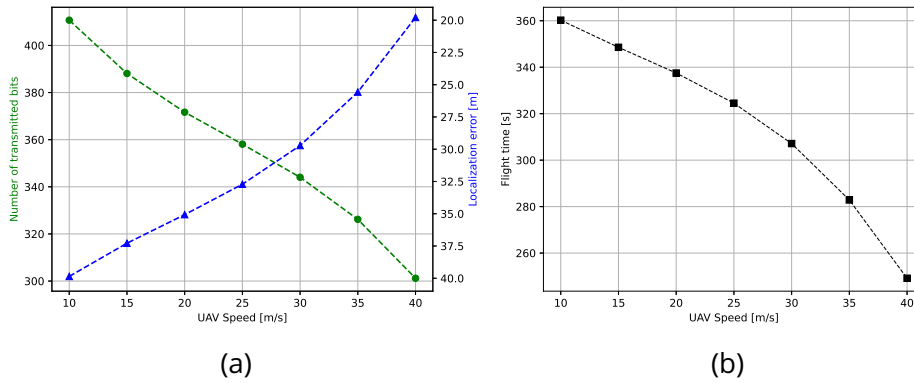


Fig. 6.7 – Impact of UAV speed on (a) Number of transmitted bits and localization error; (b) UAV flight time.

transmitted bits.

We start by illustrating the effectiveness and convergence of the proposed DDQN algorithm. The learning curve of the trained DDQN agent is shown in Fig. 6.5. The figure plots the accumulated reward versus number of training episodes. Here, the weight parameters are set to  $W_R = W_L = 1.0$ . We consider the jointly optimization of two objectives. It can be seen in Fig. 6.9 that the agent quickly learns to obtain higher expected total rewards as training progresses. And then the accumulated reward converges steadily at a high level. At first about 10000 episodes, the accumulated reward fluctuates at a very low level. It is because that the UAV is in complete experience stage. Without enough experience to learn from, the action is chosen randomly. At the same time, the loss of the network is 0 and the objectives are not optimized. When the replay memory is full, the UAV begins to sample the stored experience tuples to train networks. We can see that there is a major exploration and learning stage before about the 10000th episode.

The changing trend of two objectives as well as flight duration of corresponding results during the training are also illustrated in Fig. 6.6a, Fig. 6.6b and Fig. 6.6c. We start by examining the results obtained by training the RL agent and compare different UAV speeds on localization and communication performance. Fig. 6.6a depicts the number of transmitted bits achieved during multiple episodes for training RL agent. As it is shown in the figure, after 8000 episodes the RL agent reaches convergence. As can be seen in the figure, the UAV operating at speed  $30[m/s]$  can transmit around 330 bits, while moving at speed  $20[m/s]$  it can transmit approximately 360 bits and when traversing with  $10[m/s]$ , the UAV can transmit 400 bits during its mission duration. Fig. 6.6b illustrates the localization error obtained through training episodes. We can observe that after 8000 episodes, the RL agent reaches convergence for minimizing localization error. As the figure shows, when UAV is moving at speed  $30[m/s]$ , it can achieve the localization error of  $28[m]$  while moving at speed  $20[m/s]$ , it can reach the localization error of  $34[m]$  and when the UAV is operating with  $10[m/s]$ , it achieves  $40[m]$  error for localization. In the Fig. 6.6c, we show the flight time of UAV during training. Similar to previous figures, the UAV will return back to recharging station after reaching 70% of its battery and the RL agent reaches the convergence after 8000 episodes. From the figure we can see that the UAV has a flight time of 360, 330 and  $320[s]$  when moving at speed 10, 20 and  $30[m/s]$ , respectively.

Fig. 6.7 summarizes the comparison results for UAV speed on three performance metrics. The UAV speed is set from  $10m$  to  $40m$ . It can be seen that when the UAV operates at lower speeds, since it consumes less energy than other speed variations and the flight time is the highest, it can achieve the highest number of transmitted bits. On the other hand, when the UAV moves at higher speeds, it consumes the largest energy based on the adopted propulsion power consumption model and so lowest flight time and low number of transmitted bits. However, since it is moving at higher speed, it travels the longest path which means RSSI samples from different positions that results to better localization error. From the figures, it is clear that with limited energy the localization error reduces along with increasing the UAV speed, but the number of transmitted bits lessened.

In Fig. 6.8, we evaluate the performance of RL approach by varying the weights in the reward function from (6.25). For this purpose, we tested different weight numbers for communication rate and localization error rewards. Here, we chose two set of weights values ( $W_1$  and  $W_2$ ) that can capture the impact of reward function in the trade-off between communication and localization.  $W_1$  corresponds to the scenario when the weight of the communication reward is larger than localization, and  $W_2$  is for the case in which the weight of the localization is larger than communication. Figure.

