Wang, Ruiyu (2022) *Machine learning enabled millimeter wave cellular system and beyond.* PhD thesis.

https://theses.gla.ac.uk/82878/

Machine Learning Enabled Millimeter Wave Cellular System and Beyond

Ruiyu Wang

Submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy

School of Engineering
College of Science and Engineering
University of Glasgow

University
of Glasgow

VIA VERITAS VITA

February 2022

Abstract

Millimeter-wave (mmWave) communication with advantages of abundant bandwidth and immunity to interference has been deemed a promising technology for the next generation network and beyond. With the help of mmWave, the requirements envisioned of the future mobile network could be met, such as addressing the massive growth required in coverage, capacity as well as traffic, providing a better quality of service and experience to users, supporting ultra-high data rates and reliability, and ensuring ultra-low latency. However, due to the characteristics of mmWave, such as short transmission distance, high sensitivity to the blockage, and large propagation path loss, there are some challenges for mmWave cellular network design. In this context, to enjoy the benefits from the mmWave networks, the architecture of next generation cellular network will be more complex. With a more complex network, it comes more complex problems. The plethora of possibilities makes planning and managing a complex network system more difficult. Specifically, to provide better Quality of Service and Quality of Experience for users in the such network, how to provide efficient and effective handover for mobile users is important. The probability of handover trigger will significantly increase in the next generation network, due to the dense small cell deployment. Since the resources in the base station (BS) is limited, the handover management will be a great challenge. Further, to generate the maximum transmission rate for the users, Line-of-sight (LOS) channel would be the main transmission channel. However, due to the characteristics of mmWave and the complexity of the environment, LOS channel is not feasible always. Non-line-of-sight channel should be explored and used as the backup link to serve the users. With all the problems trending to be complex and nonlinear, and the data traffic dramatically increasing, the conventional method is not effective and efficiency any more. In this case, how to solve the problems in the most efficient manner becomes important. Therefore, some new concepts, as well as novel technologies, require to be explored. Among them, one promising solution is the utilization of machine learning (ML) in the mmWave cellular network. On the one hand, with the aid of ML approaches, the network could learn from the mobile data and it allows

the system to use adaptable strategies while avoiding unnecessary human intervention. On the other hand, when ML is integrated in the network, the complexity and workload could be reduced, meanwhile, the huge number of devices and data could be efficiently managed.

Therefore, in this thesis, different ML techniques that assist in optimizing different areas in the mmWave cellular network are explored, in terms of non-line-of-sight (NLOS) beam tracking, handover management, and beam management. To be specific, first of all, a procedure to predict the angle of arrival (AOA) and angle of departure (AOD) both in azimuth and elevation in non-line-of-sight mmWave communications based on a deep neural network is proposed. Moreover, along with the AOA and AOD prediction, a trajectory prediction is employed based on the dynamic window approach (DWA). The simulation scenario is built with ray tracing technology and generate data. Based on the generated data, there are two deep neural networks (DNNs) to predict AOA/AOD in the azimuth (AAOA/AAOD) and AOA/AOD in the elevation (EAOA/EAOD). Furthermore, under an assumption that the UE mobility and the precise location is unknown, UE trajectory is predicted and input into the trained DNNs as a parameter to predict the AAOA/AAOD and EAOA/EAOD to show the performance under a realistic assumption. The robustness of both procedures is evaluated in the presence of errors and conclude that DNN is a promising tool to predict AOA and AOD in a NLOS scenario.

Second, a novel handover scheme is designed aiming to optimize the overall system throughput and the total system delay while guaranteeing the quality of service (QoS) of each user equipment (UE). Specifically, the proposed handover scheme called O-MAPPO integrates the reinforcement learning (RL) algorithm and optimization theory. An RL algorithm known as multi-agent proximal policy optimization (MAPPO) plays a role in determining handover trigger conditions. Further, an optimization problem is proposed in conjunction with MAPPO to select the target base station and determine beam selection. It aims to evaluate and optimize the system performance of total throughput and delay while guaranteeing the QoS of each UE after the handover decision is made.

Third, a multi-agent RL-based beam management scheme is proposed, where multi-agent deep deterministic policy gradient (MADDPG) is applied on each small-cell base station (SCBS) to maximize the system throughput while guaranteeing the quality of service. With MADDPG, smart beam management methods can serve the UEs more efficiently and accurately. Specifically, the mobility of UEs causes the dynamic changes of the network environment, the MADDPG algorithm learns the experience of these changes. Based on that, the beam management in the SCBS is optimized according the reward or

penalty when severing different UEs. The approach could improve the overall system throughput and delay performance compared with traditional beam management methods.

The works presented in this thesis demonstrate the potentiality of ML when addressing the problem from the mmWave cellular network. Moreover, it provides specific solutions for optimizing NLOS beam tracking, handover management and beam management. For NLOS beam tracking part, simulation results show that the prediction errors of the AOA and AOD can be maintained within an acceptable range of $\pm 2°$. Further, when it comes to the handover optimization part, the numerical results show the system throughput and delay are improved by 10% and 25%, respectively, when compared with two typical RL algorithms, Deep Deterministic Policy Gradient (DDPG) and Deep Q-learning (DQL). Lastly, when it considers the intelligent beam management part, numerical results reveal the convergence performance of the MADDPG and the superiority in improving the system throughput compared with other typical RL algorithms and the traditional beam management method.

University of Glasgow
College of Science & Engineering
Statement of Originality

Name: Ruiyu Wang

Registration Number:

I certify that the thesis presented here for examination for a PhD degree of the University of Glasgow is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it) and that the thesis has not been edited by a third party beyond what is permitted by the University's PGR Code of Practice.

The copyright of this thesis rests with the author. No quotation from it is permitted without full acknowledgement.

I declare that the thesis does not include work forming part of a thesis presented successfully for another degree.

I declare that this thesis has been produced in accordance with the University of Glasgow's Code of Good Practice in Research.

I acknowledge that if any issues are raised regarding good research practice based on review of the thesis, the examination may be postponed pending the outcome of any investigation of the issues.

Signature: ....

Date: ....

Contents

List of Tables

List of Figures

List of Acronyms

3D	Three-dimensionality
3GPP	3rd Generation Partnership Project
4G	Fourth Generation
5G	Fifth Generation
AoA	Angle of Arrival
AoD	Angle of Departure
AAOA	AOA in Azimuth
AAOD	AOD in Azimuth
AC	Actor-Critic
AI	Artificial Intelligence
AMF	Access and Mobility Management Function
ANN	Artificial Neural Network
ANS	Adaptive Neighborhood Search
BER	Bit Error Rate
BS	Base Station
CNN	Convolutional Neural Network
CPR	Channel Pulse Response
CSI	Channel State Information
D2D	Device to Device
DDPG	Deep Deterministic Policy Gradient
DL	Deep Learning
DNN	Deep Neural Network
DP	Dynamic Programming
DPQ	Deep Q-Learning
DT	Decision Trees
DWA	Dynamic Window Approach
EAOA	AOA in Elevation

EAOD	AOD in Elevation
FCC	Federal Communication Commission
FTTC	Fiber to Cell
GAE	Generalized Advantage Estimation
gNB	gNodeB
GO	Geometrical Optic
HDTV	High Definition Television
HMBCRNA	Heterogeneous Multi-beam Cloud Radio Access Network
HO	Handover
HOF	HO Failures
HOR	Handover Rate
IoT	Internet of Things
IoV	Internet of Vehicles
IPPO	Independent Proximal Policy Optimization
KNN	K-Nearest Neighbors
LIDAR	Light Detection and Ranging
LOS	Line-of-sight
LSTM	Long Short Term Memory
LTE	Long-Term Evolution
M2M	Machine to Machine
MAC	Medium Access Control
MADDPG	Multi-agent Deep Deterministic Policy Gradient
MAPPO	Multi-agent Proximal Policy Optimization
MBS	Macro Base Station
MC	Monte Carlo
MDP	Markov Decision Process
MIMO	Multiple-input Multiple-output
ML	Machine learning
MmWave	Millimeter-waves
MOBJ	Mobile Object
MSE	Mean Square Error
NB	Naive Bayesian
NFV	Network Function Virtualization
NLOS	Non-line-of-sight
NN	Neural Network

O-MAPPO	Optimization-based MAPPO
PDF	Probability Density Functions
PLE	Path Loss Exponent
PP	Ping Pong
PPO	Proximal Policy Optimization
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Networks
ReLU	Rectified Linear Unit
RF	Radio Frequency
RL	Reinforcement Learning
RMSprop	Root Mean Squared Propagation
RNN	Recurrent Neural Network
RSRP	Reference Signal Received Power
RSS	Received Signal Strength
SARSA	State-action-reward-state-action
SC	Small Cells
SINR	Signal to Interference and Noise Ratio
SCBS	Small Cell Base Station
SN	Sequence Number
SNR	Signal-to-noise Ratio
SOM	Self-organizing Map
SQP	Sequential Quadratic Programming
SVM	Support Vector Machine
TD	Temporal Difference
ToA	Time of Arrival
UAV	Unmanned Aerial Vehicles
UE	User Equipment
URLLC	Ultra-Reliable Low-Latency Communication
UTD	Uniform Theory of Diffraction
V2X	Vehicle-to-everything
WI	Wireless Insite

List of Publications

Journals

1. **Wang, R.**, Klaine, P. V., Onireti, O., Sun, Y., Imran, M. A., and Zhang, L. (2021). Deep Learning Enabled Beam Tracking for Non-Line of Sight Millimeter Wave Communications. IEEE Open Journal of the Communications Society, 2, 1710-1720

2. **Wang, R.**, Zhang, C., Sun, Y., Imran, M. A., and Zhang, L. A Novel Handover Scheme for Millimeter Wave Network: An Approach of Integrating Reinforcement Learning and Optimization. Minor revision in Digital Communications and Networks (DCN).

3. **Wang, R.**, Sun, Y., Imran, M. A., and Zhang, L. Intelligent Beam Management for MmWave Networks. Under review in IEEE Wireless Communications Letters (WCL).

4. Zhang, B., **Wang, R.**, Xu, H., Zhang, X., and Zhang, L. DISTERNING: Distance Estimation using Machine Learning Approach for COVID-19 Contact Tracing and Beyond. Under review in IEEE JSAC Special Issue.

5. Yang, B., Mao, J., Ijaz, A. , Zhang, L., Sun, Y., **Wang, R.**, and Imran, M. A. Uplink Channel Estimation and Pilot Design for Radio-Frequency Differed Mixed-Numerology Systems. Under Revision in IEEE Transactions on Vehicular Technology

6. Yang, B., Mao, J., Zhang, L., Sun, Y., **Wang, R.**, and Imran, M. A. Power Allocation in the Presence of Mixed-numerology. Under review in IEEE WCL.

Conference Papers

1. Gou, Y., **Wang, R.**, Li, Z., Imran, M. A., and Zhang, L. Clustered Hierarchical Distributed Federated Learning. Accepted by 2022 IEEE International Conference on Communications (ICC).

2. Wen, S., Zhang, L., Quand, F., Feng, C., **Wang, R.**, Klaine, V., and Imran, M. A. Privacy-Preserving Federated Learning based on Differential Privacy and Momentum Gradient Descent. Under Review in IEEE World Congress on Computational Intelligence.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. Lei Zhang, for supporting me during this four-year PhD journey. I would also like to thank Dr. Lei Zhang for providing all the discussions, valuable feedback for all publications, and selfless help of daily life and study career. His wide knowledge, strong research enthusiasm, and hard-working attitude have inspired me during my Ph.D. period and will profoundly affect my future career. It is a great honor to have Dr. Lei Zhang as my supervisor. Lastly, thank you for always keeping me motivated, being patient and sharing your knowledge to be a better researcher.

I would also like to thank Dr. Yao Sun, Dr. Oluwakayode Onireti and Prof. Muhammad Ali Imran for sharing their knowledge on machine learning and communication technology. Thanks for their suggestions and key insights on each of my publications. I could not finish all of these works without their endless help. Thank you for all the guidance, help, and sharing your knowledge in this field, which motivated me to study and learn algorithms in this field.

I would also like to acknowledge the support and help of all my friends during the Ph.D. Without them, this journey would have been much more arduous. Thank you all for the discussions in and out of the university.

I would like to thank my girlfriend for supporting me in my Ph.D. study time, especially during the Covid period. Without her, this four-year journey would be much tough. Thanks to her care and help, I could do my research without any concern.

Last but not least, I would like to thank my family for all the support throughout this journey, especially my parents. It has been three years not going back home to visit them due to the Covid. I cannot express how much sorry for my absence in this period. I appreciate all the support from you guys and miss you so much.

Chapter 1

Introduction

The past decade has witnessed that our lives are increasingly inseparable from the mobile network, due to the wide variety of mobile applications and services in our lives. Nowadays, people worldwide are now suffering from the coronavirus (COVID-19), we have to reduce the gathering time to minimize the risk of infection. However, thanks to the mobile network, there is still a ray of sun in our lives. People can transact business by performing teleconferences when they are needed, watch their favourite videos and listen to the music to entertain their lives, talk to distant relatives, stream audio/video whenever a special event happens, share their daily lives in the social media with their friends, and so on [7].

However, with the popularity of mobile services and applications rapidly increasing and the initiation of new types of mobile devices, there is an exponential growth in network traffic, which is expected to increase thousands of times. The number of devices connecting to the mobile network is predicted to be about fifty billion by the next few years [8, 9]. In this case, the next generation mobile communication and beyond will have to address the limitations of the current mobile network and push the network performance to the next level, in terms of lower latency, larger capacity and more reliability. Therefore, there are some basic requirements which are repeated in the state-of-the-art works about next generation networks [7–10]:

1. Higher capacity: instead of 20 or even 100 MHz, the next generation network could provide devices and base stations (BSs) to use up to 800 MHz of spectrum at any one time;

2. Faster data rate: Next generation mobile networks could deliver up to 20 Gbps peak data rates, and at least 100 Mbps average data rates;

3. Ultra-Reliable Low-Latency Communication: Next generation provides ultra-high

network reliability of more than 99.999% and very low latency (of 1 millisecond) for packet transmission;

4. Massive number of connections: Next generation mobile networks are expected to support up to 1 million connected devices per 0.38 square miles, which is 500 times larger than that with the fourth generation (4G) mobile network;

5. Sustainable cost: Next generation mobile networks provide higher network energy efficiency;

6. Better quality of service (QoS) and quality of experience (QoE) to users.

7. Improved security and privacy.

As it can be seen, compared with the 4G mobile network, there is a great improvement of the network system performance in the next generation and beyond the network. However, there is an increasing prominent contradiction between capacity requirements and spectrum shortage. With the huge bandwidth in millimeter wave (mmWave), ranging from 30 to 300 GHz, mmWave communication have already become an essential part of next generation network. Further, with the requirements of next generation network are getting more stringent, making the new technologies more and more urgent. Therefore, several breakthroughs of the new technologies have been and are being discussed and developed in the literature for a couple of past years, among which the most popular and important ones are: massive multiple-input multiple-output (MIMO), network function virtualization (NFV), beamforming, implementation of self-organizing networks, and network densification - deployment of several small cells (SCs) dense cellular network [11–14]. Although all the breakthroughs are essential for the future mobile networks, the concept of network densification, as the cornerstone of the next-generation network and beyond, is the one that requires the heaviest changes [15]. Due to the characteristics of the mmWave, which is the main band of the next generation network, the dense cellular network with small cells plays a key role in the next-generation network and beyond. The deployment of a dense cellular network with mmWave could most likely solve the limitations of coverage, capacity and traffic demand, and meanwhile provide higher data rates, lower latency and better QoE/QoE with the users [16].

However, while we are enjoying the benefits of network densification and its additional technologies, some new problems are generated for the operators regarding network coordination, configuration, and management. There is no doubt that the increase of user equipments (UEs) and the deployment of a dense cellular network with mmWave would

bring a tremendous rise in the number of mobile nodes and data. They are required to be managed by the mobile operators in order to provide the users better service and maintain the basic demand of the mobile networks. Since in this case, there will be rapid growth of complex tasks to configure and manage, only applying the current techniques of network deployment, operation and management, could it not meet the demand of maintaining the network performance and stability while guaranteeing the QoS/QoE of the users [7]. One of the promising approaches to address these concerns is to design a more intelligent mobile network with artificial intelligence (AI).

On the one hand, with the massive growth of the traffic data in future mobile work, it is necessary to develop a novel method to manage the data more effectively and efficiently. In the future mobile work, there will be 10x users and 1000x traffic needed to support, which makes the design of the network considering not only the QoE/QoS of the users and the stability as well as tolerance of the network but also the energy and spectrum efficiency [17]. On the other hand, the complexity and dynamic of the communication system in the future, including ultra-dense deployments, heterogeneous nodes, networks, applications, different radio access networks (RANs) coexisting in the same setting, will be the next level [17]. Machine learning (ML), as one of the major AI algorithms, has attracted great attention in telecommunication. In the next generation network, plenty of problems are nonlinear, making the traditional mathematical too less efficient to solve the problems. To tackle these challenges mentioned above, ML could be the key technology for the future mobile network because of its considerable advantages when processing the data and learning from the experience, and environment [18, 19]. Further, with the 5G and beyond networks getting increasing complex, humongous data will be produced and generated. Data-driven and leverage AI techniques are required to manage the 5G and beyond networks efficiently [20]. Therefore, in the 3GPP Release 17 [?], a study is conducted on AI-enabled next generation radio access network, which investigates high-level principles, functional framework, potential use case, and associated solutions for AI-enabled radio access network intelligence. In addition, a 3GPP specification document is expected in Release 18 about data collection enhancements and signaling support for a set of selective AI-based use cases, such as network energy saving, load balancing, and mobility optimization. Therefore, the main objective of ML application in the mobile network could be defined as to provide intelligence to the mobile network that benefits the operators as well as users, reduce the overall network complexity, and simplify the network coordination, configuration, optimization and healing [21, 22].

