A Study on Intelligent Inter-User Distance Estimation and Antenna Beamforming for HAPS in mmWave mMIMO Communication Systems

March 2023

Yang Siyuan

A Thesis for the Degree of Ph.D. in Engineering

A Study on Intelligent Inter-User Distance Estimation and Antenna Beamforming for HAPS in mmWave mMIMO Communication Systems

Graduate School of Science and Technology
Keio University

March 2023

Yang Siyuan

Contents

List of Tables

List of Figures

A Study on Intelligent Inter-User Distance Estimation and Antenna Beamforming for HAPS in mmWave mMIMO Communication Systems

Siyuan Yang
yang@ohtsuki.ics.keio.ac.jp.
Keio University, 2023

Supervisor: Prof. Tomoaki Ohtsuki, Ph.D.
ohtsuki@ics.keio.ac.jp

Abstract

Inter-user Distance (IUD) estimation and antenna beamforming are two important techniques in Millimeter Wave (mmWave) massive Multiple-Input Multiple-Output (mMIMO) communication systems. IUD estimation can be used for location-based services and antenna beamforming can be used for improving the throughput of users. Therefore, they are two key techniques in 5G to improve the quality of communication services for a large number of users. The dissertation investigates the use of deep learning (DL)-based methods to tackle two challenging tasks. The first is to improve the estimation accuracy of the IUD. We propose a Deep Neural Network (DNN)-based method for estimating IUD with the assistance of Super Resolution (SR). The second is to improve the quality of communication by antenna beamforming in the HAPS system. We propose two novel Deep Reinforcement Learning (DRL) methods and an Equal Clustering (EC)-aided Deep Q-Network (DQN) approach to reduce the number of low-throughput users. This dissertation consists of 5 chapters.

In Chapter 1, we introduce the background of mmWave mMIMO communication systems. We discuss two challenge topics: IUD estimation and antenna beamforming of HAPS system. In Sec. 1.2 we present several existing works dealing with IUD estimation. In Sec. 1.3, we focus on dynamic antenna control in the HAPS communication system for improving the throughput of users. The relevant works will be presented as well. This chapter also summarizes the scope and contribution of this dissertation.

Chapter 2 tackles a challenging IUD estimation task that is very important to enhance the quality of communication and communication services. Using the conventional geometry-based localization approaches and the conventional fingerprint-based approaches for IUD estimation needs the Line-of-Sight (LOS) scenario or huge overhead to obtain the Channel State Information (CSI) of users. Thus, to overcome these limitations, we propose a DL approach with a novel fingerprint to address this IUD estimation problem. This novel fingerprint is easily obtained by beam sweeping and includes rich angular information of both two users. Nonetheless, we investigate the possibility of using a super-resolution network to reduce the involved beam sweeping overhead. We evaluate the proposed DNN-based IUD estimation method by using original beam images of resolution 4×4, 8×8, and 16×16. Simulation results show that our method can achieve an average distance estimation error equal to 0.13 m for a coverage area of 60×30 m^2. Moreover, our method outperforms the state-of-the-art IUD estimation methods that rely on each user's location information.

Chapters 3 and 4 focus on controlling the antenna beamforming in the HAPS communication system to reduce the number of low-throughput users. Due to wind pressure, HAPS is always subject to random rotation and shift. We need to constantly adjust the parameters of the antenna to prevent a reduction in user throughput. However, conventional intelligent antenna control algorithms (e.g. evolutionary algorithms and reinforcement learning algorithms) require significant time consumption to find the optimal solution in the face of variable environments and high-dimensional search spaces. In reality, we need to respond quickly to changes in the state of the HAPS due to the frequent and unpredictable movements of the HAPS. Therefore, traditional methods are often prone to fall into sub-optimal solutions in a limited amount of time.

To solve this problem, in Chapter 3, we first present the antenna beamforming problem as a Markov Decision Process (MDP) problem. To tackle this MDP problem, we use three conventional RL-based approaches. Besides, we use an Evolutionary Algorithm (EA) named Particle Swarm Optimization (PSO) to find the optimal antenna parameters for reducing the number of low-throughput users. However, due to the huge search

space, these approaches are difficult to obtain the optimal solutions. Motivated by this, we develop two novel DRL-based methods: MFDQN and DRLEA to search for the optimal solution as much as possible in the limited number of training. MFDQN transforms the multiple antenna array beamforming problem into a Mean Field Equilibrium (MFE) problem to find solutions that reduce the number of low-throughput users. It uses multiple DQN agents to control each antenna array in a HAPS separately instead of using only a single DQN agent to control multiple antenna arrays to reduce the action search space for each DQN agent. DRLEA is motivated by an evolutionary algorithm (EA) to overcome the suboptimal solution problem in reinforcement learning when the search space is huge. We do the simulation under four different user distribution scenarios and consider both HAPS rotation and HAPS shift cases. The simulation results show that the proposed MFDQN and DRLEA have better Cumulative Distribution Function (CDF) of users' throughput than that of conventional RL-based approaches and the PSO algorithm under all the user distribution scenarios in the HAPS rotation case. In the HAPS shift case, the proposed methods without retraining achieve a CDF of throughput performance comparable to that of the conventional methods. Besides, we compare the convergence between MFDQN and the conventional RL-based methods. The MFDQN shows that it reduces the complexity of the interactions among agents and converges faster than the conventional RL-based methods.

In Chapter 4, we decompose the problem of improving the throughput of low-throughput users into improving the bandwidth and Signal to Interference plus Noise Ratio (SINR) of users separately to reduce the complexity of each sub-problem. Our idea is to first cluster users into several high-density clusters according to the number of antenna arrays. Next, we design antenna parameters based on the geometric model to control beams toward the center of each cluster and cover each cluster separately. We use a classical K-Means clustering method and design an EC method for clustering users. Moreover, to get a better throughput performance, we use DQN to fine-tune antenna parameters after designing the antenna parameters according to the clustering result of using EC. Under the same simulation settings in Chapter 3, the simulation results show that K-Means, EC, and

EC-aided DQN are effective in improving the CDF performance of throughput over the RL-based approach in HAPS with both rotation and shift scenarios. Besides, we evaluate the clustering-based methods in a user movement scenario, we show that the clustering-based approaches can still maintain a high CDF of throughput performance, while the RL-based approach cannot.

Chapter 5 summarizes the thesis, highlighting its key points and contributions therein, and suggesting possible applications of DL and DRL in more complex localization and HAPS scenarios, to revisit the topic of DL and DRL techniques for localization and antenna control in the future.

Acknowledgments

Words cannot express my gratitude to my supervisor Prof. Tomoaki Ohtsuki for his invaluable patience and feedback. Under the guidance of Prof. Tomoaki Ohtsuki, I learn much about how to do research projects and how to upgrade and improve oneself, as well as how to be a scientific researcher. Without Ohtsuki Sensei's great support and motivation, this Ph.D. degree will not be achievable. I am very appreciative of Ohtsuki Sensei's step-by-step guidance for completing this dissertation. In addition, I would love to thank Ohtsuki Sensei for carefully revising my academic journal and conference papers.

I also could not have undertaken this journey without my defense committee members, Prof. Sanada, Prof. Shigeno, and Prof. Gui, who raise very helpful comments and suggestions for improving the technical and written quality of this dissertation.

Assistance provided by JST SPRING, Grant Number JPMJSP2123, and Keio Leading Edge Laboratory (KLL), for offering grants and scholarship to support my research during my Ph.D. study has been of great help and deserve a special thank you.

I am also grateful to all the Ohtsuki Lab members, especially Mondher Bouazizi, Yuwen Cao, and others who have always been there for me. You gave me tremendous assistance not only in my research but also in the daily life of studying abroad. I would especially like to thank Assistant Professor Mondher Bouazizi for his help in editing and coding.

I would also like to say a heartfelt thank you to my family, especially my mother. Her belief in me has kept my spirits and motivation high during this process. To my dear friends, Huanglin Cao, Hao Zheng, and Zhiqing Wu, meeting you is one of the best things that has happened to me in my life.

Last but not least, to DouDou, a fat cat, I hope you can lose some weight.

Siyuan Yang

Chapter 1

Introduction

This dissertation is concerned with the applications of deep learning (DL) and deep reinforcement learning (DRL) to millimeter-wave (mmWave) Massive Multiple-Input Multiple-Output (mMIMO) communication systems. The paper includes two topics. In the first topic, we propose a DL-based approach to address the problem of challenging inter-user Distance (IUD) estimation, as the location information of user equipment (UE) plays an increasingly important role in fifth-generation (5G) mobile networks.

In the second topic, we discuss the antenna beamforming problem in the High Altitude Platform Station (HAPS) system to reduce the number of low-throughput users caused by HAPS movement. We consider both non-user location information-assisted and user location information-assisted antenna control methods. We first implement four traditional Computational Intelligence (CI) algorithms for addressing antenna control problems in the HAPS system: Q-learning, fuzzy Q-learning, deep Q-learning (DQN), and Particle Swarm Optimization (PSO),. Then, we propose two novel DRL-based methods: Mean Field DQN (MFDQN) and DRL Evolutionary Algorithm (DRLEA). In addition, we consider using the user's location information to solve this problem. We implement the classical K-Means clustering algorithm for antenna beamforming and propose a novel Equal Clustering-aided DQN approach.

1.1 Research Background

With the rapid growth of technologies such as mobile cloud, Internet of Things (IoT), self-driving cars, and Ultra High Definition (UHD) 3D video, high data carrying capacity and low latency are required for serving a huge number of UEs [1]. In order to meet the demand for high quality communication services for large numbers of users, mmWave mMIMO systems are explored for 5G networks which can provide high-capacity wireless transmission of multi-gigabit-per-second (Gbps) data rates and significantly raise user throughput [1, 2]. Because the high frequency (30 GHz to 300 GHz), mmWave can provide wide available spectrum resources to support the user requirements of massive data throughput, high wireless bandwidth, super-fast speeds and ultra-low latency [3–9]. Due

to the high carrier frequency, there is a huge propagation loss in mmWave communications. Thus, mMIMO has become an essential technology to tackle the challenge propagation loss problem, which also gives mmWave the ability to obtain high beamforming gain, highly directional signals, narrow beams, and greater angular resolution [3,8,10,11]. Localization accuracy is greatly improved due to the highly directional signals and higher angular resolution achieved using mmWave mMIMO technology, which helps to collect accurate angular information such as angle of arrival (AoA), angle of departure (AoD) and received signal strength information (RSSI) [11]. Thus, due to these characteristics, localization become one of the key technologies in mmWave mMIMO systems that can greatly improve the quality of location-based services. Location information can be used to optimize the 5G network performance, for both indoor and outdoor environments. In indoor environments, UEs location can help reconfigure the wireless links between UEs and Base Stations (BS) to avoid obstacles like walls or other non-reflective surfaces, movement, and mutual blockage among users [12]. Once we have the user position, we can intelligently direct the mmWave beams to provide a reliable link for high data-rate communication. Outdoor localization is more challenging than indoor localization. In the outdoor scenario, the environmental noise interference is more complex compared to indoor scenes, resulting in more difficult Channel State Information (CSI) estimation. Thus, using location information-aided methods like location-aided or location-aware communications to provide high throughput is a low-power-cost way. Based on the users' location information, the BS can directly generate the transmission beams by antenna beamforming to users and also increase the coverage [13, 14].

Even though mmWave mMIMO systems can reduce the inter-cell interference by narrow beams and increase the communication by antenna array technique, due to the various obstacles, the line-of-sight (LoS) propagation of mmWave is susceptible to link blocking [9, 15, 16]. To address the problem, HAPS is a way to provide the LoS link between the ground BS and the blocked users. It is recognized as one of the hot topics for Beyond 5G (B5G) and 6G mobile communications [17–21]. As early as the 1990s, HAPS began to be paid attention to and studied through numerous research perspectives [22].

Compared with Geostationary Earth Orbit (GEO) satellites that are orbiting at a height of about 36,000 km and Low Earth Orbit (LEO) satellites that are orbiting at a height of about 1,200 km, HAPS operates at the stratosphere at heights between 20 and 50 km. Therefore, the Round Trip Time (RTT) of HAPS is much faster than that of GEO and LEO satellites. Furthermore, since HAPS is relatively close to the ground, the power density is approximately one million times that of a GEO satellite and approximately ten thousand times that of a LEO satellite, allowing HAPS to provide high-quality communication services to existing mobile devices [23]. In addition, compared to other systems such as Starlink which operates at an altitude of 340 km to 550 km [24], HAPS is much less expensive in terms of both the launch and the communication costs. Nonetheless, it does not "pollute" the upper layers of the atmosphere with the space waste it creates.

With the rapid development of DRL techniques, DRL is widely used in various fields including in 5G and B5G [25–27]. This is because ML technologies have strong applications in achieving learning from scenarios that are closer to humans. In 5G networks, there are a large number of nodes (or UEs) that need to make a decision based on local observation. Using the traditional centralized algorithms is infeasible because of the high computational complexity and enormous cost [28]. However, DL and DRL can quickly predict the right decisions based on the changes of local environments after being trained with large amounts of historical data. Thus, DL and DRL can be used to solve lots of challenges in 5G networks such as power allocation [29–31], localization [32–35], antenna array controlling [36, 37], channel estimation [38–41], and beamforming [42–44], etc.. F. B. Mismar *et al.* in [45] used Deep Q-Network (DQN) for online learning on how to maximize the users' signal-to-interference plus noise ratio (SINR) and sum-rate capacity. The authors design a binary encoding for performing multiple relevant actions at once in the DQN structure. A. Rkhami *et al.* in [46] used the DRL method to solve the virtual network embedding problem (VNEP) in 5G and B5G. The authors considered that the conventional DRL usually obtains the sub-optimal solutions of VNEP, which leads to inefficient utilization of the resources and increases the cost of the allocation process. Thus, they proposed a relational graph convolutional neural network (GCNN) combined

with DRL to automatically learn how to improve the quality of VNEP heuristics. In [47], the authors proposed a DL integrated RL which combined DL and RL. The DL is used for preparing the optimized beamforming codebook and the RL is used for selecting the best beam out of the optimized beamforming codebook based on the user movements.

1.2 Outdoor Inter-User Distance Estimation

With the rapid development of information globalization and the adoption of IoT devices, many applications emerge that require centimeter-level precision localization, such as autonomous driving, Advanced Driver Assistance Systems (ASADs), and user tracking, etc.. In the recent three years, with the Covid-19 pandemic, newer and stricter requirements have been in demand for applications built on the idea of identifying one's exposure to other individuals carrying the virus. Covid-19 Contact Confirming Application (COCOA) is a kind of smartphone app that enables users to detect nearby people who were infected with the novel coronavirus. These applications need highly accurate location information to detect the distance between self and other devices [11]. The rise of 5G cellular networks shows that mmWave mMIMO provides exceptionally accurate localization [48, 49]. In such a case, mmWave mMIMO has attracted the attention of the research community and industry alike, for several reasons, including its high data rates, energy efficiency, and low latency [50]. In particular, a major advantage that motivated us to use mmWave mMIMO for this paper is its high spatial resolution and beam-directing capability that can be used to pinpoint users very accurately [51]. While mMIMO systems are capable of high-resolution fingerprint extraction, creating a database including all locations/UEs fingerprints is a heavy burden due to their relatively high storage and matching overhead [52]. To solve these problems, deep neural networks (DNN) can extract features from the high-resolution fingerprints and match it to the users' position without high storage and high matching overhead. In [49, 53–57], the researchers have studied the application of DNN for user localization. All of these works use channel state information (CSI) as the fingerprint or need to extract the fingerprint from CSI. However, how to get the CSI with high accuracy and without high CSI estimation overhead

is another problem that needs to be addressed. Localization techniques are also a good option for determining IUDs. With the use of location techniques, users' locations can be determined up to pinpoint accuracy, at which point the distances between them (i.e., the users) can be calculated. On the other hand, the disadvantage of these systems is that they depend on accurate location determination, which can be computationally expensive or perhaps requires more than one BS to operate properly [48, 58]. The limitations stated above motivate us to find a novel method to determine the IUD with high accuracy in the communication environment with only one BS.

1.3 Dynamic Antenna Control for HAPS

In addition to IUD detection by using the user's location information, another of its applications is to assist in improving the throughput of the HAPS communication system [59]. Due to wind pressure, it is difficult for HAPS to remain stationary. Thus, the degradation of the users' throughput and handovers to UEs' end happened [60, 61] after the HAPS coverage shifting. This kind of quasi-stationary state seriously impacts the performance of the communication system [62]. To tackle this problem, reference [63, 64] researched the problem of maximization of coverage through optimization of the parameters of the HAPS antenna arrays, and proposed an optimized way to minimize both the coverage gaps between cells and the excessive cell overlap. Florin *et al*. [65] analyzed the concentric circular antenna array (CCAA) and proposed a Genetic Algorithm (GA) to minimize the maximum side-lobe level (SLL). In [66], Sun *et al*. further developed the discrete cuckoo search algorithm (IDCSA) used to reduce the maximum SLL under the constraint of a particular half-power bandwidth. In [67], the PSO GA is used for reducing the SLL to improve the carrier-to-interference ratio (CIR). However, in high-dimensional space, PSO is easy to enter a local optimum, like other GAs, and the iterative process' convergence rate is low. In [68], Wada *et al*. proposed a reinforcement learning (RL) named Fuzzy Q-learning method to dynamically control the antenna parameters based on the returned received signal power at users. However, all the works consider the ideal uniform user distribution and focus on the coverage or improving the received signal power at users. In

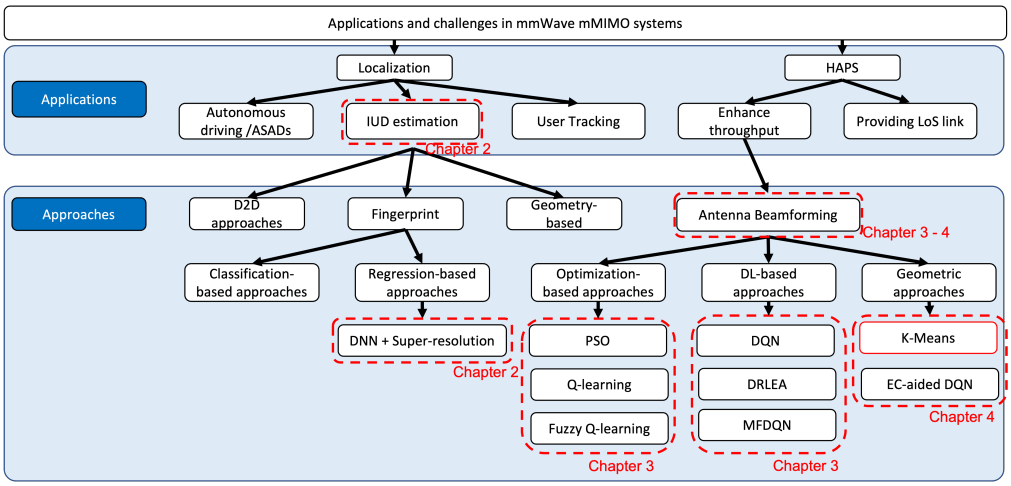


Figure 1.1. The scope of this dissertation.

reference [59], the authors first considered using the users' location information to assist in improving the throughput performance. Compared with using an EA named PSO without users' location information aided, the proposed K-Means clustering-based method effectively improves throughput performance. However, although this method can improve the throughput performance, roughly clustering users and then directing beams to each cluster does not achieve optimal throughput performance.

These limitations motivate us to develop a new method with high convergence and better throughput performance. As mentioned in Sec. 1.1, DL and DRL are two potential techniques to find the optimal solution in a complex environment. Thus, DL and DRL-based approaches have been extensively investigated and will be introduced in the following section.

1.4 Scope and Contributions of the Dissertation

1.4.1 Scope of the Dissertation

The mmWave mMIMO communication system covers a wide range of subtopics. This dissertation will focus on applications of DL and DRL in mmWave mMIMO systems. Fig. 1.1 shows our proposed approaches to address two key and potential techniques, localization and HAPS, for mmWave mMIMO systems in this dissertation.

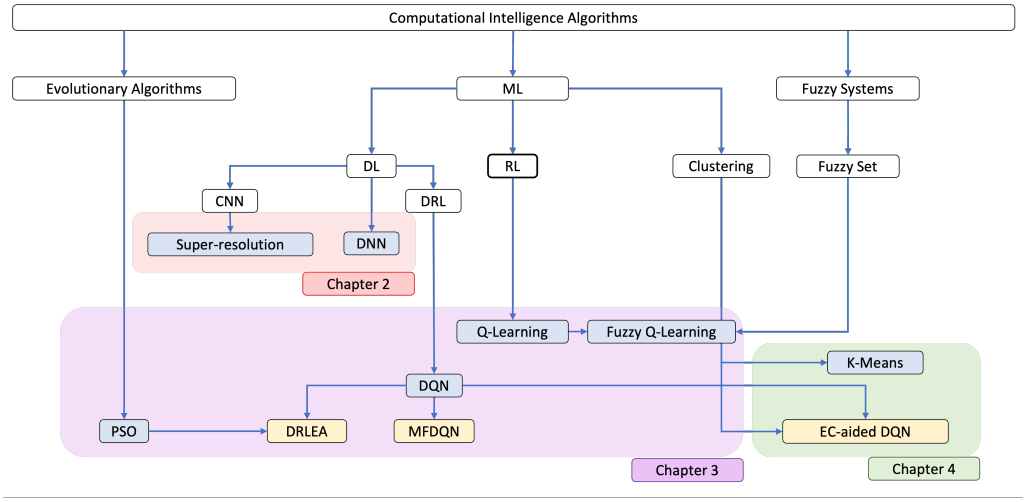


Figure 1.2. The intelligent algorithms used in this dissertation.

Because there are many location-based service scenarios such as autonomous driving/ASADs, IUDs detection for COCOA, and user tracking, the localization technique is one of the key techniques in 5G networks. In Chapter 2, we propose a DL-based approach to solve the IUD detection problem. In general, we can get the distance between different users based on users' location information obtained by the D2D-based approaches [69–72], fingerprint-based approaches [73–80], or geometry-based approach [81–86]. Compared with D2D-based approaches and geometry-based approaches, fingerprint-based approaches can achieve higher localization accuracy under the non-Line-of-Sight (NLoS) scenario. Thus, we study fingerprint-based localization approaches and introduce the proposed DL-based methods in Chapter 2.

In Chapters 3 and 4, we solve an antenna beamforming problem in the HAPS communication system to reduce the number of low-throughput users. In Chapter 3, We implement several conventional computational intelligence (CI) algorithms and propose two DRL-based approaches to design antenna arrays' parameters and generate beams for reducing the number of low-throughput users In Chapter 4, we consider utilizing users' location information to assist in beam steering for improving users' throughput. We use the K-Means clustering algorithm to assist antenna beamforming and design an interesting Equal Cluserting (EC)-aided DQN approach for antenna beamforming.

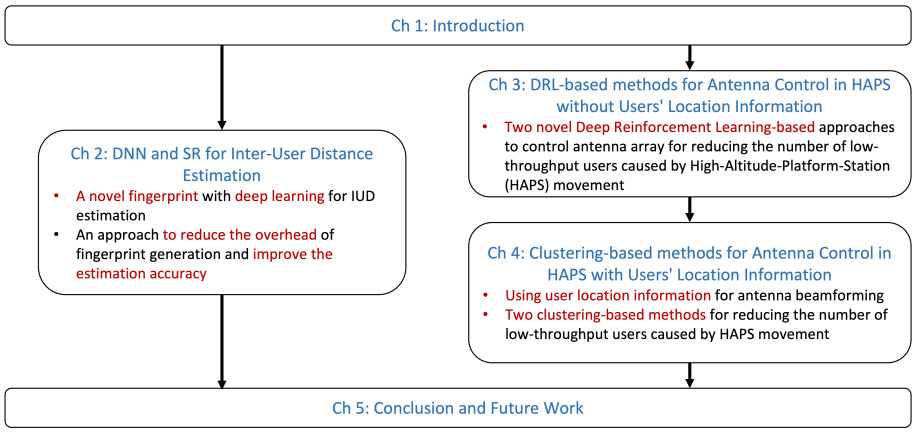


Figure 1.3. An overview of this dissertation.

In Fig. 1.2, we show the relationship of used algorithms in this dissertation for IUD estimation and antenna beamforming in the HAPS system. This dissertation includes three main kinds of CI algorithms: Evolutionary Computation Algorithm, Machine Learning (ML), and Fuzzy System. For ML, it is a wide range of sub-topics. We mainly consider three sub-algorithms of ML: DL, RL, and Clustering algorithm. In Chapter 2, we use two DL methods, Super-Resolution (SR) and Deep Neural Network (DNN) to solve the IUD detection problem. In Chapter 3, we implement three RL-related algorithms and an evolutionary computation algorithm named PSO for antenna beamforming. Based on this, we combine evolutionary computation algorithms and DRL to propose DRLEA. In addition, we combine Mean Filed Game (MFG) theory and DRL to propose MFDQN. In Chapter 4, to utilize users' location information, we use a classical K-Means clustering algorithm to assist antenna arrays to generate beams toward and cover each cluster. Moreover, we design a novel EC method that makes each cluster has the roughly same number of users. After that, we design antenna parameters based on the geometry model and use the DQN algorithm to fine-tune the antenna parameters to reduce the number of low-throughput users.

1.4.2 Summary

The objective of this research is to develop advanced DL and DRL approaches for improving the accuracy of IUD estimation and reducing the number of low-throughput users

caused by the HAPS movement. To this end, this dissertation first investigates the DL-based approach for solving the IUD estimation problem. Next, this dissertation implements several computational intelligent algorithms and proposes two novel DRL-based approaches for reducing the number of low-throughput users in the HAPS communication system. In addition, considering location information can effectively assist in providing high throughput, we propose a K-Means clustering-based method and a novel EC-aided DQN to reduce the number of low-throughput users.

Fig. 1.3 illustrates an overview of this dissertation. This dissertation consists of 5 chapters. Chapters 2, 3, and 4 contain, each, a particular problem statement, relevant related work existing in the literature, ad description of the proposed methods to handle it and an evaluation of its efficiency. In Chapter 1, we introduce the background of mmWave mMIMO communication systems. We discuss two challenge topics: localization and HAPS. In Sec. 1.2 we present several existing works dealing with IUD estimation, an application of localization. In Sec. 1.3, we focus on dynamic antenna control in the HAPS communication system for improving the throughput of users. The relevant works will be presented as well. This chapter also summarizes the scope and contribution of this dissertation.

Chapter 2 tackles a challenging localization task that is very important to enhance the quality of communication and communication services. We use DL and propose a novel fingerprint to address this problem. This novel fingerprint is easily obtained by beam sweeping and includes rich angular information of both two users. Nonetheless, we investigate the possibility of using a super-resolution network to reduce the involved beam sweeping overhead. We evaluate the proposed DNN-based IUD estimation method by using original beam images of resolution 4×4, 8×8, and 16×16. Simulation results show that our method can achieve an average distance estimation error equal to 0.13 m for a coverage area of $60 \times 30\ m^2$. Moreover, our method outperforms the state-of-the-art IUD estimation methods that rely on each user's location information.

Chapters 3 and 4 are focuses on the same task: antenna beamforming in the HAPS communication system to reduce the number of low-throughput users. In Chapter 3, we

first present the antenna beamforming problem as a Markov Decision Process (MDP) problem. To tackle this MDP problem, we use three conventional RL-based approaches. Besides, we use an Evolutionary Algorithm named PSO to find the optimal antenna parameters for reducing the number of low-throughput users. However, due to the huge search space and different user distribution scenarios, these approaches are difficult to obtain the optimal solutions. Motivated by this, we develop two novel DRL-based methods: MFDQN and DRLEA to search for the optimal solution as much as possible in the limited number of training. We do the simulation under four different user distribution scenarios and consider both HAPS rotation and HAPS shift cases. The simulation results show that the proposed MFDQN and DRLEA have better CDF of users' throughput than that of conventional RL-based approaches and PSO algorithm under all the user distribution scenarios in HAPS rotation case. In the HAPS shift case, the proposed methods without retraining achieve a CDF of throughput performance comparable to that of the conventional methods. Besides, we compare the convergence between MFDQN and the conventional RL-based methods. The MFDQN shows that it reduces the complexity of the interactions among agents and converges faster than the conventional RL-based methods.

In Chapter 4 we study user location information for assisting in tackling the antenna beamforming problem. Our idea is to first cluster users into several high-density clusters according to the number of antenna arrays. Next, we design antenna parameters based on the geometric model to control beams toward the center of each cluster and cover each cluster separately. We use a classical K-Means clustering method and design an interesting EC method for clustering users. Moreover, to get a better throughput performance, we use DQN to fine-tune antenna parameters after designing the antenna parameters according to the clustering result of using EC. Under the same simulation settings in Chapter 3, the simulation results show that K-Means, EC, and EC-aided DQN are effective in improving the CDF performance of throughput over the RL-based approach in HAPS with both rotation and shift scenarios. Besides, we evaluate the clustering-based methods in a

user movement scenario, we show that the clustering-based approaches can still maintain a high CDF of throughput performance, while the RL-based approach cannot.

Chapter 5 summarizes the thesis, highlighting its key points and contributions therein, and suggesting possible applications of DL and DRL in more complex localization and HAPS scenarios, to revisit the topic of DL and DRL techniques for localization and antenna control in the future.

1.4.3 Contributions

This dissertation investigates the applications of DL and DRL in mmWave mMIMO systems. We use DL to tackle the IUD estimation problem and use DRL to tackle antenna beamforming in the HAPS system. The contributions of this dissertation are summarized in Tab. 1.1, Tab. 1.2, and Tab. 1.3.

TABLE 1.1
LIMITATIONS OF CONVENTIONAL FINGERPRINT APPROACHES AND THE CONTRIBUTIONS OF CHAPTER 2.

Objective	• IUD estimation.
Conventional Fingerprint-based Approaches	• A single-input and SISO FMCW radar architecture that integrates two frequency scanning antennas is proposed [87]. • Positioning using a combination of measured path loss and AoA [88]. • An ADP is obtained from the CSI and using the DCNN to learn the mapping between the ADP and the users' location [53]. • An angle-delay channel amplitude matrix (ADCAM) is proposed as fingerprint. The ADCAM has rich multipath information with clear physical interpretation that can train DCNN easily [52]. • A classification-based localization method is proposed [89].
Conventional Approaches Limitations	• Only in the LoS scenarios [87]. • Needs multiple BS or anchor to obtain the multi-path angular information [88]. • The accuracy deeply depends on the CSI measurement accuracy [52, 53, 89]. • Huge dataset is needed for training [52, 53, 89].
Contributions of Chapter 2	• To improve the accuracy of the IUD estimation, we design a novel fingerprint that includes two users' location information instead of using ADP or CSI as fingerprint which only includes the location information of one user. • We propose a novel beam energy image generated by beam sweeping as the fingerprint. Compared with the conventional fingerprint-based methods, such as using the CSI or the extracted ADP from CSI as the fingerprint, the proposed fingerprint is deeply related to the horizontal and vertical angles corresponding to the user. • Using beam energy image generated by beam sweeping instead of using CSI as a fingerprint can reduce the CSI estimation overhead. • Compared with the conventional geometric methods which need multiple BSs for localization, the proposed method only needs one BS. Besides, after training, the proposed method can also work on mobile users. • In general, generating a high-resolution beam energy image of a user by beam sweeping involves relatively high time expenditure. In this sense, we utilize a super-resolution technique to improve the low-resolution beam energy images to higher resolution ones.

TABLE 1.2
LIMITATIONS OF CONVENTIONAL ALGORITHMS AND THE CONTRIBUTIONS OF CHAPTER 3.