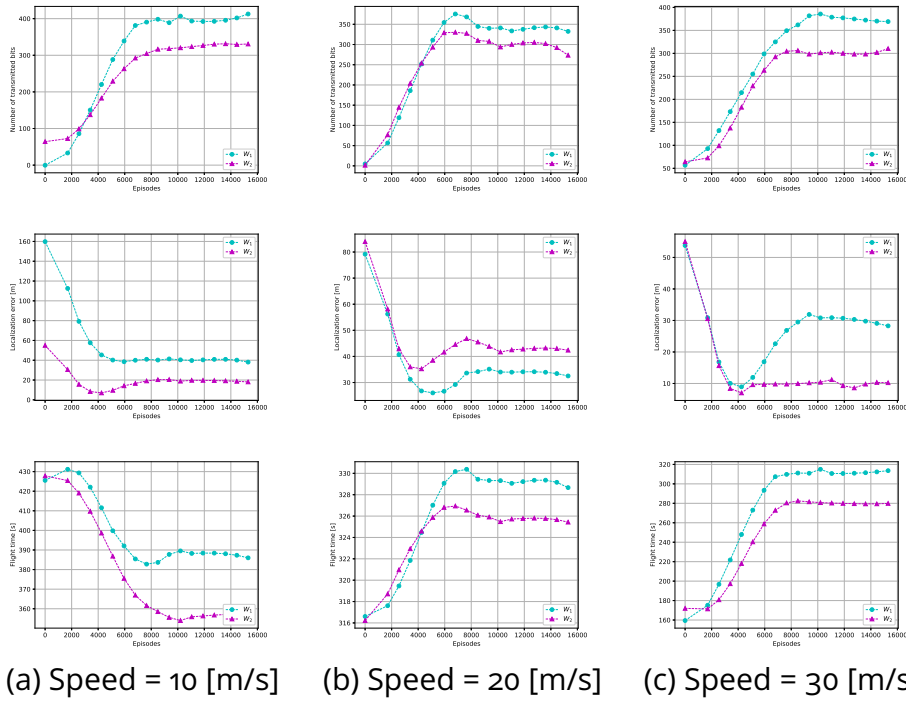


Fig. 6.8 – Performance comparison for two sets of weights in (6.25).

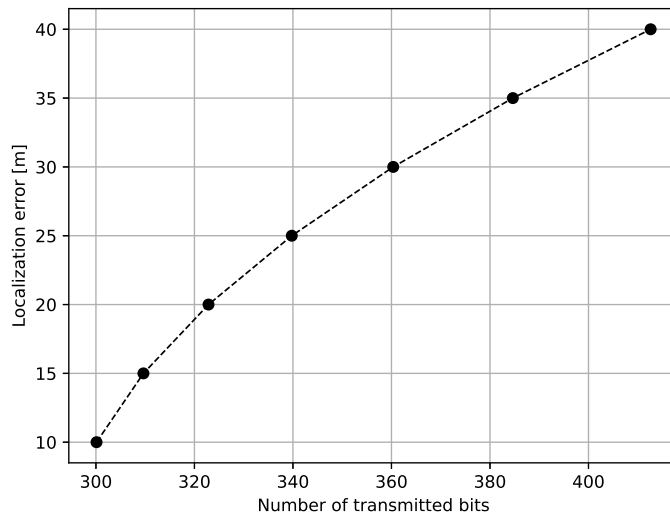


Fig. 6.9 – Localization error versus number of transmitted bits.

7a plot the number of transmitted signals, localization error and flight time during each training session when the uav speed is set to  $10[m/s]$ . It can be seen that for the case of  $W_1$ , after convergence, the UAV achieves higher transmitted bits during an episode in comparison with  $W_2$ . However, in  $W_2$

case, the UAV archives better localization performance than  $W_1$ . The flight time difference between these two cases capture the fact that when weight for communication reward is larger than localization reward, the UAV tends to hover more than moving to different spots which means that the UAV found the optimal spot for giving service to ground users and hover at that position to maximize the system throughput.

In Fig. 6.9, we show the trade-off between communication and sensing with respect to discussion in previous sections on multi objective optimization of localization and system throughput. The figure plots the localization error and number of transmitted bits resulted from different UAV speeds and multi objective weights in (6.25). In fact, whenever we increase the transmitted bits, the localization error decreases. Thus, we can see the trade-off between these two optimization objectives. To achieve any specific system performance, we can modify the weight values  $W_R$  and  $W_L$  in (6.25) and UAV speed to achieve desirable communication throughput and localization error.

## 6.7 . Conclusion

In this chapter, we studied a multi objective optimization for UAV path planing in a DFRC network, where a single UAV is employed to simultaneously serve a group of ground users for communications and localize the ground targets. we proposed a novel framework using RL to let a UAV autonomously choose a trajectory that results in finding the position of multiple targets with minimum average localization error and maximize the average number of transmitted bits to communication users under a constraint on UAV energy consumption or flying time.

## 7 - Conclusion

This chapter highlights general conclusions of the thesis and summaries possible directions for future work.

## Sommaire

---

<b>7.1</b>	<b>Conclusion . . . . .</b>	<b>86</b>
<b>7.2</b>	<b>Future Work . . . . .</b>	<b>88</b>
7.2.1	Machine Learning-aided Wireless Networks . . . . .	88
7.2.2	Federated Learning in Future Networks . . . . .	90
7.2.3	Machine Learning for Reconfigurable Intelligent Surfaces . . . . .	91

---

### 7.1 . Conclusion

In this thesis, new contributions on modeling, evaluating and optimizing the next generation of Unmanned Aerial Vehicles (UAV) communication systems have been reported. In particular the emerging technology of Machine Learning (ML), as a promising enabler for wireless communications beyond 5G, has been inspected and utilized. More specifically, the contribution of this thesis can be summarized as follows.

- In chapter 2, we provided a comprehensive study on the use of UAVs in wireless networks. We have investigated the main use cases of UAVs as aerial base stations and cellular-connected users. For each of the applications, we have explored key challenges and fundamental problems.
- In chapter 3, we have covered in detail the new research directions when ML techniques are utilized to increase the performance of UAV networks. We provided an extensive overview of ML techniques, specifically RL, that have been applied in UAV networks. Then, we discussed FL principles and advantages and where a FL approach can be used in the field of UAV networks.
- In chapter 4, we designed a new UAV-aided communication system relying on the shortest flight path of the UAV while maximizing the amount of data transmitted to mobile devices. In the considered system, we assumed that UAV does not have the knowledge of user's location except their initial position. We proposed a framework which is based on the likelihood of mobile users presence in a grid with respect to their probability distribution. Then, a deep reinforcement learning technique is developed for finding the trajectory to maximize the throughput in a specific coverage area. Numerical results were

presented to highlight how our technique strike a balance between the throughput achieved, trajectory, and the complexity.