1.1 MmWave Cellular Networks

When the previous generations of mobile networks were designed, there was a clear goal: providing connectivity to end-users and allowing them to communicate with each other. However, there are plenty of limitations of this inherited design nowadays. As the majority of the traffic in the mobile network currently has changed a lot, which is mainly based on the applications of the smartphone or other devices, such as watching the video, browsing the web, and the usage of apps [23, 24]. Further, the future generation of mobile networks is expected that people and machines will populate in. As such, the legacy design of previous networks will be more inefficient and could not meet the requirements of bandwidth, capacity, latency, reliability, new applications, and even new devices. To address these problems and design a more tolerant mobile network, much higher wave band and much dense network deployment need to be explored. Therefore, mmWave with bandwidth ranging from 30 to 300GHz, is a promising solution to improve the performance of the wireless communication system. Further, network densification is considered as a critical component of future mobile networks in order for them to cope with the expected exponential increase in traffic and capacity [7]. Thereby, the mmWave ultra-dense cellular network system will be an essential part of next generation and beyond network architecture.

1.1.1 Characteristics of MmWave Frequency

With the explosive growth of mobile traffic demand, there is an increasingly prominent contradiction between capacity requirements and spectrum shortage [25]. The capacity for wireless communication depends on spectral efficiency and bandwidth, which are also highly related to the cell size [26]. With the cell sizes becoming smaller, the physical layer technology is reaching its boundary of Shannon capacity [27], which makes more bandwidth needing to be explored. For previous generation mobile networks, almost all wireless communication spectrum ranges from 300 MHz to 3 GHz. Although the derives of this band benefits from the reliable propagation characteristics, which is over several kilometers in the different radio environment, with the rapid growth of the data traffic and UEs in the current mobile network, the efficiency and capacity in this band suffer deterioration [28]. It makes the expectation from sub-mmwave to accommodate the exploding mobile traffic and connectivity questionably. Therefore, with abundant bandwidth, ranging from 30 GHz to 300 GHz and immunity to interference, mmWave communication has been deemed a promising technology to improve network capacity dramatically in the

next generation mobile networks.
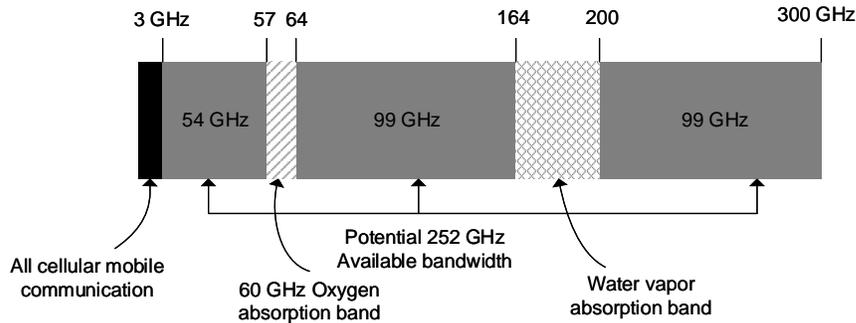
Figure 1.1: MmWave spectrum.

From the telecommunication history, the mmWave band has been utilized in many areas in the last few decades, i.e., radio astronomy, radars, airport communications, and many military applications. For example, the US Federal Communication Commission (FCC) opened the spectrum between 59 - 64 GHz and 81 - 86 GHz for unlicensed wireless and peer-to-peer communications, respectively [29]. From Fig. 1.1, within 3 GHz to 300 GHz, only 57 - 64 GHz and 164 - 200 GHz are not appropriate for telecommunication due to the absorption of oxygen and water vapors. Compared with the previous generation mobile network, hundreds of times more data and capacity could be supported with the mmWave band. The availability of most of the mmWave spectrum is opening up a new horizon for spectrum restrained future wireless communications [28].

Before benefiting from the mmWave band, the peculiar characteristics of mmWave communication should be considered. In the following sections, some main characteristics of mmWave are summarized, such as propagation characteristics, directivity, and sensitivity to blockage.

Propagation Characteristics

Due to the much higher carrier frequencies, there is a non-negligible drawback of mmWave communication: huge propagation loss compared with other communication systems using lower carrier frequencies. Thereby, the effective propagation range of mmWave is highly constrained by the rain attenuation, atmospheric absorption, and molecular absorption [30], which is shown in Fig. 1.2. However, with ultra-dense small cell cellular network deployed in the mmWave communication system, there will not be a significant additional path loss for cell sizes within 200 meters, which makes the mmWave communication mainly applicable to the indoor environment and small cell access and backhaul
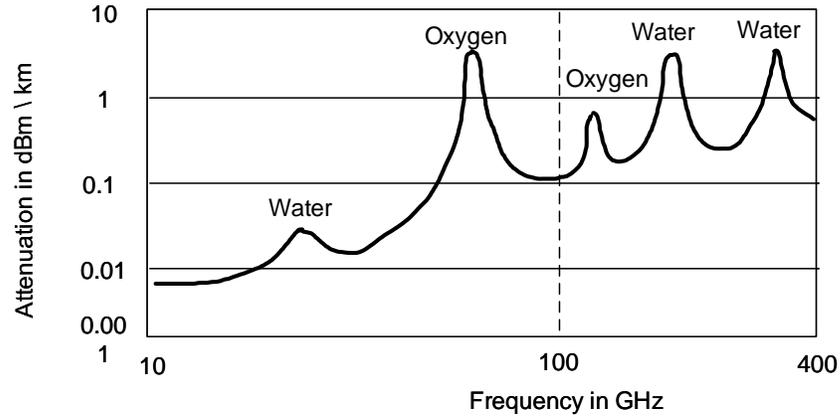
Figure 1.2: Atmospheric and molecular absorption at mmWave frequencies [6].

with the cell size smaller than 200 meters [31]. The non-line-of-sight (NLOS) channel suffers from higher attenuation than the line-of-sight (LOS) channel [1]. Table 1.1 shows the propagation characteristics of mmWave communication in different bands, including path loss exponent (PLE) with both LOS and NLOS channel [31–34]. In this table, the rain attenuation and oxygen absorption are at 200 m. As it can be seen, the mmWave propagation at 60 GHz and 73 GHz suffers more train attenuation and oxygen absorption than that at 28 GHz and 38 GHz. Further, the performance of NLOS transmission is worse than that of LOS transmission in the listed bands. To achieve the better system performance of high data rate and maximum power efficiency, LOS transmission is the main component of mmWave communication.

Table 1.1: The propagation characteristics of mmWave communication in different bands [1].

Frequency Band	PLE		Rain Attenuation@200m		Oxygen Absorption @200 m
	LOS	NLOS	5mm/h	25mm/h	
28 GHz	1.8 - 1.9	4.5 - 4.6	0.18 dB	0.9 dB	0.04 dB
38 GHz	1.9 - 2.0	2.7 - 3.8	0.26 dB	1.4 dB	0.03 dB
60 GHz	2.23	4.19	0.44 dB	2 dB	3.2 dB
73GHz	2	2.45 - 2.69	0.6 dB	2.4 dB	0.09 dB

In addition, penetrability is another considerable disadvantage for the mmWave band. Compared with the signal at lower frequencies, the penetration of mmWave signal with solid materials, e.g., brick wall and office white board is unacceptable. Table 1.2 shows

the comparison of penetration with some typical outdoor and indoor materials between mmWave at 60 GHz, and low frequency at 2.5 GHz [35]. There is a significant attenuation of drywall, whiteboard, and mesh glass while a slight attenuation of the clear glass from 2.5 GHz to 60 GHz.

However, in [36], the authors find that there is still a strong signal that can be detected within the range of small cell coverage even in a highly NLOS environment. Furthermore, spatial multiplexing and diversity can be supported at plenty of locations with multiple path clusters received. Although with dense base station deployment, signals are able to be transmitted through the LOS channel, due to the dynamics of the environment (e.g., blocking because of the UE mobility and the high deployment cost of mmWave), the LOS channel might not always be available. In this case, the NLOS link should be considered to solve the coverage issues, which can be established when a reflective path exists between the transmitter and receiver [37].

Table 1.2: Attenuation for different materials.

Material	Thickness (cm)	Attenuation 2.5 GHz [dB]	Attenuation 60GHz [dB]
Drywall	2.5	5.4	6.0
Office Whiteboard	1.9	0.5	9.6
Clear Glass	0.3 - 0.4	6.4	3.6
Mesh Glass	0.3	7.7	10.2

Directivity

Another characteristic of mmWave links is inherently directivity. The free-space propagation loss between isotropic antennas is scaled as v^2, where v is the carrier wavelength [38]. Take carrier frequency at 60 GHz and 5 GHz as an example. The propagation loss at 60 GHz is 21.6 dB, worse than that at 5 GHz for omnidirectional communication. However, when it comes to the fixed antenna aperture area, the directivity scales as $1/v^2$ and accounting for both transmit and receive antennas, the antenna gain could be given as $1/v^4$, which makes the overall antennas scaling as $1/v^2$. In this case, the propagation loss at 60 GHz is 21.5 dB better than that at 5 GHz with a directional transmitter and receiver. To take full advantage of mmWave band, beamforming technology makes the antenna array steers the beam towards one direction electronically and achieves the highest gain in this direction while providing much lower gain in other directions.

Sensitivity to Blockage

With the wavelength of electromagnetic waves increasing, its ability to diffract around obstacles is significantly limited when the size of the obstacle is larger than the wavelength [25, 38]. Taking 60 GHz mmWave as an example, its wavelength is about 5 millimeters, making the mmWave link blocked by the obstacles such as human bodies and furniture. Specifically, blockage by a human penalizes the link budget by 20 - 30 dB [38]. In the indoor environment, the mmWave channel is blocked for about 1% to 2% of the time for one or five persons, which makes the mmWave links intermittent [39]. For the outdoor environment, with the frequencies of mmWave increasing, the mmWave links suffer more from the human blockage [40]. Therefore, maintaining the seamless and reliable network connectivity and guaranteeing the QoS for some delay-sensitive applications, i.e., high definition television (HDTV), is a big challenge for mmWave communications.

1.1.2 Cellular Network on MmWave Band with its Key Techniques

Due to the exponential growth in network traffic and various new network applications, such as the Internet of Things (IoT), Internet of vehicles (IoV), Device to Device (D2D) communications, and Machine to Machine (M2M) communications, it is an extremely daunting task in 4G Long-Term Evolution (LTE) cellular network to support such data usage, and connectivity [41]. For example, considering the theoretical 150 Mbps maximum downlink data rate in 4G LTE cellular network, only $\lfloor (150/4) \rfloor$ simultaneous full HDTV at 4 Mbps rate could be supported with 2×2 MIMO [41]. Moreover, the 4G LTE standard cellular network was designed to support a maximum of 600 users that connected with radio resource control per cell. However, tens of thousands of devices are required to connect with M2M communication, and IoT [42]. Therefore, it is essential to design a new cellular network architecture that perfectly matches the mmWave band so that the benefits from mmWave could be fully enjoyed by the users.

Outdoor and Indoor Separated Architecture

Typically, wireless users staying in the indoor scenario is four times more than that in the outdoor scenario. For the previous generation cellular network, an outdoor BS is applied in the middle of the cell to serve both indoor and outdoor mobile users [43]. However, due to the peculiar characteristics of mmWave, as mentioned in the previous sections, it is no longer feasible with such a cellular network architecture. When indoor users communicate with outdoor BS, the signal has to penetrate the walls or glass, significantly reducing

the data rate, spectral efficiency, and energy efficiency with mmWave communication. Therefore, one promising solution is to design an indoor and outdoor separated cellular architecture to avoid the penetration loss through the buildings [44], which is shown in Fig. 1.3. To build such cellular network architecture, it would be very helpful with MIMO technology [45]. Specifically, some large antenna arrays would be fitted with outdoor BSs, distributed around the hexagonal cell and linked to the BSs with optical fiber cables. Further, for indoor environments, some large antenna arrays are also distributed outside of the buildings, such as large stadiums, shopping mall, and airport, since outdoor signals could not reach to all parts of such facilities [44]. There would be some wireless access points that connect with the antenna arrays applied inside the buildings to guarantee the communication of indoor users.
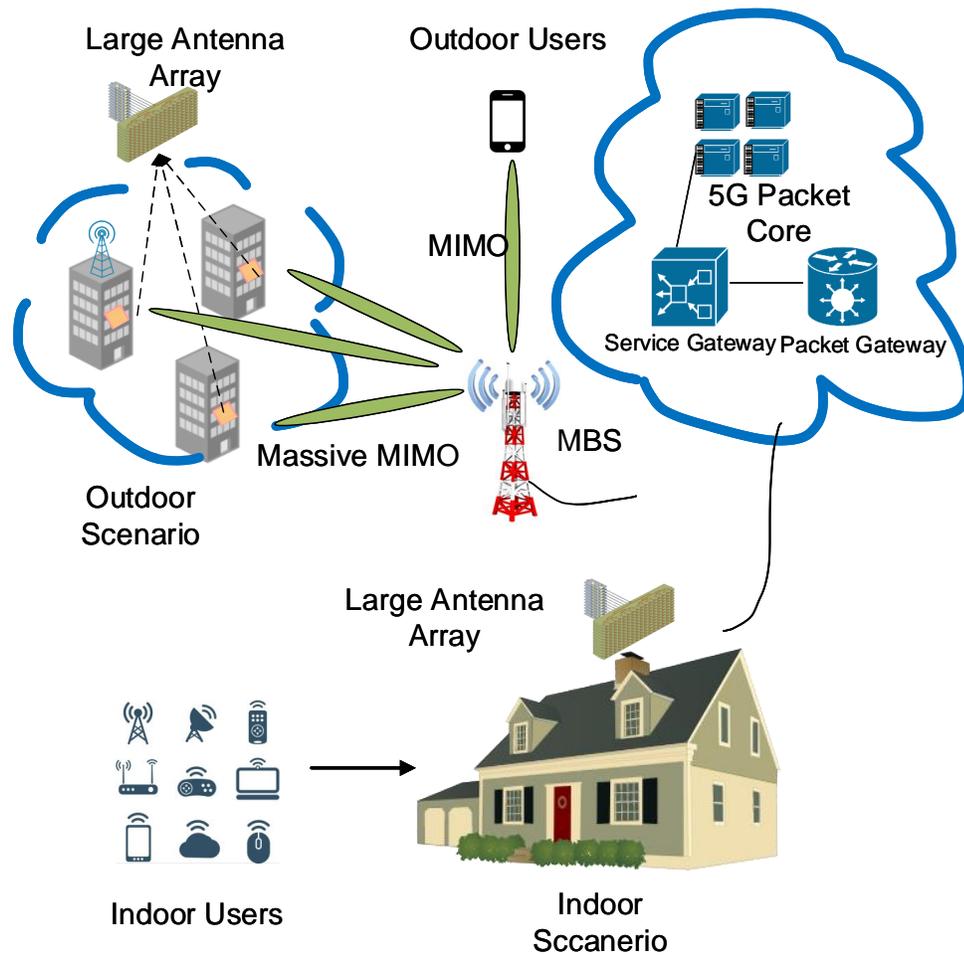
Figure 1.3: Outdoor and indoor separated mmwave cellular network architecture.

Heterogeneous Wireless Cellular Networks

To make full use of the mmWave band and overcome its shortages in terms of high path loss, directivity, and sensitivity to blockage, the next generation cellular network architecture is designed to be heterogeneous, in which there should be macrocells, mircrocells, small cells, and relays [41]. The typical heterogeneous next generation cellular network is shown in Fig. 1.4. In this cellular network, both 3.5 GHz small cell BSs and 60 GHz mmWave small cell BSs are deployed in the conventional macrocell to a multi-band heterogeneous network [46]. Further, the cloud cooperation concept is also introduced in this cellular network, connecting with all small cell BSs and the macro BS. The cloud could control the measurement and report between small cell BSs, UEs, and the macro cell. With this architecture, the limited coverage of mmWave cells could be well addressed.
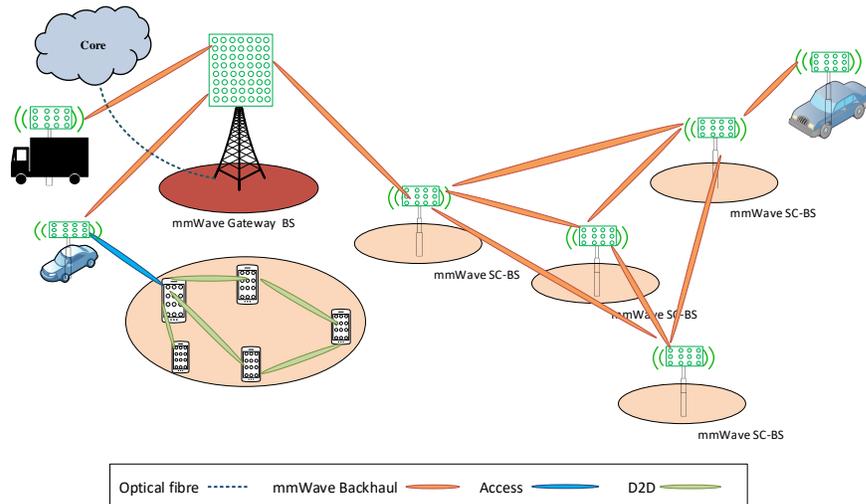
Figure 1.4: The architecture for next generation heterogeneous cellular networks.

Further, it is an integral part of the next generation cellular network with a mobile small cell concept, including mobile relay and small cells [47], which mainly serves the UEs in automobiles and high-speed trains. For example, small mobile cells are placed in the moving automobiles while massive MIMO units with large antenna arrays are set outside, which makes the UEs could communicate with each other not only inside the automobiles but also with outside BSs [47].

Massive MIMO

As the architecture of the next generation cellular network becomes ultra-dense small cells and heterogeneous, large antenna arrays will be deployed in fixed small cell BSs and

Table 1.3: Comparison of Traditional MIMO and Massive MIMO System [2].