Objective	• Controlling antenna parameters to reduce the number of low-throughput users.
Conventional Computational Algorithms	• Q-learning [90]: A classical RL algorithm which record the state-action value into Q-table for different states and actions. Thus, the optimal can be searched from this Q-table. • Fuzzy Q-learning [68]: The variable states are controlled by the Fuzzy sets, which allows multiple searches to be performed in one setup. Compared with Q-learning, it can reduce the cost of the search. • DQN [91]: Using DNN instead of Q-table to record and predict the optimal action in a certain state. • PSO [67]: A kind of evolutionary algorithm, iteratively trying to improve many randomly generated particles (candidate solutions) to a given measure of quality (optimal solution).
Limitations	• Huge searching overhead [90]. • High convergence but low performance [68]. • Easily fall into sub-optimal solutions [68, 90, 91]. • It is difficult to search the optimal solution on a high-dimensional action space [67].
Contributions of Chapter 3	• We propose two novel DRL-based methods for reducing the number of low-throughput users caused by the HAPS motion. • To reduce the search space, we decompose the single agent in the HAPS into multi-agent in each antenna array. • We model the antenna parameters control problem into the stochastic game to find the equilibrium to reduce the number of low-throughput users. We use the MFDQN to learn the transition probability and predict the Q-value to reduce the complexity of the interactions among agents and improve the convergence rate. • The proposed DRLEA combined the EA and DRL to avoid sub-optimal solutions. We design a new loss function that includes not only Q-value of the predicted optimal action, but also the historical optimal solutions obtained from previous training.

TABLE 1.3
LIMITATIONS OF K-MEANS CLUSTERING ALGORITHM AND THE CONTRIBUTIONS OF CHAPTER 4.

Objective	• Controlling antenna parameters to reduce the number of low-throughput users.
Conventional Clustering method	• K-Means Clustering: iteratively searches the center point of each cluster to make it has high density.
Conventional Approaches Limitations	• The computational complexity deeply relates to the number of data. • It cannot guarantee that the number of data in each cluster is equal.
Contributions of Chapter 4	• We design a clustering method that clusters users into several clusters of the same size and high density with low computational complexity. • Under the user movement scenario, the proposed clustering-based methods show a better Cumulative Distribution Function (CDF) of throughput performance than other learning-based methods.

Chapter 2

DNN and SR for Inter-User Distance Estimation

2.1 Introduction

With the advances and development in cellular network technology as well as artificial intelligence technology, several new apps have arisen, some of which are very reliant on the accuracy of the users' position estimation [92,93]. It still is a challenge to find a robust solution to achieve the needed degree of precision in environments with rough multi-path channel conditions. The most common approach to localizing multipath channels relies on sensing technology that mitigates multipath effects [94] or fuses multiple sources of information [95].

2.1.1 Background

More recently, with the outbreak of the COVID-19 pandemic, there is an increased demand for applications that can identify a person's exposure to others who may be carrying the virus. COVID-19 Contact Confirming Application (COCOA) is a kind of smartphone app that enables users to detect nearby people who were infected with the novel coronavirus. This app uses Bluetooth to connect with other users' smartphones directly. If a user is infected, the COCOA in his smartphone will send the notifications to other users' devices. However, this kind of Device-to-Device (D2D) communication needs the user to permit to allow other devices to connect to his/her device. If someone does not give such permission, he cannot send a COVID-19 alert notification to other users or receive the alert from others. In this sense, D2D communication alone is not enough. As such, an efficient localization method is needed. The base station (BS) can use the localization technique to know all users' locations. When someone is infected, the BS can send notifications to nearby users. A more interesting task would be collocation identification. Collocation (also referred to as co-location) refers to the task of identifying users or groups of users within a certain range from one another and estimating the distance between them. This technology, up to very recently, relies on one of two main ideas that location detection techniques similar to the ones mentioned previously or exploiting the limited range wireless connections such as Bluetooth [96] and WiFi. However, the latter set of techniques (i.e., Bluetooth and WiFi-based techniques) may have some drawbacks [97],

such as limited coverage, efficiency, and availability. Besides, there are some radar-based solutions for multi-user distance estimation [87, 98]. M. Mercuri et al. [87] proposed a single-input and single-output (SISO) frequency-modulated continuous wave (FMCW) radar architecture. The radar sensor integrates two frequency scanning antennas. Their method demonstrates that it is possible to successfully locate the human volunteers, at different absolute distances and orientations. However, in the outdoor scenario, detecting the user distance under the interference of many signals, noise, and obstacles between the users and the sensors is a challenge. In that sense, the former family of approaches (i.e., the use of cellular network-based localization) has more promise. For instance, the use of uni-directional signals brings promise to improve the methods of cellular network-based localization, which could lead to a better co-location identification. Several studies in the last few decades proposed the use of cellular networks for outdoor localization. Despite the positive results obtained using 4G Long Term Evolution (LTE) networks [99], the nature of 4G signal propagation makes it difficult to develop a highly accurate positioning system. MmWave signals have a high temporal resolution due to their propagation characteristics, making precise positioning possible. Therefore, recent works have proposed exploiting the mmWave signal of the 5G of mobile connectivity in applications related to positioning and localization [88, 100–115].

The rise of 5G cellular networks shows that mmWave mMIMO provides exceptionally accurate localization [48, 49]. In such a case, mmWave massive MIMO has attracted the attention of the research community and industry alike, for several reasons, including its high data rates, energy efficiency, and low latency [50]. In particular, a major advantage that motivated us to use mmWave massive MIMO for this Chapter is its high spatial resolution and beam-directing capability that can be used to pinpoint users very accurately [51]. While massive MIMO-OFDM systems are capable of high-resolution fingerprint extraction, creating a database including all locations/UEs fingerprints is a heavy burden due to their relatively high storage and matching overhead [52]. To solve these problems, deep neural networks (DNN) can extract features from the high-resolution fingerprints and match it to the users' position without high storage and high matching overhead.

In [49, 53–57], the researchers have studied the application of DNN for user localization. All of these works use channel state information (CSI) as the fingerprint or need to extract the fingerprint from CSI. However, how to get the CSI with high accuracy and without high CSI estimation overhead is another problem that needs to be addressed. Localization techniques are also a good option for determining inter-user Distances (IUDs). With the use of location techniques, users' locations can be determined up to pinpoint accuracy, at which point the distances between them (i.e., the users) can be calculated. On the other hand, the disadvantage of these systems is that they depend on accurate location determination, which can be computationally expensive or perhaps requires more than one BS to operate properly [48, 58]. The limitations stated above motivate us to find a novel method to determine the IUD with high accuracy in the communication environment with only one BS.

2.1.2 Related Work

In 5G mmWave networks, there are some technologies of user localization. The first kind of technique predicts the users' location by using the estimated Time of Arrival (ToA) [58, 116, 117], Angle of Arrival (AoA) [48, 88, 118–121]. For instance, the authors in [88] achieved the localization error in the range from 0.16 m to 3.25 m by using data fusion and machine learning to estimate the ToA and AoA at users. In [58], Abu-Shaban et al. investigated 3D and 2D Line-of-Sight (LoS) localization situations and provided closed-form formulations for Position Error Bound (PEB) and Orientation Error Bound (OEB). The second kind of technique designs fingerprints using for localization [52, 53, 56, 57, 89, 122–127]. In [56], Gante et al. developed a neural network architecture using the beamformed fingerprint for users localization. A short sequence of Convolutional Neural Networks (CNN) that achieves an average estimation inaccuracy of 1.78 m is proposed by them using for actual outdoor scenarios with primarily Non-Line-of-Sight (NLoS) locations. Savic et al. in [57] used received signal strengths as fingerprints for accurate users location estimation. Besides, they also gave a Gaussian process regression-based solution for fingerprint-based localization. Deep learning based technique was also

studied for solving localization problem [49, 53–55]. In [53], to improve the localization accuracy, they propose a Deep Convolutional Neural Network (DCNN) trained by Angle Delay Profiles (ADPs) as fingerprints.

When it comes to 5G mmWave cellular networks, IUD estimation has received little attention in the literature. This is most likely since it can be calculated very easily from a precise localization (if achieved). IUD estimation means detecting the distance between the users to obtain users' locations whose in a certain range from one another. This IUD estimation technique can be used in wide kinds of scenarios. In the case of the COVID-19 pandemic, for example, it is feasible to determine who has been exposed to viral carriers over lengthy periods of time.

Bluetooth and WiFi are the most common technologies used for identifying users in proximity of one another [128–135]. However, such approaches require all the users to connect to the same WiFi hotspot or allow mutual Bluetooth connection to their devices so that location information is shared, thus identifying the exposure to the virus is possible. Besides, in the outdoor scenario, noise and other interference cause the position estimation accuracy cannot be very high. Thus, the goal of this study is to employ 5G mmWave networks instead, to estimate IUD more accurately by utilizing a unique fingerprint type. The BS can collect the potential virus carrier's location information and notify other users nearby. We summarize the existing works in Table 2.1. In the next section, we will go through this in more depth.

2.1.3 Motivations and Contributions

Our goal is to utilize a single BS to estimate the distance between each pair of distinct users, allowing us to identify users who were close to one another. By beam sweeping, we obtain the difference between the beam energy images of two users as fingerprints instead of using the CSI or ADPs in this study. Then, to estimate the distance between each pair of users, we offer a unique IUD estimation technique based on deep learning.

To verify the robustness of our proposed method, we evaluate it after training in a new environment. We run the simulations on a new environment characterized by a different

Table 2.1. The summary of the existing works

Authors	**Scenario**	**Method and Advantages**	**Limitations**
Mercuri et al. [87]	Indoors	A single-input and SISO FMCW radar architecture that integrates two frequency scanning antennas is proposed. The use of dual-frequency scanning FMCW radar allows for high localization accuracy and differentiation between people and objects.	This method is valid only in the LoS scenario.
Kanhere et al. [88]	Indoors	Positioning using a combination of measured path loss and AoA. The method achieves high positioning accuracy.	The accuracy of positioning depends on the measurement accuracy.
Vieira et al. [53]	Outdoors	An ADP is obtained from the CSI and using the DCNN to learn the mapping between the ADP and the users' location. The authors were the first to use CNN for user localization and the authors using moderately deep CNNs can achieve fractional-wavelength positioning accuracies.	They extracted the ADP from the measured CSI. Thus, the accuracy of this method depends on the measurement accuracy.
Sun et al. [52]	Outdoors	An angle-delay channel amplitude matrix (ADCAM) is proposed as fingerprint. The ADCAM has rich multipath information with clear physical interpretation that can train DCNN easily.	
Sun et al. [89]	Outdoors	A classification-based localization method is proposed.	This method needs to know the CSI. Besides, this method needs a large storage overhead to save training data.

number and different locations of buildings. In the new environment, the trained model is expected to drop in the performance of IUD estimation. We analyze how much ground truth data are required to be collected for the trained model to be fine-tuned effectively in the new environment. We also study how many epochs the model is fine-tuned to achieve a distance estimation performance close to that of the model in the original environment. We identify the correlation between the fingerprint and the distance between distinct users in our technique, which reduces the estimation error significantly. Furthermore, because our technique does not rely on the individual fingerprints of individual users, but rather on how distinct they are from each other, it is more resilient to changes in the environment (for example, a change in BS). We also create a super-resolution neural network that can produce high-resolution beam energy pictures from low-resolution ones.

The network is trained by feeding the network with actual down-sampled low-resolution images along with their real high-resolution counterparts. Through training, the network learns how to recreate the high-resolution images from low-resolution ones. The main contributions of this Chapter are summarized as follows:

- To improve the accuracy of the IUD estimation, we design a novel fingerprint that includes two users' location information instead of using ADP or CSI as fingerprint which only includes the location information of one user.

- We propose a novel beam energy image generated by beam sweeping as the fingerprint. Compared with the conventional fingerprint-based methods, such as using the CSI or the extracted ADP from CSI as the fingerprint, the proposed fingerprint is deeply related to the horizontal and vertical angles corresponding to the user.

- Using beam energy image generated by beam sweeping instead of using CSI as a fingerprint can reduce the CSI estimation overhead.

- Compared with the conventional geometric methods which need multiple BSs for localization, the proposed method only needs one BS. Besides, after training, the proposed method can also work on mobile users.

- In general, generating a high-resolution beam energy image of a user by beam sweeping involves relatively high time expenditure. In this sense, we utilize a super-resolution technique to improve the low-resolution beam energy images to higher resolution ones.

The rest of the Chapter is organized as follows. Section 2.2 explains the system model and the channel model. We present in Section 2.3 the proposed method. The simulation results are shown and discussed in section 2.4. Finally, the Conclusion is shown in Section 2.5.

2.2 System Model

2.2.1 Channel Model

We consider a mmWave communication system with only one RF chain as shown in Fig. 2.1 where one BS serves K users. In this system, the BS is equipped with a Uniform Planer Array (UPA) and the users are equipped, each, with a single antenna. The received signal of the k-th user $\mathbf{r}_k$ is written as follows [132]:

$$\mathbf{r}_k = \sqrt{P}\,\mathbf{h}_k\,\mathbf{w}_{RF}^k\,\mathbf{s}_k + \mathbf{n}_k, \tag{2.1}$$

where P is the transmit power, $\mathbf{s}_k$ is the transmitted signal of the k-th user, $\mathbf{w}_{RF}^k \in \mathcal{W}$ is the analog beamformer of the k-th user. Let $\mathcal{W}$ be the Discrete Fourier Transform (DFT)-based codebook for the UPA-based transmitter. $\mathcal{W}$ is used to apply an analog beamformer to create beams of the signal. $\mathbf{n}_k \sim \mathcal{CN}\left(0, \sigma_k^2\right)$ is an additive white Gaussian noise (AWGN) with zero mean and variance σ_k^2 of the k-th user. Furthermore, the mmWave mMIMO channel $\mathbf{h}_k \in \mathbb{C}^{N_t}$ between the BS and the k-th user can be modeled by [132]:

$$\mathbf{h}_k = \sqrt{\frac{N_t}{L}} \sum_{\ell=1}^{L} \alpha_\ell \mathbf{a}(\varphi, \phi), \tag{2.2}$$

in which L denotes the total number of paths, φ and ϕ denote the horizontal and vertical angles, respectively. α_ℓ denotes the complex path gain of the ℓ-th path. $N_t = N_v \times N_h$

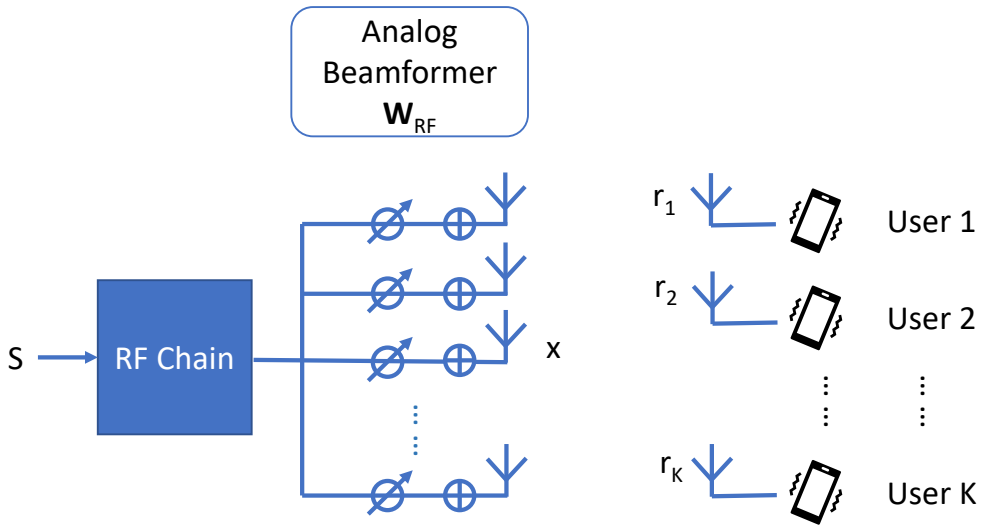


Figure 2.1. Illustration of a multi-user mmWave mMIMO system model.

corresponds to the number of transmit antennas at BS, here N_v and N_h denote the number of antennas along the vertical and horizontal, respectively. $\mathbf{a}(\varphi,\phi) = \mathbf{a}^v(\phi) \otimes \mathbf{a}^h(\varphi,\phi)$ denotes the steering vector, where $\mathbf{a}^v(\phi)$ and $\mathbf{a}^h(\varphi,\phi)$ denote the steering vector over the horizontal axis and the vertical axis, respectively. Herein, α_ℓ can be expressed as follows [136]:

$$\alpha_\ell = \frac{\lambda \cdot g}{4\pi d_p^\ell} \sqrt{N_t} \cdot e^{\frac{-2j\pi d_p^\ell}{\lambda}}, \tag{2.3}$$

where g is the complex reflection gain, d_p is the path distance, and λ is the wavelength. The vertical and horizontal axes' steering vectors can be expressed as follows [132]:

$$\mathbf{a}^v(\phi) = [1, e^{-j2\pi d \cos(\phi)/\lambda}, \cdots, e^{-j2\pi(N_v-1)d\cos(\phi)/\lambda}]^T, \tag{2.4}$$

$$\begin{aligned}\mathbf{a}^h(\varphi,\phi) = &[1, e^{-j2\pi d \sin(\phi)\sin(\varphi)/\lambda}, \cdots, \\ &e^{-j2\pi(N_h-1)d\sin(\phi)\sin(\varphi)/\lambda}]^T,\end{aligned} \tag{2.5}$$

where d is the distance between the consecutive antennas in both vertical and horizontal directions.

2.2.2 Beam Sweeping

Here we introduce a beam sweeping method based on the predefined beams in the codebook. In Fig. 2.2, we divide the coverage into $N_b = n_b^v \times n_b^h$ sub-areas, where n_b^v and n_b^h denote the number of beams at the vertical axis and horizontal axis, respectively, and perform beam sweeping for different areas. When sweeping, unlike producing beams devoted to certain users, the beams are uniformly broadcasted to predetermined places rather than being directed to specific users. The set of the analog beamformers at the n-th region is represented by the matrix $\mathbf{W}_n$ given as follows [132]:

$$\mathbf{W}_n = [\mathbf{w}_{1,n}^v \otimes \mathbf{w}_{1,n}^h, \cdots, \mathbf{w}_{K,n}^v \otimes \mathbf{w}_{K,n}^h], n \in \{1, \cdots, N_b\}, \tag{2.6}$$

where $\mathbf{w}_{k,n}^v$ and $\mathbf{w}_{k,n}^h$ are the weights on antenna elements along the vertical and horizontal directions, respectively. Our system assumes that the BS sweeps the generated beams first by broadcasting them in time slots. In other words, the BS broadcasts a signal with multiple beamformers to cover different locations for a certain number of time slots, as illustrated in Fig. 2.2. The user measures the signals during the sweeping phase and sends the results to the BS through the mmWave control channel. We utilize the data to create a fingerprint that enables precise geographical location and co-location estimates in our research.

2.2.3 Problem Description

In the context of the system described above, we assume a number of K users that are within the coverage area of a specific BS. Throughout this work, we aim to achieve an objective that identifies the IUD. Similar to localization, the estimation of the distance between users relies on the beam energy images. However, it is achieved by comparing images of different users. We will demonstrate throughout this work that, despite being similar, our approach achieves much better performance in IUD estimation than user localization. In our system, the beam codebook is defined as the set of beams (also known as "codewords") that are evenly disseminated in the downlink by the BS's UPA. We assume

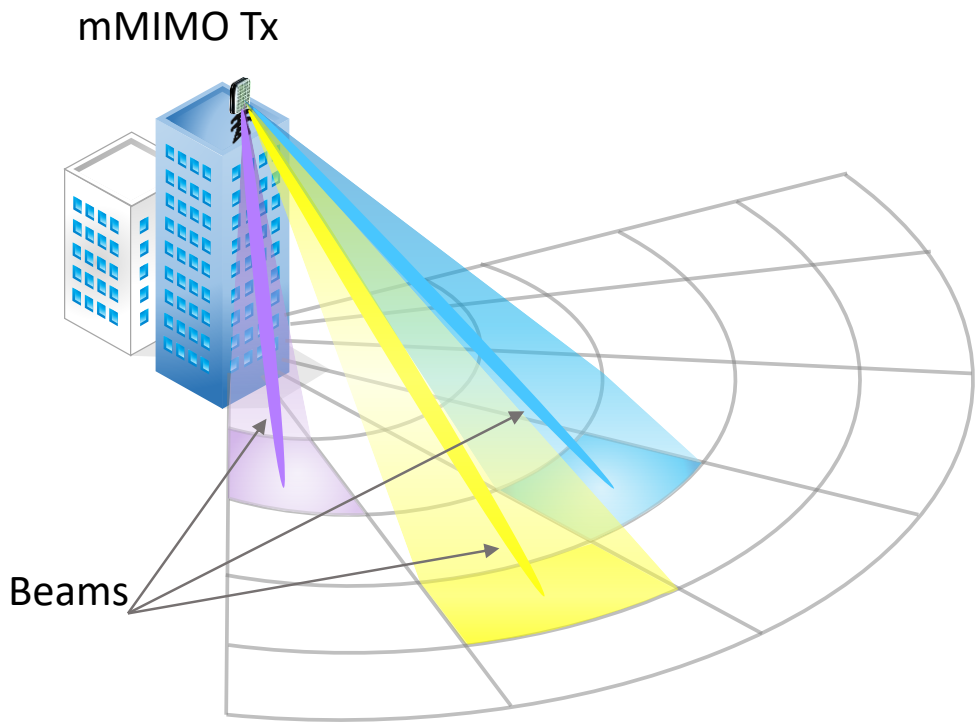


Figure 2.2. An example of beam sweeping to cover different regions in space.

that the UPA sweeps the beams in a common channel at consecutive times lots. Each user records the Received Signal Strength Indicator (RSSI) of the received beams as $m_k^{(x,y)}$ for $x \in \{0, \cdots, N_b\}$ and $y \in \{0, \cdots, N_b\}$ [121, 127, 137, 138]. The measured RSSI values, formatted in a matrix, could be seen as an image, which we refer to as the beam energy image. Each image generated by a user is used as a fingerprint for its location which a neural network will use to estimate the location relative to the BS.

About our tasks, we show in this work the limitations of such a method in general in estimating the locations of the users, and its higher potential in estimating the distance between each pair of them. However, to briefly introduce the intuition behind it, we summarize in the following the main reasons. To begin with, the fingerprint of a given location is very dependent on the channel state and varies over time. Therefore, small changes in the channel and environment (e.g., level of noise, reflections, etc.) lead to inaccuracies in the location estimation. This change does not affect the IUD measurement, as the estimation is based on the difference between images generated at the same time rather than the images themselves. Therefore, instead of relying on location detection (which is a well-studied task) to measure the IUD, we aim at directly tackling this task

(i.e., IUD estimation) using the differences in generated images by different users. The distance estimation problem could be expressed as a non-linear function of fingerprint images. Similarly, the distance between users might be represented as a non-linear and non-transitive function of the difference in the beam energy images created by various users. In other words, we formulate the problem as a mapping between the difference in beam energy images and the distance between any two users.

2.3 The IUD Estimation Approach

2.3.1 The Conventional User Localization Based IUD Estimation

Conventionally, to estimate the distance between two users, we need to first predict the location of the two users. Following are some works using the fingerprint-based method for user localization [52, 53]. Vieira et al. [53] proposed a fingerprint-based user localization method. They used the measured channel snapshots of each user as fingerprints and trained the DCNN to predict users' location in the massive MIMO system. Based on the DCNN-based method [53], Sun et al. [52] proposed a new type of fingerprint which they used for user localization in the massive MIMO system. Different from [53] which used the channel snapshots represented in the sparse domain as fingerprint, they proposed a fingerprint extraction method to extract the ADP from CSI as the fingerprint. Besides, they also propose a fingerprint compression method and clustering algorithm to reduce storage overhead and matching complexity. For the user localization problem, these methods can achieve high accuracy in estimating users' location. As shown in Fig. 2.3, we suppose that the estimation error of the user m and the user n are e_m and e_n, respectively. Thus, the IUD estimation error $0 \leq e \leq e_m + e_n$. However, if we can train a DNN for IUD estimation with the estimation error e_m or e_n, the estimation error can go up to $\max(e_m, e_n) \leq (e_m + e_n)$. Obviously, using user localization for IUD estimation is not the optimal solution.

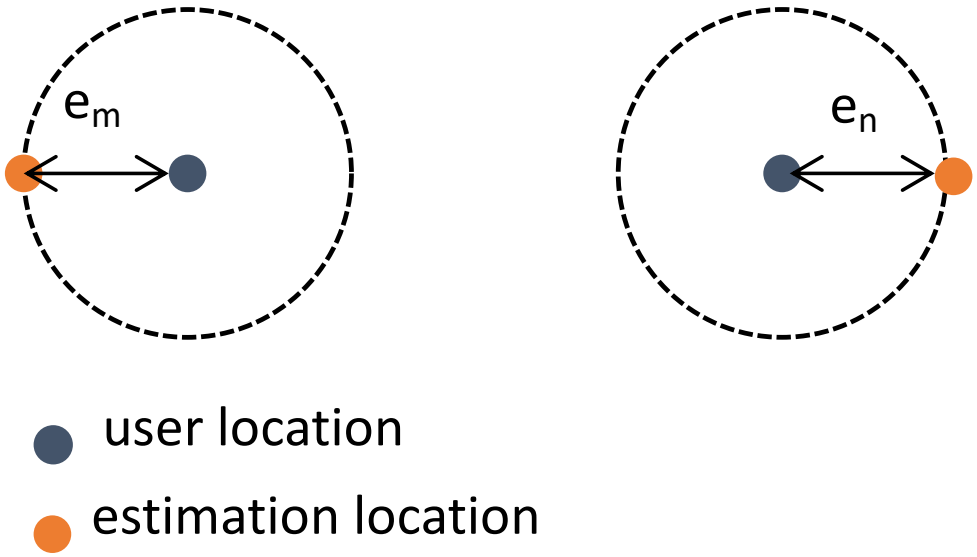


Figure 2.3. IUD estimation error of the conventional localization-based method.

2.3.2 The Proposed DNN-based IUD Estimations

For the above reasons, to improve the estimation accuracy of the distance between each pair of users, we develop a CNN to estimate the distance directly. In Fig. 2.4, we show a flowchart of the overall proposed method. As shown in the flowchart, given two users k and l, these users report the RSSI for the received beams. The BS then uses the reported RSSI values to generate two images, one for each user, and then measures the difference between them as we will describe below. Using models trained offline, the image is either enhanced or used as it is to do a non-linear regression allowing for measuring the distance between k and l. We use the beam energy difference image of each pair of users as input instead of the beam energy image of each user. The output of the proposed CNN is the estimated distance between each pair of users. Given two users k and l, we denote by $\mathbf{M}_k$ and $\mathbf{M}_l$ their respective generated power matrices/images.

$$\mathbf{M}_k = [m_k^{(x,y)}], \tag{2.7}$$

where $x, \in \{1, \cdots n_b^h\}$ and $y, \in \{1, \cdots n_b^v\}$. As stated in the previous section, we refer to the number of beams as N_b and we denote $n_b = \sqrt{N_b}$. We define the difference matrix

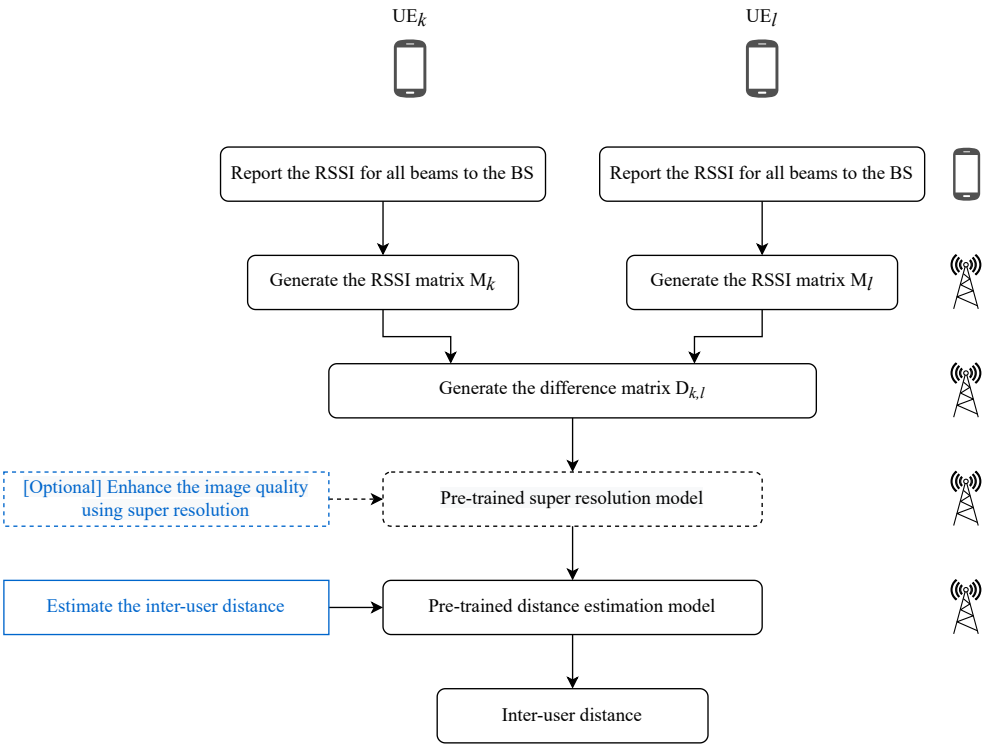


Figure 2.4. A flowchart of the proposed method for IUD estimation.

between the two users k and l as:

$$\mathbf{D}_{k,l} = \begin{bmatrix} |m_k^{(1,1)} - m_l^{(1,1)}| & \cdots & |m_k^{(1,n_b)} - m_l^{(1,n_b)}| \\ |m_k^{(2,1)} - m_l^{(2,1)}| & \cdots & |m_k^{(2,n_b)} - m_l^{(2,n_b)}| \\ \vdots & \ddots & \vdots \\ |m_k^{(n_b,1)} - m_l^{(n_b,1)}| & \cdots & |m_k^{(n_b,n_b)} - m_l^{(n_b,n_b)}| \end{bmatrix}. \tag{2.8}$$

Hereafter, we will employ a much simpler notation for the above matrix:

$$\mathbf{D}_{k,l} = |\mathbf{M}_k - \mathbf{M}_l|. \tag{2.9}$$

In Fig. 2.5, we offer an example of difference matrix visualization. The resultant matrix (the rightmost one) is sent into the neural network, which calculates the distance between the two users whose RSSI matrices are given in the leftmost part of the figure.

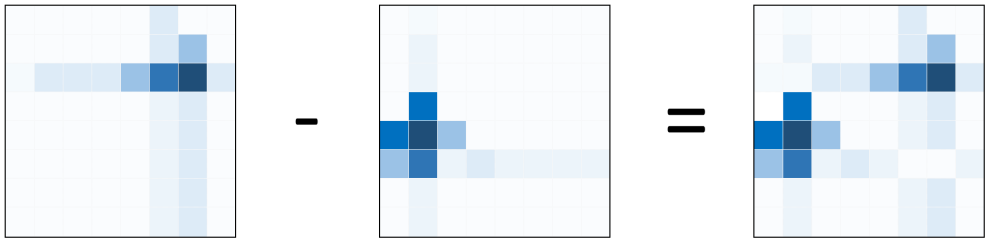


Figure 2.5. An example of visualization of the difference matrix.