- In chapter 5, we studied the localization of ground users by utilizing UAVs as aerial anchors. Specifically, we introduced a novel localization framework based on FL and RL. In contrast to the existing literature, our scenario includes multiple UAVs learning the trajectory in different environment settings which results in faster convergence of RL model for minimum localization error. Furthermore, to evaluate the learned trajectory from the aggregated model, we test the trained RL agent in a fourth environment which shows the improvement over the localization error and convergence speed. Simulation results show that our proposed framework outperforms a model trained with transfer learning by %30.
- In chapter 6, we explored the optimal trajectory for maximizing communication throughput and minimizing localization error in a DFRC in (UAV) network where a single UAV serves a group of communication users and locate the ground targets simultaneously. To balance the communication and localization performance, we formulated a multi-objective optimization problem to jointly optimize two objectives : maximization of number of transmitted bits sent to users and minimization of localization error for ground targets over a particular mission period which is restricted by UAV's energy consumption or flying time. These two objectives were in conflict with each other partly and weight parameters are given to describe associated importance. Hence, in this context, we proposed a novel framework based on RL to enable the UAV to autonomously find its trajectory that results in improving the localization accuracy and maximizing the number of transmitted bits in shortest time with respect to UAV's energy consumption. We demonstrated that the proposed method improves the average transmitted bits significantly, as well as the localization error of the network.



## 7.2 . Future Work

Many topics of interest in the field of machine learning in wireless communications and cellular networks remain open. Relying on the findings of this thesis, possible directions for research can be the following.

### 7.2.1 . Machine Learning-aided Wireless Networks

The ML-aided and learning-based wireless networks will carry out unique decision-making capacities and real-time estimation towards transforming 5G and beyond 5G networks. In this direction, the latest advancements in the ML algorithms have set up up new opportunities for the UAV-based systems and have introduced the possibility of realizing highly autonomous UAV missions while improving the system performance, safeguarding the security, and reducing human errors under complex and random scenarios. However, there exist open research problems which demand certain focus and should be investigated. In the following, we summarize a range of future research directions in ML-aided wireless networks.

In principle, the ML methods rely on massive and high-quality labeled data sets to accomplish the desired outputs. Because of the development of IoT sensor systems and cloud computing, currently the data availability is economically and technologically less expensive, but, the data collected by sensors and network equipment is generally subject to losses, redundancy, mislabeling, and class imbalance. Thus, the effectiveness of the training procedure is uncertain. Right now, the TensorFlow library which is an open-source framework developed by Google for ML inference on low-power embedded devices, can be utilized in order to successfully exploit ML and to help object recognition on unclassified data. Furthermore, Keras, a high-level neural networks application programming interface (API) written in Python, is able to run on top of TensorFlow and allow fast experimentation. Also, today's powerful multi-core central processing unit (CPU) architectures, GPUs, and broad availability of libraries for Deep Learning (DL) enable engineers for fast, parallel data processing in real-time. Nonetheless, the restricted battery capacity and on-board processing capabilities and power resources of the UAVs extremely limit the application of DL-based methods, which supports the object detection, depth prediction, target tracking and localization, and decision-making on the fly. Specifically, above all, realistic constraints regarding the computational power, parallel data processing in real-time, and power consumption restrict the design and implementation of effective DL solutions on drones. As powerful miniaturized computing devices with low-power consumption are an active working field for embedded hardware developers, these problems are envisioned to be solved in the near future. Consequently, future research efforts must be dedicated to further inspecting and proving the performance of ML-empowered aerial networks, particularly

in terms of the computation capability and hardware design. For this purpose, by adding a confidence score to predictions and a scale factor to the generated actions, the future developments can improve the processing time of the learning algorithms. Additionally, one could integrate various ML techniques in order to cooperatively complete the prediction procedure and therefore improve the computational efficiency. Aside from containing a larger number of samples for optimization, it is also worthwhile to derive the optimal parameters of the learning algorithms to achieve faster convergence. For the special scenarios of UAV swarms consisting of micro-drones with restricted capabilities, the DL methods can run on a traditional base station (BS) with high computational power that will function as a central manager connected to the UAV mesh network. To determine the best possible action, this base station will rely on the sensor data from all of the drones. However, this control approach is not the optimal choice, because it usually introduces more signaling overhead and transmission latency as a consequence of the necessary information exchange between the BS and the UAVs.

Right now, there is a gap in obtaining data from extensive measurement operations. As a future work, one can realize test-beds achieving real experiments in different propagation environments, particularly in dense, urban, skyscraper-rich settings and over sea areas for the sake of validating the accuracy of the learning algorithms. Specifically in scenarios when environment randomly changes and ground nodes moving at high velocities, while taking in to account the interference in the propagation area and examine with real-world constraints, such as the energy efficiency of the learned trajectory. However, these real-world problems typically comprise high-dimensional continuous state spaces, i.e, large number of states and/or actions, and make the corresponding problems nearly intractable with current methods. Since more measured data is necessary, ML-based algorithms can support new developments in UAV channel modeling. Additionally, ML techniques can be applied beyond channel estimation, such as the power-delay profile, correlation coefficient and matrices. Moreover, interference mitigation presents a considerable hurdle towards the effective integration of drones in future networks. ML also can improve the performance of multiple methods that have been investigated in ground networks, such as power control, UAV-user association, and seamless handover using ML techniques for predicting user mobility and network load, while the use of ML in forming the precoding matrices of massive MIMO-enhanced drones can remove the interference and can boost the quality of the transmission. Another field that can be improved by ML is the clustering of users and UAVs towards improved NOMA in the downlink and uplink, rising the chances of successful interference cancellation and maximizing the spectral efficiency of the network. Recently, the research community has been investigating the joint optimization of throughput sup-

plied by UAV BS and energy that they utilize for recharging from the grid. Consequently, in cases where multiple UAVs are installed to serve as aerial BSs, the joint consideration of physical-layer parameters and energy and the application of ML algorithms, such as DL to process heterogeneous data, can grant an increased performance, as network lifetime prolongation is a crucial feature of UAV networks.