	MIMO	Massive MIMO
Number of Antenna	≤ 8	≥ 16
Pilot Contamination	Low	High
Throughput	Low	High
Antenna Coupling	Low	High
Bit Error Rate	High	Low
Noise Resistance	Low	High
Diversity/Capacity Gain	Low	High
Energy Efficiency	Low	High
Cost	Low	High
Complexity	Low	High
Scalability	Low	High
Link Stability	Low	High
Antenna Correlation	Low	High

mobile small cell BSs. The MIMO system is already an integral part of current wireless networks to improve network performance, which mainly contributes to achieving high spectral efficiency and energy efficiency. However, with the exponential increases of users and their devices in the next generation cellular network, the current MIMO technologies related to the 4G LTE network cannot handle the huge influx in data traffic with more speed and reliability. With the introduction of concepts like IoT, machine-to-machine communication, virtual reality, and augmented reality, the current system cannot deliver the required spectral efficiency. Therefore, the massive MIMO is considered as a potential solution to solve the problem brought by the massive data traffic and users [48, 49]. With the advancement of contemporary MIMO, massive MIMO could group even thousands of antennas at the base station and serve tens of users simultaneously. With the help of the extra antennas applied by massive MIMO, the transmission energy could be focused into a smaller region of space to offer better spectral efficiency and throughput and reduce the interference to the neighboring user [50]. The massive MIMO's immense advantages compared with the traditional MIMO system are summarized in Table 1.3. Further, the benefits massive MIMO bringing to the next generation cellular network are:

1. Spectral and Energy Efficiency: With the antenna array of massive MIMO focusing narrow beams towards users, ten times higher spectral efficiency could be provided than the current MIMO system in the 4G LTE system. Further, less radiated power and energy are required since the antenna array focuses in a small specific section.

2. High Data Rate: The array gain and spatial multiplexing provided by massive MIMO

increases the data rate and capacity of wireless systems.

3. User Tracking: Since the narrow beam is provided from the BS to the users, it is more reliable and accurate for user tracking.

4. Less Fading: There is less fading with a large number of the antenna at the receiver side [51].

5. Reliability and Low Latency: More diversity gain provide by massive MIMO increases the link reliability and reduces the latency on the air interference [52].

6. Robustness: Massive MIMO systems are more robust against unintended interference, and internal jamming [53].

With the proper use of massive MIMO technology, the medium access control (MAC) layer design can be simplified by avoiding complicated scheduling algorithms. Moreover, separated signals could be sent from BS to individual users with the same time-frequency resources [54], which consequently makes massive MIMO technology be a promising candidate for a next generation cellular network.

Beamforming

As smart antennas are large-scale used with massive MIMO in next generation cellular network and the high directivity of mmWave makes directional beams also integral to emerging next generation network, beamforming technology is an essential part of the next-generation network. On the one hand, beamforming technology perfectly matches the next generation cellular network architecture. With its directional beam and powerful beam gain, beamforming could help the BS search for a suitable route to create mmWave access and backhaul links and deliver data to the users. Further, it also reduces the interference with nearby users along the route [55]. On the other hand, since the massive MIMO technology and smart antennas are indispensable parts of the next-generation cellular network, beamforming could increase spectrum efficiency and boost the data rate. Since the penetration through obstacles of mmWave is weak and the propagation loss for long-distance transmission is high, beamforming could send concentrated beams towards the users and make users receive a strong signal without interference.

With beamforming antenna used in 5G cellular network and beyond, there is great benefits of increased diversity for BS and/or UEs [56]. Specifically, the beamforming enables increase of capacity in wireless communication networks by reducing multipath

fading and channel interference. Thereby, the concentrating signal radiation could be realised in the anticipated direction. Further, it also modifies such radiation according to the signal surrounds or varying traffic situations with beamforming techniques. The main advantages of beamforming techniques is shown as follows:

1. Enhanced energy efficiency: The lower power requirements of beamforming antennas for transmitting signals to the intended user and cost reductions result in the lower power consumption and amplifier costs of massive MIMO systems [56]. Specifically, for each power consumption of each BS, the overall energy efficiency is relatively unaffected by the number of working antenna elements in the cells. Therefore, to obtain the high cost-effectiveness and overall energy efficiency, the common number of working antennas could be implemented for the entire cells in the system [57].

2. Improved spectral efficiency: For uplink and downlink signal power control, the utilization of the information for the training sequence, and the improvement of signal quality by beamforming antenna elements could enable capacity improvements. By installing beamforming antenna arrays, the wireless communication system could have potential for improving the spectral efficiency [58].

3. Increased system security: The principle of beamforming is to steer the transmitted signal power to the intended user. With this, the receiver will be the only party to recover the signal. Therefore, the probability of an eavesdropper receiving the transmitted signal will be smaller than when using conventional antennas, thereby the physical security could be achieved [59].

4. Applicability for mmWave bands: As is discussed, the main disadvantages of mmWave band are its poor propagation characteristics and short transmission distances. Therefore, highly directive antennas formed by beamforming technique must be applied to overcome this limitation.

.

When antenna scale greatly increases, it is not realistic to fully exploit MIMO gain by pure digital beamforming in baseband because of the problems on hardware cost, power consumption and standardization complexity [2]. Combining the advantages of analog and digital beamforming, hybrid beamforming is a potential architecture that simplify the architecture of mmWave antenna array [60].

1.2 Challenges in MmWave: NLOS Transmission, Handover and Beam Management

This thesis focuses on addressing some challenges of next generation network. To make full use of mmWave, the architecture of next generation and beyond trends to be more dense and heterogeneous, which makes some aspects of network more challenges. Specifically, to provide better QoS and QoE for users in the such network, how to provide efficient and effective handover for mobile users is important. The probability of handover trigger will significantly increase in the next generation network, due to the dense small cell deployment. Since the resources in the BS is limited, the handover management will be a great challenge. Further, to generate the maximum transmission rate for the users, LOS channel would be the main transmission channel. However, due to the characteristics of mmWave and the complexity of the environment, LOS channel is not feasible always. Therefore, it is worth to exploring the NLOS channel as an assistant for communication in next generation network, when LOS channel is blocked. In addition, since there is a huge demand of beamforming for mmWave communication, which makes beam management as a cornerstone of handover management and NLOS transmission. Therefore, more effective and adaptive beam management is required. In this context, the challenges of NLOS channel, handover mechanism, and beam management are introduced in the following sections.

1.2.1 Challenges of NLOS Transmission: How to Combat the NLOS Transmission in MmWave

In the mmWave cellular network, due to the characteristics of the mmWave band, LOS mmWave channels are expected, where each small cell BS will forward its traffic to its neighbors, selecting among a broad set of alternative paths to reach the core. Further, with the massive MIMO and beamforming technology applied in the system, mmWave wavelength enables higher antenna gains, resulting in highly directional links. These two reasons make the LOS channel as a main transmission channel in the next generation cellular network to achieve system performance and scalability. However, in practice, especially in the outdoor environment, mmWaves in 5G access and backhauling need to face NLOS conditions due to its sensitivity to blockage. NLOS channel has extra power loss due to scattering, diffraction and longer traveling distances, resulting in high bit error rate (BER) values, low throughput and notable performance degradation compared to LOS

propagation conditions. With the advanced antenna processing, it is possible to capture reflected signals (NLOS signals) and use them to supplement the LOS signal to increase channel capacity [61].

However, since the NLOS channel is more complex than the LOS channel, the main challenges of applying the NLOS channel focuses on NLOS identification and localization. The localization method for the LOS channel relies on different types of channel state information (CSI), including received signal strength (RSS), time of arrival (ToA), angle of arrival (AoA), and angle of departure (AoD) [62]. Nonetheless, when it comes to the NLOS channel, there is a huge positive bias in ranging [63], and spurious angular peaks for angle estimation, which induces considerable errors in NLOS localization. Moreover, with the size of next-generation network increasing and the architecture becoming complex, it is inefficient to identify and localize the NLOS channel based on the measured data approach due to a large time overhead [64]. Further, with the UEs rapidly increasing in the future network, there will be more interaction between different small cells and users. Since compared with LOS links, there are large number of NLOS links as a candidate to serve a user at each position. Therefore, effectively managing the NLOS links is a considerable challenge for NLOS transmission, especially for high mobility users.

1.2.2   Challenges of Handover

The future cellular network is going to support data-hungry applications with enhanced data rates via cell densification. As such, with more frequency movement of UEs, network densification and mmWave communication concepts will lead to more frequent handovers (HOs) [65]. It is essential to provide reliable HO mechanisms as this directly impacts the end users' QoE. However, there are some challenges of HO in next-generation networks. Firstly, since the network's size and complexity are going to increase greatly and the radio channel conditions are more dynamic, the users will be prone to more frequent HOs, which would greatly affect the QoS, especially for high mobility users and applications. Therefore, it is crucial to reduce the too frequency HOs, unnecessary HOs with Ping-Pong effect, increased HO failure rates, and increased HO delay for the HO procedure in next generation cellular network.

Second, due to the use of beamforming and massive MIMO technology, directional beams transmitted in the network will become dominant. In this case, the presence of obstacles on the path of the transmitted beam can completely stop the user from gaining access to the network, which greatly impacts the signal quality. Therefore, selecting the optimal beam when users switch from one cell to another is another important factor to

consider in the HO management process. The massive number of beams that the users need to select could be more complex.

Finally, some high mobility-based essential services for emergency scenarios, e.g., medical services to patients in ambulances, is also a crucial part of the mmWave network. In this case, how to design an effective HO optimization that helps predict the ambulance's route, determine the optimal BSs and beams, and pre-allocate the resources in BSs is another challenge of the HO procedure.

In summary, the design of the HO scheme in the next-generation network is required to select the optimal BS and beam for user connection to maximize QoS, reduce excessive or unnecessary HOs, and enable the detection of obstacles and their avoidance.

1.2.3   Challenges of Beam Management

As mentioned in the previous section, massive MIMO is one of the promising techniques for the next generation network. To maintain the complexity and implementation cost of the next-generation network low, beamforming with a large-scale antenna array is a common design to improve the performance of the next generation network system [66]. Beam management as one of the fundamental features of beamforming is an important procedure. However, there are still some challenges due to the complexity of next-generation network and the great number of UEs in the network, which is summarized as follow:

1. Beam Prediction: The excessive overhead of beam-training poses a significant challenge in mmWave communication, with massive MIMO technology applied for both transmitter and receiver. Especially in the NLOS channel, there are plenty of NLOS paths for each position of UE due to the reflection. To reduce the excessive complexity, a more efficient method needs to be explored for predicting the optimal high-resolution beam with low-resolution beam search at a low overhead [67].

2. Beam Tracking: There is a crucial challenge for beam management to handle the rapidly time-varying channels, making beam training for maintaining seamless, high-quality services more frequently. Therefore, predictability of UE's trajectory and effective beam tracking method is essential for the NLOS channel to reduce the beam management overhead.

3. Mobility of UE: Since the small cell BS will be more dense in the next-generation network and tens of thousands of UEs supposed to be served at the same time, with the UE moving from one cell to another more frequently, it is important to identify the mmWave channel effectively and accurately.

As can be seen, with more UEs joining in the network and the architecture of future generation networks becoming more dense and heterogeneous, there will be various sources of nonlinearities in the beam management procedure. To handle this, the conventional mathematical methods usually ignore these nonlinear factors and make the simplification. Therefore, more effective and efficient methods need to be explored.

1.3   The Role of Machine Learning Playing in Cellular Network

The advances of technologies nowadays, such as the miniaturized and smart electronic devices, the increased computing power of machines and the ubiquitous connectivity of the Internet, have produced trillions of bytes of information and data anytime, which is known as big data [68]. Our wireless communication network creates most of such a massive volume of data via our mobile phones, automobiles, home applications, and other communication devices. As such, it is crucial to explore the benefits of big data with data analytics and make full use of it to create a better and more profitable solution to serve the users in the network [69].

In this case, a widely-used AI algorithm named ML has attracted great attention. Different from other methods, ML is based on collecting and analyzing data to explore the intricate patterns and relationship of it and create a model related to the input and output data, rather than develop a complex and complete model of the system [70]. With future networks becoming more complex and dynamic, it is impossible to build a fixed mathematics model of a problem and find the solution. With the help of ML, the network system could learn, reason and make decisions without human interventions. Further, with the great development of the advances in electronics and cloud computing, the ability to store, process, and analyze either on the BS side or UE side has a great improvement, contributing to the feasibility of ML applied in the network system. In addition, another main advantage of ML is that it tackles complex and nonlinear tasks, which is a potential solution for most of the occasions in the next-generation network. For example, considering a task of mobile network optimization, if a model has to consider all possible situations of the network, with users in all possible positions and BSs with different configurations, it is impossible to build such a model. However, with the help of ML, it is more feasible to build a model when the data can be analyzed, and the model could be updated by learning from the environment and experience.

In this thesis, the two ML algorithms are main applied: Deep Learning (DL) and

Reinforcement Learning (RL), to solve the challenges in the mmWave cellular network, and the main advantages of them over traditional and analytical methods are summarized as follow:

1. With the massive volume and rapid data generation, it is infeasible to collect, store, and analyze data at a human scale. Applying the conventional data analysis techniques in the next-generation network is too inefficient, slow, and sub-optimal to process the data and generate adequate responses. As such, DL algorithms could enable more efficient data analysis and generate near-optimal results because they could analyze and process a large amount of previously stored historical data and learn from it.

2. Instead of relying on a fixed model or developing a model for every possible new situation, the DL algorithm enables generalization. Specifically, the model could be optimized online, which results in a model that could be synchronously and constantly improved and enhanced to suit new trends or drifts in data. With the model updated and enhanced, its robustness and reliability could be significantly improved.

3. Compared with traditional optimization methods, RL relies on temporal differences, which do not require the environment model or any previous knowledge. RL could learn from experience and interact with the environment with the inherited goal-seeking approach. As the environment of the next-generation network becomes more complex and nonlinear, this characteristic shows a great advantage to solve in the areas, such as beam management and handover management.

4. There is a clear goal designed for RL agents, which makes them tackle the whole problem rather than divide it into smaller sub-problems. It leads to an RL algorithm that is particularly suitable for long-term versus short-term reward trade-off or optimization, which is the most common problem in mobile networks.

1.4   Motivation

Based on the previous section's general introduction of how ML can benefit the next-generation network, the specific reasons why ML is necessary for NLOS, handover, and beam management are detailed discussed.

1.4.1   Why ML is Needed in NLOS Identification?

The traditional approaches for NLOS channel identification and localization are commonly based on range estimates [71] or on the channel pulse response (CPR) [72]. We summarize the several solutions for NLOS channel identification as follow:

1. Some techniques are based on that NLOS mitigation goes beyond identification and attempt to counter the positive bias which is introduced in NLOS signals [73].

2. Several schemes focus on several redundant range estimates for LOS and NLOS channel, which aims to reduce the impact of NLOS range estimates on the agent position that is underestimation [74].

3. Other methods attempt to detect the earliest path in the CPR in order to estimate the TOA in NLOS conditions more accurately [75].

However, there are some non-negligible drawbacks of traditional approaches for NLOS channel identification. First, there is a loss of information due to the direct use of ranges rather than CPRs. Second, significant latency is incurred when collecting range estimates to establish a history. Third, it is difficult to determine the joint probability distributions of the features required by many statistical approaches.

To overcome the limitation of the traditional methods, the ML algorithm could be applied. On the one hand, ML algorithms, especially DL, are well suited for addressing classification problems. Therefore, ML-based techniques are more promising to provide more accurate results since ML could extract the pattern from multiple dimensions of channel features. On the other hand, most of DL algorithms do not rely a fixed model. By inputting the data of highly related features such as ToA, AoA, RSS, and AoD, a neural network (NN) could be well trained to make the classification. With the trained NN implemented online, the channel identification could be time-varying, which significantly reduces the latency while improving efficiency and matching the dynamic environment in the cellular network.

1.4.2   Why ML is Needed in Handover Management?

The optimization of HO is essential in the next-generation network when selecting the BSs/beams that a user connects to, which could minimize frequent HO due to the small footprint of mmWave BSs and beyond. Because frequent HOs bring more HO cost, thereby reducing the network throughput. However, it is difficult to handle the HO management well with conventional methods. The classical method for HO decision is based

on specific parameter measurement [65]. For example, the targeting BS selection is determined by distance or the BS, which can provide a higher KPI such as reference signal received power, received signal strength indicator, and signal-to-noise ratio (SNR). However, this method is inefficient in the mmWave network due to severe path loss and susceptibility to LOS blockage [76].

In this case, ML techniques could assist in bringing intelligence and making the network self-optimized. On the one hand, DL algorithms can learn various network characteristics from the data produced in the network, thereby optimizing different aspects of the network, since some hidden details and patterns could be captured compared with the analytical models [77]. On the other hand, RL could play a significant role in HO management when it comes to making a smart decision, such as BS/beam selection and trigger condition, which could further reduce the HO delay, computational overhead, and frequent HOs. RL could predict target BS and beam with adequate resources before HO occurs to ensure a seamless HO and improve the QoS/QoE of the users.

1.4.3   Why ML is Needed in Beam Management?

With the next-generation network becoming more complex and dynamic, beam management requirements are increasing. To provide seamless high-quality services, beam management needs to be adopted to acquire and track the optimal BS, and UE beam pair with maximum received power [78]. Thereby, the challenges of beam management in the next-generation network could be concluded as nonlinearity from non-ideal beams, dynamic blockages, complex environment, and nonlinearity from UE's mobility. In this case, the motivation of applying ML algorithms in beam management scenario is summarized as follow:

1. To address the nonlinear problem, conventional mathematical methods simplify the real-world model and consider an ideal model with ignored nonlinear factors [78]. However, ML algorithms could offer a solution capable of accurately modeling the complex nonlinear problems and then exploring the deep relationship among them, thereby facilitating efficient beam management.

2. Due to the complexity of the network environment, it is impossible to build a promising model which includes all the related parameters and features for beam management optimization. However, with the help of the ML algorithm, a robust and reliable model could be formed with complex features, and it could also be updated when new features are added in.

3. Due to the mobility of users and the sensitivity to blockage of mmWave, the blockages in the network tend to change dynamically. Conventional beam management methods usually detect the blockage using a received power threshold and then re-sweeping, which results in excessive overhead. With the ML algorithm integrated into beam management procedure, beam management could learn from the change of environment and previous experience. After an ML model is well trained, a continued online training process could be deployed, collecting training data and updating the model based on environmental variations.

1.5   Research Contributions

This thesis aims to explore the applications of ML in the mmWave cellular network in terms of solving some challenges of NLOS, handover, and beam management scenarios. As such, three different optimization scenarios are investigated, and different ML algorithms are considered. Further, the performance of each is evaluated and compared to other state-of-the-art solutions. Based on this, the contributions of this thesis are:

1. A general literature review of mmWave cellular network and ML algorithms are presented. Specific literature review on ML algorithms in NLOS identification, handover management, and beam management is introduced.