Fig. 2.6 shows the neural network we used for distance estimation: As previously indicated, the input is the matrix $\mathbf{D}_{k,l}$ representing the difference between the RSSI of the k-th user and l-th user for the different beams of a given granularity level. The neural network is made up of four convolution layers, each with a filter size of three times three, and a max-pooling layer with a size of 2×2. 128 filters, 256 filters, 256 filters, and 128 filters are used in the convolution layers, correspondingly. The max-pooling layer is followed by four completely linked layers of 128, 512, 256, and 64 neurons, respectively. Rectified Linear Unit is the activation mode for all of the aforementioned layers (ReLU). The network's last layer is a dense layer with a single neuron and linear activation. This is since this neuron's job is to assess the distance between users. The MSE between the actual distance between users (ground truth) and the anticipated distance (prediction) is used to train the network as the loss function that the network is designed to reduce. Here, given a batch b, with a batch size S_b, the loss function of the neural network is defined as:

$$\mathrm{MSE}(X^{(b)}, y^{(b)}, F) = \frac{1}{S_b} \sum_{i=1}^{S_b} \|\hat{y}_i^{(b)}(F) - y_i^{(b)}\|^2, \tag{2.10}$$

where $X^{(b)}$ is the set of the difference matrix $\mathbf{D}_{k,l}$, $y^{(b)}$ is the ground truth distance between each pair of users, and $\hat{y}_i^{(b)}(F)$ is the estimated distance between the users by the proposed DNN network, where F denotes the weights of each layer of the proposed network. The above-mentioned network is designed to contain as few layers and parameters as feasible while yet working adequately. The performance of shallower networks suffers noticeably (MSE of the estimated distance), while the performance of deeper networks does not significantly increase over the suggested architecture. Because the network's technological

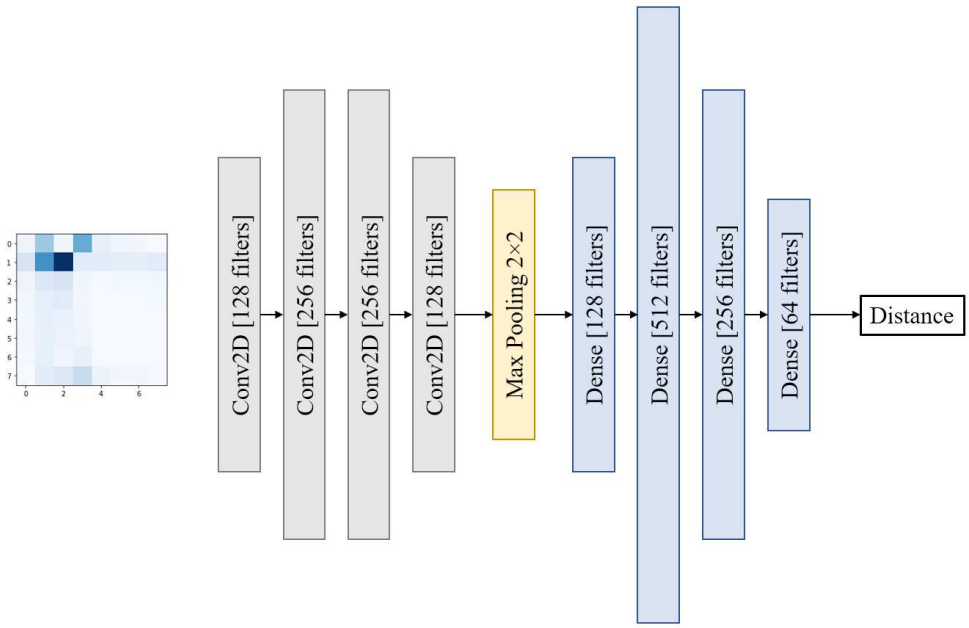


Figure 2.6. The architecture of the neural network used for distance estimation.

implementation (using Keras) necessitates defining the input shape (4×4 or 8×8), multiple networks were created. However, for clarity, we will refer to all of these network instances as if they were one.

2.3.3 Super-Resolution for High-Resolution Fingerprint Generation

As previously stated, the received power from uniformly distributed broad beams may be seen as pictures of various resolutions: 4×4 wide-beam received powers can be regarded as low-resolution 4×4 images, and 8×8 narrow-beam received powers can be regarded as higher resolution 8×8 images. The goal of super-resolution is to recover (or produce) high-resolution pictures from low-resolution ones in general. When applied to our approach, the ability to produce correct 8×8 beam images from 4×4 beam images enables the use of narrow beams to yield accurate fingerprints with co-localization accuracy comparable to that of broad beams. In our research, we used deep learning to implement a supervised approach to super-resolution. It's worth noting that we experimented with many neural network designs throughout our early trials. However, because the outcomes of these networks were so similar, we will focus on the network that produced the greatest results, which is shown in Fig. 2.7. The neural network consists of one convolution

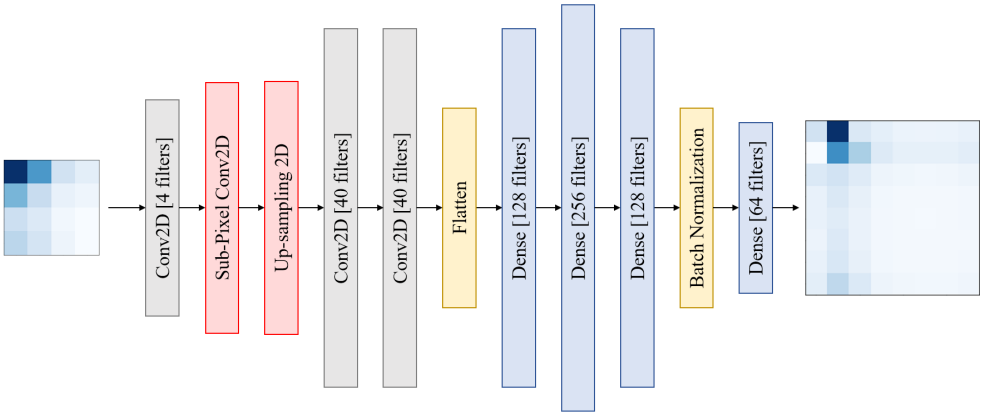


Figure 2.7. Architecture of the super-resolution neural network used.

layer with four filters, a Sub-pixel convolution layer, and an up-sampling layer, which is followed by two convolution layers with 40 filters each. The output of the second layer is flattened and linked to three dense layers of 128, 256, and 128 neurons in succession. All of the layers above have Rectified Linear Unit as their activation (ReLU). After that, batch normalization is applied, followed by a dense layer with a linear activation and a number of filters equal to the predicted output's number of pixels. The picture is supposed to be upscaled to 8×8 for 4×4 input photos. As a result, the number of neurons in the last dense layer is set to 64.

Because we're employing a supervised technique, we'll need a training set to teach the network how to build high-resolution pictures from low-resolution ones. To train the network, we employ the Mean Squared Error (MSE) as a loss function. Given a batch b with a size S_b, the MSE is defined as follows:

$$\mathrm{MSE}(X^{(b)}, y^{(b)}, \theta) = \frac{1}{S_b} \sum_{i=1}^{S_b} \|\hat{y}_i^{(b)}(\theta) - y_i^{(b)}\|^2, \tag{2.11}$$

$y_i^{(b)}$ and $\hat{y}_i^{(b)}(\theta)$ represent the ground truth high-resolution beam-quality image and the image rebuilt using the network function θ, respectively.

2.4 Experimental Results

2.4.1 Experiment Specifications

We run our experiments with collected data using Wireless Insite [139]. The channel model has been introduced in Sec. 2.2. We generate 2 groups of datasets with different environments. The part of the first group is used for training, and the remainder of the first group and the second group are used for testing. The simulation parameters' specifications are shown in Table 2.2. Keras and TensorFlow are used in all of our neural network implementations (for non-standard layers). To find suitable hyperparameters for the proposed DNN and super-resolution networks, we first try a learning rate of 0.1, the number of epochs of 10,000, and a batch size of 64. Upon obtaining a rough idea about the estimated accuracy and training time, we decided to use a smaller learning rate (i.e., a learning rate equal to 0.001), a smaller number of epochs (i.e., 1000 epochs), and a smaller batch size (i.e., a batch size equal to 32).

Table 2.2. The Simulations Parameter Settings.

Parameter description	**Value**
Carrier frequency	60 GHz
# Antennas at the BS ($N_v \times N_h$)	8×8
# Number of beams N_b	$16 \times 16/8 \times 8/4 \times 4$
user spread area	60×30 m^2
Height of BS	10 m
Total downlink power P	30 dBm
Signal to interference power ratio	10 dB
Number of paths L	25
Reflection gain g	-6 dB
Noise figure F	9.5 dB

2.4.2 Evaluation Metrics

Throughout the rest of this section, we will evaluate our proposed approach using different metrics. Therefore, we define and explain here each of these metrics.

Super Resolution To evaluate the super-resolution neural network, we use the same loss function which we defined in Equation (2.11) which refers to an average of the MSE between the predicted pixel values and their real values. We show the average loss per epoch, which reflects how far the generated image is from the ground truth at each epoch.

IUD Estimation We evaluate our approach for IUD estimation by estimating the error between the actual distance and the one reported by the neural network. In other words, given two users k and l, we refer to the real distance between them as $d(k, l)$, and the distance estimated by the neural network as $\hat{d}(k, l)$. The estimation error of the distance between the 2 users $\mathrm{e}_{\mathrm{dist}}(k, l)$ is:

$$\mathrm{e}_{\mathrm{dist}}(k, l) = \sqrt{(d(k, l) - \hat{d}(k, l))^2}. \tag{2.12}$$

Again, for visualization, we plot the CDF of $\mathrm{e}_{\mathrm{dist}}$.

2.4.3 The Training Design of the Proposed Super-Resolution Approach

The loss in terms of MSE during the training phase of the super-resolution neural network is depicted in Fig. 2.8. The neural network was still converging after 10 K epochs of training, and no over-fitting was observed: both the training and validation loss reduced constantly, indicating that training the network for more epochs can lead to higher performance. However, we stopped training at 10 K epochs and determined that the network had converged sufficiently for its output useful for our work. On the training set, the loss was 3.7×10^{-4}, and on the validation set, it was 3.8×10^{-4}. We will compare the results of distance estimation with and without super-resolution in the following sub-section to emphasize and genuinely assess the effectiveness of such procedures. However, as previously noted, various super-resolution neural network algorithms have been explored, and

the results are not far behind. This means that, regardless of the technology utilized to accomplish super-resolution, one can expect a performance improvement when compared to low-resolution photos.

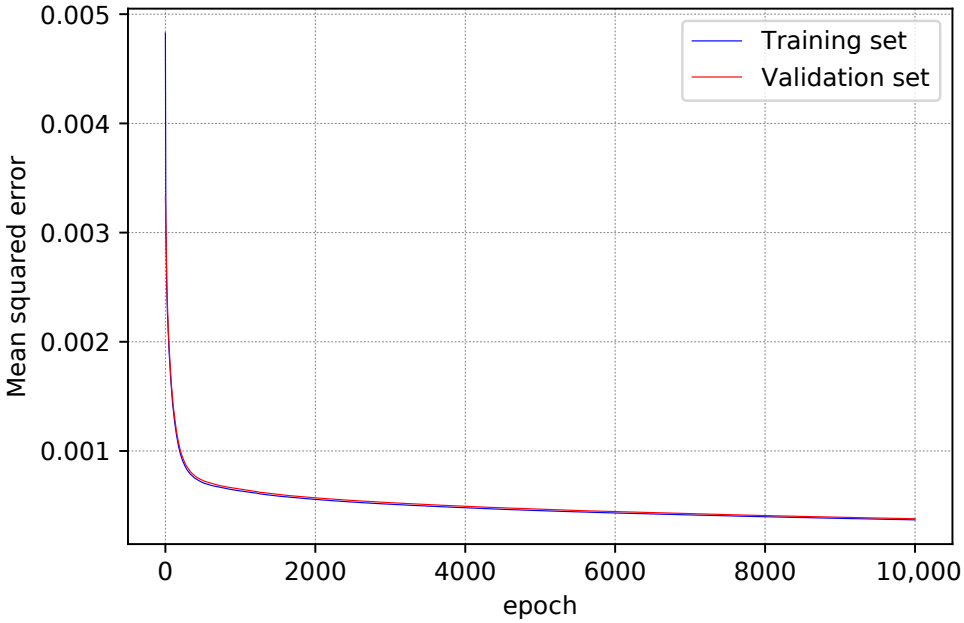


Figure 2.8. Training and validation loss during the training phase of the super-resolution neural network.

2.4.4 Distance Estimation

The results of CDF of the distance error utilizing the different approaches we presented are shown in Fig. 2.9 and Fig. 2.10. The CDF of the distance error while utilizing 4×4 photos is shown in green. The consequences of adding super-resolution to the original 4×4 photos to upscale them to 8×8 are shown in particular by a dotted line. The CDF of the distance error when utilizing 8×8 pictures is shown in light blue. The consequences of adding super-resolution to the original 8×8 photos to upscale them to 16×16 are shown in particular by a dotted line. The CDF of the distance error while utilizing 16×16 pictures is shown in blue. We compared the proposed method with the DCNN location estimation method [53], the Regression-based method [89], and the Classification-based method [52]. The simulation results show that the proposed method has a significant improvement in the reduction of the IUD estimation error. As mentioned before, [52,

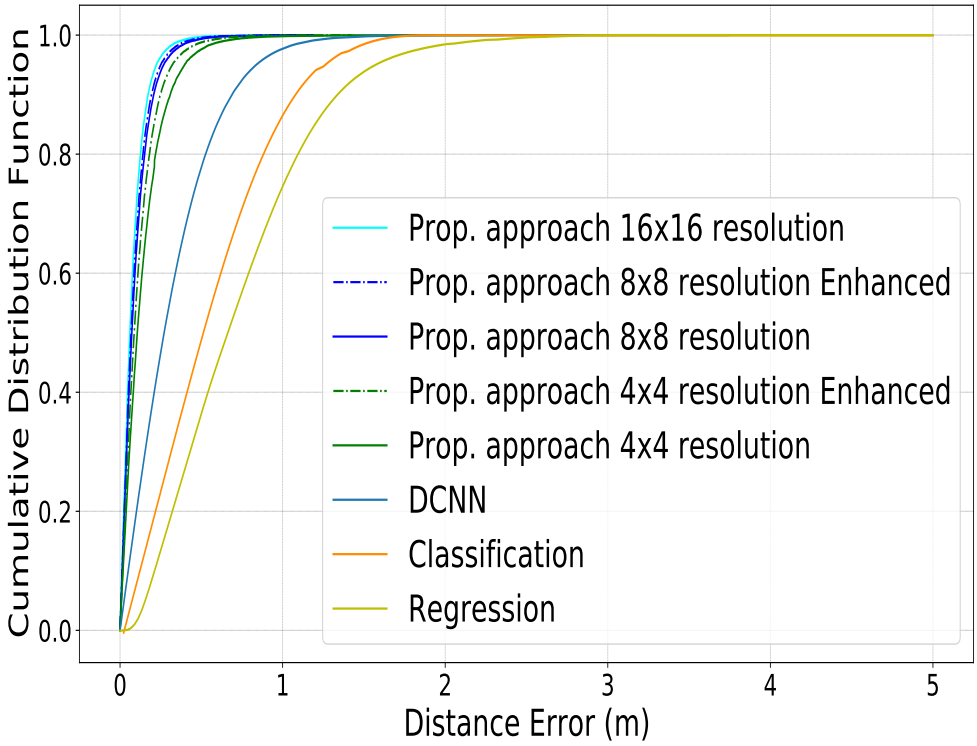


Figure 2.9. CDF of the distance estimation error for different transmit antennas configurations.

53, 89] extract the ADP from CSI as the fingerprint. However, we use beam scanning to generate beam energy images that contain more angular and positional information than ADP. Furthermore, we use the difference between the beam energy images of the two users as the fingerprint, which simultaneously contains the location information of the two users. Thus, our proposed method achieves a better inter-distance estimation performance than the other methods.

In Table 2.3, we summarize the reported values of the error at CDF = 0.5 and CDF = 0.9 for ease of comparison. As we can observe, our proposed method outperforms the conventional ones [52, 53, 89], even when using wide beams (i.e., 4×4 ones).

At CDF = 0.5, our proposed approach reaches an error equal to 0.093 m for images of size 16×16, equal to 0.097 m for images of size 8×8, and equal to 0.160 m for images of size 4×4. More interestingly, when we employ super-resolution to upscale the 8×8 and 4×4 images to 16×16 and 8×8, our proposed method reaches an error equal to 0.096 m and 0.101 m, respectively, at CDF = 0.5. On the other hand, at CDF = 0.5, the best performance of the three conventional methods reaches an error equal to 0.280 m.

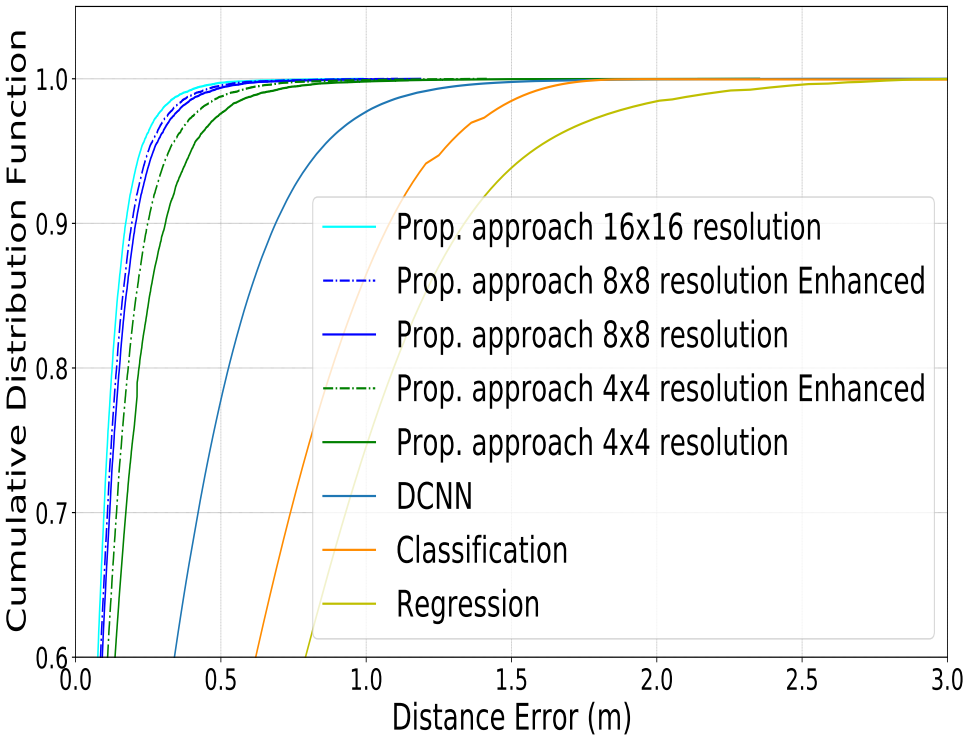


Figure 2.10. A zoom on the CDF of the distance estimation error for different transmit antennas configurations between CDF = 0.6 and CDF = 1.

TABLE 2.3
A SUMMARY OF THE DISTANCE ESTIMATION ERROR AT CDF = 0.5 AND AT CDF = 0.9.

	CDF = 0.5	**CDF = 0.9**
DCNN	0.280 m	0.703 m
Classification	0.409 m	1.018 m
Regression	0.780 m	1.304 m
4×4	0.160 m	0.304 m
4×4 enhanced	0.101 m	0.231 m
8×8	0.097 m	0.205 m
8×8 enhanced	0.096 m	0.197 m
16×16	0.093 m	0.184 m

Similarly, when measuring the error at CDF = 0.9, our proposed approach reaches an error equal to 0.184 m for images of size 16×16, 0.205 m for images of size 8×8, and 0.304 m for images of size 4×4, respectively. After applying super-resolution to images of size 8×8 and images of size 4×4, the error decreases down to 0.231 m and 0.197 m, respectively. On the other hand, at CDF = 0.9, the best performance of the three conventional methods reaches an error equal to 0.703 m at best.

As can be shown, despite falling behind when conducting location detection, the proposed technique surpasses the standard one [52,53,89] in IUD estimates. This is because, unlike the traditional method, which maps fingerprints to location, our system learns to recognize the distance between users independent of where they are. Using such uncertain data twice for the inter-user estimate, however, increases the estimation error when measuring the position of individual users. It's also worth noting that our strategy surpasses the traditional one in another way: the traditional method trains on exactly N occurrences given a set of N users in the training set, necessitating a high number of users for correct training. Our solution, on the other hand, will create a total of $\frac{N \cdot (N-1)}{2}$ instances for training for the N users, necessitating the simulation tool to generate data for fewer users to train the neural network effectively. We can observe that the narrower the beams are, the higher the detection accuracy is with our proposed technique. However, a more significant trend that we can see is that when the super-resolution approach is utilized, the results are considerably improved. This demonstrates the value of this method in terms of not only enhancing image quality but also refining distance estimates. Another advantage of our proposed strategy is that it is less likely to suffer performance degradation as a result of BS relocation. This is since it does not rely on the fingerprints themselves, but rather on the differences between them for various users. However, our strategy necessitates the training of two distinct networks: one for super-resolution and the other for distance calculation. With regards to the use of the highest resolution (i.e., 16×16) in particular, we can notice that, despite the improvement observed, this improvement is, in some sense, not justifiable: the power consumption and the time required to perform the beam sweeping are 4 times greater than when performing 8×8 sweeping, and 16 times

greater than when performing 4×4 sweeping. Nonetheless, a similar increase in computation complexity is seen when training the neural network and inferring it. Similarly, applying super-resolution to 4×4 images has led to an improvement of over 50% and 31% in the error estimation at CDF = 0.5 and CDF = 0.9, respectively. However, after applying super-resolution on 8×8 images, the improvement reaches only 1% and 4% in the error estimation at CDF = 0.5 and CDF = 0.9, respectively, leading us to believe that the computation cost might not be justifiable in this case.

2.4.5 Robustness Analysis

To verify the robustness of our proposed method, we evaluate it in a new environment. Here, we run our simulations on an environment with different building structures on Wireless Insite [139]. The difference in buildings locations and orientations leads to different reflections, thus the trained model is expected to drop in performance of IUD estimation. However, rather than training the entire network from scratch, we fine-tune the already built model (i.e., the one created in the first environment) on the new environment using a limited number of users dispersed over its area, and using a limited number of epochs. In other words, the objective of this subsection is to estimate how much ground truth data are required, and how much should the model be re-trained to provide performance close to the original one.

Fine-Tuning with Different Number of Users

As described above, since the model needs to be adjusted to fit the new environment, we need to fine-tune it with some data collected from this new environment. However, it is impractical to use the same number of users we used to first train the model. We need to evaluate how many user locations are needed to perfectly fine-tune the model. Given the different number of users (referred to as N^i_{user} where $i \in \{20, 50, 100\}$) whose location is known, we fine-tune the model using these data, and evaluate it on the entire region. The model is fine-tuned for 100 epochs using these data.

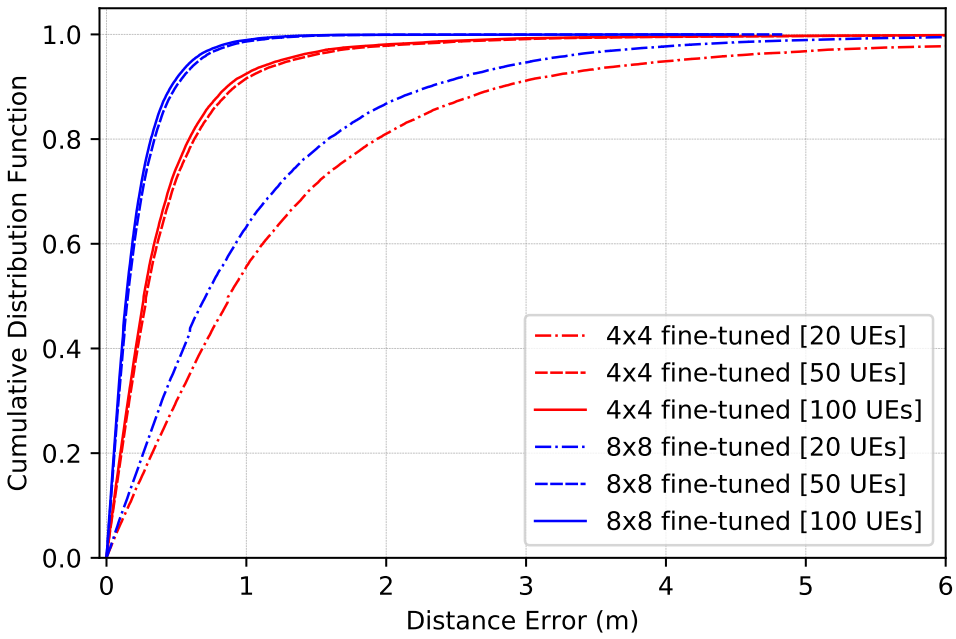


Figure 2.11. CDF of the distance estimation error on the new environment after fine-tuning the model for 50 epochs with 20, 50, and 100 users.

The CDF of the distance estimation error on the new environment for N_{user}^{20}, N_{user}^{50}, and N_{user}^{100} is given in Fig. 2.11. Here, the red color refers to the fine-tuning of the model built for 4×4 images, and there blue one refers to the fine-tuning of the model built for 8×8 images. Nevertheless, the values at CDF = 0.5 and CDF = 0.9 are given in Table 2.4. As can be seen, after training the model with only 50 users, we reach a decent localization precision. For 4×4 images, the error reaches 0.284 m at CDF = 0.5, and 0.919 m at CDF = 0.9. This is not very far from the precision when using 100 users ' data, where the error reaches for the same values of CDF 0.266 m and 0.862 m, respectively. The same behavior could be observed for images of size 8×8. When fine-tuning the network using 50 users, the error at CDF = 0.5 is equal to 0.153 m, and that at CDF = 0.9 is equal to 0.497 m for images of size 4×4. When using images of size 8×8, the error is equal to 0.142 m and 0.461 m, respectively.

Fine-Tuning with Different Number of Epochs

Here again, we need to identify the minimum number of epochs required to fine-tune the model well enough to perform as well as the model in the original environment. Since we

TABLE 2.4
A SUMMARY OF THE DISTANCE ESTIMATION ERROR AT CDF = 0.5 AND AT CDF = 0.9 ON THE NEW ENVIRONMENT USING DIFFERENT NUMBER OF USERS.

	CDF = 0.5	**CDF = 0.9**
4×4 fine-tuned [20 users]	0.873 m	2.828 m
4×4 fine-tuned [50 users]	0.284 m	0.919 m
4×4 fine-tuned [100 users]	0.266 m	0.862 m
8×8 fine-tuned [20 users]	0.710 m	2.307 m
8×8 fine-tuned [50 users]	0.153 m	0.497 m
8×8 fine-tuned [100 users]	0.142 m	0.461 m

have concluded from the previous set of experiments that 50 users is enough to fine-tune the model efficiently, we use this same number in our next set of experiments. Here, we try the different number of epochs of training of the model to perform the overfitting. We refer to the number of epochs as N^i_{epoch} where $i \in \{10, 50, 100\}$.

The CDF of the distance estimation error on the new environment for N^{10}_{epochs}, N^{50}_{epochs}, and N^{100}_{epochs} is given in Fig. 2.12. Here, the red color refers to the CDF of the fine-tuning of the model built for 4×4 images, and the blue one refers to the CDF of the fine-tuning of the model built for 8×8 images. In addition, the values at CDF = 0.5 and CDF = 0.9 are given in Table 2.5. Here, we can observe that after training the model for 10 epochs, the performance of the model is very poor, both when using 4×4 images and 8×8 images. In the case of 4×4 images, the error reaches 0.873 m at CDF = 0.5, and 2.826 m at CDF = 0.9. In the case of 8×8 images, the error reaches 0.655 m at CDF = 0.5, and 2.129 m at CDF = 0.9.

The precision improves when training for more epochs. For instance, after training the model for 50 epochs, and using images of size 4×4, the error at CDF = 0.5 and

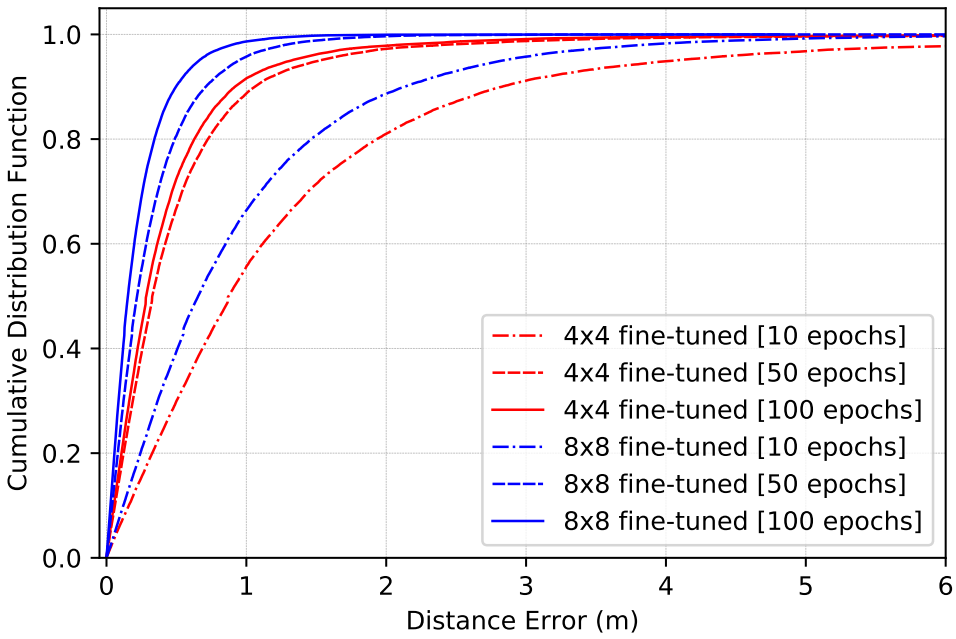


Figure 2.12. CDF of the distance estimation error on the new environment after fine-tuning the model for 10, 50, and 100 epochs.

CDF = 0.9 reach 0.327 m and 1.060 m, respectively. Using images of size 8×8, these values reach 0.218 m and 0.710 m, respectively. When using 100 epochs, the precision improves even further: Using images of size 4×4, the error at CDF = 0.5 and CDF = 0.9 reach 0.284 m and 0.919 m, respectively. Using images of size 8×8, these values reach 0.153 m and 0.497 m, respectively

2.4.6 Complexity Analysis

To estimate the complexity of our proposed method, we use the total number of parameters of the neural networks as an indicator. We have a set of convolutions, a set of dense layers, and a single max pooling layer. To that, we add the number of ReLU parameters. To recall, every convolutional layer is followed by a ReLU layer. The total number of parameters P of a given convolutional layer c is given by:

$$P(c) = ((m \cdot n \cdot p) + 1) \cdot k \tag{2.13}$$

where m and n are the width and height of each filter (3×3 in our case), p is the number of channels and k is the number of filters in the layer. The total number of parameters P

TABLE 2.5
A SUMMARY OF THE DISTANCE ESTIMATION ERROR AT CDF = 0.5 AND AT CDF = 0.9 ON THE NEW ENVIRONMENT FOR DIFFERENT EPOCHS.

	CDF = 0.5	**CD = 0.9**
4×4 fine-tuned [10 epochs]	0.873 m	2.826 m
4×4 fine-tuned [50 epochs]	0.327 m	1.060 m
4×4 fine-tuned [100 epochs]	0.284 m	0.919 m
8×8 fine-tuned [10 epochs]	0.655 m	2.129 m
8×8 fine-tuned [50 epochs]	0.218 m	0.710 m
8×8 fine-tuned [100 epochs]	0.153 m	0.497 m

of a given ReLU layer a is given by:

$$P(a) = h \cdot w \cdot k \tag{2.14}$$

where h and w are the height and width of the input image, respectively, and k is again the number of filters. In addition to Equations (2.13) and (2.14), we use the following equation to calculate the total number of parameters P of a given dense layer d:

$$P(d) = (s \cdot t) + 1 \tag{2.15}$$

where s is the size of the dense layer (the number of neurons) and t is the number of neurons in the previous layer.

As shown in Table 2.6, compared to the conventional methods, our neural network has a larger number of parameters than that of the other three methods. The total number of parameters of the classification neural network, when using input images of size 4×4, is about 1.4 M. When using input images of size 8×8, it is equal to 1.65 M, and when using

input images of size 16×16, it is equal to 2.44 M. When applying the super-resolution technique, another network is being used, leading to a total number of parameters of about 2.86 M and 3.06 M for input images of size 4×4 and 8×8, respectively.

TABLE 2.6
COMPLEXITY OF THE PROPOSED METHOD AND THAT OF DCNN.