### 7.2.2 . Federated Learning in Future Networks

One important factor that should not be overlooked is the fact that FL is not necessarily applied only for UAV or mobile user networks, instead, it is being used successfully in many daily applications. For instance, Google implements the FL to learn a RNN to predict our next word when we start typing on the keyboard. Nonetheless, it should be pointed out that it is not certain how to select specific parameters in the FL algorithm. For instance, the client selection process has been defined as random, which raises the question of whether there is a superior approach to assign clients in each round of the FL algorithm. The aforementioned issue requires more investigation in depth for UAV networks where several parameters can affect the client selection process. From a wireless communication point of view, channel quality, LoS/NLoS link, available data, and battery state are important factors that can substantially impact the client selection process. In particular, those parameters can make a subgroup of users more suitable to be chosen for the FL training. Moreover, although a great part of the research community argue that the main goal of FL is data confidentiality, others question this assumption and state that even sharing only updates over the wireless network is not secure. In some part that is true, since the FL can be subject to a virulent attack threatening the integrity of the model. These kinds of attacks are popular in the ML community by backdoor attacks and are generally executed either by a single node or by a group of nodes infuse wrong data into the model to negatively influence it. Above all, even FL stays vulnerable to this category of attacks not by sending wrong data but by infecting the model itself by some malicious clients. In future, as a advanced solution to the unreliability of FL systems, we suggest to aid drone networks using Blockchain methods to increase the integrity of local models at each UAV. The combination of Blockchain and FL is examined as a major breakthrough and a number of recent research works have begun to study this topic.

It has been stated that in addition to the increased level of stability and integrity, the Blockchain method can boost the users motivation to participate in training by precisely rewarding them for their contribution. Recently, the research community has begun to implement the concept of a Blockchain combined with FL to propose solutions for drone networks. For instance, secure FL framework can be applied to mobile crowdsensing

aided by a UAV-network. The local model exchanges of the FL algorithm can be secured with respect to a Blockchain architecture. In summary, we emphasize the potential of coupling Blockchain and FL in future works. Aside from the security issues, more focus should be given to the convergence of a FL algorithm which is not always guaranteed. Convergence depends on the particular class of problem, such as the convexity of the loss function and the number of updates performed on the model. For instance, the optimization of the overall model will fail if we pick wrong clients that are not available or do not have enough data. It should be noted that this problem overlaps with the client selection issue mentioned previously and it is associated with client selection and also to the type of the loss function.

Furthermore, as we proposed FL as a solution to train ML model on multiple UAVs in different environment settings, we should also mention that the massive exchange of updates across the network will bring in a huge amount of communication loads in the training phase, specifically for neural networks, which will induce a scalability problem for FL. Many CNN architectures demand a large number of parameters to be updated at each round. In fact drone networks are generally characterized by a restricted battery capacity and limited bandwidth, which makes the UAVs unable to support all these communication loads. To solve this problem, many researchers have been working on alternatives and approaches that could improve memory consumption and communication efficiency by proposing compression techniques and reducing the number of communication rounds. However, a drawback of FL starts to appear when operating in a heterogeneous UAV network formed by various types of UAVs, rotary or fixed-wing, with different processing capabilities and different GPUs. These dissimilarities mean that some drones will have fast response times while others will experience severe delays. Consequently, these induced delays will cause an important issue by significantly slowing down convergence since the FL algorithm is anticipated to receive the required model updates at each communication round. In future, a distributed computation scheme can be introduced to reduce the influence of slow nodes on convergence for gradient methods. Additionally, the quality of connectivity can impact the convergence of the FL algorithm due to the fact that several network nodes may encounter an unexpected failure when transmitting their local updates. These interruptions can also reduce the overall performance of the FL by slowing the convergence speed which should be investigated.

### **7.2.3 . Machine Learning for Reconfigurable Intelligent Surfaces**

Next-generation wireless networks should deal with a growing density of mobile users while accommodating a swift rise in mobile data traffic flow and a wide range of services and applications. In future networks, high-frequency waves will act as an crucial role, but these signals are regularly obstructed

by objects and diminish over long distances. Reconfigurable intelligent surfaces (RISs) is a promising solution due to their potential to improve wireless network capacity and coverage by intelligently changing the wireless propagation environment. Therefore, RISs carry out a potential technology for the sixth generation of communication networks. In fact, for maximizing the possible advantages of RIS-assisted communication systems, ML is an effective method when the computational complexity of operating and deploying RIS increases rapidly as the number of interactions between the user and the infrastructure starts to expand. Considering the fact that ML is a promising approach for improving the network and its performance, the application of ML in RISs is anticipated to open new paths for interdisciplinary studies as well as practical applications.