2. A deep learning enabled method to prediction the AOA and AOD in NLOS channel for mmWave communication is proposed in this thesis. Specifically, a procedure to predict the AoA and AoD both in azimuth and elevation in NLOS mmWave communications based on deep neural network (DNN) is proposed. Moreover, to make the AoA/AoD prediction more reliable, a UE trajectory prediction based on the dynamic window approach (DWA) is employed. To ensure the simulation data is more closed to that in practice, the simulation scenario is built in the three-dimensionality (3D) model with ray tracing technology. The robustness of the method is evaluated, and the comparison in terms of computation complexity between the proposed method and the traditional one is discussed.

3. A novel handover scheme aiming to optimize the overall system throughput as well as the total system delay while guaranteeing the QoS of each UE is proposed. Specifically, a multi-agent RL algorithm is designed to optimize the handover trigger conditions. Further, an optimization problem in conjunction with the RL algorithm is designed to select the target base station and determine beam selection. It

aims to evaluate and optimize the system performance of total throughput and delay while guaranteeing the QoS of individual UE after handover decisions are made. The performance of the proposed scheme and other RL-based handover method is compared and discussed.

4. A multi-agent RL-based beam management scheme is proposed. Compared with other research based on the UE side, the RL algorithm is applied on each small cell BS to maximize the system throughput while guaranteeing the quality of service. Numerical results reveal the algorithm's convergence performance and the superiority in improving the system's overall throughput compared with other typical RL algorithms and the traditional beam management method.

1.6 Thesis Outline

The rest of this thesis is organized as follows. Chapter 2 starts with a comprehensive background of mmWave cellular network and ML. Then, the detailed background and literature review of NLOS channel identification, handover management and beam management are introduced, respectively. Lastly, the state-of-the-art of mmWave cellular network and ML are summarized.

Chapter 3 is based on the procedure to predict the AoA and AoD in azimuth and elevation in NLOS mmWave communications based on DNN. It starts with the system model of the simulation design, including simulation environment design based on the 3D ray-tracing model, DNN structure, data collection, and trajectory prediction. Then, the procedure and basic principle of predicting the AOA/AOD of potentially best NLOS beam and UE possible trajectory is proposed. Finally, the discussion and analysis of the results and system robustness as well as computation complexity evaluation are introduced.

Chapter 4 is based on a novel handover scheme for mmWave network: an approach of integrating RL and optimization theory. It starts with the system model design of mmWave network topology and channel. Then the basic framework of method is proposed and discussed. After that, the design of an intelligent handover trigger condition scheme based on the RL algorithm for the handover decision of UEs is detailed. The design of optimal handover decision to manage the resources in BSs, improve the system performance, and guarantee each UE's QoS is proposed. Finally, the simulation results and the performance of the proposed scheme and other RL-based handover methods are compared and discussed.

Chapter 5 is based on a multi-agent RL-based beam management scheme. It starts

with the system model design of mmWave network topology and channel. Then we formulate the problem and design the algorithm to solve it. After that, simulation settings and some benchmark algorithms settings used to evaluate this method are proposed. Then, the numerical results and the evaluation of the proposed method are discussed.

Chapter 6 concludes the thesis and discusses the future trends. When ML applied in the next-generation network and beyond, it is essential to consider the privacy as well as security of the user, the data set for ML training, the deployment of ML and offline or online learning.

Chapter 2

Background and Literature Review

Before the specific challenges mentioned in previous chapter are addressed, this chapter presents the general background of cellular network. After that, the general background of handover and beam management in the 5G cellular network and beyond is presented, followed by a brief overview of ML. Finally, the state-of-the-art ML-based schemes to solve the problems for handover and beam management are discussed.

2.1   Background

2.1.1   The Evolution of Cellular Network: From 1G to 4G

Since early 1970s, the cellular network has been initiated and evolved for five generations. First generation only provided the voice communication facility while there are voice and data services offered by second generation cellular network. After that, there is a great evolution of third generation cellular network, which could provide the users image and video services. Then, in fourth generation cellular network, ultra-broadband internet access service meets the demand of increasing data traffic. Nowadays, for fifth generation cellular network, compared with the previous generation cellular network, high capacity, higher data rate, low End to End latency, massive device connectivity, reduced cost and consistent QoE provisioning are the main requirements [79]. Therefore, before the specific technologies are discussed, in this section, there is a general evolution of cellular network among different generations.

1. First generation - analog system: The first generation network was first launched in 1980s [80], which is an analog based system. Simultaneous listening and talking are not allowed at the beginning, therefore, Improved Mobile Telephone System

was developed then. In this system, there are two channels. One of them is used for sending while the other one is for receiving. Since it is the analog system, there is a significant security issues. For example, an unknown receiver with all-band radio could listen into the conversation. Further, paging networks was used at that time. The significant problems for paging network is that the data transmission rate is very low, thereby it caused large overloading and delay.

2. Second generation - digital system: Based on the digital transmission, the second generation cellular network was first introduced in late 1980s [81]. Compared with the first generation system, digital multiple access technologies such as time division multiple access (TDMA) and code division multiple access (CDMA) is used. Consequently, higher spectrum efficiency, better data services, and more advanced roaming could be offer by second generation cellular network. Further, another magnificent evolution of it is the global system for mobile services, which partly overcame the distance limitation for that in first generation system.

3. Third generation: The wide brand wireless network is applied in the third generation cellular network, which highly increases the clarity. Further, the transfer rate of third generation networks could be at least 2 Mbps, which made high-volume movement of data possible. However, still the packet transfer on the air-interface behaves link a circuit switches call, which leads to part of the packet connection less efficiency. Moveover, one of the big problem for third generation network is that the standards for developing the networks were different for different parts of the world, which makes the global communication extremely inefficient.

4. Fourth generation: The fourth generation cellular network is successor of the second and third generation families of standards. Mobile web access, IP telephony, gaming services, high-definition mobile TV, and streaming services are developed with the fourth generation network. Highly heterogeneous and time varying quality of service from the underlying protocol layers are required for the emergence of the applications for 3G and 4G wireless systems. The long term evolution (LET) standard has been commercially deployed since 2009 [82]. The target values of peak spectrum efficiency for LTE Advanced systems were set to 30 bps/Hz and 15 bps/Hz in downlink and uplink transmission respectively. Further, the enhanced MIMO channel transmission techniques and coordinated multipoint (CoMP) transmission/reception were considered as the key techniques for LTE.

with the popularity of mobile services and applications rapidly increasing and the initiation of new types of mobile devices, there is an exponential growth in network traffic, which is expected to increase thousands of times. The number of devices connecting to the mobile network is predicted to be about fifty billion by the next few years. In this case, the fifth generation mobile communication and beyond will have to address the limitations of the current mobile network and push the network performance to the next level, in terms of lower latency, larger capacity and more reliability. The general background has already been introduced in the Chapter 1. In the following section, there will be a specific introduction of the key techniques, such as NLOS channel, handover management, and beam management in cellular networks.

2.1.2 Handover Management

In the cellular network, handover management is the system ensuring that there is a continues connection from the UE to the network during mobility of user, which is one of the key detail for mobility management [83]. This section will first introduce some basic concepts of HO and then propose some new requirements for HO in a 5G cellular network and beyond.

Handover Procedure

In the cellular network, the general HO procedure is introduced as follows. The UE performs the measurements, such as signal strength or signal quality over a specific downlink reference signal from serving BS or the neighboring BSs. After the measurement is processed, if a certain condition is fulfilled, a measurement report is transmitted to the serving BS. Once the measurement report is correctly received at the serving BS, the HO preparation procedure from targeting BS to serving BS starts, and the request of HO is transmitted to the targeting BS. After HO command is successfully received, the HO execution phase starts, where the UE gets access to the target cell with cell synchronization and a random access procedure. In the HO completion phase, the downlink data path is switched to the target site by the user data gateway. As a result, the targeting BS starts receiving packets from it. Finally, the targeting BS transmits a HO complete message to the serving BS and serving BS releases some allocated resources to the UE. Based on that, one of the main objectives of the 5G cellular network and beyond is to provide the user seamless communications, which ensures that while the users are in movement, no communication interruptions are perceived at the application level. To achieve this, a reliable

HO mechanism that provides high data rates for moderate-to-high speed users in an urban environment is essential in the 5G cellular network and beyond.

Handover Types

HO events could be classified with different perspectives [84]. One of them is based on triggering events: QoS-based, load balancing, and coverage-based. Specifically, for QoS-based, HO is initiated if there is any neighboring cell that could provide better signal quality than the existing condition. In the load balancing-based scheme,, HO is initiated by the network to balance traffic load of cells. For the coverage-based HO triggering, HO is initiated to mitigate connection loss when the serving BS cannot provide coverage to the UE.

Further, HO types also could be classified by network types. In this case, it can be typed as horizontal or vertical HO, which is based on whether a handover takes place linking a single type of network interface or a variety of different network interfaces [85]. Specifically, vertical HO is the process of a mobile terminal among access points supporting different network technologies, while horizontal HO is for supporting the same network technology.

In addition, from the frequencies perspective, the HO types could be classified as inter-frequency HO and intra-frequency HO. Inter-frequency HO is the HO process of a mobile terminal over access points operating on different frequencies, while intra-frequency HO is the mobile terminal over access points operating on the same frequencies [86]. This type of HO is significantly common in the 5G cellular network and beyond since the architecture of it is heterogeneous.

Moreover, from the number of connections perspective, the HO types could be differentiated as hard HO and soft HO. In the hard HO, the radio link to the previous BS is released at the same time when a radio link to the new BS is accepted. Contrary to hard HO, a mobile node accepts a radio connection with no less than two BS in an overlapping HO region and does not release any signal until it falls below a specified threshold value. Note that soft HO is only available when the mobile node shifts in cells with the same frequency. Although soft HO prevents connection interruptions, it increases signaling overheads.

Handover Performance Metrics

To evaluate the HO performance, different performance indicators are applied. In this section, there is a brief introduction of some performance indicators summarized as follow:

1. HO Failures (HOF) Rate: HOF failures may occur at different stages in the HO process, which includes failures at radio link due to poor radio conditions for both uplink and downlink, failure to convey particular messages after a given number of re-transmission attempts, synchronization failures, and others [87]. HOF rate is one of the most important performance indicators to evaluate the HO strategies.

2. HO Rate: HO rate is the number of HOs per second. Typically, the HO rate grows up with the increase of UE speed and decreasing cell size. In a 5G cellular network and beyond, with the number of cells increasing and the coverage of cells decreasing, there will be more HO rates. However, too many HO rates will lead to huge power consumption as well as large signaling overhead. Further, the increasing of HO rates highly relate to the HOF rates thereby there is a trade-off between reducing the HO rate and HOF rates [86].

3. Ping Pong (PP) Rate: is the number of ping pong events during a given period of time. The ping pong event is the occurrence of HO between a serving BS and targeting BS, which is followed by another HO to the original serving BS [88]. All these events happen under a predefined and generally short time.

4. HO Energy Consumption: is the amount of energy consumed in a HO process [89].

5. Data Latency: is the period between the last data packet sent from the serving BS and the first package received by the targeting BS [90].

6. HO Interruption Time: is the time period in a HO process when the UE cannot exchange user plane packets with any BSs [91].

In a 5G cellular network and beyond, to optimize the HO procedure, HOF rate, HO delay, HO rate, PP rate, energy consumption, overhead, latency, packet loss, and HO interruption time are the major indicators that should be minimized. Meanwhile, the gain in average throughput should be obtained as large as possible.

HO Management in 5G Cellular Network and Beyond

The specific HO procedures in the 5G cellular network and beyond are shown in Fig. 2.1 [86]. Step 1-4 is the HO preparation stage, Step 6-7 is the HO execution stage, and Step 8-10 is the HO completion stage.

1. Step 1: In the serving gNodeB (gNB), the UE measurement procedure is configured based on the access restriction and roaming information. In this process, the UE sends a measurement report to the serving gNB.

2. Step 2: The serving gNB makes a HO decision based on the measurement report. After that, a HO Request message is sent from the serving gNB to targeting gNB.

3. Step 3: The targeting gNB performs Admission Control procedure when its resources could be granted.

4. Step 4: The targeting gNB sends the HO Request Acknowledgement message. When the serving gNB receives it, the data forwarding is initiated.

5. Step 5:Serving gNB sends the HO command and Sequence Number (SN) Status transfer message to the UE and target gNB.

6. Step 6: UE detaches from the old cell and synchronizes with the target cell.

7. Step 7: The Path Switch Request message from targeting gNB informs that UE has changed the cell.

8. Step 8: The downlink path switches towards the target side.

9. Step 9: The Access and Mobility Management Function (AMF) acknowledges the Path Switch Request.

10. Step 10: The success of HO is informed by targeting gNB to serving gNB. Targeting gNB triggers the release of resources by the serving gNB. Finally, the serving gNB releases the radio resources associated with the UE.

As mentioned before, the main challenges in 5G cellular network and beyond are providing mobility robustness and minimizing service interruption, which leads to some big challenges on HO management [92]. The first one is the frequent HO resulting dramatically increase in the HOF rate. The second one is the growing number of intra/inter-frequency measurements that reduce the mobile user's battery life. The last one is the increased overheads due to frequent HO at mmWave frequency, limiting the frequency resources for static users. Therefore, finding an efficient and effective method to manage the HO in 5G cellular network and beyond is essential. The ML-based HO management will be introduced in the later section.
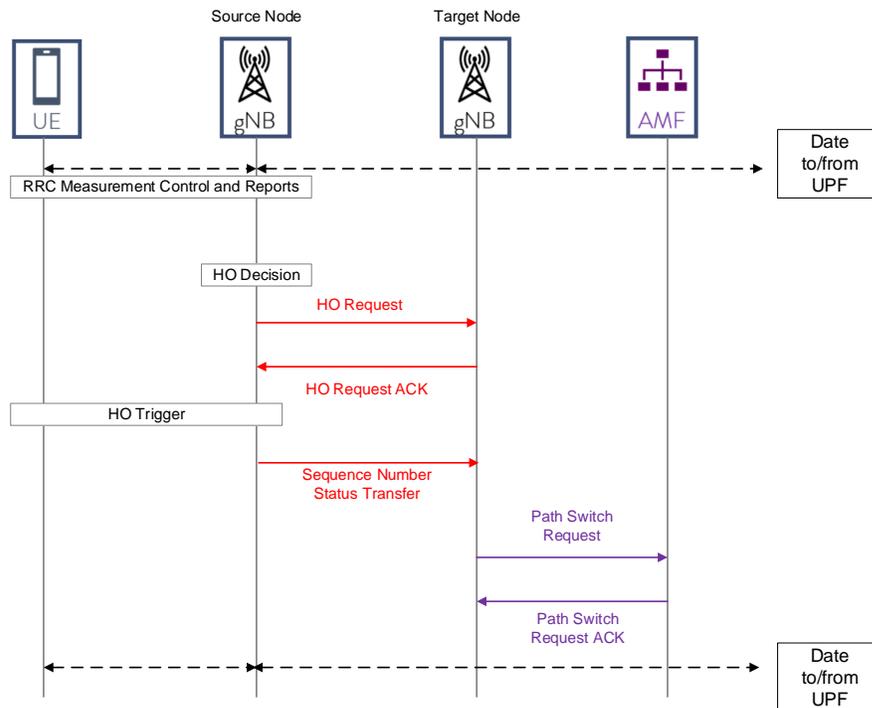
Figure 2.1: General handover procedure in 5G and beyond.

2.1.3 Beam Management in Next Generation Cellular Network

To make full use of the mmWave and provide the best QoS/QoE to the users, next-generation cellular networks must provide a set of mechanisms with which UE and mmWave gNB could perform highly directional transmission link [93]. It usually uses high-dimensional phased arrays to benefit from the resulting beamforming gain. In this case, directional links require a good alignment of the transmitter and receiver beams. To achieve this, the operation known as beam management is necessary to the 5G cellular network and beyond. Since in the mmWave cellular network, the beam pair link could be used for downlink and uplink transmission/reception, the beam management procedures contain the following six aspects [94].

1. Beam Sweeping: An operation of covering a spatial area, with beam transmitted and/or received during a time interval in a predetermined way.

2. Beam Measurement: An evaluation of the quality of the received signal at the gNB or UE. Different indicators could be used for evaluation, i.e., the SNR.

3. Beam Reporting: A procedure for UE to report information of beamforming signal based on beam measurement.

4. Beam Determination: The selection of matchable beams at the gNB or UE based on the beam measurement.

5. Beam Maintenance: A procedure for UE to maintain the candidate beams by beam tracking or refinement to adapt to the channel changes because of UE movement or blockage.

6. Beam Recovery: A procedure for UE to detect new candidate beam after detecting beam failure.

The procedures are periodically repeated to update the optimal transmitter and receiver beam pair over time. A detailed introduction of these procedures is proposed in the following sections.

Beam Sweep

Specifically, beam sweeping is a process in which the BS or the UE covers a spatial area by sequentially with different analog beams when the reference signals are transmitted or received [95]. In this procedure, there are pre-determined analog beam codebooks for BS and UE, from which the BS and UE sequentially use beams to find suitable Tx-Rx beam pairs for the data and control channels. On the transmitter side, beam sweeping is applied by sending beamformed reference signals, while on the receiver side, it is done by the received beams when measuring beamformed reference signals. There are three types of beam sweeping shown in Fig. 2.2.

1. Procedure 1 (P1) - Beam Selection: In this type, BS as well as UE perform beam sweeping. For example, in the downlink, the BS periodically sweeps its TX beams with synchronization signal. Meanwhile the UE applies a different Rx beam for each synchronization signal burst set [96]. P1 is the most common type of beam sweeping in particular.

2. Procedure 2 (P2) - Beam Refinement for the BS: Only BS performs beam sweeping, and the UE fixes the beam. This type is usually used to refine the BS beam after coarse beam alignment has been achieved [93].