Model	**Total Params**
4×4	1,461,121
4×4 enhanced	2,861,537
8×8	1,657,729
8×8 enhanced	3,058,145
16×16	2,444,161
DCNN [53]	41,401
Classification [52]	85,332
Regression [89]	61,231

Compared to conventional methods that use shallow networks such as DCNN [53], our method might seem much more complex. However, it is important to keep in mind that conventionally, image classification techniques are much more expensive, computation-wise. For instance, typical network architectures such as ResNet34 [140] and VGG16 [141] have a total number of parameters that is about 21 M and 138 M, respectively. That being said, compared to conventional methods, our proposed method can extract more information-rich features from the beam energy images, thus achieving a significant improvement in the reduction of the IUD estimation error. Finally, once the network is fully trained, the total number of basic operations (addition and multiplication) to be performed is constant and grows linearly with the number of users. The such number of operations could be justified for the sake of achieving an estimation error that is of the order of a few centimeters.

2.5 Conclusion

In this Chapter, we proposed a novel approach for IUD estimation using low-resolution beam energy images. The approach relies on the difference between the user-generated beam energy images to estimate the distance between each pair of users. We then applied a super-resolution technique to improve the IUD estimation accuracy with low-resolution beam energy images. Our experiments show that our method can achieve a distance estimation error equal to 0.13 m for a coverage area of 60×30 m^2. Our method outperforms the conventional methods based on user location to measure the IUD. Besides, applying super-resolution to images of resolution 4×4 and 8×8, improving their resolution to 8×8 and 16×16, respectively, has led to a further improvement in the estimation of the distance between the users. Compared with the original 4×4 and 8×8 images, the enhanced versions of these images by super-resolution exhibit better performance in the estimation of IUD. In fact, they achieve an estimation accuracy comparable to the original 8×8 and 16×16 images, respectively.

Chapter 3

DRL-Based Methods for Antenna Control in HAPS without Users' Location Information

3.1 Introduction

High Altitude platform station (HAPS) provides extremely broad coverage regions and a powerful line-of-sight (LoS) connectivity to terrestrial user equipment (UE) at the ground. As early as the 1990s, HAPS began to be paid attention to and studied through numerous research perspectives [22].

Compared with Geostationary Earth Orbit (GEO) satellites that are orbiting at a height of about 36,000 km and Low Earth Orbit (LEO) satellites that are orbiting at a height of about 1200 km, HAPS operates at the stratosphere at heights between 20 and 50 km. Therefore, the Round Trip Time (RTT) of HAPS is much faster than that of GEO and LEO satellites. Furthermore, since HAPS is relatively close to the ground, the power density is approximately one million times that of a GEO satellite and approximately ten thousand times that of an LEO satellite, allowing HAPS to provide high-quality communication services to existing mobile devices [23]. In addition, compared to other systems such as Starlink, which operates at an altitude of 340 km to 550 km [24], HAPS is much less expensive in terms of both the launch and the communication costs. Nonetheless, it does not "pollute" the upper layers of the atmosphere with the space waste it creates.

It is recognized as one of the hot topics for Beyond 5G (B5G) and 6G mobile communications [17–21]. With the necessary advanced materials and technological leaps, HAPS has been discussed as a viable technique.

3.1.1 Background

With their potential and with the decrease in the cost of the technology behind it, they are expected to be massively deployed for consumer usage in the coming years as candidates for cellular coverage to provide service or to augment the capacity of other broadband service providers [142].

S. Karapantazis et al. [18] and A. K. Widiawan et al. [143] summarized the essential technical aspects of HAPS systems and the current and potential applications of HAPS. In [144], the authors studied the potential HAPS system architectures and deployment strategies in order to achieve global connectivity.

In [145], White et al. studied the possibility of using HAPS to provide high data rate communications simultaneously to a number of trains in motion. According to the estimated and tracked Direction-of-Arrival (DoA) at the HAPS, they can control the parameters of the antenna array at the HAPS to transmit the beam to the UEs.

3.1.2 Related Work

However, due to wind pressure, it is difficult for HAPS to remain stationary. Thus, the degradation of the users' throughput and handovers to UEs' end happened [60, 61] after the HAPS coverage shifting. This kind of quasi-stationary state seriously impacts the performance of the communication system [62]. Dessouky et al. [63, 64] researched the problem of maximization of coverage through optimization of the parameters of the HAPS antenna arrays, and proposed an optimized way to minimize both the coverage gaps between cells and the excessive cell overlap. Yasser et al. [146] studied the influence of handover performance when the HAPS is moved or rotated by winds. He et al. [147] examined the swing state modeling of the cellular coverage geometry model and the influence of swing on handover. Many studies [65–67, 144, 148–153] on antenna control of HAPS proposed employing antenna control methods to prevent interference between surrounding cells and HAPSs to alleviate the decrease in received signal power caused by HAPS shifting or rotation. Kenji et al. [148] proposed a beamforming method to reduce the impact of the degradation of system capacity caused by handover between two cells. Florin et al. [65] analyzed the concentric circular antenna array (CCAA) and proposed a Genetic Algorithm (GA) to minimize the maximum side-lobe level (SLL). In [66], Sun et al. further developed the discrete cuckoo search algorithm (IDCSA) used to reduce the maximum SLL under the constraint of a particular half-power bandwidth. To increase the system capacity, Dib et al. [149] also researched SLL reduction. In contrast to existing approaches, they proposed a Symbiotic Organism Search (SOS) algorithm. The SOS algorithm requires no tuning of parameters, which makes it an attractive optimization method. In [67], the particle swarm optimization (PSO) GA is used for reducing the SLL to improve the carrier-to-interference ratio (CIR). However, in high-dimensional

space, PSO is easy to enter a local optimum, such as other GAs, and the iterative process' convergence rate is low. These limitations motivate us to develop a new method with high convergence and better throughput performance.

With the rapid development of deep learning techniques, reinforcement learning (RL) is widely used in various fields including in 5G and B5G [25–27]. F. B. Mismar et al. in [45] used Deep Q-Network (DQN) for online learning on how to maximize the users' signal-to-interference plus noise ratio (SINR) and sum-rate capacity. The authors design a binary encoding for performing multiple relevant actions at once in the DQN structure. A. Rkhami et al. in [46] used the RL method to solve the virtual network embedding problem (VNEP) in 5G and B5G. The authors considered that the conventional Deep Reinforcement Learning (DRL) usually obtains the sub-optimal solutions of VNEP, which leads to inefficient utilization of the resources and increases the cost of the allocation process. Thus, they proposed a relational graph convolutional neural network (GCNN) combined with DRL to automatically learn how to improve the quality of VNEP heuristics. In [47], the authors proposed a deep learning integrated RL, which combined deep learning (DL) and RL. The DL is used for preparing the optimized beamforming codebook and the RL is used for selecting the best beam out of the optimized beamforming codebook based on the user movements.

According to those early works, we can find that the DRL technique can train the neural network with feedback from the environment. Thus, such as solving the beamforming problem in [45] that using the DRL solution does not require the CSI to find the SINR-optimal beamforming vector. Moreover, different from model-free RL methods that need a huge optimal solution searching overhead for solving a complex problem, the DRL approach uses the DNN to predict the optimal solution without searching overhead. The DRL solution can be trained by the SINR feedback from users. Based on this motivation, we study the DRL and propose a novel DRL approach.

3.1.3 Motivations and Contributions

The movement and rotation of the HAPS due to wind pressure can cause the cell range to shift, which in turn causes the degradation of users' throughput and handovers between cells [60, 61]. With the development of Global Positioning System (GPS) technology, HAPS can control itself to recover its original position state according to GPS positioning technology and thus restore the user signal quality. That being said, the position and rotation state of HAPS will always fluctuate within a certain range due to the unpredictable wind direction and wind force. However, we can improve throughput with antenna beamforming by designing antenna parameters. Thus, to solve the degradation of received power at UEs, we propose two DRL-based methods, MFDQN and DRLEA. In our previous work [68], we proposed a Fuzzy Q-learning method. This method used Fuzzy logic in the model-free RL method named Q-learning to control multiple searches in a single training step. Compared with the Q-learning, the proposed Fuzzy Q-learning method has a lower cost of action searching. However, we found that the throughput performance of the proposed Fuzzy Q-learning method under non-uniform user distribution scenarios needs to be improved. Besides, however, assuming that there are f selections of each antenna parameter, F antenna parameters for each antenna array, and N antenna arrays in each HAPS, then the number of actions is f^{F^N}. Thus, to search and learn the optimal actions, it will cost a huge searching overhead and need a large number of datasets to train the DQN. Therefore, we proposed MFDQN and DRLEA for dynamic antenna control in the HAPS system to reduce the number of low-throughput users under the non-uniform user distribution scenarios. In MFDQN, to reduce the searching overhead, we decompose the antenna parameters control in each HAPS into each antenna array. Each antenna array as an agent, therefore, the number of actions is f^F of each agent. We model the multi-agent antenna parameters control problem into stochastic game equilibrium. To address this issue, we use the proposed MFDQN to reduce the complexity of the interactions among all agents.

The proposed DRLEA considers that each iteration in the conventional DRL is a new generation. Starting from the first generation (the first iteration), DRLEA searches for the

optimal solution in the current generation and trains the DNN to learn this evolutionary process. The DRLEA records the best result of the current generation as the historical optimal solution for guiding the next generation. If the current generation cannot find better solutions to reduce low throughput users, i.e. users whose throughput is lower than the median throughput prior to antenna parameters adjustment, the mutation happens (randomly selecting an action). The process is repeated until the specified execution time is reached. Compared with conventional RL methods, such as Q-learning-based and DQN-based methods, the proposed method can avoid the sub-optimal solution as much as possible. This is because, in every iteration a random initial set of parameters is used, and is then optimized, leading to less chances of falling into the local optimal.

In addition, the performance of user throughput is closely related to the user's SINR and bandwidth. We can improve the user's SINR by gradually adjusting the antenna parameters based on the user's feedback. As for the user's dedicated bandwidth, it is only a function of the coverage of the antenna array in which the user is located as well as the number of users within that coverage area. Therefore, we do not necessarily need to know the channel state information (CSI) to design the beamforming matrix. Compared with obtaining an accurate CSI, obtaining the user location information, such as using the GPS technique, is less costly in terms of resources and time. It is nonetheless easier and computationally less expensive to adjust the antenna array parameters to account for the changes to the footprint of the users' locations than that of their CSI.

Moreover, we implement three conventional methods, PSO, Q-learning, and DQN as benchmarks to evaluate the proposed MFDQN and DRLEA methods This Chapter's contributions can be outlined as follows:

- We propose the MFDQN method that decomposes the single agent in the HAPS into multi-agent in each antenna array to reduce the action space. Therefore, we can model the antenna array control problem into the stochastic game to find the equilibrium to reduce the number of low-throughput users. To solve this stochastic game problem, we use the proposed MFDQN to learn the transition probability and

predict the Q-value to reduce the complexity of the interactions among agents and improve the convergence rate.

- We propose a novel DRLEA approach that combined the EA and DRL to avoid sub-optimal solutions. We design a new loss function that includes not only Q-value of the predicted optimal action, but also the historical optimal solutions obtained from previous training.

- We use the realistic user distribution instead of using the sample uniform user distribution. The simulation results show that the proposed MFDQN and DRLEA effectively reduce the much more number of low throughput users than that of using four conventional algorithms.

- To evaluate the robustness of the proposed methods, we train the proposed methods under the HAPS with rotation of 30 degrees scenarios and evaluate the proposed methods under the HAPS with a left shift of 5 km scenarios. The simulation results show that the proposed MFDQN and DRLEA without retraining under the shift scenarios achieve a cumulative distribution function (CDF) of throughput performance comparable to that of the conventional methods trained or learned under the shift scenarios.

3.2 System Model

3.2.1 Model of HAPS

in Fig. 3.1, we show a typical HAPS communication system model. We think about the scenario of M HAPSs serving multiple users in the mMIMO mmWave networks. In more detail, each HAPS is equipped with N antenna arrays to generate N beams. M HAPSs assist the BS to serve multiple users. HAPSs relay the signal transmitted from the BS to users to improve the throughput of users. Moreover, we also consider the wind-caused HAPS movement, as shown in Fig. 3.2. Figure 3.2a shows the HAPS with shifting and

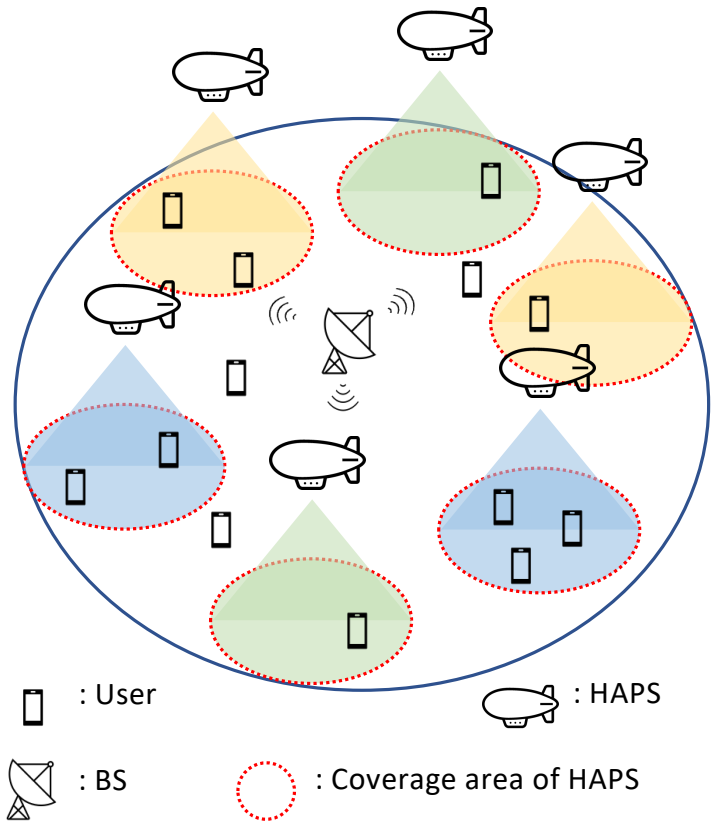
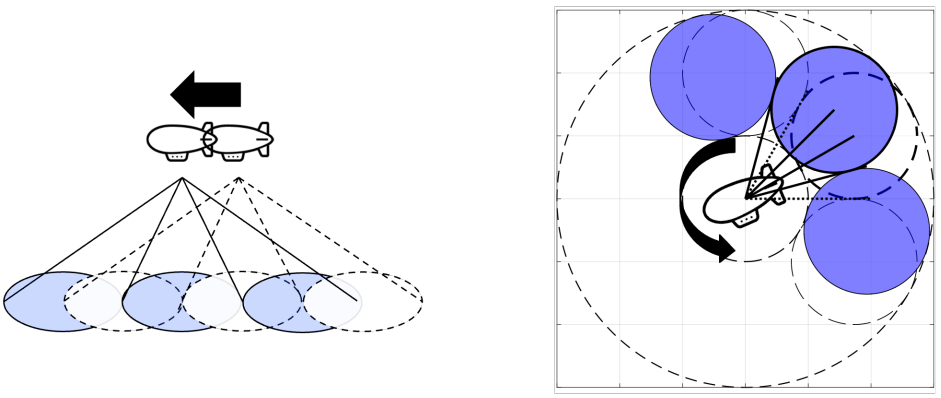


Figure 3.1. A HAPS system model.

Figure 3.2 (b) shows the HAPS with rotation. Whatever shifting or rotation, it will cause the degradation of users' throughput.

3.2.2 Antenna Pattern

Planar patch antennas [154] used in each antenna array is considered. To obtain the vertical and horizontal antenna gains, for each antenna array, four antenna parameters are considered: Fig. 3.3a shows the vertical beam half power beam width (HPBW) θ_{3dB} and the horizontal beam HPBW ϕ_{3dB}, Fig. 3.3b shows the vertical tilt θ_{tilt} and the horizontal tilt ϕ_{tilt}. In Fig. 3.3b, $\Delta\phi_{tilt}$ and $\Delta\theta_{tilt}$ denote the changing of ϕ_{tilt} and θ_{tilt}, respectively. From there, we could derive the expression of the horizontal or vertical antenna gains for an angle Ψ from the main beam direction as [155]:



(a) HAPS with shifting (b) HAPS with rotation

Figure 3.2. Movement scenarios.

$$G_h(\Psi) = G_v(\Psi) \begin{cases} -3(\Psi/\Psi_b)^2, & (0^\circ \leq \Psi \leq \Psi_1) \\ L_N, & (\Psi_1 < \Psi \leq \Psi_2) \\ X - 60\log_{10}(\Psi), & (\Psi_2 < \Psi \leq \Psi_3) \\ L_F, & (\Psi_3 < \Psi \leq 90^\circ) \end{cases} \tag{3.1}$$

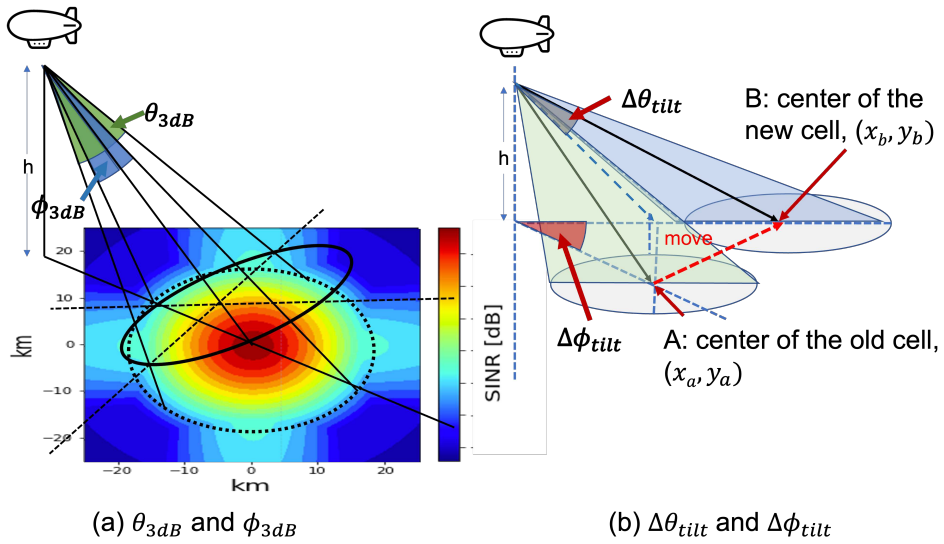
where Ψ_b is one-half the 3 dB beamwidth in the plane of interest, $\Psi_1 = \Psi_b\sqrt{-L_N/3}$, $\Psi_2 = 3.745\Psi_b$, $X = L_N + 60\log_{10}(\Psi_2)$, and $\Psi_3 = 10^{(X-L_F)/60}$ [156], L_N denotes the near-in-side-lobe level, and L_F denotes the far-side-lobe level.

Therefore, we could express the combined gain G as:

$$G = \max(G_v + G_h, L_F) + G_p, \tag{3.2}$$

where G_v and G_h are the vertical and horizontal antenna gains, respectively, and G_p denotes the maximum antenna gain as shown in Equation (3.3).

$$G_p = 10\log_{10}(\frac{{B_w}^2}{\theta_{3dB}\phi_{3dB}}) + G, \tag{3.3}$$



(a) θ_{3dB} and ϕ_{3dB} (b) $\Delta\theta_{tilt}$ and $\Delta\phi_{tilt}$

Figure 3.3. Antenna parameters.

where B_w denotes the beamwidth. Fig. 3.4 shows an example of an antenna pattern for vertical or horizontal polarization.

3.3 Problem Formulation

To reduce the number of users whose with low throughput, we formulate the dynamic antenna control problem into maximizing the throughput of users. The throughput $\mathcal{T}$ of a user can be obtained by the following equation:

$$\mathcal{T} = b \times \log_2(1 + \gamma), \tag{3.4}$$

where γ denotes the SINR at the user, $b = B/K$ denotes the bandwidth assigned to the user, B denotes the bandwidth of each antenna array and K denotes the number of users in an antenna array coverage. In the dissertation, we address the case of a single HAPS movement. Our objective is to control only the HAPS in question (which has supposedly moved) to reduce the number of low-throughput users in the area it was serving before it moved. This means that no overlap or exchange of areas with surrounding HAPS will occur.

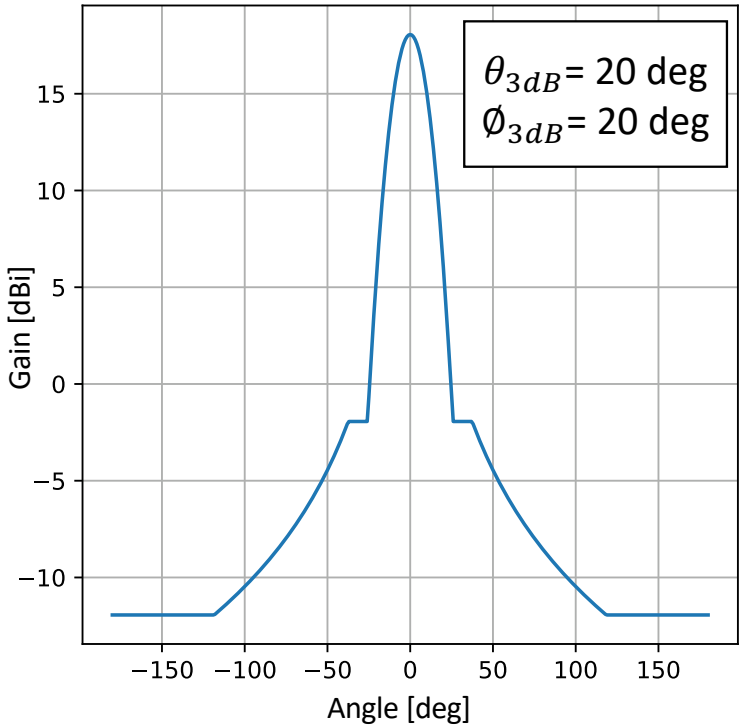


Figure 3.4. Example of an antenna pattern for vertical or horizontal polarization.

Therefore, we can simplify the system model as follows. We consider $M - 1$ fixed-position HAPS around the HAPS in question that is moved due to wind pressure. Our target is to optimize the antenna parameters of this HAPS to reduce the number of users with low throughput caused by the HAPS movement. No handovers between the different HAPS is accounted for, and no new users are introduced to the coverage area of the HAPS in question.

3.3.1 Markov Decision Process

To solve this antenna control problem, we convert it into a Markov Decision Process (MDP) problem. We define the state at time t as s_t, and the selected action under the state s_t is A_j. Moreover, we consider using the four antenna parameters $[\phi_{tilt}, \theta_{tilt}, \phi_{3dB}, \theta_{3dB}]$ at time t as the state s_t. The action $A_j \in \mathcal{A}$ is defined as the change of one set of antenna parameters, where $\mathcal{A}$ denotes the action set of the HAPS. To reduce the computational complexity, we use discrete antenna parameters to reduce the number of actions. Moreover, to perform the all actions at once, we design the action mapping list as shown in

Table 3.1. We assume that the number of antenna parameters of each antenna array is $P = 4$, and the number of values of each antenna parameter is $V = 3$. Thus, the number of actions of each antenna array is $L = V^P$, and the number of actions of a HAPS is $J = L^N$. $A_1(a_{(1,1)}, \ldots, a_{(N,1)})$ in Table 3.1 indicates that antenna array 1 to antenna array N performs action index 0.

Table 3.1. Actions mapping list.

Actions of the HAPS	Actions of the antenna array i	Action index	ϕ_{3dB}	θ_{3dB}	ϕ_{tilt}	θ_{tilt}
$A_1(a_{(1,1)}, \ldots, a_{(N,1)})$	$a_{(i,1)}(0,0,0,0)$	0	$+\Delta\phi_{3dB}$	$+\Delta\theta_{3dB}$	$+\Delta\phi_{tilt}$	$+\Delta\theta_{tilt}$
$A_2(a_{(1,1)}, \ldots, a_{(N-1,2)})$	$a_{(i,2)}(0,0,0,1)$					
$\vdots$	$\vdots$	1	$-\Delta\phi_{3dB}$	$-\Delta\theta_{3dB}$	$-\Delta\phi_{tilt}$	$-\Delta\theta_{tilt}$
$A_{J-1}(a_{(1,L)}, \ldots, a_{(N,L-1)})$	$a_{(i,L-1)}(2,2,2,1)$	2	$0\,^\circ$	$0\,^\circ$	$0\,^\circ$	$0\,^\circ$
$A_J(a_{(1,L)}, \ldots, a_{(N,L)})$	$a_{(i,L)}(2,2,2,2)$					

The reward $\mathcal{R}(s_t, A)$ with the state s_t and the action A is represented by the Equation (3.5).

$$\mathcal{R}(s_t, A) = \frac{\sum_{i,j}^{\frac{K}{2}} \mathcal{T}_i^{'} - \mathcal{T}_j}{\frac{K}{2}}, \tag{3.5}$$

where K denotes the number of users, $\sum_i^{\frac{K}{2}} \mathcal{T}_i^{'}$ denotes the sum of the throughput of the 50 percent users with the least throughput after performing the selected action A, and $\sum_j^{\frac{K}{2}} \mathcal{T}_j$ denotes the sum of the throughput of the 50 percent users with the least throughput under the initial state. Thus, the antenna parameters control problem is represented by an MDP:

$\mathcal{M} \triangleq (\mathcal{S}, \mathcal{A}, \mathcal{R}, p, \alpha)$, where $\mathcal{S}$ denotes an infinite state space, p denotes the transition probability that characterizes the stochastic evolution of states in time, with the collection of probability distributions over the state space $\mathcal{S}$, and $\alpha \in [0, 1)$ is the reward discount factor.

The goal is to find a deterministic optimal policy $\pi^* : \mathcal{S} \to \mathcal{A}$, such that:

$$\pi^* := \arg\max_{\pi \in \Phi} \mathbb{E}\left[\sum_{t=0}^{\infty} \alpha^t r\left(s_t, \pi\left(s_t\right)\right)\right], \tag{3.6}$$

where Φ is the set of all admissible deterministic policies. At time step t, HAPS selects an action simultaneously based on the policies π. The Q function is shown as follows:

$$Q^{\pi}(s, A) := \mathbb{E}\left[\sum_{t=0}^{\infty} \alpha^t \mathcal{R}\left(s_t, \pi\left(s_t\right)\right) \mid s_0 = s, A_0 = A\right]. \tag{3.7}$$

Thus, the optimal policy π^* can be obtained by:

$$\begin{aligned} \pi^*(s) &= \arg\max_{A} Q^*(s, A) \\ &= \mathbb{E}_{s' \sim p(\cdot|s,A)}\left[\mathcal{R}(s, A) + \max_{A'} \alpha Q^*\left(s', A'\right)\right] \end{aligned} \tag{3.8}$$

where s' denotes the next state, $\max_{A'} Q^*(s', A')$ denotes performing the action A' that can obtain the maximum Q value under the state s'.

3.4 Conventional Approaches for Antenna Beamforming

3.4.1 Q-Learning

Q-learning is a kind of classical RL algorithm [90]. To search the optimal Q function, Q-learning builds a Q table to record the sum of existing Q value $Q(s, A)$ of the action A for the current state s. The Q value is obtained by Equation (3.9).

$$Q(s, A) \leftarrow \mathrm{Q}(s, A) + \beta\left[\mathcal{R}(s, A) + \alpha \max_{j} Q\left(s', A_j\right) - Q(s, A)\right]. \tag{3.9}$$

Algorithm 1 Q – *learning* method for HAPS antenna control.

Input: s_t, α, ϵ, $\mathcal{A}$.
Output: $Q(s_t, A_1), Q(s_t, A_2), \ldots, Q(s_t, A_J)$.

Build a Q table QT.
for epoch = 0 to E **do**
    Initialize the antenna parameters.
    for step t = 0 to T **do**
        Obtain the antenna parameters as state s_t.
        Randomly generates r in the range of (0,1).
        if $r > \epsilon$ **then**
            Randomly select an action $A \in \mathcal{A}$
        else
            $A = \arg\max_j QT(s_t, A_j)$
        end if
        Perform the action A.
        Get reward $\mathcal{R}(s_t, A)$ by Equation (3.5).
        Update s_t to s_{t+1}.
        Update the Q table QT:
        $QT(s_t, A) = QT(s_t, A) + Q(s_t, A)$
    end for
end for

Here, β denotes the learning rate and $\max_j Q(s', A_j)$ denotes performing the action A_j that can obtain the maximum Q value under the state s'. The details of the Q-learning algorithm are shown in Algorithm 1.

3.4.2 Fuzzy Q-Learning

By using Fuzzy inference systems (FIS), Fuzzy logic can be used to represent actions and Q functions. Lookup tables indexing states and actions are used to store the Q function. When a state achieves a high reward, the reward is propagated to its neighboring states, and learning is facilitated. $Q_t(s_t, a_t)$ is the Q-value at the time step t. After we select and perform the action a_t at time step t, we can obtain a reward r_t from the environment and then update the state s_t to s_{t+1}. Therefore, if there are a large number of states and actions, it will take much longer to find an optimal action in a lookup table for the Q-values. To reduce the searching overhead, we consider using the FIS for storing Q-values [157]. Since it is possible to embed prior in the FIS, the Fuzzy set can provide approximate values for states and actions, which reduces both the lookup table sizes and the search

times. We define the Fuzzy set as M_i and $\boldsymbol{a}^{\dagger}_{i(=1,\ldots,n)}$ as the action selected by the Fuzzy set M_i. $Q(M_i, \boldsymbol{a}_j)$ denotes the Q-value with the state S_i and the action $\boldsymbol{a}_{j(=1,\ldots,n)}$. To update the Q-value, we use the Temporal Difference (TD) method [158]. The TD learning methods are capable of learning by contact with the environment on a step-by-step basis without the need for an environment model. The TD methods just need the observed reward and an estimate of the value of the following state to perform an update to the state value function. The TD algorithms can also be applied to multi-step backup procedures. By doing so, several states which have been previously visited are updated via a single experience, thus, the speed of the algorithm highly improves. The details are shown in the following (Eqs. (3.10) – (3.13)). α is the learning rate and γ is the discount rate. $\boldsymbol{a}^{*}_{i} = \arg\max_{j(=1,\ldots,n)} Q_t(S_i, \boldsymbol{a}_i)$ denotes the action which maximizes the Q-value in the state S_i. The memory of which state-action values is represented by the eligibility traces $e(S_i, \boldsymbol{a}_j)$. When the selected action $\boldsymbol{a}_j$ same as the previous state, we add the eligibility traces $e(S_i, \boldsymbol{a}_j)$ to sum the temporary Fuzzy membership value. In order to learn the appropriate Q-values for different actions at the same state, the action $\boldsymbol{a}^{\dagger}_{i}$ is selected for each Fuzzy set M_i using the ϵ-greedy method. We weight the action $\boldsymbol{a}_{i(=1,\ldots,n)}$ of each selected Fuzzy set by the FIS agreement $\mu_i(x)$ and add the action together to get the action $\boldsymbol{a}$, as shown in Eq. (3.14).

$$\begin{aligned} Q_{t+1}(s_t, \boldsymbol{a}_t) &= Q_t(s_t, \boldsymbol{a}_t) \\ &+ \alpha[r_t + \gamma \max_{\boldsymbol{a}} Q_t(s_{t+1}, \boldsymbol{a}) - Q_t(s_t, \boldsymbol{a}_t)], \end{aligned} \tag{3.10}$$

$$Q_t(M_i, \boldsymbol{a}^{\dagger}{}_i) = Q_t(M_i, \boldsymbol{a}^{\dagger}{}_i)_t + \varepsilon \times \Delta Q \times e(M_i, \boldsymbol{a}^{\dagger}{}_i), \tag{3.11}$$

$$\begin{aligned} \Delta Q = r + \gamma \times & \frac{\sum_{i=1}^{n} \mu_i(x) \times Q(M_i, \boldsymbol{a}^{*}{}_i)}{\sum_{i=1}^{n} \mu_i(x)} \\ - & \frac{\sum_{i=1}^{n} \mu_i(x) \times Q(M_i, \boldsymbol{a}^{\dagger}{}_i)}{\sum_{i=1}^{n} \mu_i(x)}, \end{aligned} \tag{3.12}$$

$$e(S_i, \boldsymbol{a}_j) = \begin{cases} \lambda\gamma e(M_i, \boldsymbol{a}_j) + \dfrac{\sum_{i=1}^{n}\mu_i(x)}{\sum_{i=1}^{n}\mu_i(x)}, & (\boldsymbol{a}_j = \boldsymbol{a}_i^{\dagger}) \\ \lambda\gamma e(M_i, \boldsymbol{a}_j), & \text{otherwise} \end{cases} \tag{3.13}$$

$$\boldsymbol{a}(x) = \frac{\sum_{i=1}^{n}\mu_i(x) \times \boldsymbol{a}^{\dagger}{}_i}{\sum_{i=1}^{n}\mu_i(x)}. \tag{3.14}$$

When the HAPS moved by the wind-pressure, each antenna beam can be controlled by the proposed Fuzzy Q-learning method to cover the target cell range. We use the KPI that is the coverage ratio of the beam for each antenna i in a predetermined cell C_i as the state, as shown in Eq. (3.15).

$$KPI_i = \frac{|\{u(\in U_i)|\gamma_i(u) \geq \Gamma\}|}{|U_i|} \tag{3.15}$$

$\boldsymbol{a}_t^{\dagger} \in \mathcal{A}$ is the action at each time step t, where $\mathcal{A} = [\Delta\phi_{3dB}, \Delta\phi_{tilt}, \Delta\theta_{3dB}, \Delta\theta_{tilt}]$. Here, $\Delta\phi_{3dB}$, $\Delta\phi_{tilt}$, $\Delta\theta_{3dB}$, and $\Delta\theta_{tilt}$ denote the changes of the horizontal HPBW and tilt, and the vertical HPBW and tilt, respectively. Thus, we can improve the KPI of each cell by adjusting the values of these 4 antenna parameters. Besides, to obtain the reward of each action, we define the lower and upper bounds of KPI, KPI_{bad}, and KPI_{good}. When $KPI \geq KPI_{good}$, we define the reward is 1. If $KPI \leq KPI_{bad}$ or the number of iterations for action selection and Q-table update exceeds the maximum number of iterations, the reward is -1. And if $KPI_{good} \geq KPI \geq KPI_{bad}$, the reward is 0. Based on the fed back reward, the proposed method can adjust the antenna parameters to reduce the number of outage users. When we perform an action and obtain a reward, the Q-table will be updated. After enough learning, the proposed method can search for the optimal action under different states from the Q-table. We define the Fuzzy set $M_{k(=1,\ldots,n)}$ as shown in Fig. 3.5 and the membership function F_k for each Fuzzy set M_k as shown in Eq. (3.16). Compared with the conventional Q-learning [159], using the Fuzzy sets reduce the cost of searching the optimal action from the Q-table, resulting in fast convergence to the optimal solution.