It should be noted that some certain challenges must be addressed before obtaining the advantages of RISs. Accurate channel state information (CSI) for optimum reflection on the RIS is mandatory. It is very demanding for a realistic RIS-empowered wireless network to obtain a precise value for CSI on a continuous basis because of capacity in flexibility of the served client and the obstruction-prone character of the signal. Thus, the issues of CSI evaluation and optimization of network performance under weak CSI must be accordingly addressed to permit a real-time and effective RIS-assisted transmission. Channel assessment complexity is high in RIS-assisted wireless networks due to the considerable number of components been used, which is a major challenge. Furthermore, gaining channel knowledge may need a extensive training overhead. Moreover, the phase shift of the reflecting elements complicates the designing of an ideal passive beamforming system, and the conventional methods demand complicated procedures for the configuration of the RIS which is both power and time consuming. As a consequence of their ability to learn and the requirement of operating over wider search areas, ML approaches have attracted attention in wireless communications, particularly in the field of RISs. In the future, scholars must attempt to overcome these obstacles. They can utilize various ML algorithms for the communication sector so that the infrastructure can independently solve all challenges. The majority of ML techniques function by learning the parameters and constructing an optimization model from the input information for the goal function. In our present time, since a large amount of data must be handled, the efficiency and effectiveness of mathematical optimization procedures substantially affect the popularity and application of ML models.

## Bibliographie

- [1] A. Shahbazi and M. Di Renzo, "Analysis of optimal altitude for uav cellular communication in presence of blockage," in *2021 IEEE 4th 5G World Forum (5GWF)*, pp. 47–51, IEEE, 2021.
- [2] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Efficient deployment of multiple unmanned aerial vehicles for optimal wireless coverage," *IEEE Communications Letters*, vol. 20, no. 8, pp. 1647–1650, 2016.
- [3] M. Mozaffari, W. Saad, M. Bennis, Y.-H. Nam, and M. Debbah, "A tutorial on uavs for wireless networks : Applications, challenges, and open problems," *IEEE communications surveys & tutorials*, vol. 21, no. 3, pp. 2334–2360, 2019.
- [4] J. Lyu, Y. Zeng, and R. Zhang, "Uav-aided offloading for cellular hotspot," *IEEE Transactions on Wireless Communications*, vol. 17, no. 6, pp. 3988–4001, 2018.
- [5] A. Merwaday and I. Guvenc, "Uav assisted heterogeneous networks for public safety communications," in *2015 IEEE wireless communications and networking conference workshops (WCNCW)*, pp. 329–334, IEEE, 2015.
- [6] Y. Zeng, R. Zhang, and T. J. Lim, "Wireless communications with unmanned aerial vehicles : Opportunities and challenges," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 36–42, 2016.
- [7] Y.-H. Nam, M. S. Rahman, Y. Li, G. Xu, E. Onggosanusi, J. Zhang, and J.-Y. Seol, "Full dimension mimo for lte-advanced and 5g," in *2015 Information Theory and Applications Workshop (ITA)*, pp. 143–148, IEEE, 2015.
- [8] T. Lagkas, V. Argyriou, S. Bibi, and P. Sarigiannidis, "Uav iot framework views and challenges : Towards protecting drones as "things"," *Sensors*, vol. 18, no. 11, p. 4015, 2018.
- [9] M. Chen, M. Mozaffari, W. Saad, C. Yin, M. Debbah, and C. S. Hong, "Caching in the sky : Proactive deployment of cache-enabled unmanned aerial vehicles for optimized quality-of-experience," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 5, pp. 1046–1061, 2017.
- [10] M. Mozaffari, A. T. Z. Kasgari, W. Saad, M. Bennis, and M. Debbah, "Beyond 5g with uavs : Foundations of a 3d wireless cellular network," *IEEE Transactions on Wireless Communications*, vol. 18, no. 1, pp. 357–372, 2018.

- [11] A. Al-Hourani, S. Kandeepan, and A. Jamalipour, "Modeling air-to-ground path loss for low altitude platforms in urban environments," in *2014 IEEE global communications conference*, pp. 2898–2904, IEEE, 2014.
- [12] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et al., "Human-level control through deep reinforcement learning," *nature*, vol. 518, no. 7540, pp. 529–533, 2015.
- [13] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, "Deterministic policy gradient algorithms," in *International conference on machine learning*, pp. 387–395, PMLR, 2014.
- [14] M. Riedmiller, "Neural fitted q iteration—first experiences with a data efficient neural reinforcement learning method," in *European conference on machine learning*, pp. 317–328, Springer, 2005.
- [15] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5g be?," *IEEE Journal on selected areas in communications*, vol. 32, no. 6, pp. 1065–1082, 2014.
- [16] V. W. Wong, R. Schober, D. W. K. Ng, and L.-C. Wang, *Key technologies for 5G wireless systems*. Cambridge university press, 2017.
- [17] I. Valiulahi and C. Masouros, "Multi-uav deployment for throughput maximization in the presence of co-channel interference," *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3605–3618, 2020.
- [18] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Unmanned aerial vehicle with underlaid device-to-device communications : Performance and tradeoffs," *IEEE Transactions on Wireless Communications*, vol. 15, no. 6, pp. 3949–3963, 2016.
- [19] M. Zhang, S. Fu, and Q. Fan, "Joint 3d deployment and power allocation for uav-bs : A deep reinforcement learning approach," *IEEE Wireless Communications Letters*, 2021.
- [20] C. Zhao, J. Liu, M. Sheng, W. Teng, Y. Zheng, and J. Li, "Multi-uav trajectory planning for energy-efficient content coverage : A decentralized learning-based approach," *IEEE Journal on Selected Areas in Communications*, 2021.
- [21] X. Liu, Y. Liu, and Y. Chen, "Reinforcement learning in multiple-uav networks : Deployment and movement design," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 8036–8049, 2019.
- [22] C. Bettstetter, "Smooth is better than sharp : A random mobility model for simulation of wireless networks," in *Proceedings of the 4th ACM international workshop on Modeling, analysis and simulation of wireless and mobile systems*, pp. 19–27, 2001.