3. Procedure 3 (P3) - Beam refinement for the UE: The UE performs beam sweeping while BS fixes its beam [93].
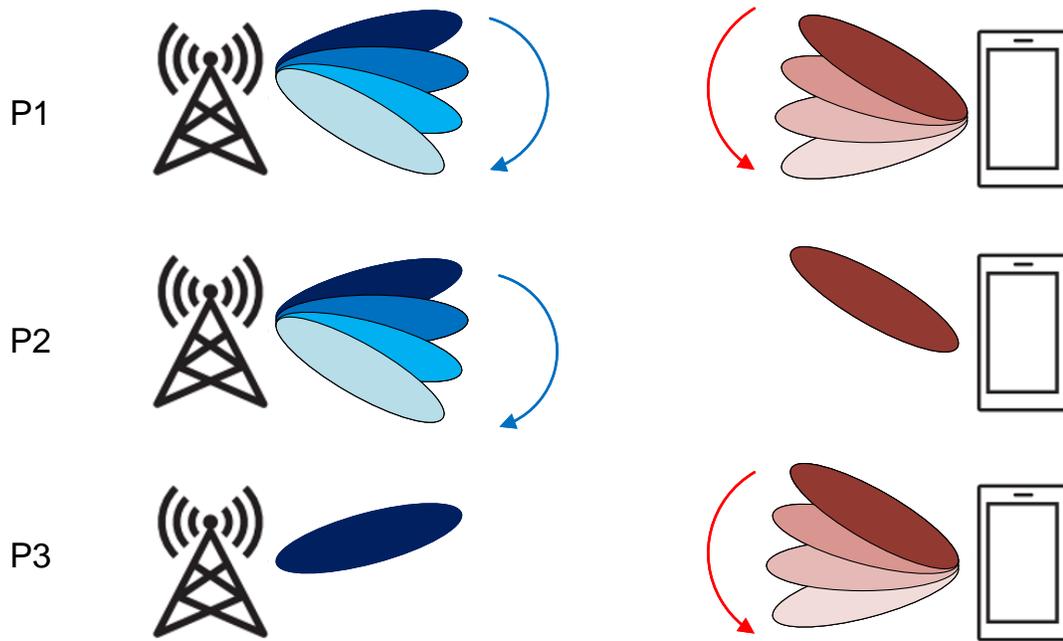
Figure 2.2: Beam sweeping procedure.

Beam Measurement and Reporting

In the uplink or downlink, the BS or UE measures the reference signals and reports related to reference signal received power (RSRP). Beam report by the UE requires to be configured by the BS. Based on the capability of different UEs, it could report the identifiers and measurements for up to fours reference signals per report. In addition, the beam report should be scheduled to be transmitted on either the physical uplink control channel or the physical uplink shared channel [93].

Beam Determination and Indication

Based on the downlink Tx beams, measurement quantities and group information reported by the UEs, the beams are determined by a transmit-receive point which is used for data transmission [66]. The most common case of beam determination by transmit-receive point is to follow UE recommendation and use the beam with the best-reported RSRP. However, considering the multiple perspectives such as multi-user transmission, interference coordination and channel reciprocity do not happen all the time. In this case, beam indication is required for the transmit-receive point to inform UE which beams are suitable for data transmission.

Beam Maintenance

Beam maintenance contains beam tracking and refinement, which is to handle the issues of beam misalignment due to the unexpected UE mobility and to support beam refinement for wide to narrow beams [66]. By detecting neighboring beams, beam tracking could efficiently track and compensate for optimal transmission direction change. For example, when the link qualities are lower than expected, the transmit-receive point and UE could directly detect alternative beams before switching and subsequently determine whether to switch its transmission beam group to an alternative group, reducing the outage of the beam failure probability.

Beam Failure Recovery

In the 5G cellular network and beyond, the sensitivity of blockage for mmWave and the dense UE with mobility will significantly lose beamforming gain when beam tracking fails. If beam maintenance is not functional and beam link failure occurs, the beam failure recovery procedure would be initiated for identifying new potential beams. In this case, first of all, there is a beam failure detection to evaluate whether the beam failure occurs. If so, the new candidate beam is identified. Then, the beam failure recovery request is transmitted to the BS, and UE will monitor the transmit-receive point response.

2.1.4 Background of Machine Learning

As is discussed in Chapter 1, ML algorithms could be a promising solution in modeling various technical problems of the next-generation network. In Fig. 2.2, the family-tree of ML techniques and their potential applications as a 5G cellular network and beyond. The following sections propose a discussion on different ML techniques, including supervised learning, unsupervised learning, and reinforcement learning, their working principles, advantages, and limitations in the wireless communication system.

Supervised Learning

Supervised learning is a ML technique which uses labeled data sets and explores a mapping of input function and output function, which is extremely helpful when address the data-driven problems of estimation, prediction, classification, and regression [17, 19]. There are training samples and feature in the labeled data set, which is usually divided in two sub sets: the training set and testing set. The training set is used to train the model
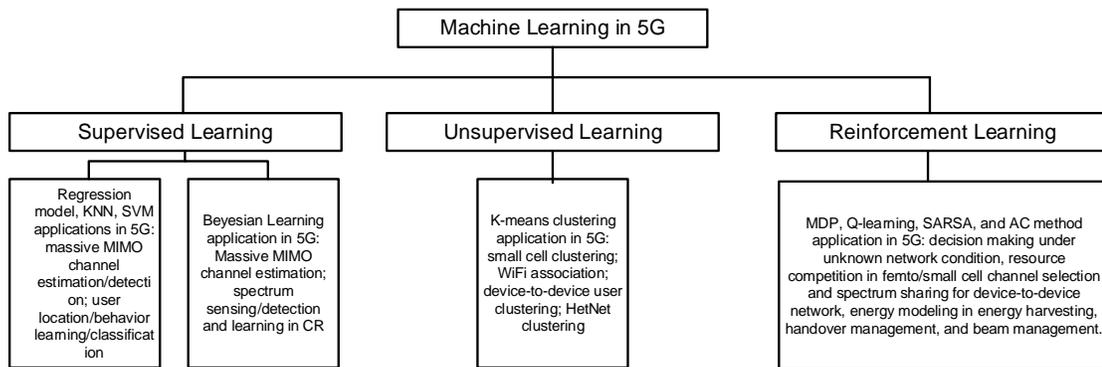
Figure 2.3: ML techniques categories and their applications.

while testing set is for evaluating the model and ensuring the accuracy of predictions. With a well-trained model, the error between predictions and actual values could be minimized [97].

There are two major applications of supervised learning, classification and regression. The classification is used when the output value we want to predict is discrete while regression problem is applied when the output value is a real number [98]. There are huge amount of supervised algorithms which can be found in literature. In the following, the most common of them will be introduced according to [99].

1. K-Nearest Neighbors (KNN) is a nonlinear method in which the input consists of k closest training samples in the input space. The predicted output is the average of the values of its k nearest neighbours. The Euclidean distance is the normally used distance metric for continuous variables. KNN could be helpful when addressing both classification and regression problems, with the advantages of being easy to interpret, fast in training, and the amount of parameter tuning in minimal. For example, in the next generation network, KNN could be used for finding the optimal handover solutions [100]. However, the accuracy of the prediction is generally limited and the output of KNN should be continuous.

2. Naive Bayesian (NB) is the method based on Bayes theorem, i.e., the probabilities is calculated based on the priors probability, which makes it a posteriori distribution calculation. It commonly is used for classification, which classify new data points when they arrive [101]. The core idea of NB is that it assume all attributes are conditionally independent, and is recommended when the dimensionality of the input is high. For example, authors in [102] design a channel estimation scheme for massive MIMO system. The advantages of applying this algorithm is only a small amount

of training data is required to extimate the means and variances of the variables.

3. Support Vector Machine (SVM) are inspired by statistical learning theory, known as a promising tool for estimating multidimensional functions [103]. The SVMs is based on nonlinear mapping. First, the raw data is transformed into a higher dimension where it changes into separable variables. Then, SVM searches for the optimal linear separating hyperplane which is capable of separating one class from another in this higher dimension. As such, the hyperplane with the main minimum distance from the sample points is generated. Finally, the sample points is called support vectors and create the final model. The advantages of SVM is that it could achieve high accuracy in prediction problems and nonlinear problems with appropriate kernel method applied. For example, the physical locations of nodes in an indoor wireless network and the channel noise in a MIMO wireless network are well estimated by SVM algorithm [104]. However, to maintain the high performance of SVM, the correct choice of kernel parameters is crucial, which is usually selected by exhaustive search, and thereby the model training of SVM is complicated [105].

4. Decision Trees (DT) is based on the flow chart model, where each internal node represents a test on an attribute, each leaf node is for a response, and each branch contributes to the outcome of the test [106]. DT could solve both classification and regression problems, with its configurable parameters, such as desired depth and the number of leaves in the tree. The advantages of DT is that it could be robust enough without any prior knowledge of the data. Meanwhile, it could achieve good results even on noisy data. However, like many classifiers, DT relies on the coverage of the training data and suffers with overfitting. Authors in [107] design a method to improve the handover trigger accuracy with DT algorithm.

5. Artificial Neural Network (ANN) is a statistical learning model which is inspired by the human brain. In interconnected nodes in the ANN is similar to the neurons in the human brain to produce appropriate responses [108]. ANN has the good performance for classification and regression problems, with its basic idea that efficiently train and validate the neural network. Then the well-trained neural network is used to predict or classify on the test data set. The main drawbacks of ANN is that it requires parameters or distribution model derived from data set, which means before input data in the neural network, there must be a data processing procedure. Further, due to the simply structure of neural network, the forward loop and feedback loop might stop working with training process, which leads to the neural network

stucking in local optimum and thereby reduce the accuracy of the results.

Further, as the derivative of ANN, Deep Learning is new trend in ML, which is a method of AI which understands the function of human brains and with this understanding to create patterns based on ANN [19]. It extends the typical ML algorithm by adding more 'depth' or 'complexity' into the model while transforming the data with various functions that allows data representation in a hierarchical way [109]. The most common DL algorithms are DNN, Recurrent Neural Network (RNN), Long Short Term Memory (LSTM) and Convolutional Neural Network (CNN). The strong advantage of DL is feature learning, i.e. the automatic feature extraction from raw data, and the features of higher level are formed by composition of the features at lower level [110]. Compared with classic supervised learning methods, DL could solve more complex problem with more accuracy and faster computation speed [111]. In this thesis, applying DNN to NLOS beam tracking problem is considered, thereby the detailed DNN introduction will be stated in Chapter 3.

Unsupervised Learning

Different from supervised learning, unsupervised learning receives unlabelled input patterns intending to explore a pattern in it [112]. In other words, when facing the unknown data set, unsupervised learning learns the difference by itself, and there are no correct answers to the problem provided. It is extremely useful for some problems requiring identifying anomalous behaviors, recognizing patterns or reducing the dimensionality of the data [78]. The aim of clustering is to identify groups of data to create a representation of the input. Non-overlapping, hierarchical and overlapping clustering methods are the typical category of it [113]. Specifically, overlapping clustering is when an observation can exist in more than one cluster simultaneously, of which Fuzzy C-means and Gaussian mixture models are the typical algorithms. Further, when the clusters at one level are joined as a cluster at the next level (cluster-tree), it is a hierarchical cluster method. Apart from these two categories, there is a non-overlapping clustering method. K-means [114] and Self-organizing Maps (SOMs) [115] are the two main algorithms of the non-overlapping clustering method.

The advantages and drawbacks of unsupervised learning are summarized as follows. Advantages:

1. Unsupervised Learning is less complex since there is no labeled of data required;

2. It is functioned in real-time;

3. It is significantly helpful related to the cluster analysis problems to track the hidden patterns.

Drawbacks:

1. The accuracy of unsupervised learning is less compared with other ML algorithms;

2. The computation complex is high.

Reinforcement Learning

RL is a method that learns from the interactions with the environment to achieve a certain goal [116]. In practice, it is impossible to provide explicit supervision to the training for some sequential decision and control problems. However, RL is good at making suitable decisions by mapping the situations to actions and evaluating which actions are necessary for maximizing a long-term reward. In RL, the learner or decision maker called an *agent*
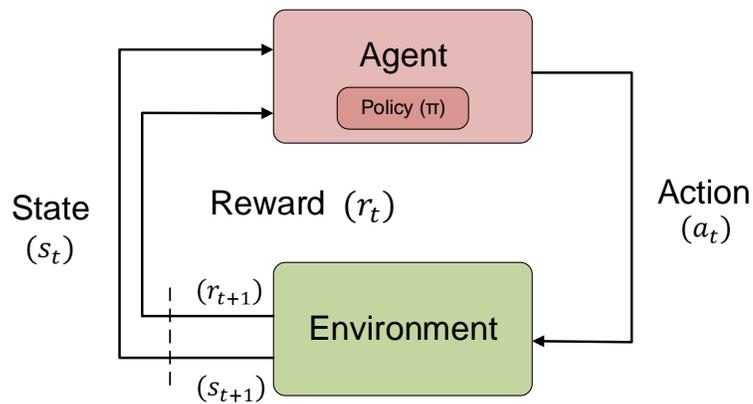
Figure 2.4: The Block diagram of a RL system.

interacts with its surroundings, the *environment*. The *agent* takes different *actions* and the *environment* responds to those *actions* and evolves into new situations called *state* by evaluating the *actions* with a *reward*, which is shown in Fig 2.3. Either positive or negative *reward* is stored as the learning experience. With the training processing, the *agent* learns from the experiences and take the optimal *actions* to obtain the positive *reward* for most of time.

The RL problem usually is defined as a Markov Decision Process (MDP) $\{\mathbf{S}, \mathbf{A}, \mathbf{T}, \mathbf{R}, \gamma\}$, where $\mathbf{S}$ is the set of possible states in the environment, $\mathbf{A}$ is the set of possible actions, $\mathbf{T}$ is the transition function, $\mathbf{R}$ is the reward function, and γ is a discount factor. At each time step, the agent explores a mapping from states to probabilities of selecting each possible

action. This mapping is the agent's policy. The goal of the RL learning process is to find an optimal policy for each state and thereby maximize some cumulative measure of the reward received over time.

There are two approaches to solve MDPs in RL algorithms: model-based and model-free. First, the model-based method includes the Dynamic Programming (DP) and Monte Carlo (MC) method.

1. DP is based on the knowledge of the state transition probability between two states after executing a certain action to address the MDPs. Further, DP breaks them into sub-problems and stores their results to avoid repeated computation when solving complex problems. Moreover, DP relies on updating the rules derived from the Bellman equation. Finally, DP is good at solving some problems, such as scheduling, graph algorithms, and bio-informatics.

2. MC algorithm relies on the experiences, such as sample sequences of states, actions and rewards. The application of the MC algorithm in practical cases is limited.

Second, in model-free method, Temporal Difference (TD) algorithm is the cornerstone of other algorithms, a combination of MC and DP ideas. In the TD algorithm, the learning agents could learn directly from their experience with the environment, without complete environment dynamics. Further, the estimates of TD are updated based on other learned estimates, which are similar to that of DP. TD methods have some common RL algorithms: Q-Learning, SARSA, and Actor-Critic (AC).

1. Q-learning: Q-Learning is one of the most popular RL algorithms, which is first proposed in [117]. It is a TD learning method that learns based on the action-value function $Q(s,a)$, representing the expected value of the agent being in a certain state and taking a specific action. On each time step, Q-Learning chooses an action to maximize its value function $Q(s_t, a_t)$, which indicates how good is taking action at a specific state according to the reward r. Typically, Q-Learning can be defined as:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_{t+1} + \gamma \cdot \max_a Q(s_{t+1}, a) - Q(s_t, a_t)], \qquad (2.1)$$

where $Q(s_t, a_t)$ is the current action-value function, α is the learning rate, r_{t+1} is the expected reward at the next time step, γ is the discount factor and $\max_a Q(s_{t+1}, a)$ is an estimate of the optimal future action-value function at the next time step.

Since there are two policies in the Q-Learning algorithm: one for generating its behavior ($\varepsilon - greedy$) and the other is to evaluate and improve, Q-Learning is an

off-policy algorithm. The advantages of this are that the estimated policy could be deterministic, while the other policy that controls the agent's behavior could continue to sample all possible actions.

2. SARSA (state-action-reward-state-action): SARSA also learns from the action-value function, $Q(s,a)$. However, different from Q-Learning, which chooses the next action based on the maximum expected value, SARSA chooses the next action according to the same policy used to choose the current action [118]. Further, SARSA utilizes the same policy to generate the behavior as well as evaluate and estimate the value of the action-function, and thereby SARSA is an on-policy learning algorithm. The definition of SARSA is:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_{t+1} + \gamma \cdot Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)]. \qquad (2.2)$$

As it can be seen, the difference between Q-Learning and SARSA is the target value, where the next action is chosen following a greedy policy in Q-Learning while it is determined by the current policy being followed in SARSA.

3. AC method: There is a separate memory structure to represent the policy independently of the value function in AC method [119]. Specifically, the policy structure is called actor, which is used to select actions, while the estimated value function is known as a critic, which learns and critiques whatever policy the actor follows and takes the form of a TD error. After each action is selected, the critic evaluates the new state to determine whether the results are better or worse than expected.

The RL algorithms mentioned above are some fundamental and common algorithms, and some new and derived RL algorithms used in this thesis, especially for the multi-agent scenario, such as multi-agent proximal policy optimization (MAPPO) and multi-agent deep deterministic policy gradient (MADDPG) will be detailed in Chapter 4 and Chapter 5.

2.2   Literature Review and State-of-the-Art

2.2.1   NLOS Propagation Channel identification with ML

One of the biggest challenges in wireless communications for mmWave is the NLOS propagation channel identification. At the frequency of mmWave, the propagation loss is more

severe than 2 GHz, or 5 GHz channels [120]. Therefore, the channel quality is highly dependent on the presence of the direct path of LOS components. Authors in [121] study the application of ML to channel classification for identifying whether a channel belongs to the LOS or NLOS classes. The existing LOS/NLOS classification approach is based on a computation of some metrics, followed by binary hypothesis testing. However, the ability of ML to work on multi-dimensional features and to identify patterns is shown to achieve better classification results than the single metric classifier. Moreover, the proposed approach only relies on the received preamble sequence compared with existing approaches, making the channel estimates using the received preambles. The authors use the Random Forest, which has higher accuracy than the other classifiers, and the learning algorithm performances are very good both on idea scenario and practical scenario.