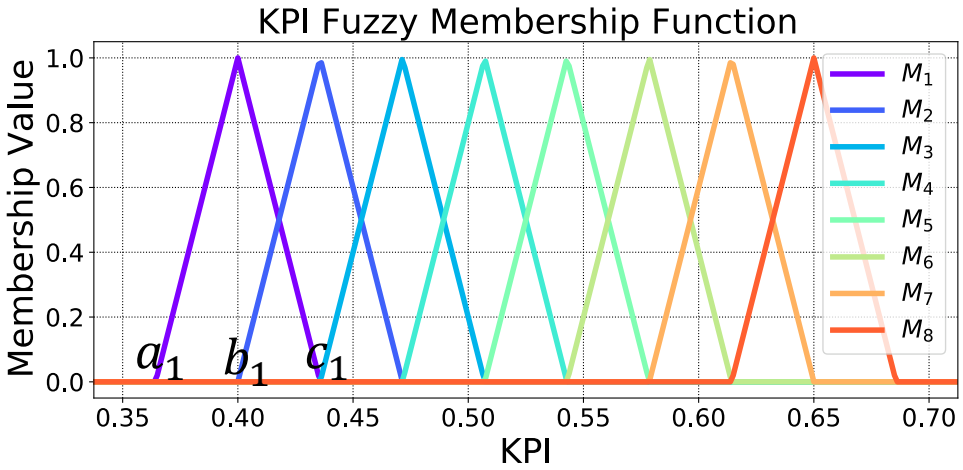


Figure 3.5. KPI Fuzzy membership sets.

$$F_k(KPI) = \begin{cases} 0, (KPI < a_k, KPI > c_k) \\ \frac{KPI - a_k}{b_k - a_k}, (a_k < KPI < b_k) \\ -\frac{KPI - b_k}{c_k - b_k}, (b_k < KPI < c_k) \end{cases} \tag{3.16}$$

$$(\Delta\phi_{3dB}, \Delta\phi_{tilt}, \Delta\theta_{3dB}, \Delta\theta_{tilt}) = \sum_k F_k(KPI) \times \boldsymbol{a}^{\dagger}{}_k \tag{3.17}$$

3.4.3 Deep Q-Network

DQN is another classical RL [91]. It calculates the Q value of each action based on the reward returned from the environment states and uses the DNN instead of Q table in Q-learning method to predict the optimal Q value. Therefore, we can obtain the Q value with the state s and the action A based on Equation (3.18).

$$Q(s, A) \leftarrow \mathcal{R}(s, A) + \alpha \max_j Q(s', A_j). \tag{3.18}$$

Algorithm 2 Fuzzy Q-Learning Algorithm

for $i = 0$ to N **do**
  Initialize antenna parameters
  $loop = 1$
  while $loop <= Loop_{max}$ **do**
    for M_1 to M_k **do**
      Select action $\boldsymbol{a}^{\dagger}{}_{\boldsymbol{k}}$ based on ε-greedy;
    end for
    Get the antenna control value:

$$(\Delta\phi_{3dB}, \Delta\phi_{tilt}, \Delta\theta_{3dB}, \Delta\theta_{tilt}) = \sum_{k} F_k(KPI) \times \boldsymbol{a}^{\dagger}{}_{\boldsymbol{k}}$$

    Update antenna parameters:

$$\begin{aligned} \phi_{3dB} &+= \Delta\phi_{3dB} \\ \phi_{tilt} &+= \Delta\phi_{tilt} \\ \theta_{3dB} &+= \Delta\theta_{3dB} \\ \theta_{tilt} &+= \Delta\theta_{tilt} \end{aligned}$$

    $loop++$
    if $KPI < KPI_{bad}$ **then**
      $reward = -1$
      Break
    else if $KPI > KPI_{good}$ **then**
      $reward = 1$
      Break
    end if
    if $loop == Loop_{max}$ **then**
      $reward = -1$
      Break
    end if
    for M_1 to M_k **do**
      Update Q value:

$$Q_t(M_i, \boldsymbol{a}^{\dagger}{}_i) = Q_t(M_i, \boldsymbol{a}^{\dagger}{}_i)_t + \varepsilon \times \Delta Q \times e(S_i, \boldsymbol{a}^{\dagger}{}_i)$$

    end for
  end while
end for

Algorithm 3 DQN-based method for HAPS antenna control.

Input: s_t, α, ϵ, $\mathcal{A}$.
Output: $Q(s_t, A_1), Q(s_t, A_2), \ldots, Q(s_t, A_J)$.

Initialize main network $\mathcal{F}(\Theta)$ and target network $\mathcal{F}(\Theta^{'})$.
Initialize experience reply memory $\mathcal{D}$.
for epoch = 0 to E **do**
    Initialize the antenna parameters.
    for step t = 0 to T **do**
        Obtain the antenna parameters as state s_t.
        Randomly generates r in the range of (0,1).
        if $r > \epsilon$ **then**
            Randomly select an action $A \in \mathcal{A}$
        else
            $A = \arg\max_j Q(s_t, A_j; \Theta)$
        end if
        Perform the action A.
        Get reward $\mathcal{R}(s_t, A)$ by Equation (3.5).
        Update s_t to s_{t+1}.
        Store $(s_t, A, \mathcal{R}(s_t, A), s_{t+1})$ in the $\mathcal{D}$.
        Get the output of the main network $\mathcal{F}(\Theta)$: $Q(s_t, A) = \mathcal{F}(s_t, A; \Theta)$.
        Generate target Q value:

        $Q^{'}(s_{t+1}, A') = \max \mathcal{F}(s_{t+1}; \Theta^{'})$.
        Update the main network $\mathcal{F}(\Theta)$ to minimize the loss function:

$$\begin{aligned} \mathcal{L}(\Theta) =& MSE(Q(s_t, A), \mathcal{R}(s_t, A) + \\ & \alpha Q^{'}(s_{t+1}, A')) \end{aligned} \tag{3.19}$$

    end for
    Update the target network: $\mathcal{F}(\Theta^{'}) \leftarrow \mathcal{F}(\Theta)$.
end for

Different from Q-learning searching Q table to obtain the optimal solution, the DQN method builds a deep neural network to learn the Q value of each possible action corresponding to the input environment state. The details of the proposed DQN-based antenna control method are shown in Algorithm 3.

In Algorithm 3, E denotes the maximum number of epochs, T denotes the maximum number of steps, and Θ denotes the weights of the deep neural network. We build two DQNs, one is the main network used for evaluating the Q value of the action obtained by the ϵ-greedy policy. This main network is trained during each step to estimate the approximate optimal action in the current state. The target network is updated with a copy

of the latest learned parameters of the main network after each epoch. In other words, using a separate target network helps keep runaway bias from dominating the system numerically causing the estimated Q values to diverge. Thus, using two DQNs instead of only one DQN can avoid the DQN algorithm to overestimate the true rewards [160]. We calculate the reward $\mathcal{R}(s_t, A)$ with the state s_t and the action A by Equation (3.5). In each step, we select an action based on the ϵ-greedy method, and store the current state s_t, the selected action A, the reward $\mathcal{R}(s_t, A)$, and the next state s_{t+1} into the experience replay memory $\mathcal{D}$ for training the main network $\mathcal{F}(\Theta)$. Next, we train the main network $\mathcal{F}(\Theta)$ to minimize the loss function $\mathcal{L}(\Theta)$, as shown in Equation (3.19). After enough training, we input the current state s_t into the main network, then we can observe the predicted Q value of all actions under the state s_t. Thus, the optimal action $A = \arg\max_j Q(s_t, A_j; \Theta)$ can be obtained.

3.4.4 Particle Swarm Optimization

In this dissertation, we modify the PSO antenna control method in [67] to reduce the number of low throughput users, as shown in Algorithm 4.

Where X_i^k denotes the antenna parameters of the k-th particle in the i-th iteration, V_i^k denotes the k-th particle's velocity (the change of antenna parameters) in the i-th iteration, $pbest_i^k$ denotes the best X of k-th particle until the i-th iteration, $gbest$ denotes the best X of all particles in all the iterations, ω_1 and ω_2 denote two independent learning rates, $rand_1$ and $rand_2$ are two independent random numbers for increasing randomness. The PSO algorithm randomly generates P particles to search the optimal antenna parameters X in each iteration and record it into $pbest$. After each iteration, if the $pbest$ of the current iteration is larger than the previous $pbest$, record it into $gbest$ as the optimal solution.

Algorithm 4 PSO based algorithm for HAPS antenna control.

Initialize the particles X and V.
for $i = 0$ to E **do**
  for $k = 0$ to P **do**
    Update V:

$$V_i^k = V_i^{k-1} + \omega_1 * rand1_i^k * (pbest_i^k - X_i^{k-1}) + \omega_2 * rand2_i^k * (gbest_i^k - X_i^{k-1}), \tag{3.20}$$

    Update X:

$$X_{i+1}^k = X_i^k + V_i^k \tag{3.21}$$

    Calculate $\mathcal{R}$ by Equation (3.5) according to X_{i+1}^k
    if $\mathcal{R}_{i+1}^k \geq \mathcal{R}_i^k$ **then** then
      $pbest_i^k \leftarrow X_{i+1}^k$
    end if
    if $\mathcal{R}_{i+1}^k \geq \mathcal{R}_{i+1}$ of other particles **then**
      $gbest \leftarrow X_{i+1}^k$
    end if
  end for
end for
Output the best antenna parameters $gbest$

3.5 Proposed Mean Field Deep Q-Network for Antenna Beamforming

Considering the interference between each antenna array and also the interference between each HAPS, we model the antenna control problem as an N agents mean field game. The core idea of mean-field game theory is to use the average value of the rewards and actions of an agent's neighbors to transform the complex relationship between an agent and other agents into the relationship between an agent and the average of all its neighbors. Thus, the antenna parameters control problem is formulated by the tuple $\Gamma \triangleq \left(\mathcal{S}, \mathcal{A}^1, \ldots, \mathcal{A}^N, r^1, \ldots, r^N, p, \gamma\right)$, where $\mathcal{S} = \{\mathrm{s}_1, \ldots, \mathrm{s}_i, \ldots, \mathrm{s}_N\}$ denotes the state space, $\mathrm{s}_i = [\phi_{tilt}, \theta_{tilt}, \phi_{3dB}, \theta_{3dB}]$ denotes the state of each antenna array, p denotes the transition probability that characterizes the stochastic evolution of states in time, with the collection of probability distributions over the state space $\mathcal{S}$, and $\gamma \in [0, 1)$ is the reward discount factor [161]. $\mathcal{A}^j = [\Delta\phi_{tilt}, \Delta\theta_{tilt}, \Delta\phi_{3dB}, \Delta\theta_{3dB}]$ is the action space of agent

$j \in \{1, \ldots, N\}$. r^j denotes the reward of the agent j as shown follows [161]:

$$r^j = \frac{1}{K^j} \sum_{k=1}^{K^j} b_k \times \log_2(1 + \gamma_k), \tag{3.22}$$

where K^j denotes the number of users of the agent j. At time step t, all agents select their action simultaneously based on their policies π. We use $\boldsymbol{\pi} \triangleq \left[\pi^1, \ldots, \pi^N\right]$ denoting the joint policy of all the agents. The Q-function of agent j is shown as follows [161, 162]:

$$Q^j(\mathcal{S}, \boldsymbol{a}) = \frac{1}{\mathcal{N}^j} \sum_{k=1}^{\mathcal{N}^j} Q^j(\mathcal{S}, a^j, a^k), \tag{3.23}$$

$$\begin{aligned} &Q^j_{t+1}(\mathcal{S}_t, a^j, a^k) = \\ &(1-\alpha) Q^j_t(\mathcal{S}_t, a^j, \bar{a}^j) + \alpha[r^j + \gamma v^j_t(\mathcal{S}_{t+1})], \end{aligned} \tag{3.24}$$

$$a^k = \bar{a}^j + \delta a^{j,k}, \tag{3.25}$$

$$\text{where } \bar{a}^j = \frac{1}{N^j} \sum_{k=1}^{N^j} a^k, a^k \sim \pi^k_t\left(\cdot \mid s, \bar{a}^k_{t-1}\right) \tag{3.26}$$

$$\begin{aligned} &v^j_t(S_{t+1}) = \\ &\sum_{a^j} \pi^j_t\left(a^j \mid \mathcal{S}_{t+1}, \bar{a}^j\right) \mathbb{E}_{a^j(\bar{a}^j) \sim \bar{\pi}^j_t}\left[Q^j_t\left(\mathcal{S}_{t+1}, a^j, \bar{a}^j\right)\right], \end{aligned} \tag{3.27}$$

$$\pi^j_t\left(a^j \mid s, \bar{a}^j\right) = \frac{\exp\left(\beta Q^j_t\left(s, a^j, \bar{a}^j\right)\right)}{\sum_{a^{j'} \in s^j} \exp\left(\beta Q^j_t\left(s, a^{j'}, \bar{a}^j\right)\right)}, \tag{3.28}$$

where Q^j is the Q value of the agent j, a^j denotes the selected action of the agent j. $\mathcal{N}^j$ denotes the set of the neighboring agents of the agent j, and $\bar{a}^j$ is the mean action of the j's neighbors. $\delta a^{j,k} = a^j - a^k$ is a small fluctuation, and $v^j(\mathcal{S})$ is the value function of the agent j at the state $\mathcal{S}$. α denotes the learning rate and β denotes the exploration rate. Usually, in the standard literature on the Mean Field Equilibrium (MFE) on stochastic games [163–165], we need to know the transition probabilities for obtaining the stochastic evolution of states in time. Differently, here we use a double DQN structure to solve the MFE. We use two networks: the main network and the target network. The main network

is used to evaluating the mean field Q value $Q(\mathcal{S}_t, a^j, \bar{a}^j)$. The target network is used to generate the target mean field value $y^j = r^j + \gamma v_t(\mathcal{S}_{t+1})$. In other words, the target network learns the transition probability for predicting the potential reward of the different actions under the next state s_{t+1}. Thus, the loss function in the agent j is:

$$\mathcal{L}(\phi^j) = \text{MSE}(y^j, Q^j(\mathcal{S}, a^j, \bar{a}^j)), \tag{3.29}$$

where ϕ^j denotes the weight of the main network. The details of the proposed mean field DQN algorithm are shown in Algorithm 5.

Algorithm 5 Mean field DQN

Input: $\mathcal{S}_t$, γ, ϵ, $\boldsymbol{a}_t$.
Output: $Q_{\phi^j}(\mathcal{S}_t, a^j, \bar{a}^j)$ for all $j \in \{1, \ldots, N\}$.

Initialize the weights of main network ϕ_M^j and target network ϕ_T^j.
Initialize experience reply memory $\mathcal{H}$.
for epoch = 0 to E **do**
    Initialize the antenna parameters.
    for step $t = 0$ to T **do**
        for $j = 1, \ldots, N$ **do**
            if $r > \epsilon$ **then**
                Randomly select an action $a_t^j \in \mathcal{A}^j$
            else
                $a^j = \arg\max_{a_t^j} Q_{\phi_M^j}(\mathcal{S}, a_t^j, \bar{a}^j)$
            end if
        end for
        The joint action is $\boldsymbol{a} = [a^1, \ldots, a^N]$.
        Perform the action a_t^j for all $j \in \{1, \ldots, N\}$ in each agent.
        Get reward r^j, $j \in [0, \ldots, N]$ by Eq. (3.5).
        The reward of joint action is $\boldsymbol{r} = [r^1, \ldots, r^N]$)
        Store $(\mathcal{S}_t, \boldsymbol{a}, \boldsymbol{r}, \mathcal{S}_{t+1})$ in $\mathcal{H}$.
        for $j = 1, \ldots, N$ **do**
            Training the ϕ_M^j to minimizing the loss function Eq. (3.29).
        end for
    end for
    Update the weights of the target network:
    $\phi_T^j \leftarrow \alpha\phi_M^j + (1 - \alpha)\phi_T^j$.
end for

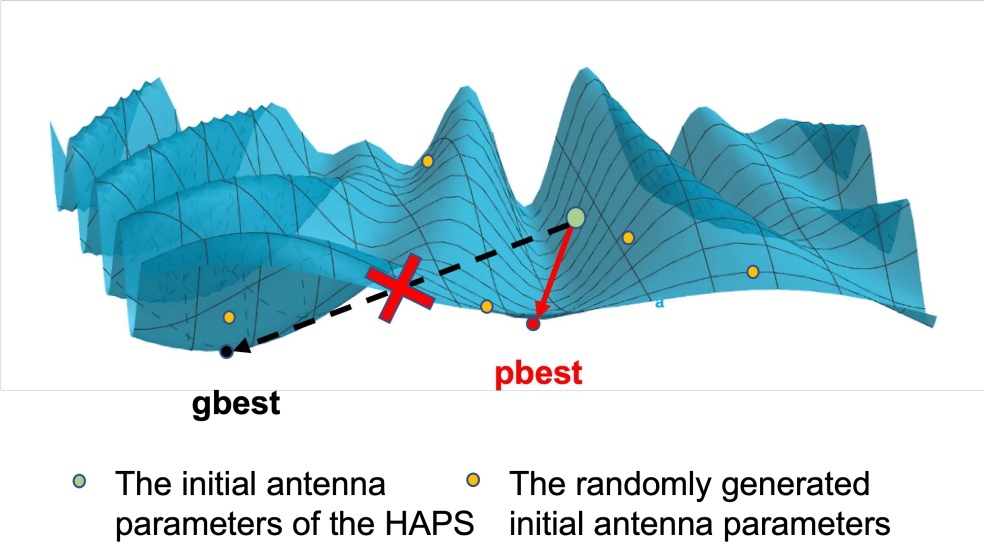


Figure 3.6. An example of how to find the optimal solution.

3.6 Proposed Deep Reinforcement Learning Evolutionary Algorithm for Antenna Beamforming

In the DQN method, the Q value of each action is calculated based on the reward returned from the environment states. During the repeating of the experimental process, DQN can learn how to adjust antenna parameters to reduce the number of low-throughput users. Same with DQN approach, we obtain the Q value by using the Equation (3.18). However, using the DQN method for the HAPS system is difficult to search for the optimal solution. We use Fig. 3.6 to explain the reason. In Fig. 3.6, the 'pbest' is one of the sub-optimal solutions, the 'gbest' is the optimal solution. The DQN agent will adjust the antenna parameters step-by-step to find the optimal solution by using gradient descent. Nonetheless, if the initial state is located in the green point shown in Fig. 3.6, the DQN agent cannot obtain the optimal solution even using the Epsilon-greedy method for action selection. When the DQN agent searches for the suboptimal solution 'pbest', subsequent randomly selected actions can only search for worse solutions. Thus, it will cause DQN to stop the search along the direction of the 'gbest'. To address this problem, we design a novel DRLEA algorithm as shown in Algorithm 6. The workflow of the proposed method is shown in Fig. 3.7. Different from the DQN method with the same initial state at the

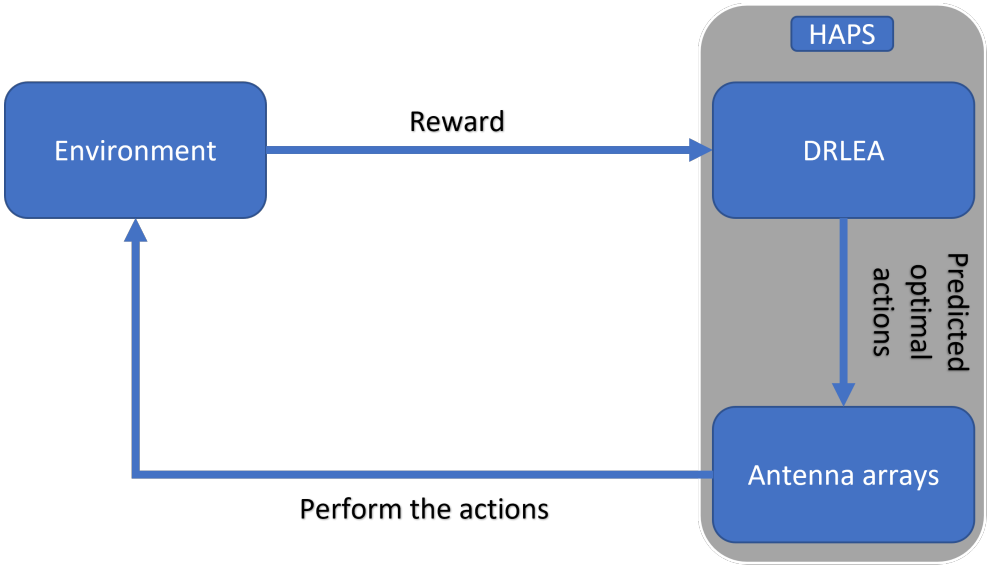


Figure 3.7. The pipeline of the proposed DRLEA method.

beginning of each training epoch, the DRLEA will randomly generate a different initial state for each training epoch, such as the yellow points shown in Fig. 3.6. For each epoch, the DRLEA performs many steps to search for the 'optimal' solution (actually is a sub-optimal solution). After a training epoch, the DRLEA compares the 'optimal' solution with the historical optimal solution and then keeps the better one. The details of the proposed DRLEA are shown in Algorithm 6.

Same with Algorithm 3, in Algorithm 6, we build two DNNs, the main network and the target network. We calculate the reward $\mathcal{R}(s_t, A)$ with the state s_t and the action A by Equation (3.5). In step t, we first select an action A to maximize the Q-value. If the reward $\mathcal{R}(s_t, A)$ is lower than 0 or the reward of the previous step, the DRLEA will re-select and perform a random action from $\mathcal{A}$. Next, we train the main network $\mathcal{F}(\Theta)$ to minimize the loss function $\mathcal{L}(\Theta)$, as shown in Equation (3.30). In Equation (3.30), with the increase in the epochs, the influence of the current optimal solution is increased. After enough training, we input the current state s_t into the main network; then, we can obtain the predicted Q value of all actions under the state s_t. Thus, the optimal action $A = \arg\max_j Q(s_t, A_j; \Theta)$ can be obtained. After E iterations, the optimal antenna parameters are recorded in $gbest$.

Algorithm 6 The Proposed DRLEA for Antenna Beamforming in HAPS

Input: s_t, γ, ϵ, actions $[a_1, \ldots, a_l]$.
Output: $Q(s_t, a_1), Q(s_t, a_2), \ldots, Q(s_t, a_J)$.

Initialize main network $\mathcal{F}(\Theta)$ and target network $\mathcal{F}(\Theta^{'})$.
Initialize experience reply memory $\mathcal{D}$.
Initialize $\mathcal{X}$ with random matrices.
$\mathcal{X} = (x_0, \ldots, x_i, \ldots, x_E)$ denotes the E different randomly generated antenna parameters.
$x_i = [\phi_{3dB}, \phi_{tilt}, \theta_{3dB}, \theta_{tilt}]$.
Initialize *pbest* and *gbest* with zero matrices.
for $i = 0$ to E **do**
    Initialize the antenna parameters.
    Initialize $s_t = x_e$, *pbest*.
    for step $t = 0$ to T **do**
        $A = \arg\max_j Q(s_t, A_j; \Theta)$
        Perform the action A.
        Get reward $\mathcal{R}(s_t, A)$ by Equation (3.5).
        if $\mathcal{R}(s_t, A) \leq \max(0, \mathcal{R}(s_{t-1}, A^{t-1}))$ **then**
            Randomly select an action $A \in \mathcal{A}$
            Perform the action A.
        end if
        if $\mathcal{R}(s_t, A) \geq \mathcal{R}(s_{t-1}, A^{t-1})$ **then**
            Update *pbest*: $pbest = pbest + A$
        end if
        Update s_{t+1}: $s_{t+1} = s_t + A$.
        Store $(s_t, A, \mathcal{R}(s_t, A), s_{t+1})$ in the $\mathcal{D}$.
        Get the output of the main network $\mathcal{F}(\Theta)$:
        $Q(s_t, A) = \mathcal{F}(s_t, A; \Theta)$.
        Generate target Q value:
        $Q^{'}(s_{t+1}, A^{t+1}) = \max \mathcal{F}(s_{t+1}; \Theta^{'})$.
        Update the main network $\mathcal{F}(\Theta)$ to minimize the loss function:

$$\mathcal{L}(\Theta) = \mathrm{MSE}\Bigg(Q(s_t, A), \mathcal{R}(s_t, A) + \frac{\alpha}{i+1} Q^{'}\left(s_{t+1}, A^{t+1}\right) + \left(\alpha - \frac{\alpha}{i+1}\right) * \left(1 - \frac{|s_t - gbest|}{|gbest|}\right)\Bigg) \tag{3.30}$$

    end for
    if $\mathcal{R}(s_t, A) \geq \mathcal{R}_{\text{best}}$ **then**
        $\mathcal{R}_{\text{best}} = \mathcal{R}(s_t, A)$
        $gbest \leftarrow pbest$
    end if
    Update the target network: $\mathcal{F}(\Theta^{'}) \leftarrow \mathcal{F}(\Theta)$.
end for

3.7 Simulation Results

3.7.1 Simulation Setting

To evaluate our proposed methods, we generate the four different non-uniform UEs distribution datasets obtained by [166]. The four different user distributions are Tokyo, Osaka, Sendai, and Nagoya, as shown in Fig. 3.8.

We assume that the HAPS works at a height of 20 km. Each HAPS can cover an area within a 20 km radius. The transmit power and bandwidth of each antenna array are 43 dBm and 20 Mhz, respectively. We consider that the transmission frequency is 2 GHz. Considering the interference between HAPSs, we set 18 HAPSs to surround 1 HAPS. We set the updated value of horizontal tilt, horizontal HPBW, vertical tilt, and vertical HPBW to 20 deg, 4 deg, 10 deg, and 4 deg, respectively. We use python to implement all simulation programs. We define the median value after users' throughput drops due to HAPS rotation or movement as the low throughput threshold.

We set 100 epochs and 100 steps in each epoch, in the four RL-based approaches (Q-learning, DQN, MFDQN, and DRLEA). In the DQN, MFDQN, and DRLEA, we set the learning rate of the neural network as 0.0001. The discount factor α is 0.65 and the learning rate β for Q value calculation is 0.75 in the three RL-based approaches. The α and β have been chosen empirically. It is true that for each method there is an optimal α that can make the method easier to find the optimal solution and faster to converge to the optimal solution. However, in this dissertation, we do not focus on finding the optimal α of different methods. We evaluate the performance of different methods under the same value of α. In the PSO approach, the number of iterations is 100, the number of particles is 100, and the $\omega_1 = 0.5$ and $\omega_2 = 0.5$.

3.7.2 CDF of UE Throughput Performance

Figure 3.9 shows the CDF of the users' throughput under the rotation scenario in the user distribution cases of Tokyo, Osaka, Sendai, and Nagoya, respectively. In all the user distribution cases, the proposed MFDQN and DRLEA reduce the number of low-throughput

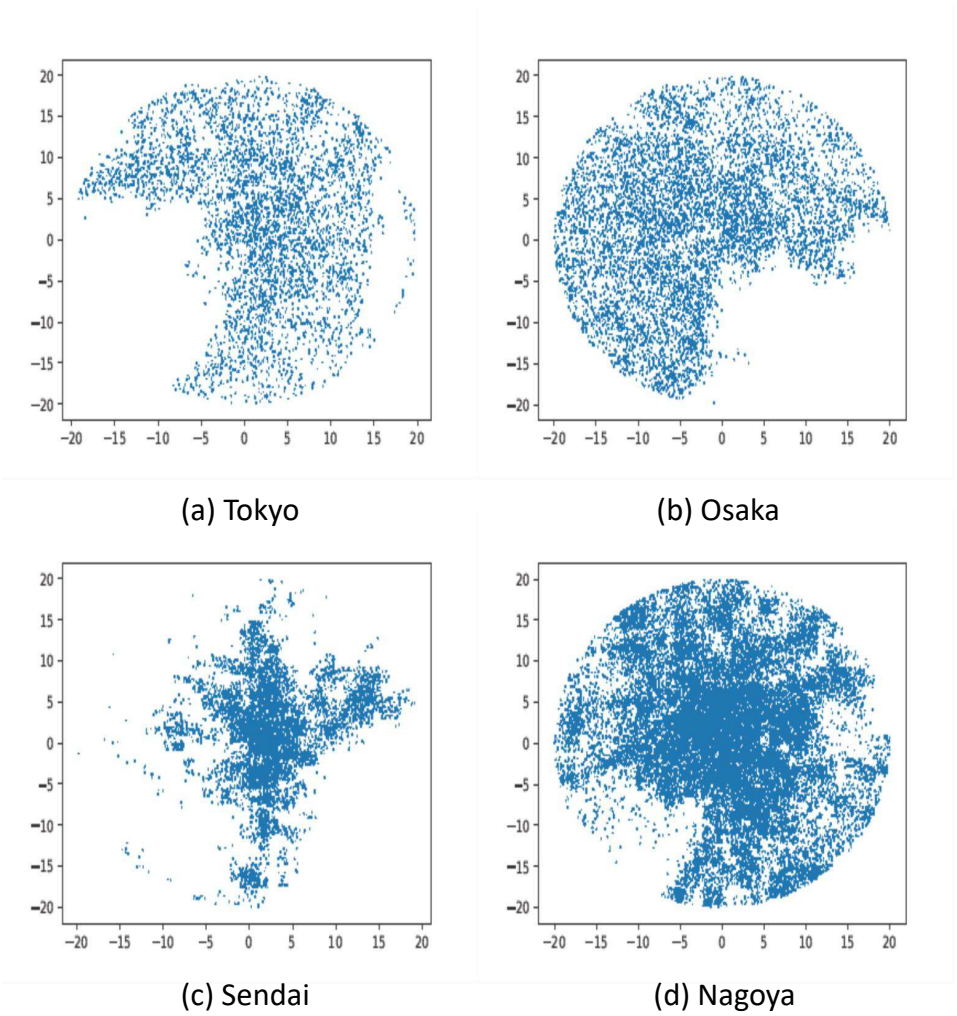


Figure 3.8. The four user distribution scenarios: (**a**) Tokyo, (**b**) Osaka, (**c**) Sendai, and (**d**) Nagoya.

users in the throughput range of [0.0, 16.0] kbps. To easily compare the performance of different algorithms, we use Tab. 3.2 to record the percentage of low throughput users after using different algorithms. In Tab. 3.2, we can find that the proposed DRLEA achieves the lowest number of low-throughput users under all the user distribution cases. The proposed MFDQN achieves the second-lowest number of low-throughput users in the cases of Tokyo, Sendai, and Nagoya. In the case of Osaka, the number of low-throughput users of the proposed MFDQN is slightly worse than the Q-learning.