- [23] A. Asahara, K. Maruyama, A. Sato, and K. Seto, "Pedestrian-movement prediction based on mixed markov-chain model," in *Proceedings of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems*, pp. 25–33, 2011.
- [24] J. Krumm and E. Horvitz, "Predestination : Inferring destinations from partial trajectories," in *International Conference on Ubiquitous Computing*, pp. 243–260, Springer, 2006.
- [25] J. J.-C. Ying, W.-C. Lee, T.-C. Weng, and V. S. Tseng, "Semantic trajectory mining for location prediction," in *Proceedings of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems*, pp. 34–43, 2011.
- [26] R. S. Sutton and A. G. Barto, *Reinforcement learning : An introduction*. MIT press, 2018.
- [27] H. Hasselt, "Double q-learning," *Advances in neural information processing systems*, vol. 23, pp. 2613–2621, 2010.
- [28] D. P. Kingma and J. Ba, "Adam : A method for stochastic optimization," *arXiv preprint arXiv :1412.6980*, 2014.
- [29] A. Dammann, G. Agapiou, J. Bastos, L. Brunelk, M. Garcia, J. Guillet, Y. Ma, J. Ma, J. J. Nielsen, L. Ping, *et al.*, "Where2 location aided communications," in *European Wireless 2013 ; 19th European Wireless Conference*, pp. 1–8, VDE, 2013.
- [30] S. Kuutti, S. Fallah, K. Katsaros, M. Dianati, F. Mccullough, and A. Mouzakitis, "A survey of the state-of-the-art localization techniques and their potentials for autonomous vehicle applications," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 829–846, 2018.
- [31] D. Ebrahimi, S. Sharafeddine, P.-H. Ho, and C. Assi, "Autonomous uav trajectory for localizing ground objects : A reinforcement learning approach," *IEEE Transactions on Mobile Computing*, vol. 20, no. 4, pp. 1312–1324, 2020.
- [32] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal lap altitude for maximum coverage," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 569–572, 2014.
- [33] S. Niknam, H. S. Dhillon, and J. H. Reed, "Federated learning for wireless communications : Motivation, opportunities, and challenges," *IEEE Communications Magazine*, vol. 58, no. 6, pp. 46–51, 2020.
- [34] U. Challita, A. Ferdowsi, M. Chen, and W. Saad, "Machine learning for wireless connectivity and security of cellular-connected uavs," *IEEE Wireless Communications*, vol. 26, no. 1, pp. 28–35, 2019.



- [35] B. Li, Z. Fei, and Y. Zhang, "Uav communications for 5g and beyond : Recent advances and future trends," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2241–2263, 2018.
- [36] A. Ryan, M. Zennaro, A. Howell, R. Sengupta, and J. K. Hedrick, "An overview of emerging results in cooperative uav control," in *2004 43rd IEEE Conference on Decision and Control (CDC)(IEEE Cat. No. 04CH37601)*, vol. 1, pp. 602–607, IEEE, 2004.
- [37] F. Liu, Y. Cui, C. Masouros, J. Xu, T. X. Han, Y. C. Eldar, and S. Buzzi, "Integrated sensing and communications : Towards dual-functional wireless networks for 6g and beyond," *arXiv preprint arXiv :2108.07165*, 2021.
- [38] A. Hassanien, M. G. Amin, Y. D. Zhang, and F. Ahmad, "Dual-function radar-communications : Information embedding using sidelobe control and waveform diversity," *IEEE Transactions on Signal Processing*, vol. 64, no. 8, pp. 2168–2181, 2015.
- [39] F. Liu, L. Zhou, C. Masouros, A. Li, W. Luo, and A. Petropulu, "Toward dual-functional radar-communication systems : Optimal waveform design," *IEEE Transactions on Signal Processing*, vol. 66, no. 16, pp. 4264–4279, 2018.
- [40] F. Liu, L. Zhou, C. Masouros, A. Li, W. Luo, and A. Petropulu, "Dual-functional cellular and radar transmission : Beyond coexistence," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 1–5, IEEE, 2018.
- [41] F. Liu, C. Masouros, A. Li, H. Sun, and L. Hanzo, "Mu-mimo communications with mimo radar : From co-existence to joint transmission," *IEEE Transactions on Wireless Communications*, vol. 17, no. 4, pp. 2755–2770, 2018.
- [42] X. Wang, Z. Fei, J. A. Zhang, J. Huang, and J. Yuan, "Constrained utility maximization in dual-functional radar-communication multi-uav networks," *IEEE Transactions on Communications*, vol. 69, no. 4, pp. 2660–2672, 2020.
- [43] F. Liu, Y.-F. Liu, A. Li, C. Masouros, and Y. C. Eldar, "Cramr-rao bound optimization for joint radar-communication beamforming," *IEEE Transactions on Signal Processing*, 2021.
- [44] C. Sturm, T. Zwick, and W. Wiesbeck, "An ofdm system concept for joint radar and communications operations," in *VTC Spring 2009-IEEE 69th Vehicular Technology Conference*, pp. 1–5, IEEE, 2009.
- [45] J. A. Zhang, X. Huang, Y. J. Guo, J. Yuan, and R. W. Heath, "Multi-beam for joint communication and radar sensing using steerable analog antenna arrays," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 671–685, 2018.