Similarly, authors in [122] do the research using deep learning to identify the NLOS channel with commodity WLAN devices in an indoor scenario. Specifically, they design an indoor scenario, based on which they propose a RNN structure consisting of a LSTM block for identifying channel conditions to make the use of CSI more efficient. Then, they generate practical data from their indoor scenario for training and testing the neural network. Finally, they verify the performance of the proposed RNN model with conventional schemes to prove their work valuable.

In [123], the authors investigate the viability of using ML techniques for estimating user-channel features at a large-array BS. Specifically, they use user-channel features to predict the AOD of the dominant propagation paths in the user channel. The data is real data generated from a square of Tokyo. The neural network exploits different combinations of measured features at the BS to infer the unknown parameters at the users. From their preliminary results, deep learning could prove a valuable tool in allowing big data to predict unobserved channel parameters and improve network resource utilization and user experience.

Further, in [124], the authors find that since mmWave signals are blocked by many materials, small changes in the position or orientation of the handset relative to objects in the environment can cause large swings in the channel quality. They address the issue of tracking the SNR, which is an essential procedure for rate prediction hand and radio link failure detection. Specifically, they proposed a novel method for estimating the channel quality using synchronization signals and directional scanning. They derive an unbiased estimate for the instantaneous wide-band SNR in a particular pointing direction. Then, they evaluate the SNR tracking through real measurements using a novel high-speed measurement system. They experimentally measure the dynamics of the channel in various common

blockage scenarios using a high-speed channel sounder at 60GHz. Then they combine the measured channel traces with the statistical models to evaluate the SNR tracking algorithms. The result of their simulation shows that the SNR can be mostly tracked within a few dB of error, even when measurements are in very low SNR.

Lastly, we summarize the literature review based on the work proposed in Chapter 4, which states how ML applies to solve the NLOS channel beam management problems. The channel feasibility of mmWave NLOS outdoor mobile communication was demonstrated in [125] via an experimental measurement campaign. They found that although some well-known lossy objects, such as the human body, have poor penetration, they can be treated as good reflectors at mmWave frequency. In [126], the authors proposed an AI-enabled procedure to predict statistical channel characteristics based on CNN to obtain the mapping relationship between the location information of transmitter and receiver antennas. In [127], a channel condition identification method using a recurrent neural network structure with a long short-term memory block was proposed. Specifically, the authors classified NLOS and LOS channels and compared the ML method with the traditional method. An efficient deep learning model to predict optimal mmWave beam and blockage status was presented in [128]. Their method can predict mmWave beams and blockages with success probabilities and predict the optimal mmWave beams to approach the upper bounds while requiring no beam training overhead. Authors in [129] discussed and evaluated typical neural network architectures that are suitable to the beamformed fingerprint positioning problem in NLOS positions. Regarding trajectory prediction, an efficient vehicle trajectory prediction framework based on a recurrent neural network was proposed in [130]. In the framework, ML was employed to analyze the temporal behavior and predict the future coordinates of the surrounding vehicles. On the other hand, the dynamic window approach to reactive collision avoidance was proposed in [131]. Authors in [132], provide some use cases of beam management in vehicle-to-everything (V2X) scenario. Among them, the initial beam alignment, beam tracking, and beam recovery cases are also considerable in mmWave beam management in the urban scenario. There is a field experiment on the downlink throughput performance of a beam tracking in small cell BS presented in [133]. The authors prove that in the NLOS scenario, although the signal quality is reduced due to the reflection, it is still possible for the UE to connect to the access point through the reflected paths. A practical experiment is conducted in [134] to prove that in the NLOS mobility scenario, a connection between BS and UE can be maintained. However, the throughput is limited because of low effective scatters.

2.2.2   Handover Management with ML

When it comes to ML-based HO management, two broad categories are encompassed: visual data-based and wireless network-based HO management techniques. One application of visual data for HO optimization is the prediction of obstacles that might affect the magnitude of the received power or data rate at the user end. Authors in [135] propose a cooperative sensing scheme for proactive HO in mmWave networks with the images captured from multiple cameras and received power. The method is to map the images with HO decision with RL, which makes a proactive HO decision initiating before the received signal is blocked. With the help of multiple cameras, a complete view of the network environment could be generated. Similarly, the authors in [136] designed an RL scheme to optimize HO timing with camera images. To ensure the proactive HO is performed before data rate degradation occurs due to the blockage, they predict the future data rate of the mmWave link and select a potential link to serve the users. However, these two methods are all based on images. Although ML could handle images well, it takes significant computation complexity, bringing great latency to the HO management and reducing the HO accuracy.

Another use-case of visual data for HO optimization is to predict select the optimal beam to serve the users under the consideration of the mobility of the user and the blockage from the obstacles. Authors in [137] present a CNN-based decentralized architecture to reduce the overhead related to mmWave beam selection and LOS detection with data generated from Light Detection and Ranging (LIDAR) sensors and thereby optimize the HO procedure. Further, authors in [138] develop a DL-based centralized method to optimize the beam selection and LOS detection in vehicular networks with location information and LIDAR data. In this case, the HO management of high mobility users is discussed and optimized. In addition, the authors in [139] for the first time, design a realistic image data set for ML-based mmWave network optimization with rich environment dynamics. Based on the data set, they propose a vision-aided beam tracking framework to predict mobile users' future beams, thereby improving the HO performance.

For the wireless network-based HO management, a proactive HO framework that enables the user to switch connection to another BS before link disconnects is proposed in [140]. The method predicts obstacles and triggers HO before the link's disconnection occurs with ML, which ensures the reliability of the link and eliminates the HO delay due to the link disconnection. Further, authors in [141] propose an optimal BS selection approach based on RL to ensure the BS could serve the UE for a longer time after HO occurs. In their work, the post-HO trajectory of users and the blockages along the LOS is

considered to predict the next moment HO, which makes their method more promising. In addition, authors in [142] propose an intelligent HO decision scheme with double RL to optimize the BS selection. The proposed scheme aims to learn the optimal BS for user association to minimize the HO frequency and optimize the average throughput on the users' route. In [143], the authors proposed a DL model for user localization and proactive HO management; meanwhile, the behaviors of users in the network are considered. In the model, the received signal measurements are applied to reduce unnecessary HOs. Moreover, to ensure the system performance, the user's location is predicted, and the network's throughput is optimized. In [144], the authors develop a learning-based framework that jointly optimizes the HO and beamforming for mmWave networks with RL algorithms. Their framework determines the optimal backup BSs on mobile users' routes, thereby contributing to reducing the overall signaling during channel estimation and minimizing the HO frequency.

Lastly, the literature review based on the work proposed in Chapter 5 is summarized as follows. In order to improve the performance of handover in the mmWave system, some research work starts to exploit reinforcement learning with the consideration of different factors, including RSRP, QoS of UEs, UE mobility characteristics, BS load, etc. In [145], an algorithm with predicted channel information is designed to help UE decide whether to make the handover or not based on information such as UE speed and location. Further, a handover algorithm is proposed in [146], which focuses on context parameters, such as UE velocity, channel gain and cell load information. To maximize the average capacity of UE, MDP Process model is applied to help make the selection of BS. Similarly, in [147], the authors provide a handover algorithm with MDP, where the authors combine the handover overhead, cell load and channel condition in the reward function to achieve high throughput while decreasing the handover rate. Further, in [148], the authors propose a Q-learning based handover policy, in which the decision is learned with optimal policy without prior knowledge of the environment. The results show that the significant quality of experience performance is improved in heterogeneous mmWave networks.

However, the above research are focused on a single UE handover scenario. In practice, especially in the mmWave system, the handover rate is more frequent, and the cost of handover failure is inestimable with the UE increasing. In this case, authors in [3] design a smart handover policy for multi-UEs with different UE densities to reduce the handover rate while maintaining the QoS of the UE. Further, the authors in [149] propose a multi-agent handover algorithm with the actor-critic (AC) method. Specifically, the handover decisions are made by individual candidates' RSRPs and current connection information

with a shred artificial neural network. With the reward function's penalty added, redundant handover is significantly reduced. In [150], the authors propose a handover management and power allocation scheme to maximize the overall throughput and reduce the handover frequency. To achieve this, the authors develop a multi-agent RL algorithm based on the PPO method, which separates the learning process to centralized training and decentralized execution. In this case, the global information generated from BS could be used to train the UE at the initial stage. After that, the UEs make their decision and take action based on the local observation.

2.2.3   Beam Management with ML

MmWave bands are usually above 10G Hz with short wavelengths. Those short wavelengths with high-frequency bands enable implementations with many more antenna elements per system within a supersmall space [151]. However, it also increases the signal path and propagation challenges associated with operating at these frequencies. For example, due to the gas absorption, the attenuation for a 60 GHz waveform is more than 10 dB/km, while a 700 MHz waveform experiences an attenuation on the order of 0.01 dB/km. These losses can be compensated with the elaborate array design and the application of spatial signal-processing techniques, including beamforming. Beamforming can be enabled by large antenna arrays and can be applied directly to provide higher transmit gains to cope with the path loss and harmful interference signals. To achieve desirable flexibility and controllability with beamforming in the antenna array design, adopting an independent weighting control over each antenna-array element. This requires a transmit or receive component dedicated to each antenna-array element. Thus, beam management is very important. The goal of beam management is to choose the best pair of beams for analog beamforming, where both transmitter and receiver have antenna arrays with only one radio frequency (RF) chain and fixed beam codebooks.

Authors in [152] propose a novel framework to leverage machine learning tools with the availability of situational awareness in predicting mmWave beam power. Specifically, moving vehicles might be the most important mobile reflectors in the urban canyons. The authors use the vehicle locations as features to predict the received power of any beam in the beam codebook, with low or almost-zero feedback overhead. The vehicle location can be generated through the method mentioned in section 3 or obtained from the basic safety message in dedicated short-range communication. The whole process can be treated as a regression problem. The advantage of their method is that it gives extra information to the order of the beam pair power and helps to select the beam pair for data transmission

meanwhile, the accuracy of distinguishing between the "good" and "bad" beam can be very high.

Further, authors in [153] focus on multi-beam concurrent transmission, which is one of the promising solutions for a mmWave network to provide seamless handover, robustness to blockage, and continuous connectivity. Their work is to optimize the beam pair selection, which is essential to improving the mmWave network performance. They propose a novel heterogeneous multi-beam cloud radio access network (HMBCRAN) architecture that provides seamless mobility and coverage for mmWave networks. They also design a novel acquirement method for candidate beam pair links in HMBCRANs architecture to reduce user power consumption, signaling overhead, and overall consumptions. Specifically, a decentralized algorithm based on HMBCRANs architecture and binary log-linear learning is proposed to obtain the optimal pure strategy Nash equilibrium of the proposed game, in which a concurrent multi-player selection scheme and a piece of information exchange protocol among players are developed. The main contributions of these authors are that they solve the multi-beam management problem in multi-beam concurrent transmission mmWave networks to maximize network sum rate and prove that this method is more efficient. Secondly, they designed a new beam management scheme based on interference distance and interference measurement to reduce the complexity of the exiting schemes. Thirdly, they analyze the effects of multiple concurrent beams of a UE on its incomplete blockage probability and achievable rate and evaluate the performance of the proposed scheme with different system configurations.

In addition, the authors in [154] draw their attention to lens-based mmWave massive MIMO, which is considered a key technique for 5G wireless communications and beyond. However, most existing beam management schemes are designed for time-invariant channels. Although these schemes can achieve satisfying performance, they usually incur high complexity with a fast time-variant mmWave channel since the large range search will be executed frequently. Thus, they propose an adaptive neighborhood search (ANS) beam management to solve the problems. Specifically, by exploiting the mmWave channel property that the angles of departure of channel paths are slowing varying [155]. The process of this method is shown as follows. Firstly, they propose to use the beam selector in the previous time slot as the initial solution and perform the neighborhood search developed from machine learning to select the beam in the current time slot with low complexity. Then, they utilize the correlation between two neighboring beam selectors to simplify the neighborhood search computation further. Finally, they propose adjusting the neighborhood range to avoid the local optimum. Their simulation also showed promising results.

Hybrid beamforming is a promising low-cost solution for large MIMO systems, where the BS is equipped with fewer radio frequency chains. The selection of code words in the system is essential to optimize the uplink sum rate. Thus, authors in [156] proposed a data-driven method of analog beam management with ML to achieve a near-optimal sum rate with low complexity. Specifically, they take the beam management problem as a multiclass-classification problem, where the training data set consists of a large number of samples of the mmWave channel. The training data applies the SVM algorithm to obtain a statistical classification model, maximizing the sum rate. With the derived classification model, the low complexity of the optimal analog beam for each user can be selected for real-time transmissions.

In [157], authors make some effort on gathering the data for ML in the typical urban canyon scenario with ray tracing technology, which can provide very accurate results. In the configuration stage, the user provides information to enable the conversion of coordinates between the two main software. To facilitate the interaction with the traffic simulator, the orchestrator associates each mobile transmitter or receiver to a mobile object (MOBJ), which can also play the role of a blocker or scatterer, with no associated transceiver. In the simulation stage, the orchestrator written in Python code invokes the traffic simulator and then positions the MOBJs to compose the scene. Based on the output of the traffic simulator, some files of the base scenario are modified and stored in order to allow reproducing the ray tracking simulation of that scene. Their method provides an efficient way to generate the data from the urban canyon, which is important for ML utilized in the 5G system and beyond.

Lastly, the literature based on Chapter 5 is summarized as follows. Authors in [158] propose Q-learning-based single UE non-line-of-sight beam selection scheme in a mmWave system with good beam alignment. In [159], authors present a novel beam tracking solution that is based on a multi-agent Q-learning algorithm, which generates better spectral efficiency than the beam sweeping technique for multi-user MIMO cases. However, these two works do not consider the performance of Q-learning compared with other RL algorithms. Further, the authors in [160], apply the deep RL and clustering to optimize the beam management and radio resources allocation for the Ultra-Reliable Low-Latency Communication (URLLC) users. The results show that the proposed scheme outperforms the baseline algorithm regarding latency, reliability, and rate of URLLC users. In addition, in [161], the authors propose a MADDPG based beam management scheme to maximize the secure capacity by jointly optimizing the trajectory of Unmanned aerial vehicles (UAVs), the transmit power from the UAV transmitter and the jamming power from

the UAV jammers. To improve the learning efficiency and convergence, the authors also propose a continuous action attention MADDPG (CAA-MADDPG) method, and the results show that rewards performance of CAA-MADDPG is better than MADDPG. Moreover, the authors in [162] propose a Learning-based Cost-efficient Resource Allocation (LCRA) algorithm employing the deep RL to learn policies from experiences to ensure system performance while achieving cost efficiency. The results show the superiority of their method in improving the cost-efficiency in hybrid energy powered mmWave backhaul HetNet compared with other typical RL algorithms, i.e., proximal policy optimization (PPO) and actor-critic algorithms.

Chapter 3

Deep Learning-Enabled NLOS Beam Tracking

3.1 Introduction

The explosive demand in users' mobile data experience makes an increasing strain on the network's use of the available wireless spectrum. In order to solve this issue, one of the most important missions for the telecommunications industry is to explore higher frequency in wireless communication networks [163]. As such, in the fifth-generation of wireless communication networks, mmWave frequencies, ranging from 30-300 GHz, are being explored to overcome the spectrum shortage. With its rich spectrum resources, mmWave can support high data rate transmissions, which makes mmWave one of the most promising technologies in future wireless networks [163–167]. However, mmWave also faces some challenges, such as high propagation loss, resulting in short propagation distances, and signal blockage caused not only by building materials and foliage, but also human body and high oxygen absorption [58].

To address the path loss issues of mmWave communications, one effective solutions is beamforming [168], which brings plenty of benefits, such as better coverage at a cell's edge, improved signal quality, tracking the UE, and allowing cooperation among BSs. Although directional beamforming helps compensate for the significant path loss incurred by mmWave signals, it comes up with a complex beam alignment issue. More specifically, it is essential for a BS to know the AOA and the AOD of its users in order to determine the beamforming direction[1]. A natural approach to perform beamforming training to improve

---

[1]Note that the state of the art algorithms typically use precoding vectors for beamforming, which is essentially a function of the AOA and DOA.

the alignment accuracy is to exhaustively search for all possible pairs to identify the best beam alignment [169]. When there are only LOS channels in the mmWave communication system, the exhaustive search procedure has a calculation complexity of exponential growth [170]. With dense base station deployment, signals are able to be transmitted through LOS channel, however, due to the dynamics of the environment (e.g., blocking because of the UE mobility, and the high deployment cost of mmWave) the LOS channel might not be always available. In this case, the NLOS link should be considered to solve the coverage issues , which can be established when a reflective path exists between the transmitter and receiver [171, 172].

With NLOS propagation, multiple copies of the transmitted signals arrive at different times at the receiver, each with a different amplitude [173]. Due to the nature of narrow beams in mmWave communication, only limited number of angles can be covered by a beam. In this case, to identify the most suitable NLOS channel, the BS can search the obstacles surrounding the UE, such as buildings, and pick up the possible buildings as the reflector to form the NLOS channel. However, there could be many reflectors in the urban city scenario around UEs, leading to high complexity and latency, especially when UE has a mobility thus discontinuous angle change is expected due to the blockage. Moreover, these surfaces may have significantly different reflecting factors, which could affect the received signal power. To identify which surface is the best reflector, an efficient way is to find the AOA and AOD of the strongest received power beam of the NLOS path [174]. Therefore AOA and AOD for mmWave beam in NLOS scenario are the key parameters required for determining the suitable NLOS propagation path for a UE.

However, if current techniques, such as exhaustive beam search are applied, a significant overhead and a heavy computational burden can be imposed on the system. In order to solve this issue, ML [175] is a potential method, which has received great attention due to its capability of finding valuable and hidden patterns from huge unknown datasets, such as in channel information [126]. On one hand, ML is extremely flexible and accurate in making predictions. On the other hand, massive data in the communication system is easy be to obtained. Thus, communication systems can benefit a lot with plenty of data [176]. Further, when compared to traditional methods, ML can learn complex relationships between raw input and output data through a training process [127]. Based on the intrinsic parameters of the collected data found by ML, predictions can be done with a trained model. This brings some advantages to ML over mathematical methods, such as, not relying on a specific mathematical models, resulting in flexible and adaptable algorithms, and being able to learn just from data.