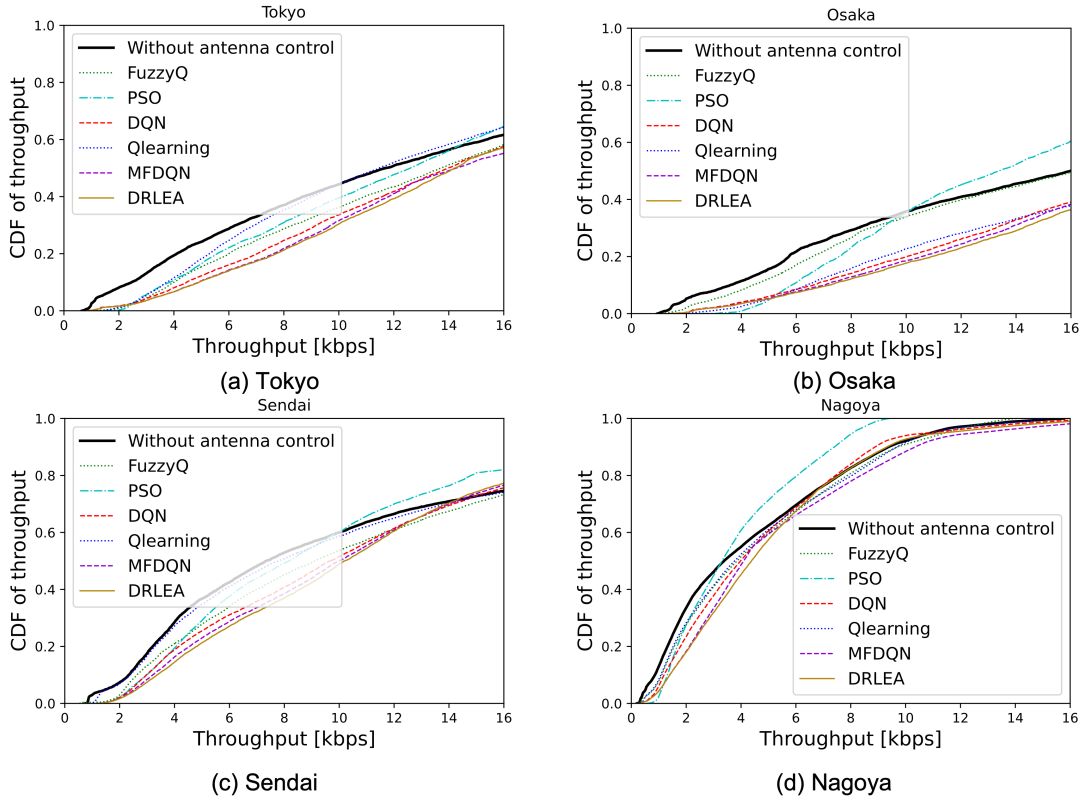


Figure 3.9. CDF of UE throughput performance under the HAPS with rotation of 30 degrees.

Moreover, we evaluate the proposed method under the shifting scenario. Here, we use the DQN, MFDQN, and DRLEA methods, which are trained under the rotation scenario, and the other three methods are trained under the shifting scenario. As shown in Fig. 3.10, we can find that the DQN and DRLEA without retraining achieve comparable throughput performance to that of Q-learning and Fuzzy Q-learning with training in the HAPS with a left shift of 5 km scenario. Since throughput is closely related to the user distribution,

TABLE 3.2
THE PERCENTAGE OF LOW-THROUGHPUT USERS AFTER USING DIFFERENT ALGORITHMS UNDER HAPS WITH ROTATION OF 30 DEGREES

	Tokyo	Osaka	Sendai	Nagoya
Q-learning	50.8 %	37.7 %	48.0 %	46.3 %
Fuzzy Q-learning	42.5 %	49.4 %	41.8 %	44.2 %
DQN	40.6 %	39.1 %	37.3 %	47.3 %
PSO	46.4 %	60.4 %	45.7 %	48.1 %
MFDQN	39.4 %	38.3 %	35.2 %	48.6 %
DRLEA	37.9 %	36.4 %	33.7 %	41.9 %

DNNs trained under the same user distribution know how to adjust the antenna parameters to reduce the number of low throughput users, even if HAPS moves again.

In Tab. 3.3, we can find that the proposed DRLEA achieves the lowest number of low-throughput users in cases of Tokyo and Nagoya. The MFDQN method achieves the fourth-lowest, fifth-lowest, fifth-lowest, and third-lowest number of low throughput users for four different user distributions, respectively

TABLE 3.3
THE PERCENTAGE OF LOW-THROUGHPUT USERS AFTER USING DIFFERENT ALGORITHMS UNDER HAPS WITH LEFT SHIFT OF 5 KM

	Tokyo	Osaka	Sendai	Nagoya
Q-learning	45.2 %	59.6 %	47.7 %	45.9 %
Fuzzy Q-learning	47.6 %	39.4 %	54.0 %	45.1 %
DQN	44.7 %	43.3 %	55.6 %	43.0 %
PSO	59.6 %	43.2 %	64.2 %	51.3 %
MFDQN	46.4 %	44.4 %	59.7 %	39.0 %
DRLEA	40.9 %	42.4 %	49.6 %	36.9 %

3.7.3 Convergence Analysis

Here we compare the convergence performance of the Q-learning method, the Fuzzy Q-learning method, the proposed MFDQN method, and the proposed DRLEA method. We

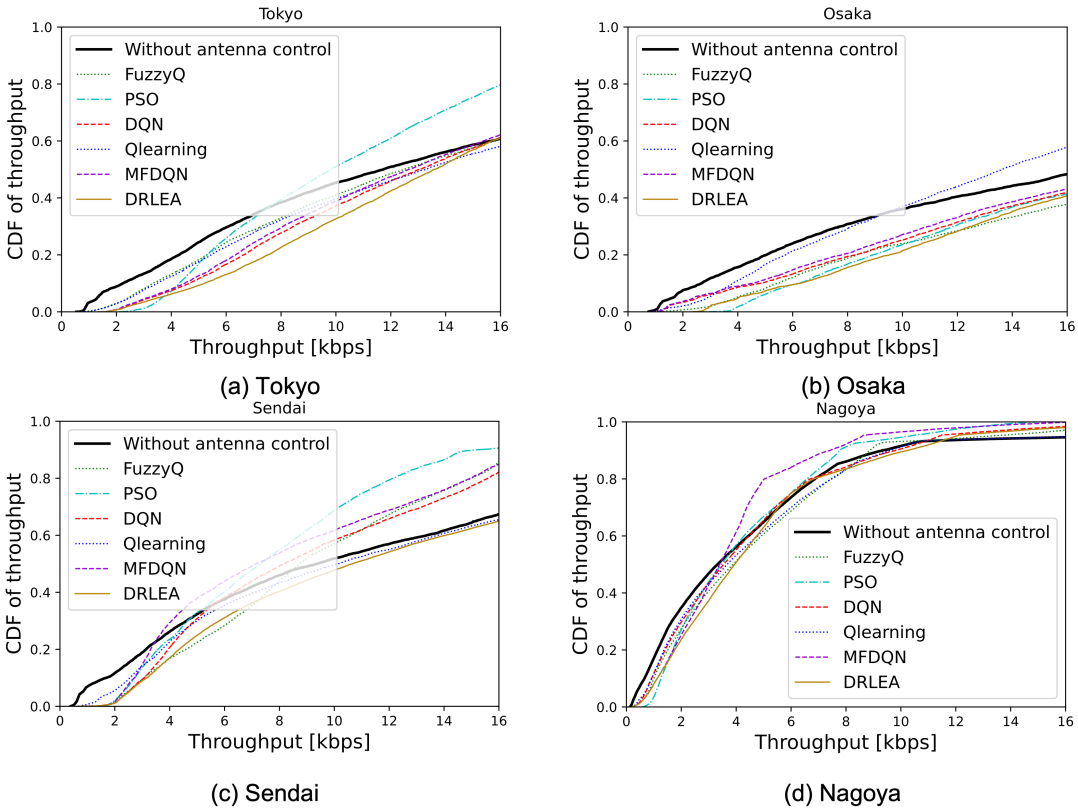


Figure 3.10. CDF of UE throughput performance under the HAPS with a left shift of 5 km.

use the KPI [68] mentioned in Sec. 3.4.2 to show the convergence performance of each method. The KPI represents the percentage of users u whose received signal power $\mathcal{T}_i(u)$ is greater than or equal to the received signal power threshold τ among the set of users U_i located in the coverage as shown in Eq. (3.15). In Fig. 3.11, the Fuzzy Q-learning method has the highest convergence rate. However, the KPI performance of the Fuzzy Q-learning method is worse than that of the other methods. The proposed MFDQN method achieves a higher KPI performance and higher convergence rate than the Q-learning and DQN methods. The proposed DRLEA method has the worst convergence performance but the best KPI performance.

3.7.4 Discussion

In Sec. 3.7.2, we show the CDF performance under both HAPS with rotation of 30 degrees and left shift of 5 km scenarios. In the rotation scenario, the proposed DRLEA

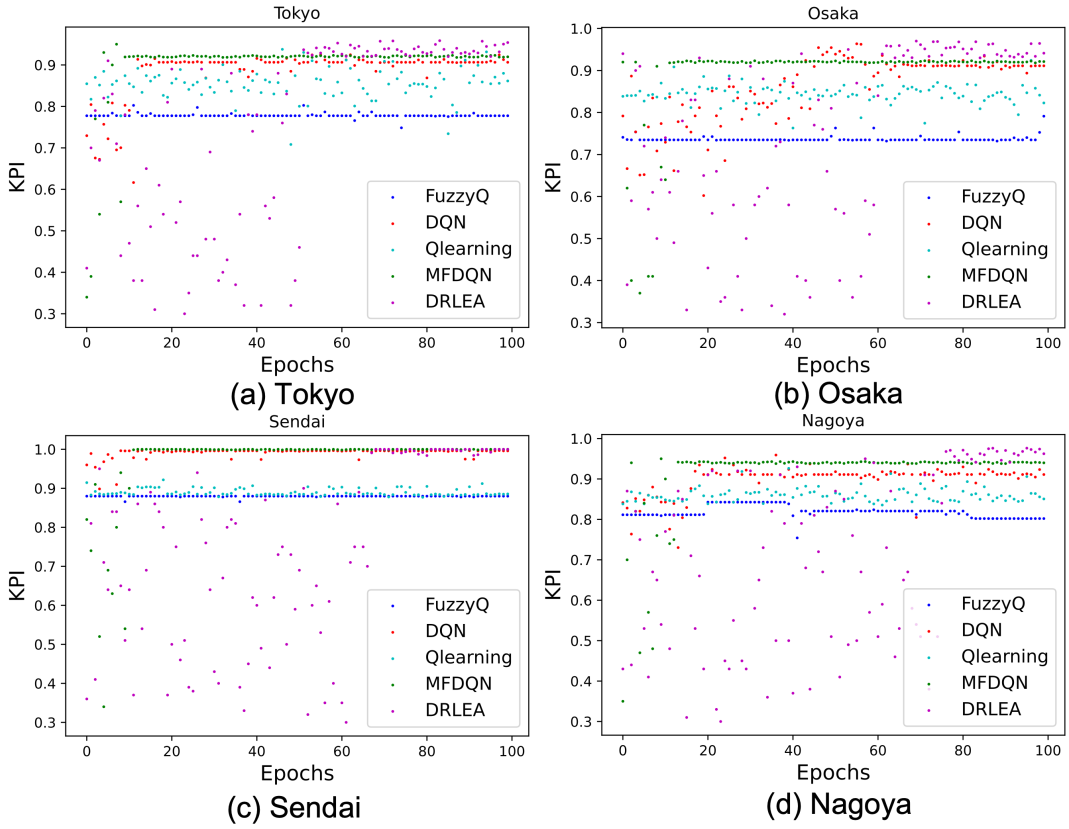


Figure 3.11. Convergence performance of HAPS rotation scenarios for four different user distribution scenarios.

achieves the best CDF performance under all the user distribution cases and the proposed MFDQN achieves the second-best CDF performance (except the Osaka case). That is because, in the proposed DRLEA, we search the wide range of candidate solutions by randomly initializing the antenna parameters at the beginning of each training epoch. We design a novel loss function that includes not only Q-value of the predicted optimal action, but also the historical optimal solutions obtained from previous training. However, other RL-based methods (Q-learning, Fuzzy Q-learning, DQN, and MFDQN), which only use the simple ϵ-greedy method for searching the optimal solution, are difficult to escape the local optimal. When these methods fall into a suboptimal solution, the randomly selected actions can only get a worse reward. It makes the RL agent stop to search for the optimal solution toward the optimal solution. The PSO algorithm also randomly generates many particles to search for the optimal solution, however, the performance is worse

than the proposed methods. The high-dimensional search space is the main problem. In the simulation, we set the number of antenna arrays in each HAPS as 3, thus, the search space is 12-dimensional space. in Fig. 3.9, we can find that in the $[0.0, 2.0]$ kbps throughput range, the PSO method has comparable or even better CDF of throughput performance than the other methods. In particular, in Fig. 4.5 (b), PSO achieves the best CDF performance in the $[0.0, 4.0]$ kbps throughput range. However, due to the high-dimensional space, the number of particles we set in our experiments is not enough for PSO to search for the optimal solution. It is still easy to fall into the local optimum. In the proposed MFDQN, we decompose the high-dimensional search space into several lower-dimensional spaces to search for the optimal solution separately. Thus, it achieves better CDF of throughput performance and a faster convergence rate than the conventional methods.

In the shifting scenario, the DQN method, the proposed MFDQN method, and the proposed DRLEA method still can reduce the number of low-throughput users without retraining under the same user distribution case. in Fig. 3.3, we show the relationship between antenna parameters and footprint. In other words, we use the footprint as state and train the DNN to find the optimal footprint over the user distribution. Thus, under the same user distribution, DRL-based methods still have the ability to reduce the number of low-throughput users without retraining.

3.8 Conclusion

In this Chapter, we addressed the problem of the reduction of the number of users with low throughput caused by the movement of HAPS. To do so, we first present the antenna control problem as a Markov Decision Process (MDP) problem. To tackle this MDP problem, we use three conventional RL-based approaches. Besides, we use an Evolutionary Algorithm named PSO to find the optimal antenna parameters for reducing the number of low-throughput users. However, due to the huge search space and different user distribution scenarios, these approaches are difficult to obtain the optimal solutions. Motivated by this, we develop two novel DRL-based methods: MFDQN and DRLEA to

search for the optimal solution as much as possible in the limited number of training. We do the simulation under four different user distribution scenarios and consider both HAPS rotation and HAPS shift cases. The simulation results show that the proposed MFDQN and DRLEA have better CDF of users' throughput than that of conventional RL-based approaches and PSO algorithm under all the user distribution scenarios in the HAPS rotation case. In the HAPS shift case, the proposed methods without retraining achieve a CDF of throughput performance comparable to that of the conventional methods. Besides, we compare the convergence between MFDQN and the conventional RL-based methods. The MFDQN shows that it reduces the complexity of the interactions among agents and converges faster than the conventional RL-based methods.

Chapter 4

Clustering-Based Methods for Antenna Control in HAPS with Users' Location Information

4.1 Problem Re-formulation

In Sec. 3.3, to reduce the number of users whose with low throughput, we formulate the antenna control problem into maximizing the throughput of users. For easier reading, we show the formula of throughput below:

$$\mathcal{T} = b \times \log_2(1 + \gamma), \tag{4.1}$$

As obvious from Equation (4.1) and the x-z plane in Fig. 4.1, we can know that the rising of the throughput with the increase in the SINR is very slow when the bandwidth is very small. Thus, considering making users' bandwidth as large as possible is necessary, particularly in the non-uniform user distribution scenario. Thus, in this chapter, we re-formulate the antenna control problem to two sup-problems: maximizing the SINR of users and maximizing the bandwidth of users. To maximize the SINR of users, one idea is to cluster users into several clusters with high density and then control the beams toward and cover each cluster respectively. To maximize the bandwidth of users, we can cluster users into several clusters with the same number of users. Thus, in this chapter, we design an Equal Clustering (EC)-aided approach to address the antenna control problem.

4.2 Contributions

In this chapter, we design a novel clustering method to cluster the users into several clusters with the same number of users. Moreover, we implement a conventional clustering method, the K-Means clustering algorithm, as benchmarks to evaluate the proposed method and show that the proposed method is still reliable and efficient. This Chapter's contributions can be outlined as follows:

- We propose a novel clustering-aided DQN that addresses the problem of dynamic control of the HAPS antenna parameters to decrease the number of users with low throughput.

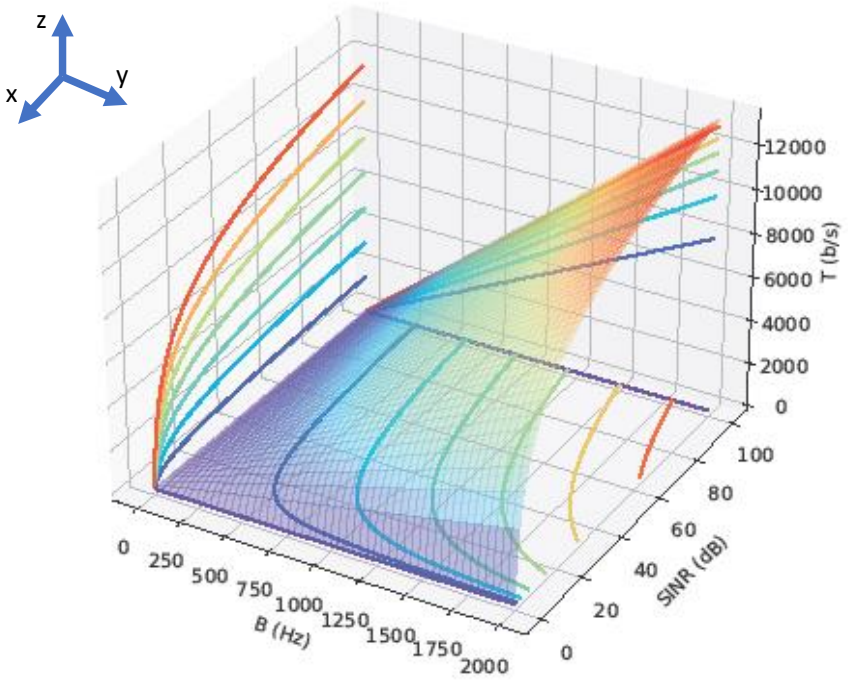


Figure 4.1. Throughput.

- We design a clustering method that clusters users into several clusters of the same size and high density with low computational complexity.

- The conventional DRL approaches need to collect lots of datasets for training. However, in practice, collecting the SINR of users needs huge time and power consumption. To address this problem, we propose an EC-aided DQN approach that reduces the number of training datasets required.

- Under the user movement scenario, the proposed method shows a better Cumulative Distribution Function (CDF) of throughput performance than other methods.

4.3 Conventional Clustering-Based Approach for Antenna Beamforming

K-means algorithm is a typical clustering algorithm widely used for clustering a large set of data [167]. The algorithm consists of two separate steps. In the first step, we need to select k centers of clusters randomly. For a more concise explanation of this algorithm, here we assume the number of antenna arrays $N = 3$ on each HAPS. Thus, we need to divide users into 3 clusters. In the second step, we take each data object to the nearest center according to the Euclidean distance between each data object and the cluster centers. The detail of the K-Means algorithm is shown in Algorithm 7:

Algorithm 7 K-Means UE clustering algorithm

Input: k, U
Output: $C_1, C_2, \ldots, C_k$

repeat
  Randomly select k UEs from U as cluster center.
  for $i = 0$ to n **do**
    for $j = 0$ to k **do**
      $l_{i,j} = d(u_i, c_j)$
    end for
    $m = \arg\min(l_{i,1}, \ldots, l_{i,k})$
    Assign u_i to c_m
  end for
  for $j = 0$ to k **do**

$$c_j^{'} = \text{minimize} \frac{1}{n_j} \sum_{i=1}^{n_j} \|u_i - c_j\|^2, \quad \text{subject to } u_i \in U_j \tag{4.2}$$

  end for
until $c_j^{'} = c_j, j \in (1, \ldots, k)$

Here we define the U to represent the UEs' locations and n denotes the number UEs in a HAPS coverage area. $C_j = \{u_1, \cdots, u_n\}, j \in (1, \ldots, k)$ denotes the set of UEs which belongs to the cluster j. $c_j, j \in (1, \ldots, k)$ denotes the central of C_j. $l_{i,j}$ denotes the Euclidean distance between the UE i and the center c_j. For each iteration, we calculate the $l_{i,j}$ between the i-th UE u_i and the j-th central of the cluster c_j. Next, for UE u_i, we can get the closest cluster center c_m, where m is the index of the minimum of the Euclidean

distance set between the UE u_i and all clusters' center. Thus, the UE u_i belongs to the cluster c_m. Finally, after we assign all UEs to k clusters, we recalculate the center of each cluster by Eq. (4.2) and repeat this flow again until no change in the center of clusters. According to the UE locations, we first redesign the cell configuration and divide UEs into three new cells by the K-Means clustering method. We denote the three new cells as $C = \{c_1, c_2, c_3\}$. For each cell, we can obtain the radius r_j as follows:

$$r_{i.j} = \max d(u_i, c_j), u_i \in U_i \tag{4.3}$$

where, u_i denotes the u_i-th UE location in the cell j, U_i denotes the set of UEs in the cell j. $d(u_i, c_j)$ denotes the distance between the u_i-th UE and the center of the cell j. We select the maximum distance as the radius of the cell. According to the beam coverage geometry model in [62], we can obtain the horizontal tilt, vertical tilt, horizontal beam HPBW, and vertical beam HPBW after we obtain the radius and center of each cell to make beams toward and cover each cluster respectively.

4.4 The Proposed Equal Clustering-aided Deep Q-Learning for Antenna Beamforming

As known in Sec. 4.1, improving the bandwidth of each user is a very important thing for reducing the number of low throughput users. Considering the limitation and same bandwidth of each antenna array, thus, making each antenna array server the same number of users can maximize the bandwidth of each user. Thus, to improve the throughput of users, we design a novel clustering method for maximizing the bandwidth of users. Different from the K-means algorithm variation with equal cluster size which has high computational complexity for iteratively searching the centroid of each cluster, the proposed clustering method only needs to search the sparsest area to locate a boundary of a sector and then divide users into several clusters by the same number of users. We use an

example of (0, 5) 1-layer cell configuration as shown in Fig. 4.2 to explain the proposed clustering algorithm. The details of the proposed clustering algorithm are in Algorithm 8

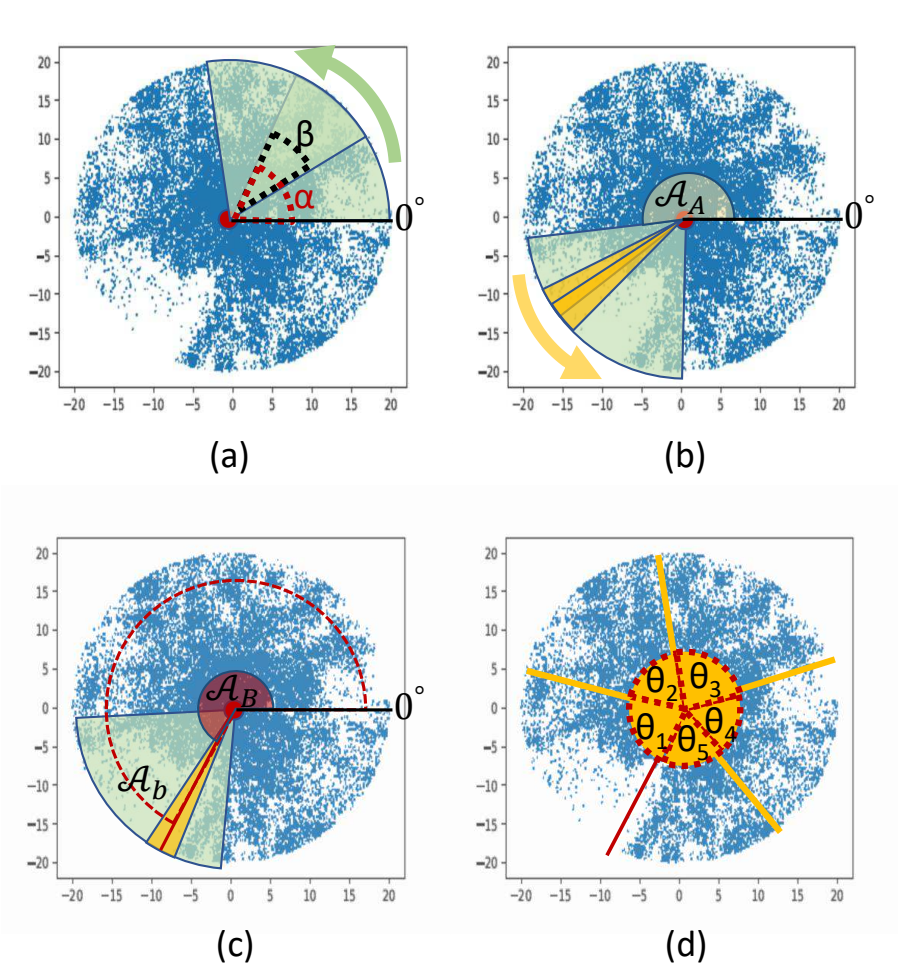


Figure 4.2. The flow of the proposed equal clustering algorithm.

Here, ρ and ρ_{min} denote the density of a sector and the minimum density, respectively. Angle_A and Angle_B denote the angle of a side of sector A and sector B, respectively. k_A and k_B denote the number of users in sector A and sector B, respectively. S_A and S_B denote the area of sector A and sector B, respectively. Angle_b denotes the angle of a boundary of a cluster located in the sparsest area. Fig. 4.3 shows an example of the extension of the proposed algorithm in a 2-layer cell configuration. In the 2-layer cell configuration, the difference is that we need to search the boundary located in the sparsest area for each layer.

Algorithm 8 The proposed Equal clustering algorithm

Input: $\alpha, \beta, \alpha^{'}, \beta^{'}, K, N$.
Output: $\text{Angle}_b, \theta_1, \ldots, \theta_N$.

Generating a sector A with the angle of α rotate around the HAPS.
Initializing ρ_{min} with a large number, $\text{Angle}_A = 0$.
Searching the sparsest area of the user distribution (Fig. 4.2(a)).
for $a = 0°$ to $360°$, angle step = β **do**
    Calculating the density of sector A by Eq. (4.4).

$$\rho = \frac{k_A}{S_A} \tag{4.4}$$

    if $\rho < \rho_{min}$ **then**
        $\rho_{min} = \rho$
        $\text{Angle}_A = a$
    end if
end for
Generating a sector B with as much smaller angle $\alpha^{'}$ rotate around the HAPS with a much smaller angle $\beta^{'}$.
Initializing ρ_{min} with a large number, $\text{Angle}_B = 0$.
Searching the sparsest area of sector A (Fig. 4.2(b)).
for $a = \text{Angle}_A$ to $\text{Angle}_A + \beta$, angle step = $\beta^{'}$ **do**
    Calculating the density of sector B by Eq. (4.5).

$$\rho = \frac{k_B}{S_B} \tag{4.5}$$

    if $\rho < \rho_{min}$ **then**
        $\rho_{min} = \rho$
        $\text{Angle}_B = a$
    end if
end for
$\text{Angle}_b = \text{Angle}_B + \frac{\alpha^{'}}{2}$ (Fig. 4.2(c)).
Adjusting $[\theta_1, \ldots, \theta_N]$ so that the number of users in each sector equals $\frac{K}{N}$ (Fig. 4.2(d)).

After clustering, we can get the new cell configuration. Based on the Geometric coverage model, as shown in Fig. 4.4, we can design the antenna parameters by equations in the following:

$$\theta_{3dB} = \arctan\left(\frac{g+r}{h}\right) - \arctan\left(\frac{g-r}{h}\right), \tag{4.6}$$

$$\phi_{3dB} = 2\arctan\left(\frac{r}{\sqrt{h^2+g^2}}\right), \tag{4.7}$$

$$\Delta\theta_{tilt} = \arctan\left(\frac{g_B}{h}\right) - \arctan\left(\frac{g_A}{h}\right), \tag{4.8}$$

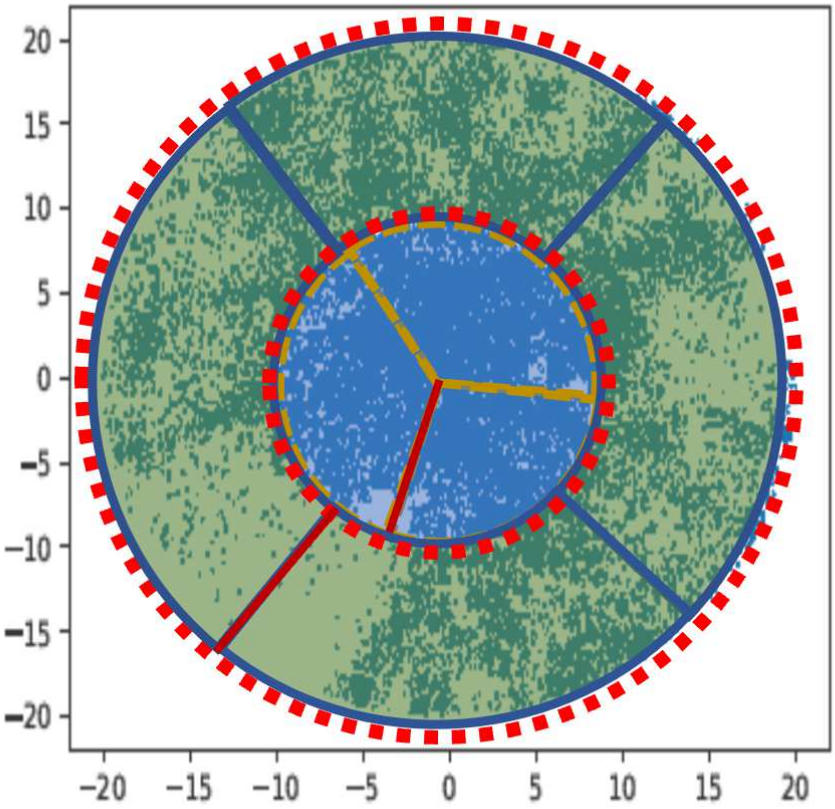


Figure 4.3. An example of the proposed clustering method in 2-layer cell config.

$$\Delta\phi_{tilt} = \arctan\left(\frac{x_b - x_a}{y_b - y_a}\right), \tag{4.9}$$

where, g_A and g_B denote the horizontal distance between the HAPS and the center of the old cell and the center of the new cell (as shown in Fig. 4.4 (b)), respectively.

After the antenna parameters design, we roughly get the sub-optimal solution. Next, we use the DQN approach which is mentioned in Sec. 3.4.3 to fine-turn the antenna parameters obtained before.