- [46] Y. Luo, J. A. Zhang, W. Ni, J. Pan, and X. Huang, "Constrained multibeam optimization for joint communication and radio sensing," in *2019 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, IEEE, 2019.
- [47] X. Wang, A. Hassanien, and M. G. Amin, "Dual-function mimo radar communications system design via sparse array optimization," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 3, pp. 1213–1226, 2018.
- [48] C. Shi, F. Wang, S. Salous, and J. Zhou, "Joint subcarrier assignment and power allocation strategy for integrated radar and communications system based on power minimization," *IEEE Sensors Journal*, vol. 19, no. 23, pp. 11167–11179, 2019.
- [49] A. Zanella, "Best practice in rss measurements and ranging," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, pp. 2662–2686, 2016.
- [50] A. Wang, X. Ji, D. Wu, X. Bai, N. Ding, J. Pang, S. Chen, X. Chen, and D. Fang, "Guideloc : Uav-assisted multitarget localization system for disaster rescue," *Mobile Information Systems*, vol. 2017, 2017.
- [51] T. Tomic, K. Schmid, P. Lutz, A. Domel, M. Kassecker, E. Mair, I. L. Grix, F. Ruess, M. Suppa, and D. Burschka, "Toward a fully autonomous uav : Research platform for indoor and outdoor urban search and rescue," *IEEE robotics & automation magazine*, vol. 19, no. 3, pp. 46–56, 2012.
- [52] C. Liu, D. Fang, Z. Yang, H. Jiang, X. Chen, W. Wang, T. Xing, and L. Cai, "Rss distribution-based passive localization and its application in sensor networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 4, pp. 2883–2895, 2015.
- [53] S. Tomic, M. Beko, and R. Dinis, "Rss-based localization in wireless sensor networks using convex relaxation : Noncooperative and cooperative schemes," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 5, pp. 2037–2050, 2014.
- [54] T. Stoyanova, F. Kerasiotis, C. Antonopoulos, and G. Papadopoulos, "Rss-based localization for wireless sensor networks in practice," in *2014 9th International Symposium on Communication Systems, Networks & Digital Sign (CSNDSP)*, pp. 134–139, IEEE, 2014.
- [55] D. Koutsonikolas, S. M. Das, and Y. C. Hu, "Path planning of mobile landmarks for localization in wireless sensor networks," *Computer Communications*, vol. 30, no. 13, pp. 2577–2592, 2007.
- [56] J. Rezazadeh, M. Moradi, A. S. Ismail, and E. Dutkiewicz, "Superior path planning mechanism for mobile beacon-assisted localization in wireless sensor networks," *IEEE Sensors Journal*, vol. 14, no. 9, pp. 3052–3064, 2014.

- [57] J. Jiang, G. Han, H. Xu, L. Shu, and M. Guizani, "Lmat : Localization with a mobile anchor node based on trilateration in wireless sensor networks," in *2011 IEEE Global Telecommunications Conference-GLOBECOM 2011*, pp. 1–6, IEEE, 2011.
- [58] R. Sumathi and R. Srinivasan, "Rss-based location estimation in mobility assisted wireless sensor networks," in *Proceedings of the 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems*, vol. 2, pp. 848–852, IEEE, 2011.
- [59] X. Zhang, Z. Duan, L. Tao, and D. K. Sung, "Localization algorithms based on a mobile anchor in wireless sensor networks," in *2014 23rd International Conference on Computer Communication and Networks (ICCCN)*, pp. 1–6, IEEE, 2014.
- [60] Z. Gong, C. Li, F. Jiang, R. Su, R. Venkatesan, C. Meng, S. Han, Y. Zhang, S. Liu, and K. Hao, "Design, analysis, and field testing of an innovative drone-assisted zero-configuration localization framework for wireless sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 11, pp. 10322–10335, 2017.
- [61] P. Perazzo, F. B. Sorbelli, M. Conti, G. Dini, and C. M. Pinotti, "Drone path planning for secure positioning and secure position verification," *IEEE Transactions on Mobile Computing*, vol. 16, no. 9, pp. 2478–2493, 2016.
- [62] C. M. Pinotti, F. Betti Sorbelli, P. Perazzo, and G. Dini, "Localization with guaranteed bound on the position error using a drone," in *Proceedings of the 14th ACM International Symposium on Mobility Management and Wireless Access*, pp. 147–154, 2016.
- [63] F. B. Sorbelli, S. K. Das, C. M. Pinotti, and S. Silvestri, "Precise localization in sparse sensor networks using a drone with directional antennas," in *Proceedings of the 19th International Conference on Distributed Computing and Networking*, pp. 1–10, 2018.
- [64] F. B. Sorbelli, S. K. Das, C. M. Pinotti, and S. Silvestri, "Range based algorithms for precise localization of terrestrial objects using a drone," *Pervasive and Mobile Computing*, vol. 48, pp. 20–42, 2018.
- [65] F. Demiane, S. Sharafeddine, and O. Farhat, "An optimized uav trajectory planning for localization in disaster scenarios," *Computer Networks*, vol. 179, p. 107378, 2020.
- [66] G. Afifi and Y. Gadallah, "Autonomous 3-d uav localization using cellular networks : Deep supervised learning versus reinforcement learning approaches," *IEEE Access*, vol. 9, pp. 155234–155248, 2021.
- [67] M. Atif, R. Ahmad, W. Ahmad, L. Zhao, and J. J. Rodrigues, "Uav-assisted wireless localization for search and rescue," *IEEE Systems Journal*, 2021.

- [68] N. A. Alrajeh, M. Bashir, and B. Shams, "Localization techniques in wireless sensor networks," *International journal of distributed sensor networks*, vol. 9, no. 6, p. 304628, 2013.
- [69] H. Sallouha, M. M. Azari, and S. Pollin, "Energy-constrained uav trajectory design for ground node localization," in *2018 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–7, IEEE, 2018.