It is quite challenging to identify the AOA and AOD of NLOS channels in wireless networks, mainly due to the user mobility, since the surrounding environment is constantly changing. Especially, it is more practical to assume that the UE location is unknown in such an estimation [125]. In this case, a UE trajectory prediction algorithm could be a utilized. By generating the channel information of the whole NLOS area and training the DNN with part of the channel information, the trajectory of the UE to obtain its location information from the trajectory prediction algorithm (TPA) could be predicted. With the estimated location information as an input, the trained neural network is utilized to predict the AOA and AOD of the potentially best NLOS beam for each position on the UE's predicted path[2].

This chapter creates a NLOS simulation model to generate the datasets, consisting of received power, location, and the number of clusters from raw data obtained by K-means, which is used to train a DNN without UE monility prediction. This trained DNN is then used to estimate the AOA and AOD in given positions. After the training, a new dataset is formed, where the position of the UE is unknown to the trained DNN. In this case, the TPA is applied to predict the UE and generating the location information. Based on that, the DNN and estimate both the next position as well as AOA and AOD of NLOS channel is tested. Lastly, a comparison in terms of predicting AOA/AOD between CNN and the proposed networks is performed. The main contributions of this chapter are summarized as follows,

1. A procedure for predicting the AOA/AOD of the potentially best NLOS beam based on a DNN for a 3D mmWave outdoor scenario is proposed. With the dataset including, received power, location, and the number of clusters from raw data obtained by K-means, the trained neural network can predict the AOA/AOD of NLOS beams in the azimuth and elevation.

2. In order to make the simulation scenario more practical, it is assumed that the location information is unknown to the trained DNN. In this case, a robot path plan is utilized to design the TPA for UE to generate the location information. With the new location information predicted by TPA as an input of the trained DNN, AOA and AOD are estimated. Results show that the trained neural network can predict the AOA/AOD with very low loss around 0.02%. Moreover, the proposed DNN algorithm is compared with CNN in the case of training and predicting AOA and AOD with the dataset consisting of known location information to prove the

---

[2]Note that when the predicted UE location information is input into the trained DNN, the data is completely unfamiliar for the DNN.

proposed model is more suitable for predicting and estimating AOA and AOD in NLOS channel.

The rest of this chapter is organized as follows. The system model of the proposed simulation, including simulation environment design, DNN structure, data collection, and trajectory prediction are stated in Section 3.2. The procedure and basic principle of predicting the AOA/AOD of potentially best NLOS beam and UE possible trajectory is proposed in Section 3.3. Section 3.4, discusses and analyzes the results and the robust check on the proposed system is performed. Conclusion and future works are summarized in Section 3.5.

3.2   System Model

In this section, the channel model based on the research is proposed, specifically the received power and AOA/AOD.

In the system model, it is considered that there is a single BS and a single UE. The ray tracing software, Wireless Insite is utlized to build the simulation environment. Ray tracing is a classical deterministic method used or modeling radio propagation. By tracing paths in the simulation environment, the received power can be obtained as [177]

$$P_R = \sum_{i=1}^{N_P} P_i, \tag{3.1}$$

where N_P is the number of paths and P_i is the time averaged power in watts of the ith path. P_i is given as

$$P_i = \frac{\lambda^2}{\pi \eta_0} |E_{(\theta,i)} \, g_\theta \, (\theta_i, \phi_i) + E_{(\phi,i)} \, g_\phi \, (\theta_i, \phi_i)|^2, \tag{3.2}$$

where λ is the wavelength, η_0 is the impedance of free space, $E_{(\theta,i)}$ and $E_{(\phi,i)}$ are θ and ϕ components of the electric field of the ith path at the receiver point, respectively, and g_θ and g_ϕ are the direction of arrival of path i from the θ and ϕ directions, is given by

$$g_\theta \, (\theta_i, \phi_i) = \sqrt{(|G_\theta \, (\theta_i, \phi_i)| e^{j\phi_\theta})}, \tag{3.3}$$

where G_θ is the θ component of the receiving antenna gain, ψ_θ is the relative phase of the θ component of the far zone electric field.

The way to calculate AOA and AOD in azimuth and elevation angles are related to the antenna in the Wireless Insite (WI) software [178]. The location, orientation, and
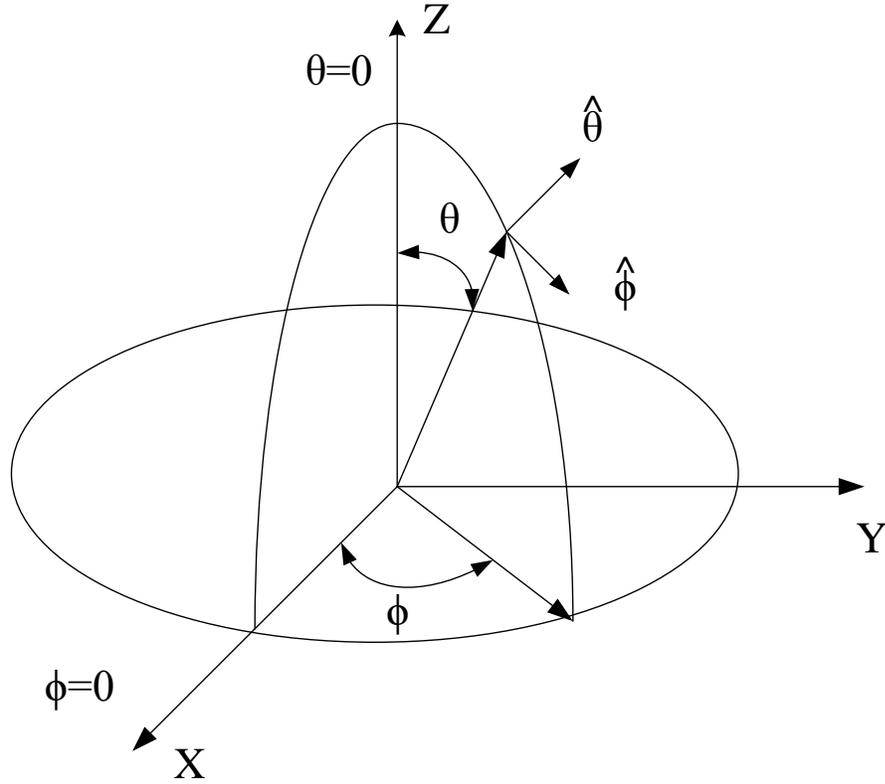
Figure 3.1: The Wireless InSite spherical coordinate system.

polarization of the antenna are set by the location of the associated transmitter or receiver and the rotation angles about the X, Y, and Z axis for each association of the antenna with a transmitter or receiver. The coordinate system used for singular rotation is shown in Fig. 3.1. In this case, the angles θ_A and ϕ_A, with reference to the spherical coordinate system, give the direction from which the propagation path arrives at a receiver point. From Fig. 3.1 the AOA in azimuth and elevation angles can be obtained as

$$\hat{a} = \sin(\theta_A)\cos(\phi_A)\hat{x} + \sin(\theta_A)\cos(\phi_A)\hat{y} + \sin(\theta_A)\cos(\phi_A)\hat{z}, \qquad (3.4)$$

Similarly, the AOD in azimuth and elevation angles can be obtained as

$$\hat{h} = \sin(\theta_H)\cos(\phi_H)\hat{x} + \sin(\theta_H)\cos(\phi_H)\hat{y} + \sin(\theta_H)\cos(\phi_H)\hat{z}, \qquad (3.5)$$

Regarding the UE position, the whole considered area consists of a rectangle, divided into N grids of one meter squared. The position coordinate from the first position on the lower left corner of the area to the last position on the top right corner is numbered. The
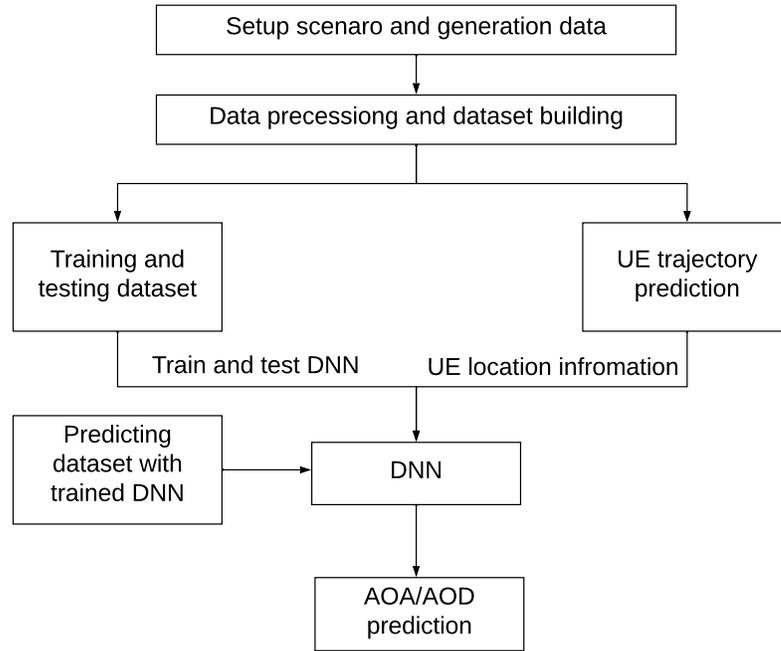
Figure 3.2: Flowchart showing the procedure for DNN enabled beam tracking in millimeter wave communications.

UE can move to the nearest four grids in four different directions (up, down, left, and right) with the velocity of one meter per second. Further, it is also considered that the UE is able to avoid obstacles. In other words, if there is an obstacle in front of the UE, it has to find another way to go around the obstacles. More specifically, the UE and obstacles have their own radius. While UE is moving, for every step the algorithm calculates the distance between the UE and the nearby obstacle. If the distance is smaller than the UE's radius, the algorithm will find another direction for the UE's next move, else the UE will keep moving on the previous direction. The details are stated in Section 3.3.3.

3.3 Deep Learning Based Beam Tracking Approach

This section presents the procedure for the ML based AOA/AOD prediction based on the analysis of UE trajectory in NLOS millimeter wave communications. The flow chart of the procedure is shown in Fig. 3.2. With the simulation environment described in Section 3.1, The raw network datasets [179] is obtained, including the received signal power, AOA,

AOD, and actual UE location information. Further, the raw data is processed into the right data, which is suitable for the training of the neural network.

With the processed data, the DNN to predict the AOA/AOD in azimuth and elevation is trained. The DNN is trained by 70% data with the received power, location information and the number of clusters from raw data obtained by K-means as the input. In the following, the main procedures is proposed, which is involved in the DNN enabled beam tracking, namely data processing and database building, AOA/AOD prediction and the UE trajectory prediction.

When training the DNN, apart from AOA and AOD, other parameters such as received power, real UE location information, and a cluster by K-means are also considered as inputs of the DNN. With the trained DNN, only predicted UE locations is input into it, which is predicted by the UE trajectory prediction algorithm so that the DNN can predict the AOA and AOD with the predicted UE locations. To evaluate the performance of the proposed AOA and AOD prediction, the error between the predicted AOA and AOD with the real AOA and AOD on different locations is calculated.

3.3.1 Data Processing and Database Building

The simulation environment, which is based on the University of Glasgow Gilmorehill campus. The ray tracing software, Wireless InSite (WI), is used to build the simulation environment. Ray tracing is based on geometrical optic (GO) and the uniform theory of diffraction (UTD). The interactions between rays and objects can be classified as reflection, transmission, scattering, and diffraction. An area (shaded in blue in Fig. 3.3) of $X \times Y$ meters, was considered in WI, with N grid positions for the UE to move. In this simulation scenario, There is one receiver and one transmitter, with $X \times Y$ available positions. The scattering which is caused by surface roughness is not considered, due to the complexity of the simulation environment. The hybrid precoding technique is applied for beamforming when generating the channel information. After collecting the raw network datasets from the simulated environment, The received power, AOA in azimuth (AAOA), AOD in azimuth (AAOD), AOA in elevation (EAOA), AOD in elevation (EAOD), and user location are labeled as features. The whole dataset is divided into training and test datasets based on the ratio 7 : 3. After that, K-means clustering is applied to the raw data, which creates another feature that improves the DNN training accuracy. K-means clustering is a method that partitions n observations into k clusters in which each observation belongs to the cluster with the nearest mean [180]. The K-means algorithm classifies the raw data in different classes. Thus, there is a metrics containing the different classes divided by
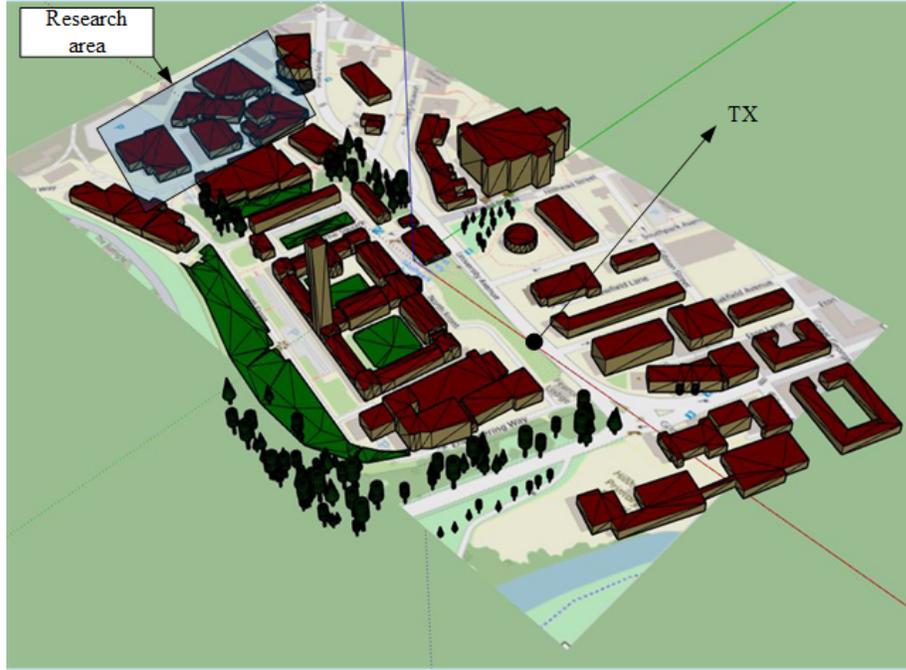
Figure 3.3: Simulation Environment (Based on The University of Glasgow campus, Gilmore Hill, UK).

the K-means method. In this case, there are different features for DNN training: received power, AOA in azimuth (AAOA), AOD in azimuth (AAOD), AOA in elevation (EAOA), AOD in elevation (EAOD), and user location. However, due to the different units among these features, first the input data is normalized before input the data into neural network. By normalization, the deviation of the data from the mean is computed and divided by the standard deviation. The transformed value of the input value x_{norm} after standardization can be expressed as

$$x_{norm} = \frac{(high - low) \times (x - \min X)}{\max X - \min X},$$
(3.6)

where $high$ and low are the range of data after scaling. $\max X$ and $\min X$ are the minimum and maximum value of the attribute X of input dataset [181].

However, it is impossible to use the transformed data to calculate the error between the real AOA/AOD and the predicted AOA/AOD. In this case, after AOA/AOD is predicted with the trained neural network, the inverse-normalization is applied to transform the data into its real form.
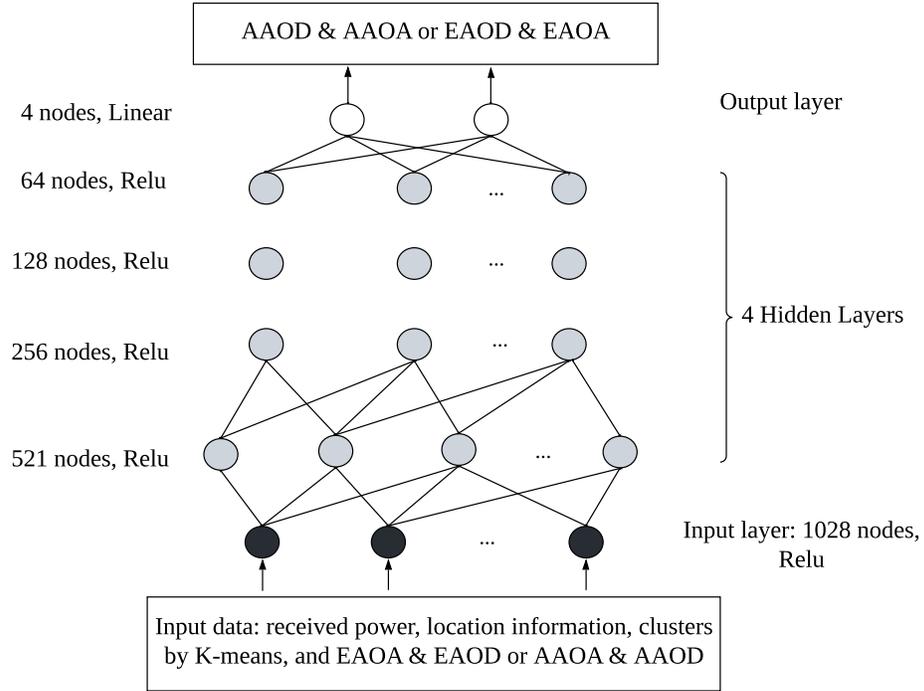
Figure 3.4: Basic DNN structure.

3.3.2   AOA/AOD Prediction Based on DNN

DNN Design

ANN is a computing system which is inspired by biological neural networks [182]. Generally, DNNs are deeper version of ANNs with more hidden layers to improve its representation or recognition ability [109]. The basic structure of a DNN is shown in Fig. 3.4. The input layer is at the bottom. Each node on the input layer, shown in the figure refers to the number of inputs inserted in the DNN. The output layer is at the top, and the number of nodes stands for the number of outputs coming out from the DNN. In the middle of the DNN, there are some hidden layers, which have strong relevance with the design of DNN. Each neuron on the hidden layer in the network is actually a non-linear transform. For example, the relu function is a non-linear transform, which can be defined as $f(x) = \max(0, x)$. Relu has some advantages, such as fast convergence, less required data, and sparse activation, which are very important for short response time systems like the wireless communications system [183]. Therefore, the output of the network z is a cascade of nonlinear transformations of the input data I, which can be expressed as

$$z = f(I) = f^{(L-1)}(f^{(L-2)}(\cdots f^{(1)}(I))), \qquad (3.7)$$

Table 3.1: Parameter of Neural Network

Layers	Nodes	Activation
Input	1028	Relu
Dropout	N/A	N/A
Hidden Layer 1	512	Relu
Hidden Layer 2	256	Relu
Hidden Layer 3	128	Relu
Hidden Layer 4	64	Relu
Output	4	Linear

where L is the number of layers and α are the weights of neural network.