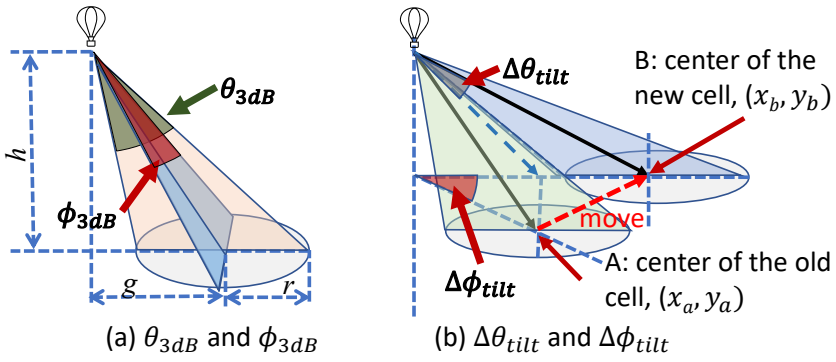


Figure 4.4. Geometric coverage model.

4.5 Simulation Results

4.5.1 CDF of UE Throughput Performance

Figure 4.5 shows the CDF of the users' throughput under the rotation scenario in the user distribution cases of Tokyo, Osaka, Sendai, and Nagoya, respectively. In all the user distribution cases, the K-Means clustering method the proposed EC methods and the proposed EC-aided DQN method reduce the number of low-throughput users in the throughput range of $[0.0, 16.0]$ kbps. To easily compare the performance of different algorithms, we use Tab. 4.1 to record the percentage of low throughput users after using different algorithms. In Tab. 4.1, we can find that the proposed EC-aided DQN achieves the lowest number of low-throughput users under all the user distribution cases. The proposed EC without DQN achieves a number of low-throughput users' performances comparable to that of the K-Means clusters methods.

Moreover, we evaluate the proposed method under the shifting scenario. Here, we use the MFDQN and DRLEA methods, which are trained under the rotation scenario, and the DQN part of the EC-aided DQN method is also trained under the rotation scenario. As shown in Fig. 4.6, we can find that the proposed EC and EC-aided DQN without retraining achieve a throughput performance comparable to that of K-Means clustering method in the HAPS with a left shift of 5 km scenario.

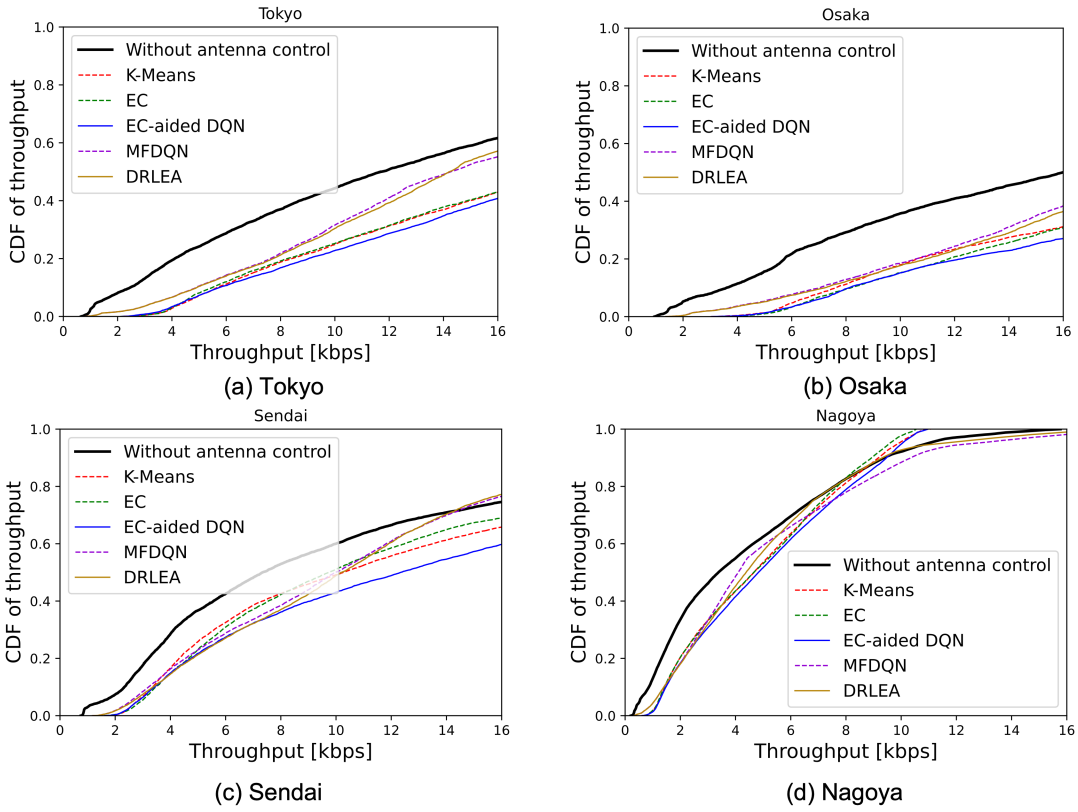


Figure 4.5. CDF of UE throughput performance under the HAPS HAPS with rotation of 30 degrees.

In Tab. 3.3, we can find that the proposed EC-aided DQN achieves the lowest number of low-throughput users under the user distribution cases of Osaka and Sendai. In the case of Tokyo, the proposed EC-aided DQN is slightly worse than the K-Means clustering method. In the case of Nagoya, the proposed EC-aided DQN is slightly worse than the EC without DQN.

In addition, we also evaluate the proposed method with user movement under both rotation and shift scenarios. We show the CDF of throughput performance under the rotation and shift scenario in Fig. 4.7 and Figure 4.8, respectively. We show the percentage of number of low-throughput users after antenna control by using different methods in Tab. 4.3 and Tab. 4.4, respectively. We can find that because the user distribution is changed, the MFDQN and DRLEA methods cannot reduce the number of low-throughput users, but clustering-based methods still can work.

TABLE 4.1
THE PERCENTAGE OF LOW-THROUGHPUT USERS AFTER USING DIFFERENT ALGORITHMS UNDER HAPS WITH ROTATION OF 30 DEGREES

	Tokyo	Osaka	Sendai	Nagoya
MFDQN	39.4 %	38.3 %	38.3 %	39.0 %
DRLEA	37.9 %	36.4 %	37.1 %	36.9 %
K-Means	30.3 %	31.2 %	40.0 %	37.3 %
EC	30.4 %	30.8 %	38.8 %	37.3 %
EC-aided DQN	27.6 %	27.1 %	33.5 %	34.8 %

TABLE 4.2
THE PERCENTAGE OF LOW-THROUGHPUT USERS AFTER USING DIFFERENT ALGORITHMS UNDER HAPS WITH LEFT SHIFT OF 5 KM

	Tokyo	Osaka	Sendai	Nagoya
MFDQN	46.4 %	44.3 %	59.7 %	48.6 %
DRLEA	40.9 %	42.4 %	45.6 %	41.9 %
K-Means	30.8 %	40.9 %	48.9 %	42.4 %
EC	32.1 %	38.4 %	44.6 %	40.9 %
EC-aided DQN	30.9 %	30.9 %	38.7 %	41.2 %

4.5.2 Discussion

In Sec. 4.1, we mentioned that the rising of the throughput with the increase in the SINR is very slow when the bandwidth is very small. Thus, considering making users' bandwidth as large as possible is necessary. If a cluster is not high-density, the SINR of the users in this cluster will be not high. If a cluster has a larger number of users than other clusters, the bandwidth per user in this cluster will be very small. In the proposed EC method, we cluster users into several clusters with both high density and roughly the same number of users. The accuracy of the EC deeply affects the bandwidth and SINR of each user. That is why the proposed EC method only is slightly worse than the K-Means clustering method in the cases of Tokyo and Nagoya, and is better than the K-Means clustering method in the cases of Osaka and Sendai. Besides, we use DQN to fine-tune the antenna parameters after the EC method to find the optimal solution. Thus, in all the cases (both

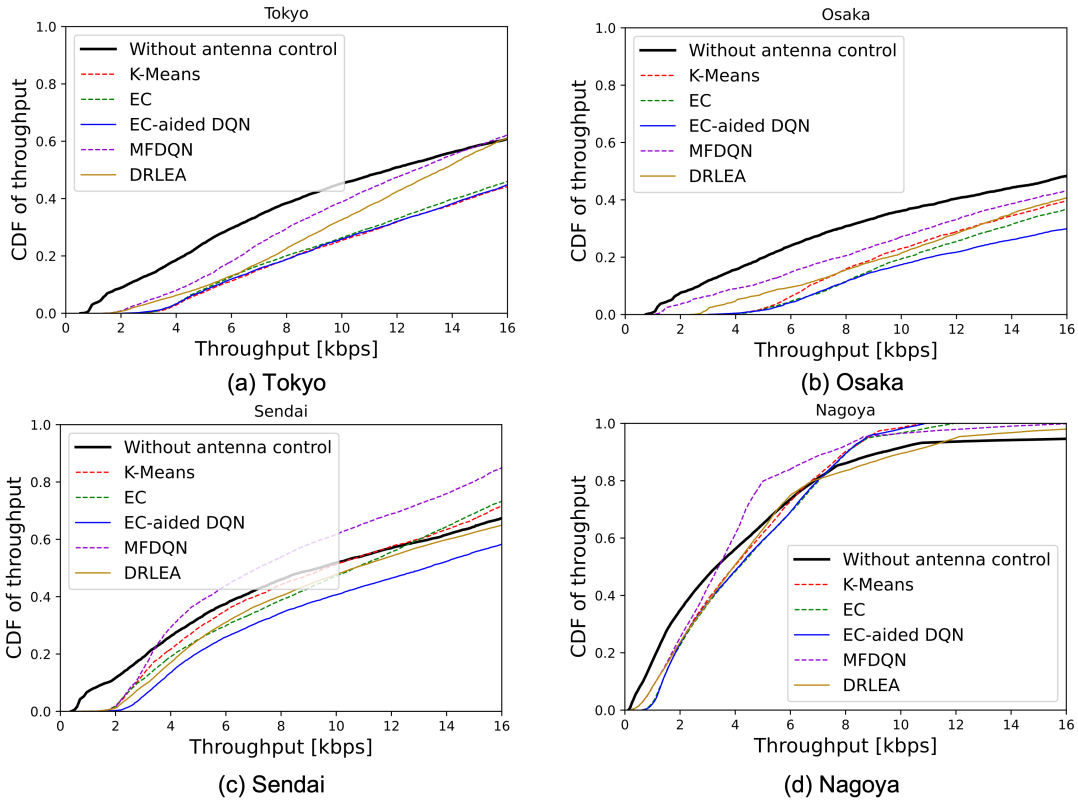


Figure 4.6. CDF of UE throughput performance under the HAPS with left shift of 5 km.

rotation and shifting scenarios), the EC-aided DQN can achieve the best CDF of throughput performance. For the proposed DRLEA and MFDQN, the main reason of the worse CDF of throughput performance of them is the change in user distribution. When the distribution of users changes substantially, DRLEA and MFDQN need to be retrained under the new distribution to learn how to reduce the number of low throughput users under the new distribution.

4.6 Conclusion

In this chapter, we study user location information for assisting in tackling the antenna control problem. Our idea is to first cluster users into several high-density clusters according to the number of antenna arrays. Next, we design antenna parameters based on the geometric model to control beams toward the center of each cluster and cover each

TABLE 4.3
THE PERCENTAGE OF LOW-THROUGHPUT USERS AFTER USING DIFFERENT ALGORITHMS UNDER HAPS WITH ROTATION OF 30 DEGREES IN THE USER MOVEMENT SCENARIO

	Osaka	Sendai	Nagoya
MFDQN	67.6 %	65.2 %	58.3 %
DRLEA	51.6 %	52.7 %	51.6 %
K-Means	31.2 %	40.0 %	37.3 %
EC	30.8 %	38.8 %	37.3 %
EC-aided DQN	27.1 %	33.5 %	34.8 %

TABLE 4.4
THE PERCENTAGE OF LOW-THROUGHPUT USERS AFTER USING DIFFERENT ALGORITHMS UNDER HAPS WITH LEFT SHIFT OF 5 KM IN USER MOVEMENT SCENARIO

	Osaka	Sendai	Nagoya
MFDQN	49.5 %	71.6 %	58.0 %
DRLEA	67.2 %	59.3 %	55.3 %
K-Means	40.9 %	48.9 %	42.4 %
EC	38.4 %	44.6 %	40.9 %
EC-aided DQN	30.9 %	38.7 %	41.2 %

cluster separately. We use a classical K-Means clustering method and design an interesting EC method for clustering users. Moreover, to get a better throughput performance, we use DQN to fine-tune antenna parameters after designing the antenna parameters according to the clustering result of using EC. Under the same simulation settings in Chapter 3, the simulation results show that K-Means, EC, and EC-aided DQN are effective in improving the CDF performance of throughput over the RL-based approach in HAPS with both rotation and shift scenarios. Besides, we evaluate the clustering-based methods in a user movement scenario, we show that the clustering-based approaches can still maintain a high CDF of throughput performance, while the RL-based approach cannot.

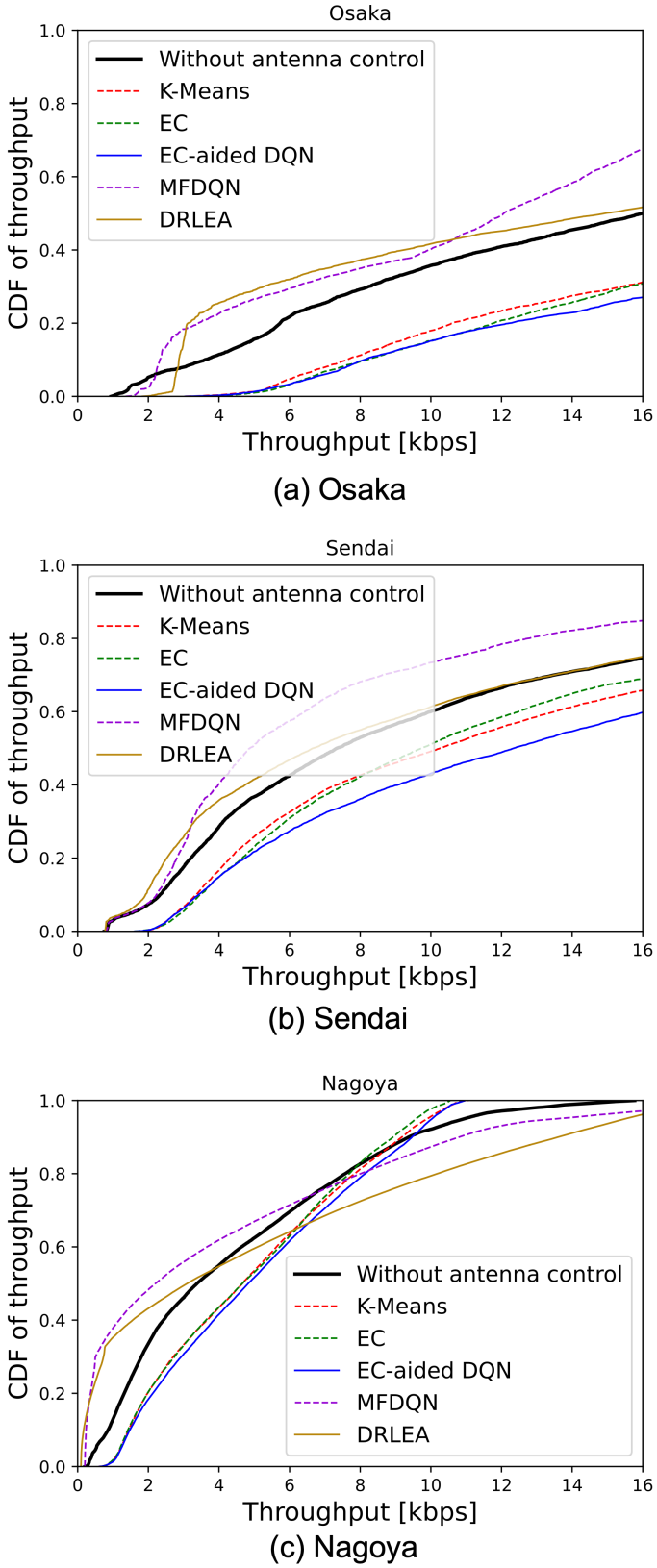
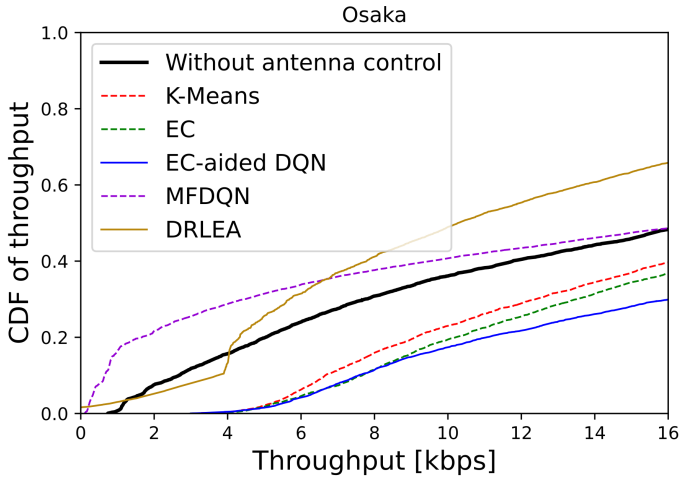
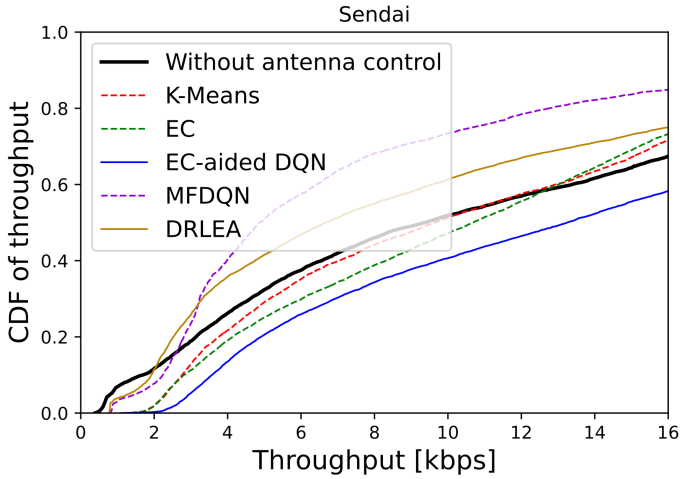


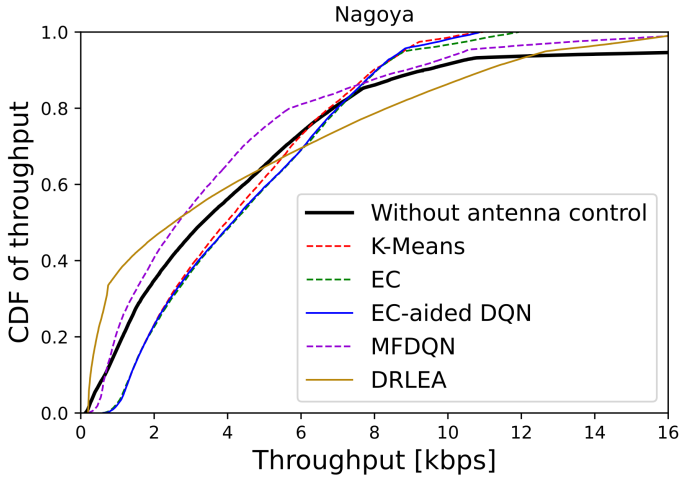
Figure 4.7. CDF of UE throughput performance under the HAPS with rotation of 30 degrees in the user movement scenario.



(a) Osaka



(b) Sendai



(c) Nagoya

Figure 4.8. CDF of UE throughput performance under the HAPS with left shift of 5 km in user movement scenario

Chapter 5

Conclusions and Future Work

In this paper, we use DL- and DRL-based algorithms to address two key techniques in mmWave mMIMO systems to achieve high-quality communication services. We first propose a DL-based fingerprint and super-resolution approaches to improve the accuracy of IUD estimation. Besides, we propose three novel DRL-based methods to reduce the low-throughput users in the HAPS system.

5.1 Contributions

Chapter 2 tackles a challenging localization task that is very important to enhance the quality of communication and communication services. We use DL and propose a novel fingerprint to address this problem. This novel fingerprint is easily obtained by beam sweeping and includes rich angular information of both two users. Nonetheless, we investigate the possibility of using a super-resolution network to reduce the involved beam sweeping overhead. We evaluate the proposed DNN-based IUD estimation method by using original beam images of resolution 4×4, 8×8, and 16×16. Simulation results show that our method can achieve an average distance estimation error equal to 0.13 m for a coverage area of $60 \times 30\ m^2$. Moreover, our method outperforms the state-of-the-art IUD estimation methods that rely on each user's location information.

Chapters 3 and 4 are focuses on the same task: antenna beamforming in the HAPS communication system to reduce the number of low-throughput users. In Chapter 3, we first present the antenna beamforming problem as a Markov Decision Process (MDP) problem. To tackle this MDP problem, we use three conventional RL-based approaches. Besides, we use an Evolutionary Algorithm named PSO to find the optimal antenna parameters for reducing the number of low-throughput users. However, due to the huge search space and different user distribution scenarios, these approaches are difficult to obtain the optimal solutions. Motivated by this, we develop two novel DRL-based methods: MFDQN and DRLEA to search for the optimal solution as much as possible in the limited number of training. We do the simulation under four different user distribution scenarios and consider both HAPS rotation and HAPS shift cases. The simulation results show that the proposed MFDQN and DRLEA have better CDF of users' throughput than

that of conventional RL-based approaches and PSO algorithm under all the user distribution scenarios in HAPS rotation case. In the HAPS shift case, the proposed methods without retraining achieve a CDF of throughput performance comparable to that of the conventional methods. Besides, we compare the convergence between MFDQN and the conventional RL-based methods. The MFDQN shows that it reduces the complexity of the interactions among agents and converges faster than the conventional RL-based methods.

In Chapter 4 we study user location information for assisting in tackling the antenna beamforming problem. Our idea is to first cluster users into several high-density clusters according to the number of antenna arrays. Next, we design antenna parameters based on the geometric model to control beams toward the center of each cluster and cover each cluster separately. We use a classical K-Means clustering method and design an interesting EC method for clustering users. Moreover, to get a better throughput performance, we use DQN to fine-tune antenna parameters after designing the antenna parameters according to the clustering result of using EC. Under the same simulation settings in Chapter 3, the simulation results show that K-Means, EC, and EC-aided DQN are effective in improving the CDF performance of throughput over the RL-based approach in HAPS with both rotation and shift scenarios. Besides, we evaluate the clustering-based methods in a user movement scenario, we show that the clustering-based approaches can still maintain a high CDF of throughput performance, while the RL-based approach cannot.

5.2 Future Works

In the IUD estimation problem, our proposed method still has some limitations. Even though the proposed method is usable even in scenarios where the users are moving, we can only get the outdated beam energy image of users. Thus, the accuracy of detection will be highly affected by the frequency of beam sweeping: if the sweeping is very frequent, a high accuracy can be obtained, however, it will lead to a huge overhead. On the other hand, if the sweeping is not very frequent, the overhead is reduced, however, a drop in the accuracy is expected. Identifying a good balance between the accuracy and the

sweeping frequency is yet to be identified. Nonetheless, we could use a different type of neural network (ConvLSTM) to account for the change over time of the user's location and predict the beam energy image of the users based on their history [132] to reduce the frequency of the sweeping, while accurately predicting their distances.

For antenna control in the HAPS system, in this dissertation, we only consider the single HAPS movement scenario. The more complex scenario should be considered such as multiple HAPS movement scenarios in the future. In the case of multiple HAPS and users moving at the same time, increased cost of obtaining user location information due to more frequent changes in user distribution. Therefore, we still need to explore DRL- and DL-based algorithms to tackle antenna control problems in more complex cases.

Appendix A

List of Author's Publications and Awards

A.1 Journals

1. **S. Yang**, M. Bouazizi, Y. Cao, and T. Ohtsuki, "Inter-User Distance Estimation Based on a New Type of Fingerprint in Massive MIMO System for COVID-19 Contact Detection," *Sensors*, vol. 22, no. 16, p. 6211, Aug. 2022, doi: 10.3390/s22166211.

2. **S. Yang**, M. Bouazizi, T. Ohtsuki, Y. Shibata, W. Takabatake, K. Hoshino, A. Nagate, "Deep Reinforcement Learning Evolution Algorithm for Dynamic Antenna Control in Multi-Cell Configuration HAPS System," in *Future Internet*, vol. 15, no. 1, p. 34, Jan. 2023, doi: 10.3390/fi15010034.

A.2 Full Articles on International Conferences Proceedings

1. K. Wada, **S. Yang**, M. Bouazizi, T. Ohtsuki, Y. Shibata, W. Takabatake, K. Hoshino, A. Nagate, "Dynamic Antenna Control for HAPS Using Fuzzy Q-Learning in Multi-Cell Configuration," in *ICC 2022 - IEEE International Conference on Communications*, Seoul, Korea, Republic of, 2022, pp. 1-6, doi: 10.1109/ICC45855.2022.9838271.

2. **S. Yang**, M. Bouazizi, T. Ohtsuki, Y. Shibata, W. Takabatake, K. Hoshino, A. Nagate, "Dynamic Antenna Control for HAPS Using Geometry-based Method in Multi-Cell Configuration," in *2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring)*, Helsinki, Finland, 2022, pp. 1-6, doi: 10.1109/VTC2022-Spring54318.2022.9860609.

3. **S. Yang**, M. Bouazizi, T. Ohtsuki, "Dynamic Antenna Control for HAPS Using Mean Field Reinforcement Learning in Multi-Cell Configuration," in *ICC 2023 - IEEE International Conference on Communications*, Rome, Italy, Accepted, 2023.

A.3 Articles on Domestic Conference Proceedings

1. **S. Yang**, L. Wang, M. Bouazizi, T. Ohtsuki, "Denoising-aided Channel Estimation and DNN-based Joint Beamforming and Phase Shift Design in IRS-aided MISO System", *IEICE General Conf*, Jul. 2021.

2. **S. Yang**, M. Bouazizi, T. Ohtsuki, " Dynamic Antenna Control for HAPS Using Geometric Based Method in Multi-Cell Configuration", *IEICE General Conf*, Mar. 2022.

A.4 Technical Reports

1. **S. Yang**, T. Ohtsuki, " Deep Learning Based Joint Design of Beamforming and Phase Shifts in Intelligent Reflecting Surface-aided MISO Systems", *IEICE RCS*, Mar. 2021.

2. **S. Yang**, M. Bouazizi, T. Ohtsuki, "GAN-based Denoising for Channel Estimation and DNN-based Joint Design of Beamforming and Phase Shifts in the IRS-aided MISO System", *IEICE RCS*, Sep. 2021.

3. **S. Yang**, M. Bouazizi, T. Ohtsuki, "Dynamic Antenna Control for HAPS Using Geometric Based Method in Multi-Cell Configuration", *IEICE RCS*, Mar. 2022.

4. **S. Yang**, M. Bouazizi, T. Ohtsuki, Y. Shibata, W. Takabatake, K. Hoshino, A. Nagate, “Dynamic Antenna Control for HAPS Using DQN-based Method in Multi-Cell Configuration”, *MIKA*, Oct. 2022.

References

[1] A. N. Uwaechia and N. M. Mahyuddin, "A comprehensive survey on millimeter wave communications for fifth-generation wireless networks: Feasibility and challenges," *IEEE Access*, vol. 8, pp. 62 367–62 414, 2020.

[2] S. A. Busari, K. M. S. Huq, S. Mumtaz, L. Dai, and J. Rodriguez, "Millimeter-wave massive mimo communication for future wireless systems: A survey," *IEEE Communications Surveys and Tutorials*, vol. 20, no. 2, pp. 836–869, 2018.

[3] L. Zhang, H. Zhao, S. Hou, Z. Zhao, H. Xu, X. Wu, Q. Wu, and R. Zhang, "A survey on 5G millimeter wave communications for uav-assisted wireless networks," *IEEE Access*, vol. 7, pp. 117 460–117 504, 2019.

[4] A. Ghosh, T. A. Thomas, M. C. Cudak, R. Ratasuk, P. Moorut, F. W. Vook, T. S. Rappaport, G. R. MacCartney, S. Sun, and S. Nie, "Millimeter-wave enhanced local area systems: A high-data-rate approach for future wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1152–1163, 2014.

[5] P. Wang, Y. Li, L. Song, and B. Vucetic, "Multi-gigabit millimeter wave wireless communications for 5G: from fixed access to cellular networks," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 168–178, 2015.

[6] N. Panwar, S. Sharma, and A. K. Singh, "A survey on 5G: The next generation of mobile communication," *Physical Communication*, vol. 18, pp. 64–84, 2016, special Issue on Radio Access Network Architectures and Resource Management for 5G. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1874490715000531

[7] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, 2014.

[8] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications (mmwave) for 5G: Opportunities and challenges," *Wirel. Netw.*, vol. 21, no. 8, p. 2657–2676, nov 2015. [Online]. Available: https://doi.org/10.1007/s11276-015-0942-z

[9] M. Xiao, S. Mumtaz, Y. Huang, L. Dai, Y. Li, M. Matthaiou, G. K. Karagiannidis, E. Björnson, K. Yang, C.-L. I, and A. Ghosh, "Millimeter wave communications for future mobile networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 1909–1935, 2017.

[10] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[11] J. A. del Peral-Rosado, R. Raulefs, J. A. López-Salcedo, and G. Seco-Granados, "Survey of cellular mobile radio localization methods: From 1g to 5G," *IEEE Communications Surveys and Tutorials*, vol. 20, no. 2, pp. 1124–1148, 2018.

[12] J. Palacios, P. Casari, H. Assasa, and J. Widmer, "Leap: Location estimation and predictive handover with consumer-grade mmwave devices," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, 2019, pp. 2377–2385.

[13] Z. Xiao, Y. Zeng, and Song, "An overview on integrated localization and communication towards 6g," *Science China Information Sciences*, vol. 65, no. 3, pp. 1869–1919, 2021.

[14] K. Hoshino, S. Sudo, and Y. Ohta, "A study on antenna beamforming method considering movement of solar plane in haps system," in *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, 2019.

[15] L. Li, H. Ren, Q. Cheng, K. Xue, W. Chen, M. Debbah, and Z. Han, "Millimeter-wave networking in the sky: A machine learning and mean field game approach for joint beamforming and beam-steering," *IEEE Transactions on Wireless Communications*, vol. 19, no. 10, pp. 6393–6408, 2020.

[16] H. Shokri-Ghadikolaei, F. Boccardi, C. Fischione, G. Fodor, and M. Zorzi, "Spectrum sharing in mmwave cellular networks via cell association, coordination, and beamforming," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 11, pp. 2902–2917, 2016.

[17] N. Gao, S. Jin, X. Li, and M. Matthaiou, "Aerial ris-assisted high altitude platform communications," *IEEE Wireless Communications Letters*, vol. 10, no. 10, pp. 2096–2100, 2021.

[18] S. Karapantazis and F. Pavlidou, "Broadband communications via high-altitude platforms: A survey," *IEEE Communications Surveys Tutorials*, vol. 7, no. 1, pp. 2–31, 2005.

[19] C. Ding, J.-B. Wang, H. Zhang, M. Lin, and G. Y. Li, "Joint optimization of transmission and computation resources for satellite and high altitude platform assisted edge computing," *IEEE Transactions on Wireless Communications*, vol. 21, no. 2, pp. 1362–1377, 2022.

[20] J. Ye, S. Dang, B. Shihada, and M.-S. Alouini, "Space-air-ground integrated networks: Outage performance analysis," *IEEE Transactions on Wireless Communications*, vol. 19, no. 12, pp. 7897–7912, 2020.

[21] H.-B. Jeon, S.-H. Park, J. Park, K. Huang, and C.-B. Chae, "An energy-efficient aerial backhaul system with reconfigurable intelligent surface," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2022.

[22] T. Tozer and D. Grace, "High-altitude platforms for wireless communications," *Electronics & Communications Engineering Journal (ECEJ)*, vol. 13, pp. 127–137, 07 2001.

[23] "Haps: Why softbank is looking to the stratosphere," Softbank, https://www.softbank.jp/en/sbnews/entry/20190826_01 Accessed Aug. 26, 2019.