## 8 - Synthèse en français

Ces dernières années, des progrès rapides ont été réalisés dans la conception et l'amélioration des véhicules aériens sans pilote (drone) de différentes tailles, formes et leurs capacités de communication. Les drones peuvent se déplacer de manière autonome grâce à des microprocesseurs connectés ou peuvent être exploités à distance sans nécessiter de personnel humain. En raison de leur adaptabilité, de leur installation facile, de leurs faibles coûts de maintenance, de leur polyvalence et de leurs coûts d'exploitation relativement faibles, l'utilisation de drones prend en charge de nouvelles voies pour les applications commerciales, militaires, civiles, agricoles et environnementales telles que la surveillance des frontières, le relais pour les réseaux ad hoc, la gestion des incendies de forêt, la surveillance des catastrophes, l'estimation du vent, la surveillance du trafic, la télédétection et les opérations de recherche et de destruction. Beaucoup de ces applications nécessitent un seul système drone et d'autres comme la surveillance de zone pour les environnements dangereux exigent des systèmes multi- drone. Bien que les systèmes de drones uniques soient utilisés depuis des décennies, en fonctionnant et en développant un grand drone, l'exploitation d'un ensemble de petits drone présente de nombreux avantages. Chaque drone agit comme un nœud isolé dans les systèmes drone uniques, il ne peut communiquer qu'avec le nœud au sol. Par conséquent, le système de communication drone est établi uniquement via une communication drone -infrastructure, et la communication entre les drone peut être basée sur l'infrastructure. La capacité d'un seul système drone est limitée par rapport au système multi drone qui présente de nombreux avantages. D'abord et avant tout, les tâches sont principalement accomplies à moindre coût avec les systèmes multi- drone. De plus, le travail collaboratif des drones peut améliorer les performances du système. De plus, si drone échoue dans une mission dans un système multi- drone, l'opération peut continuer à exister avec les autres drone, et les tâches sont généralement terminées plus rapidement et efficacement avec les systèmes multi- drone.

Dans cette thèse, de nouvelles contributions sur la modélisation, l'évaluation et l'optimisation de la prochaine génération de systèmes de communication de drone ont été rapportées. En particulier, la technologie émergente de Apprentissage Automatique (AA), en tant que catalyseur prometteur pour les communications sans fil au-delà de la 5G, a été inspectée et utilisée. Plus précisément, la contribution de cette thèse peut être résumée comme suit. Dans les premiers chapitres, nous avons fourni une étude approfondie sur l'utilisation des drone dans les réseaux sans fil. Nous avons étudié les principaux cas d'utilisation des drones en tant que stations de base aériennes

et utilisateurs connectés au cellulaire. Pour chacune des applications, nous avons exploré les défis clés et les problèmes fondamentaux. De plus, nous avons couvert en détail les nouvelles directions de recherche lorsque les techniques AA sont utilisées pour augmenter les performances des réseaux drone. Nous avons fourni un aperçu complet des techniques AA, en particulier Apprentissage par renforcement (AR), qui ont été appliquées dans les réseaux drone. Ensuite, nous avons discuté des principes et des avantages Apprentissage Fédéré (AF) et où une approche AF peut être utilisée dans le domaine des réseaux drone.

Dans l'un de nos principaux travaux, nous avons conçu un nouveau système de communication assisté par drone reposant sur la trajectoire de vol la plus courte de l' drone tout en maximisant la quantité de données transmises aux appareils mobiles. Dans le système considéré, nous avons supposé que l' drone n'a pas connaissance de l'emplacement de l'utilisateur à l'exception de sa position initiale. Nous avons proposé un cadre basé sur la probabilité de présence d'utilisateurs mobiles dans une grille par rapport à leur distribution de probabilité. Ensuite, une technique d'apprentissage par renforcement profond est développée pour trouver la trajectoire afin de maximiser le débit dans une zone de couverture spécifique. Des résultats numériques ont été présentés pour mettre en évidence comment notre technique établit un équilibre entre le débit atteint, la trajectoire et la complexité. Contrairement aux travaux précédents, nous avons étudié la localisation des utilisateurs au sol en utilisant des drones comme ancres aériennes. Plus précisément, nous avons introduit un nouveau cadre de localisation basé sur AF et AR. Contrairement à la littérature existante, notre scénario comprend plusieurs drone apprenant la trajectoire dans différents environnements, ce qui se traduit par une convergence plus rapide du modèle AR pour une erreur de localisation minimale. De plus, pour évaluer la trajectoire apprise à partir du modèle agrégé, nous testons l'agent AR formé dans un quatrième environnement qui montre l'amélioration de l'erreur de localisation et de la vitesse de convergence. Les résultats de la simulation montrent que notre cadre proposé surpasse un modèle formé avec l'apprentissage par transfert de %30.

Enfin, nous avons exploré la trajectoire optimale pour maximiser le débit de communication et minimiser les erreurs de localisation dans un réseau drone où un seul drone dessert un groupe d'utilisateurs de communication et localise les cibles au sol simultanément. Pour équilibrer les performances de communication et de localisation, nous avons formulé un problème d'optimisation multi-objectifs pour optimiser conjointement deux objectifs : la maximisation du nombre de bits transmis envoyés aux utilisateurs et la minimisation de l'erreur de localisation pour les cibles au sol sur une période de mission particulière qui est limitée par l'énergie du drone. consommation ou temps de vol. Ces deux objectifs étaient partiellement en conflit l'un avec

l'autre et des paramètres de pondération sont donnés pour décrire l'importance associée. Par conséquent, dans ce contexte, nous avons proposé un nouveau cadre basé sur AR pour permettre au drone de trouver sa trajectoire de manière autonome, ce qui améliore la précision de localisation et maximise le nombre de bits transmis dans les plus brefs délais par rapport à la consommation d'énergie du drone. Nous avons démontré que la méthode proposée améliore significativement la moyenne des bits transmis, ainsi que l'erreur de localisation du réseau.