[24] P. Tereza and H. Elizabeth, "Starlink satellites: Everything you need to know about the controversial internet megaconstellation," SPACE, https://www.space.com/spacex-starlink-satellites.html Accessed Nov. 23, 2022.

[25] L. Zhang, S. Lai, J. Xia, C. Gao, D. Fan, and J. Ou, "Deep reinforcement learning based irs-assisted mobile edge computing under physical-layer security," *Physical Communication*, vol. 55, p. 101896, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1874490722001732

[26] W. Zhan, B. Huang, A. Huang, N. Jiang, and J. Lee, "Offline reinforcement learning with realizability and single-policy concentrability," in *Proceedings of Thirty Fifth Conference on Learning Theory*, ser. Proceedings of Machine Learning Research, P.-L. Loh and M. Raginsky, Eds., vol. 178. PMLR, 02–05 Jul 2022, pp. 2730–2775. [Online]. Available: https://proceedings.mlr.press/v178/zhan22a.html

[27] A. Lazaridou and M. Baroni, "Emergent multi-agent communication in the deep learning era," 2020. [Online]. Available: https://arxiv.org/abs/2006.02419

[28] J. Kaur, M. A. Khan, M. Iftikhar, M. Imran, and Q. Emad Ul Haq, "Machine learning techniques for 5G and beyond," *IEEE Access*, vol. 9, pp. 23 472–23 488, 2021.

[29] X. Hu, S. Xu, L. Wang, Y. Wang, Z. Liu, L. Xu, Y. Li, and W. Wang, "A joint power and bandwidth allocation method based on deep reinforcement learning for v2v communications in 5G," *China Communications*, vol. 18, no. 7, pp. 25–35, 2021.

[30] H. T. Huong Giang, P. D. Thanh, H. Ko, and S. Pack, "Deep reinforcement learning-based power allocation for downlink rsma system," in *2022 13th International Conference on Information and Communication Technology Convergence (ICTC)*, 2022, pp. 775–777.

[31] L. Zhang, S. Lambotharan, and G. Zheng, "A countermeasure against adversarial attacks on power allocation in a massive mimo network," in *2022 IEEE Symposium on Wireless Technology and Applications (ISWTA)*, 2022, pp. 13–16.

[32] A. Poulose and D. S. Han, "Feature-based deep lstm network for indoor localization using uwb measurements," in *2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2021, pp. 298–301.

[33] S. Roy, W. Menapace, S. Oei, B. Luijten, E. Fini, C. Saltori, I. Huijben, N. Chennakeshava, F. Mento, A. Sentelli, E. Peschiera, R. Trevisan, G. Maschietto, E. Torri, R. Inchingolo, A. Smargiassi, G. Soldati, P. Rota, A. Passerini, R. J. G. van Sloun, E. Ricci, and L. Demi, "Deep learning for classification and localization of covid-19 markers in point-of-care lung ultrasound," *IEEE Transactions on Medical Imaging*, vol. 39, no. 8, pp. 2676–2687, 2020.

[34] J. Akbar, M. Shahzad, M. I. Malik, A. Ul-Hasan, and F. Shafait, "Runway detection and localization in aerial images using deep learning," in *2019 Digital Image Computing: Techniques and Applications (DICTA)*, 2019.

[35] P. Klinwichit, K. Chinnasarn, A. Onuean, S. Limchareon, S.-H. Lee, and J.-S. Jang, "The radiographic view classification and localization of lumbar spine using deep learning models," in *2022 13th International Conference on Information and Communication Technology Convergence (ICTC)*, 2022, pp. 1316–1319.

[36] B. Zhang, C. Jin, K. Cao, Q. Lv, and R. Mittra, "Cognitive conformal antenna array exploiting deep reinforcement learning method," *IEEE Transactions on Antennas and Propagation*, vol. 70, no. 7, pp. 5094–5104, 2022.

[37] J. Peng, H. Cao, Z. Ali, X. Wu, and J. Fan, "Intelligent reflecting surface-assisted interference mitigation with deep reinforcement learning for radio astronomy," *IEEE Antennas and Wireless Propagation Letters*, vol. 21, no. 9, pp. 1757–1761, 2022.

[38] R. Q. Shaddad, E. M. Saif, H. M. Saif, Z. Y. Mohammed, and A. H. Farhan, "Channel estimation for intelligent reflecting surface in 6g wireless network via deep learning technique," in *2021 1st International Conference on Emerging Smart Technologies and Applications (eSmarTA)*, 2021.

[39] M. H. Rahman, M. Shahjalal, M. O. Ali, S. Yoon, and Y. M. Jang, "Deep learning based pilot assisted channel estimation for rician fading massive mimo uplink communication system," in *2021 Twelfth International Conference on Ubiquitous and Future Networks (ICUFN)*, 2021, pp. 470–472.

[40] R. Yao, Q. Qin, S. Wang, N. Qi, Y. Fan, and X. Zuo, "Deep learning assisted channel estimation refinement in uplink ofdm systems under time-varying channels," in *2021 International Wireless Communications and Mobile Computing (IWCMC)*, 2021, pp. 1349–1353.

[41] Z. Ma, P. Jia, D. Han, M. Zhang, Z. Ghassemlooy, and L. Wang, "Deep-learning-based channel estimation for multi-wavelength visible light communication system," in *2022 4th West Asian Symposium on Optical and Millimeter-wave Wireless Communications (WASOWC)*, 2022, pp. 01–04.

[42] W. Son and D. S. Han, "Deep learning approach for improving spectral efficiency in mmwave hybrid beamforming systems," in *2022 27th Asia Pacific Conference on Communications (APCC)*, 2022, pp. 66–69.

[43] Y. Ahn and B. Shim, "Deep learning-based beamforming for intelligent reflecting surface-assisted mmwave systems," in *2021 International Conference on Information and Communication Technology Convergence (ICTC)*, 2021, pp. 1731–1734.

[44] T. T. Hien Pham, T. V. Nguyen, and S. Cho, "Intelligent beamforming design in mmwave mmimo: A reinforcement learning approach," in *2022 13th International Conference on Information and Communication Technology Convergence (ICTC)*, 2022, pp. 1033–1036.

[45] F. B. Mismar, B. L. Evans, and A. Alkhateeb, "Deep reinforcement learning for 5G networks: Joint beamforming, power control, and interference coordination," *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1581–1592, 2020.

[46] A. Rkhami, Y. Hadjadj-Aoul, and A. Outtagarts, "Learn to improve: A novel deep reinforcement learning approach for beyond 5G network slicing," in *2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC)*, 2021.

[47] G. Eappen, J. Cosmas, S. T, R. A, R. Nilavalan, and J. Thomas, "Deep learning integrated reinforcement learning for adaptive beamforming in b5G networks," *IET Communications*, vol. n/a, no. n/a. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/cmu2.12501

[48] D. He, X. Chen, L. Pei, F. Zhu, L. Jiang, and W. Yu, "Multi-bs spatial spectrum fusion for 2-d doa estimation and localization using uca in massive mimo system," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–13, 2021.

[49] Y. Pan, S. De Bast, and S. Pollin, "Indoor direct positioning with imperfect massive mimo array using measured near-field channels," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.

[50] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proceedings of the IEEE*, vol. 102, no. 3, pp. 366–385, 2014.

[51] N. Garcia, H. Wymeersch, E. G. Larsson, A. M. Haimovich, and M. Coulon, "Direct localization for massive mimo," *IEEE Transactions on Signal Processing*, vol. 65, no. 10, pp. 2475–2487, 2017.

[52] X. Sun, C. Wu, X. Gao, and G. Y. Li, "Fingerprint-based localization for massive mimo-ofdm system with deep convolutional neural networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 11, pp. 10 846–10 857, 2019.

[53] J. Vieira, E. Leitinger, M. Sarajlic, X. Li, and F. Tufvesson, "Deep convolutional neural networks for massive mimo fingerprint-based positioning," in *2017 IEEE*

28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), 2017.

[54] S. J. Patel and M. J. Zawodniok, "3d localization of rfid antenna tags using convolutional neural networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–11, 2022.

[55] J. Liu and G. Guo, "Vehicle localization during gps outages with extended kalman filter and deep learning," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–10, 2021.

[56] J. Gante, G. Falcão, and L. Sousa, "Deep learning architectures for accurate millimeter wave positioning in 5G," *Neural Processing Letters*, vol. 51, pp. 487–514, 2020.

[57] V. Savic and E. G. Larsson, "Fingerprinting-based positioning in distributed massive mimo systems," in *2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall)*, 2015.

[58] Z. Abu-Shaban, X. Zhou, T. Abhayapala, G. Seco-Granados, and H. Wymeersch, "Performance of location and orientation estimation in 5G mmwave systems: Uplink vs downlink," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, 2018.

[59] S. Yang, M. Bouazizi, T. Ohtsuki, Y. Shibata, W. Takabatake, K. Hoshino, and A. Nagate, "Dynamic antenna control for haps using geometry-based method in multi-cell configuration," in *2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring)*, 2022.

[60] B. El-Jabu and R. Steele, "Effect of positional instability of an aerial platform on its cdma performance," in *Gateway to 21st Century Communications Village. VTC 1999-Fall. IEEE VTS 50th Vehicular Technology Conference (Cat. No.99CH36324)*, vol. 5, 1999, pp. 2471–2475 vol.5.

[61] J. Thornton and D. Grace, "Effect of lateral displacement of a high-altitude platform on cellular interference and handover," *Trans. Wireless. Comm.*, vol. 4, no. 4, p. 1483–1490, jul 2005. [Online]. Available: https://doi.org/10.1109/TWC.2005.850282

[62] H. Panfeng, C. Naiping, and N. Shuyan, "Coverage model of multi beam antenna from high altitude platform in the swing state," in *2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, 2016, pp. 750–753.

[63] M. I. Dessouky, H. A. Sharshar, and Y. A. Albagory, "Geometrical analysis of high altitude platforms cellular footprint," *Progress in Electromagnetics Research-pier*, vol. 67, pp. 263–274, 2007.

[64] M. Dessouky, M. Nofal, H. Sharshar, and Y. Albagory, "Optimization of beams directions for high altitude platforms cellular communications design," in *Proceedings of the Twenty Third National Radio Science Conference (NRSC'2006)*, 2006, pp. 1–8.

[65] F. Enache, D. Depărăţeanu, and F. Popescu, "Optimal design of circular antenna array using genetic algorithms," in *2017 9th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, 2017.

[66] G. Sun, Y. Liu, Z. Chen, Y. Zhang, A. Wang, and S. Liang, "Thinning of concentric circular antenna arrays using improved discrete cuckoo search algorithm," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, 2017.

[67] A. Ismaiel, E. Elsaidy, Y. Albagory, H. Atallah, A. Abdel-Rahman, and T. Sallam, "Performance improvement of high altitude platform using concentric circular antenna array based on particle swarm optimization," *AEU - International Journal of Electronics and Communications*, vol. 91, 05 2018.

[68] K. Wada, S. Yang, M. Bouazizi, T. Ohtsuki, Y. Shibata, W. Takabatake, K. Hoshino, and A. Nagate, "Dynamic antenna control for haps using fuzzy q-learning in multi-cell configuration," in *ICC 2022 - IEEE International Conference on Communications*, 2022.

[69] J. Zhang, F. Yang, Z. Deng, X. Fu, and J. Han, "Research on D2D co-localization algorithm based on clustering filtering," *China Communications*, vol. 17, no. 8, pp. 121–132, 2020.

[70] Y. Li, L. Xie, M. Zhou, and Q. Jiang, "D2D cooperative localization approach based on euclidean distance matrix completion," in *2020 IEEE/CIC International Conference on Communications in China (ICCC)*, 2020, pp. 52–56.

[71] Z. Luo, Q. Zhang, W. Wang, and T. Jiang, "Single-antenna device-to-device localization in smart environments with backscatter," *IEEE Internet of Things Journal*, vol. 9, no. 12, pp. 10 121–10 129, 2022.

[72] N. Chukhno, S. Trilles, J. Torres-Sospedra, A. Iera, and G. Araniti, "D2D-based cooperative positioning paradigm for future wireless systems: A survey," *IEEE Sensors Journal*, vol. 22, no. 6, pp. 5101–5112, 2022.

[73] J. Liu, G. Chuai, and W. Gao, "Outdoor location fingerprint based on optimized k-means clustering," in *2021 IEEE/CIC International Conference on Communications in China (ICCC Workshops)*, 2021, pp. 88–93.

[74] T. Jingwangsa, S. Soonjun, and P. Cherntanomwong, "Comparison between innovative approaches of RFID based localization using fingerprinting techniques for outdoor and indoor environments," in *2010 The 12th International Conference on Advanced Communication Technology (ICACT)*, vol. 2, 2010, pp. 1511–1515.

[75] Sutiyo, R. Hidayat, Sunarno, and I. W. Mustika, "Regression analysis for estimated distance in fingerprinting-based wlan outdoor localization system," in *2018 4th International Conference on Science and Technology (ICST)*, 2018.

[76] P. Cherntanomwong, J.-i. Takada, H. Tsuji, and D. Gray, "New radio source localization using array antennas based on fingerprinting techniques in outdoor environment," in *2007 Asia-Pacific Microwave Conference*, 2007.

[77] S. Hu, L.-X. Guo, and Z.-Y. Liu, "A wireless outdoor fingerprint locating method based on ray-tracing model," in *2020 9th Asia-Pacific Conference on Antennas and Propagation (APCAP)*, 2020.

[78] Q. Wu, G. Chuai, and W. Gao, "A fingerprint database construction method based on universal kriging interpolation for outdoor localization," in *2020 IEEE/CIC International Conference on Communications in China (ICCC)*, 2020, pp. 46–51.

[79] R. Wang, Z. Li, J. Tang, and H. Wen, "Rf fingerprint identification of commercial uav in outdoor environment," in *2022 International Conference on Computing, Communication, Perception and Quantum Technology (CCPQT)*, 2022, pp. 367–371.

[80] X. Sun, C. Wu, X. Gao, and G. Y. Li, "Fingerprint-based localization for massive mimo-ofdm system with deep convolutional neural networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 11, pp. 10 846–10 857, 2019.

[81] G. Sun, J. Chen, W. Guo, and K. R. Liu, "Signal processing techniques in network-aided positioning: a survey of state-of-the-art positioning designs," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 12–23, 2005.

[82] H. Deng and A. Sayeed, "Mm-wave mimo channel modeling and user localization using sparse beamspace signatures," in *2014 IEEE 15th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2014, pp. 130–134.

[83] S. D. Regani, C. Wu, B. Wang, M. Wu, and K. J. R. Liu, "mmwrite: Passive handwriting tracking using a single millimeter-wave radio," *IEEE Internet of Things Journal*, vol. 8, no. 17, pp. 13 291–13 305, 2021.

[84] K. A. Gotsis, K. Siakavara, and J. N. Sahalos, "On the direction of arrival (DoA) estimation for a switched-beam antenna system using neural networks," *IEEE Transactions on Antennas and Propagation*, vol. 57, no. 5, pp. 1399–1411, 2009.

[85] F. Lemic, J. Martin, C. Yarp, D. Chan, V. Handziski, R. Brodersen, G. Fettweis, A. Wolisz, and J. Wawrzynek, "Localization as a feature of mmwave communication," in *2016 International Wireless Communications and Mobile Computing Conference (IWCMC)*, 2016, pp. 1033–1038.

[86] A. J. Weiss, "Direct position determination of narrowband radio frequency transmitters," *IEEE Signal Processing Letters*, vol. 11, no. 5, p. 513 – 516, 2004, cited by: 317. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-2442459959&doi=10.1109%2fLSP.2004.826501&partnerID=40&md5=1ff085cfe53fae82e4c2a9ab2055180a

[87] M. Mercuri, G. Sacco, R. Hornung, P. Zhang, H. J. Visser, M. Hijdra, Y.-H. Liu, S. Pisa, B. van Liempd, and T. Torfs, "2-d localization, angular separation and vital signs monitoring using a siso fmcw radar for smart long-term health monitoring environments," *IEEE Internet of Things Journal*, vol. 8, no. 14, pp. 11 065–11 077, 2021.

[88] O. Kanhere and T. S. Rappaport, "Position locationing for millimeter wave systems," in *2018 IEEE Global Communications Conference (GLOBECOM)*, 2018, pp. 206–212.

[89] X. Sun, X. Gao, G. Y. Li, and W. Han, "Single-site localization based on a new type of fingerprint for massive mimo-ofdm systems," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 7, pp. 6134–6145, 2018.

[90] B. Jang, M. Kim, G. Harerimana, and J. W. Kim, "Q-learning algorithms: A comprehensive classification and applications," *IEEE Access*, vol. 7, pp. 133 653–133 667, 2019.

[91] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015. [Online]. Available: http://www.nature.com/articles/nature14236

[92] I. A. Junglas and R. T. Watson, "Location-based services," *Commun. ACM*, vol. 51, no. 3, p. 65–69, mar 2008. [Online]. Available: https://doi.org/10.1145/1325555.1325568

[93] I. Guvenc and C.-C. Chong, "A survey on toa based wireless localization and nlos mitigation techniques," *IEEE Communications Surveys Tutorials*, vol. 11, no. 3, pp. 107–124, 2009.

[94] S. Maranò, W. M. Gifford, H. Wymeersch, and M. Z. Win, "Nlos identification and mitigation for localization based on uwb experimental data," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 7, pp. 1026–1035, 2010.

[95] Y. Shen, S. Mazuelas, and M. Z. Win, "Network navigation: Theory and interpretation," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 9, pp. 1823–1834, 2012.

[96] X. Yang, Z. Wu, and Q. Zhang, "Bluetooth indoor localization with gaussian–bernoulli restricted boltzmann machine plus liquid state machine," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–8, 2022.

[97] F. Zafari, A. Gkelias, and K. K. Leung, "A survey of indoor localization systems and technologies," *IEEE Communications Surveys Tutorials*, vol. 21, no. 3, pp. 2568–2599, 2019.

[98] M. Mercuri, I. R. Lorato, Y.-H. Liu, F. Wieringa, C. V. Hoof, and T. Torfs, "Vital-sign monitoring and spatial tracking of multiple people using a

contactless radar-based sensor," vol. 2, no. 6, pp. 252–262. [Online]. Available: https://doi.org/10.1038/s41928-019-0258-6

[99] X. Ye, X. Yin, X. Cai, A. Pérez Yuste, and H. Xu, "Neural-network-assisted ue localization using radio-channel fingerprints in lte networks," *IEEE Access*, vol. 5, pp. 12 071–12 087, 2017.

[100] M. Koivisto, A. Hakkarainen, M. Costa, P. Kela, K. Leppanen, and M. Valkama, "High-efficiency device positioning and location-aware communications in dense 5G networks," *IEEE Communications Magazine*, vol. 55, no. 8, pp. 188–195, 2017.

[101] Y. Shen and M. Z. Win, "On the use of multipath geometry for wideband cooperative localization," in *GLOBECOM 2009 - 2009 IEEE Global Telecommunications Conference*, 2009, pp. 1–6.

[102] R. Di Taranto, S. Muppirisetty, R. Raulefs, D. Slock, T. Svensson, and H. Wymeersch, "Location-aware communications for 5G networks: How location information can improve scalability, latency, and robustness of 5G," *IEEE Signal Processing Magazine*, vol. 31, no. 6, pp. 102–112, 2014.

[103] P. Ferrand, A. Decurninge, and M. Guillaud, "Dnn-based localization from channel estimates: Feature design and experimental results," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, 2020, pp. 1–6.

[104] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up mimo: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, 2013.

[105] H. Wymeersch, S. Marano, W. M. Gifford, and M. Z. Win, "A machine learning approach to ranging error mitigation for uwb localization," *IEEE Transactions on Communications*, vol. 60, no. 6, pp. 1719–1728, 2012.

[106] Y. Wang, Y. Wu, and Y. Shen, "Multipath effect mitigation by joint spatiotemporal separation in large-scale array localization," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, 2017.

[107] M. Vari and D. Cassioli, "mmwaves rssi indoor network localization," in *2014 IEEE International Conference on Communications Workshops (ICC)*, 2014, pp. 127–132.

[108] Y. Shen, H. Wymeersch, and M. Z. Win, "Fundamental limits of wideband localization— part ii: Cooperative networks," *IEEE Transactions on Information Theory*, vol. 56, no. 10, pp. 4981–5000, 2010.

[109] R. M. Buehrer, H. Wymeersch, and R. M. Vaghefi, "Collaborative sensor network localization: Algorithms and practical issues," *Proceedings of the IEEE*, vol. 106, no. 6, pp. 1089–1114, 2018.

[110] M. Z. Win, W. Dai, Y. Shen, G. Chrisikos, and H. Vincent Poor, "Network operation strategies for efficient localization and navigation," *Proceedings of the IEEE*, vol. 106, no. 7, pp. 1224–1254, 2018.

[111] Y. Cao, T. Ohtsuki, and T. Q. S. Quek, "Dual-ascent inspired transmit precoding for evolving multiple-access spatial modulation," *IEEE Transactions on Communications*, vol. 68, no. 11, pp. 6945–6961, 2020.

[112] M. Luan, B. Wang, Y. Zhao, Z. Feng, and F. Hu, "Phase design and near-field target localization for ris-assisted regional localization system," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 2, pp. 1766–1777, 2022.

[113] Z. Wang, H. Zhang, T. Lu, and T. A. Gulliver, "Cooperative rss-based localization in wireless sensor networks using relative error estimation and semidefinite programming," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 483–497, 2019.

[114] K.-H. Lam, C.-C. Cheung, and W.-C. Lee, "Rssi-based lora localization systems for large-scale indoor and outdoor environments," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 12, pp. 11 778–11 791, 2019.

[115] Z. Abu-Shaban, H. Wymeersch, T. Abhayapala, and G. Seco-Granados, "Single-anchor two-way localization bounds for 5G mmwave systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 6, pp. 6388–6400, 2020.

[116] Z. Ma and K. Ho, "Toa localization in the presence of random sensor position errors," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 2468–2471.

[117] S. Sadowski and P. Spachos, "Rssi-based indoor localization with the internet of things," *IEEE Access*, vol. 6, pp. 30 149–30 161, 2018.

[118] M. Zane, M. Rupp, and S. Schwarz, "Performance investigation of angle of arrival based localization," in *WSA 2020; 24th International ITG Workshop on Smart Antennas*, 2020.

[119] S. Zhao, X.-P. Zhang, X. Cui, and M. Lu, "Optimal two-way toa localization and synchronization for moving user devices with clock drift," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 8, pp. 7778–7789, 2021.

[120] Z. Dai, G. Wang, X. Jin, and X. Lou, "Nearly optimal sensor selection for tdoa-based source localization in wireless sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 10, pp. 12 031–12 042, 2020.

[121] Q. Zheng, L. Luo, H. Song, G. Sheng, and X. Jiang, "A rssi-aoa-based uhf partial discharge localization method using music algorithm," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, 2021.

[122] C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: A survey," *IEEE Communications Surveys Tutorials*, vol. 21, no. 3, pp. 2224–2287, 2019.

[123] Z. Shen, J. Li, and Q. Wu, "Data-driven interference localization using a single satellite based on received signal strength," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 8657–8669, 2020.

[124] C. Zhou, J. Liu, M. Sheng, Y. Zheng, and J. Li, "Exploiting fingerprint correlation for fingerprint-based indoor localization: A deep learning based approach," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 6, pp. 5762–5774, 2021.

[125] L. Ma, Y. Zhang, and D. Qin, "A novel indoor fingerprint localization system based on distance metric learning and ap selection," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–15, 2022.

[126] Q. Pu, J. K.-Y. Ng, and M. Zhou, "Fingerprint-based localization performance analysis: From the perspectives of signal measurement and positioning algorithm," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–15, 2021.

[127] V. Bianchi, P. Ciampolini, and I. De Munari, "Rssi-based indoor localization and identification for zigbee wireless sensor networks in smart homes," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 2, pp. 566–575, 2019.

[128] R. Canetti, A. Trachtenberg, and M. Varia, "Anonymous collocation discovery: Harnessing privacy to tame the coronavirus," *ArXivorg*, vol. 2003, no. 13670. [Online]. Available: https://par.nsf.gov/biblio/10156173

[129] M. Dmitrienko, A. Singh, P. Erichsen, and R. Raskar, "Proximity inference with wifi-colocation during the COVID-19 pandemic," *CoRR*, vol. abs/2009.12699, 2020. [Online]. Available: https://arxiv.org/abs/2009.12699

[130] P. M. Varela, J. Hong, T. Ohtsuki, and X. Qin, "Igmm-based co-localization of mobile users with ambient radio signals," *IEEE Internet of Things Journal*, vol. 4, no. 2, pp. 308–319, 2017.

[131] M. Cudak, T. Kovarik, T. A. Thomas, A. Ghosh, Y. Kishiyama, and T. Nakamura, "Experimental mm wave 5G cellular system," in *2014 IEEE Globecom Workshops (GC Wkshps)*, 2014, pp. 377–381.

[132] H. Echigo, Y. Cao, M. Bouazizi, and T. Ohtsuki, "A deep learning-based low overhead beam selection in mmwave communications," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 1, pp. 682–691, 2021.

[133] Y. Yu, R. Chen, W. Shi, and L. Chen, "Precise 3d indoor localization and trajectory optimization based on sparse wi-fi ftm anchors and built-in sensors," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 4, pp. 4042–4056, 2022.

[134] R. C. Luo and T.-J. Hsiao, "Indoor localization system based on hybrid wi-fi/ble and hierarchical topological fingerprinting approach," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 11, pp. 10 791–10 806, 2019.

[135] M. Zhou, Y. Li, M. J. Tahir, X. Geng, Y. Wang, and W. He, "Integrated statistical test of signal distributions and access point contributions for wi-fi indoor localization," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 5, pp. 5057–5070, 2021.

[136] D. Tse and P. Viswanath, *MIMO I: spatial multiplexing and channel modeling*. Cambridge University Press, 2005, p. 290–331.

[137] C. Li, E. Tanghe, D. Plets, P. Suanet, J. Hoebeke, E. De Poorter, and W. Joseph, "Reloc: Hybrid rssi- and phase-based relative uhf-rfid tag localization with cots devices," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 10, pp. 8613–8627, 2020.

[138] B. Mukhopadhyay, S. Srirangarajan, and S. Kar, "Rss-based localization in the presence of malicious nodes in sensor networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–16, 2021.

[139] "Wireless em propagation software - wireless insite." [Online]. Available: http://www.remcom.com/wireless-insite

[140] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[141] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[142] D. Zhou, S. Gao, R. Liu, F. Gao, and M. Guizani, "Overview of development and regulatory aspects of high altitude platform system," *Intelligent and Converged Networks*, vol. 1, no. 1, pp. 58–78, 2020.

[143] A. K. Widiawan and R. Tafazolli, "High altitude platform station (haps): A review of new infrastructure development for future wireless communications," *Wirel. Pers. Commun.*, vol. 42, no. 3, p. 387–404, aug 2007. [Online]. Available: https://doi.org/10.1007/s11277-006-9184-9

[144] A. Mohammed, A. Mehmood, F.-N. Pavlidou, and M. Mohorcic, "The role of high-altitude platforms (haps) in the global wireless connectivity," *Proceedings of the IEEE*, vol. 99, no. 11, pp. 1939–1953, 2011.

[145] G. P. White and Y. V. Zakharov, "Data communications to trains from high-altitude platforms," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 4, pp. 2253–2266, 2007.

[146] Y. Albagory, M. Nofal, and A. Ghoneim, "Handover performance of unstable-yaw stratospheric high-altitude stations," *Wireless Personal Communications*, vol. 84, pp. 1–13, 05 2015.

[147] P. He, N. Cheng, and J. Cui, "Handover performance analysis of cellular communication system from high altitude platform in the swing state," in *2016 IEEE International Conference on Signal and Image Processing (ICSIP)*, 2016, pp. 407–411.

[148] K. Hoshino, S. Sudo, and Y. Ohta, "A study on antenna beamforming method considering movement of solar plane in haps system," in *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, 2019.

[149] N. Dib, "Design of planar concentric circular antenna arrays with reduced side lobe level using symbiotic organisms search," *Neural Computing and Applications*, vol. 30, 12 2018.

[150] D. Mandal, H. Kumar, S. Ghoshal, and R. Kar, "Thinned concentric circular antenna array synthesis using particle swarm optimization," *Procedia Technology*, vol. 6, pp. 848–855, 2012, 2nd International Conference on Communication, Computing & Security [ICCCS-2012]. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2212017312006482

[151] S. C. Arum, D. Grace, P. D. Mitchell, and M. D. Zakaria, "Beam-pointing algorithm for contiguous high-altitude platform cell formation for extended coverage," in *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, 2019.

[152] D. Grace, J. Thornton, G. Chen, G. White, and T. Tozer, "Improving the system capacity of broadband services using multiple high-altitude platforms," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 700–709, 2005.

[153] J. Thornton, D. Grace, M. Capstick, and T. Tozer, "Optimizing an array of antennas for cellular coverage from a high altitude platform," *IEEE Transactions on Wireless Communications*, vol. 2, no. 3, pp. 484–492, 2003.

[154] *Minimum Performance Characteristics and Operational Conditions For High Altitude Platform Stations Providing IMT-2000 in the Bands 1885- 1980 MHz, 2010-2 025MHz and 2110-2170 MHz in Regions 1 and 3 and 1885- 1980 MHz and 2110-2160 MHz in Region 2, document ITU-R M.1456*, 2000.

[155] Y. Shibata, N. Kanazawa, M. Konishi, K. Hoshino, Y. Ohta, and A. Nagate, "System design of gigabit haps mobile communications," *IEEE Access*, vol. 8, pp. 157 995–158 007, 2020.

[156] ——, "Correction to 'system design of gigabit haps mobile communications'," *IEEE Access*, vol. 9, pp. 618–618, 2021.

[157] P. Glorennec, "Fuzzy q-learning," in *Proceedings of 6th International Fuzzy Systems Conference*, 1997, pp. 659–662 vol.2.

[158] K. Dalamagkidis, D. Kolokotsa, K. Kalaitzakis, and G. Stavrakakis, "Reinforcement learning for energy conservation and comfort in buildings," *Building and Environment*, vol. 42, pp. 2686–2698, 07 2006.

[159] V. François-Lavet, P. Henderson, R. Islam, M. G. Bellemare, and J. Pineau, 2018.

[160] H. v. Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double q-learning," in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, ser. AAAI'16. AAAI Press, 2016, p. 2094–2100.

[161] Y. Yang, R. Luo, M. Li, M. Zhou, W. Zhang, and J. Wang, "Mean field multi-agent reinforcement learning," in *Proceedings of the 35th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, J. Dy and A. Krause, Eds., vol. 80. PMLR, 10–15 Jul 2018, pp. 5571–5580. [Online]. Available: https://proceedings.mlr.press/v80/yang18d.html

[162] M. Agarwal, V. Aggarwal, A. Ghosh, and N. Tiwari, "Reinforcement learning for mean field game," 2019. [Online]. Available: https://arxiv.org/abs/1905.13357

[163] S. Adlakha and R. Johari, "Mean field equilibrium in dynamic games with complementarities," in *49th IEEE Conference on Decision and Control (CDC)*, 2010, pp. 6633–6638.

[164] G. Y. Weintraub, C. L. Benkard, and B. Van Roy, "Markov perfect industry dynamics with many firms," *Econometrica*, vol. 76, no. 6, pp. 1375–1411, 2008. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA6158

[165] S. Adlakha, R. Johari, and G. Y. Weintraub, "Equilibria of dynamic games with many players: Existence, approximation, and market structure," *CoRR*, vol. abs/1011.5537, 2010. [Online]. Available: http://arxiv.org/abs/1011.5537

[166] "Dynamic population data," Agoop Corporation, https://www.agoop.co.jp/service/dynamic-population-data/ Accessed Dec. 6, 2021.

[167] S. Na, L. Xumin, and G. Yong, “Research on k-means clustering algorithm: An improved k-means clustering algorithm,” in *2010 Third International Symposium on Intelligent Information Technology and Security Informatics*, 2010, pp. 63–67.