In the DNN, the weights for the neurons are required to be optimized while training. Usually, in DNN, the number of hidden layers, and the number of nodes on hidden layers are large and thus it causes the DNN to be more complex [184]. However, there is a trade-off between the number of hidden nodes and the accuracy. In our case, the basic DNN structure of predicting AAOA and AAOD is shown in Table I, where four hidden layers are considered. The input features in the DNN are received signal power, location information of UE, clusters by K-means and AOA/AOD in elevation when predicting AOA/AOD in azimuth. The number of inputs inserted in the input layer for training DNN is 3, while the number of input nodes for each hidden layer are 1028, 512, 256, and 64. The number of outputs coming out from output layer is 4. The desired output in this DNN is AAOA, AAOD, EAOA, and EAOD.

The performance of the DNN can be improved by using some hyper parameters to address challenges such as overfitting and learning rate selection. Overfitting results in the model learning the statistical noise in the training data, and this causes poor performance when the model is evaluated on new data. One approach to reduce overfitting is to fit all possible neural networks on the same dataset and average the predictions from each model [185]. However, this is not feasible in practice because of the low efficiency [185]. Dropout is a regularization method that approximates the training of a large number of neural network neurons with different architectures in parallel [185]. While training, some number of layer outputs are randomly dropout with dropout rate (one of the hyper-parameters) 0.4 for predicting AOA and AOD in azimuth and elevation. This makes the layer to be treated-like a layer with a different number of nodes and connectivity to the prior layer.

Secondly, an initializer is added for the DNN on the input layer to initialize its weight. The aim of the initializer is to prevent layer activation outputs from exploding or vanishing

during the course of a forward pass through DNN. If exploding or vanishing happens, loss gradients will either be too large or too small to flow backward beneficially, and this makes the neural network to converge slower. The initializer is Xavier [186], which could maintain the variance of activation and back-propagated gradients all the way up or down the layers of the network. Xavier initialization sets a layer's weights to values ranging from a random uniform distribution to

$$\beta_{layer} = \pm \frac{\sqrt{6}}{\sqrt{n_i + n_{i+1}}} \qquad (3.8)$$

where n_i is the number of incoming network connections to the layer, and n_{i+1} is the number of outgoing network connections from a given layer.

Thirdly, the performance of DNN predicting AOA and AOD in azimuth and elevation with different learning rates is also tested. The learning rate is another hyper-parameter in neural network training, which controls how much change is made to the model in response to the estimated error each time the model weights are updated. It has a huge influence on the speed of the training process. Large learning rates may lead to a sub-optimal set of weight or an unstable training process. On the other hand, small learning rates may result in a long training process or the system could even get stuck [187]. To find a suitable learning rate in each stage of training, the adaptive learning rate gradient descent is applied to the DNN. Because each stage adapts the learning rate, often some configurations are required in each stage. The specific activation function in hidden layers is Relu. Further, the activation function of output layer is linear activation function, which is defined as $y = cx$. It thus creates an output signal proportional to the input.

Adam optimization algorithm is used, which has been used for the classical stochastic gradient descent procedure to update network weights based on training data [186]. The Adam optimization algorithm has a number of benefits, such as low computational complexity, having little memory requirements, and its high suitability to problems with very noisy/or sparse gradients [188].

AOA and AOD Prediction

The procedure of AOA/AOD prediction is presented in Algorithm 1. In this scenario, the location of UE is known and the actual UE location is used as the input when we train the DNN to predict. Specifically, to predict the AOA and AOD, we input the features generated with the method in Section 3.3.1 and standardize all the features in order to have them on the same scale. Then the DNN is configured via the method stated in Section 3.3.2

Algorithm 1: DNN enabled AOA and AOD prediction.

Input: Received power, location information, and clusters by K-means in the training dataset

Output: Errors between real AOA/AAOD and predicted AOA/AAOD initialization;

1. Normalize the input data;
2. Input the data into neural network;
3. Add dropout layer and initializers into neural network;
4. Train the neural network;
5. Input the location information in testing dataset into trained neural network;
6. Generate predicted AOA and AOD;
7. Calculate the absolute errors between real AOA/AOD and predicted AOA/AOD;

---

(1) to improve the performance of the proposed system.

After training the DNN, the same features from a test dataset are input into the trained neural networks to predict the AOA/AOD using the actual UE location as input. the absolute error between the predicted AOA/AOD and the real values is calculated to evaluate the prediction performance. Further, some errors with truncated normal distribution is generated, which is added to the location feature input in the DNN in order to evaluate the performance of the proposed system in the presence of errors. The reason is that, in practice, there might be some errors when generating the data. If the system retains good performance, it means that the proposed system is robust enough to measurement errors. The upper bound and lower bound of truncated normal distribution range from ± 10, ± 7, ± 5, and ± 2 (meters). According to the experience, the threshold for the mmWave beam signal is $\pm 7°$. The system is expected to retain a high AOA/AOD prediction accuracy for a degree of error below $\pm 7°$.

3.3.3 UE Trajectory Prediction Design

The DWA proposed in [189] is used here for the reactive collision avoidance for the UE. DWA is executed with a fixed frequency, and only a set of velocities can be applied to the UE due to its acceleration and velocity limits. Among the set of velocities, a reward function is proposed to select the best velocities to follow [190]. The approach is directly from the motion dynamics of the UE. The motion can be obtained as follows

$$x(t_n) = x(t_0) + \int_{t_0}^{t_n} v(t) \cdot \cos \theta(t) dt, \tag{3.9}$$

$$y(t_n) = y(t_0) + \int_{t_0}^{t_n} v(t) \cdot \sin\theta(t) dt, \tag{3.10}$$

where $x(t)$ and $y(t)$ are the UE's coordinate at time t in the cartesian coordinate system, while the UE's orientation is dictated by $\theta(t)$ and t_0 is the initial time while t_n can be any time when the UE is moving.

The motion of the UE is constrained in a way that the translational velocity v always leads in the steering direction θ of the UE, which is called a non-holonomic constraint [191]. In the DWA, the search for commands controlling the UE is carried out directly in the space of velocities. The dynamics of the UE are incorporated into the method by reducing the search space to those velocities that are reachable under the dynamic constraints. Due to these constraints, only velocities which are safe with respect to the obstacles are considered. Then by substituting the corresponding initial kinematic and dynamic configuration $v(t_0)$, $\theta(t_0)$, and $\omega(t_0)$ into (9) and (10), it is obtained

$$x(t_n) = x(t_0) + \int_{t_0}^{t_n} (v(t_0) + \int_{t_0}^{t} (\dot{v}(\hat{t}) d\hat{t})) \cdot$$

$$\tag{3.11}$$

$$\cos(\theta(t_0) + \int_{t_0}^{t} (\omega(t_0) + \int_{t_0}^{\hat{(t)}} \dot{\omega}(\tilde{t}) d\tilde{t}) d\tilde{t}) dt$$

Equation (11) is now in the form that the trajectory of the UE depends exclusively on its initial dynamic configuration at time t_0 and its accelerations. However, in our case, the angular velocity θ is discrete, and θ is in a set $\theta \in \{\theta_1, \theta_2, \theta_3, \theta_4\}, \forall\ 0 \leq \theta \leq \pi$. The values are evenly spaced by a θ_{step}, which will create a different number of directions for a single UE. To take into account the limited accelerations exertable by the UE, the overall search space is reduced to the DWA. It contains only the velocities that can be reached within the next time interval. In this case, t is the time interval during which accelerations $\dot{v}$ and $\dot{\omega}$ will be applied, considering (v_a, ω_a) as the actual velocity of a given UE, the dynamic window V_d is defined as

$$V_d = (v, \omega) | v \in [v_a - \dot{v} \cdot t, v_a + \dot{v} \cdot t]$$

$$\tag{3.12}$$

$$\wedge \omega \in [\omega_a - \dot{\omega} \cdot t, \omega_a + \dot{\omega} \cdot t]$$

The dynamic window is centered around the actual velocity and the extensions of it depend on the accelerations that can be exerted. The alignment of the UE with the target direction is measured by target heading (v, ω). It is given by $180 - \theta$, where θ is the angle of the
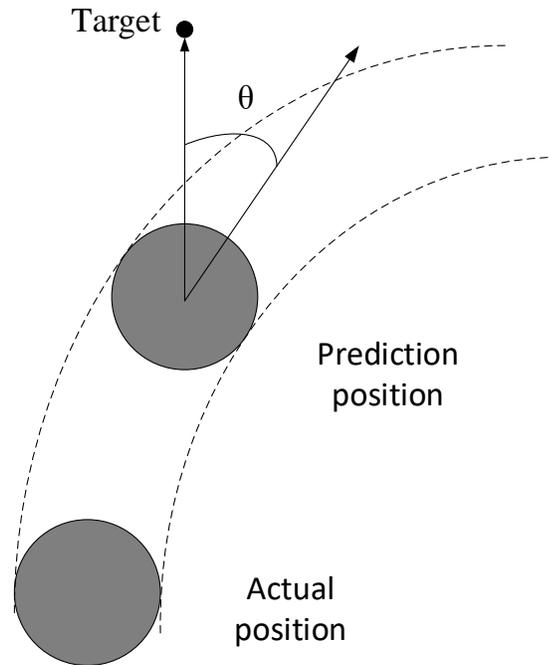
Figure 3.5: Target direction within DWA.

target relative to the UE's heading direction, as shown in Fig. 3.5.

In order to make the UE trajectory prediction scenario closer to the simulation environment, the simulation area is zoom in and shown in Fig. 3.3 and create a new scenario for trajectory prediction, as shown in Fig. 3.6.

It is assume that a UE (red cross), with the limited radius, randomly appears at the location indicated by a yellow cross within the area of interest and that the UE has a destination point indicated with a blue cross. For every step, the UE detects the obstacles (black points) in eight different directions and it calculates the distance between itself and the nearest obstacle or destination point. The UE finally stops at the destination point when the distance is smaller than its radius. Before the UE stops, the DWA will predict the possible directions on every step of the UE, which is shown as a green line. When the UE arrives at its destination, its path in the areas of interest is shown as a red line. The coordinate on the predicted path will be recorded as the location information. The whole procedure is based on collision avoidance. The destination of the UE is set in the scenario, which the UE tries to reach while avoiding the blockages on its path, such as buildings and trees.

In this procedure, a total of 30 location information with a random starting point of the UE is generated and the value of location information in the coordinate in the scale of
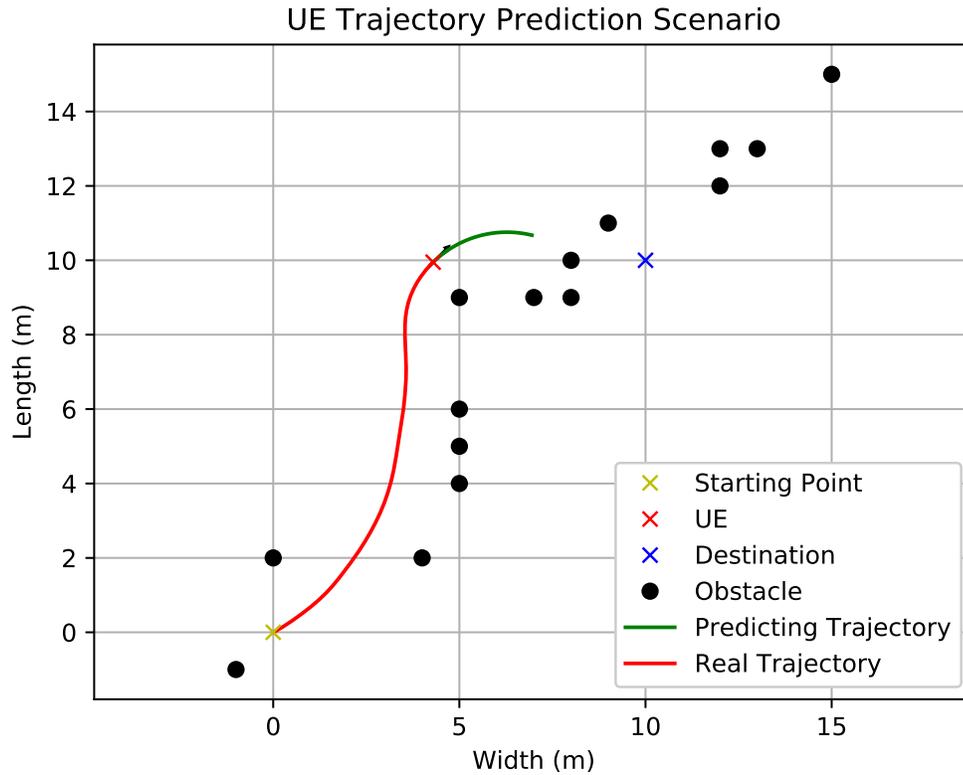
Figure 3.6: UE trajectory prediction.

the campus simulation environment is inversed. The AOA and AOD prediction with UE trajectory prediction is the second simulation scenario. Based on trained DNN, the actual UE location information generated from the campus simulation environment is replaced with the predicted UE location information and input into the trained DNN in order to predict the AOA/AOD. After that, there is another simulation to evaluate the performance of the proposed method when the location information are with some errors. Similar to the simulation in Section 3.3.2, some errors with truncated normal distribution are generated, in which the upper limit and lower limit range from ± 10, ± 5, and ± 2 meters. The errors are added on the predicted location information, which is input into the same trained DNN to evaluate the performance of the method. Note that ± 10 meters are large errors for the prediction. On the other hand, ± 5, and ± 2 are allowable errors for trajectory prediction. Although there are such errors, the proposed system still can make most of UE position access to the BS. This simulation shows the robustness of the system to prediction errors. With this simulation, the error tolerance of the system is proven. Detailed results are presented in the next section.

3.4   Results and Discussion

There are some experiments to evaluate the performance of the AOA/AOD estimation with deep learning. First, the AOA and AOD are predicted with the trained DNN using some of given UE locations. After that, It is assumed that the UE locations are unknown, with the AOA and AOD angles in azimuth and elevation being estimated based on the UE trajectory prediction proposed in Section 3.3.3. The performance of the predication algorithm is evaluated in the presence of errors showing its robustness. Secondly, the performance of the proposed method with another typical ML method – CNN is compared, in terms of accuracy in the prediction of EAOA and EAOD. The parameters of the simulation settings are shown in Table II.

Table 3.2: Parameters in simulation environment.

Transmit power	43 dBm
Carrier Frequency	30 GHz
Noise	-88 dBm
Effective Bandwidth	200 MHz
TX Height	40 m
Area Length (*X*)	223 m
Area Width (*Y*)	328 m
Number of Grid Positions (*N*)	73696

3.4.1   AOA and AOD Prediction with Perfect Testing Data

The first simulation is the AOA and AOD prediction by training the DNN. In this work, the features used to train the DNN consists of received power, location information, and clusters by K-means. The training performance is shown in Fig. 3.7, where the loss function is given by the mean square error (MSE), which is defined as:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} e_i^2, \tag{3.13}$$

where e_i is the training error by the i-th sample and n is the total number of samples.

As it can be seen, the loss curve converges after 300 epochs and the testing loss curve fluctuates slightly. The reason is that when the UE position changes, there is few changes on AOA and AOD. However, both training loss and testing loss maintain a very low level
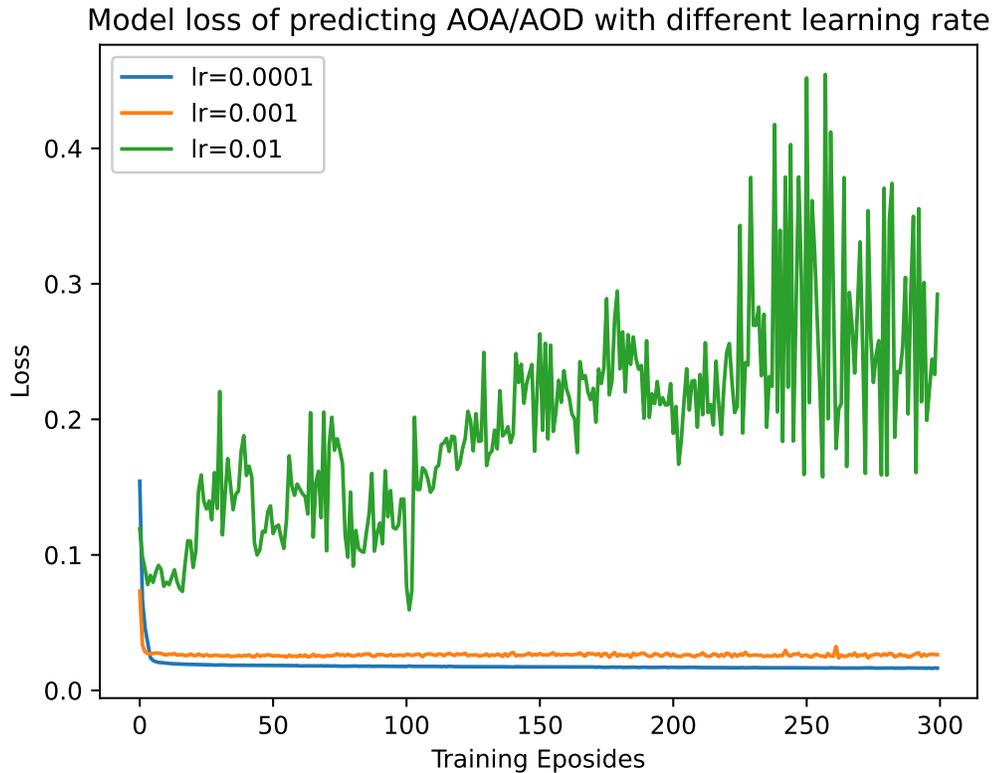
Figure 3.7: Performance of training DNN to predict AOA/AOD.

(0.02), which is acceptable for DNN training.

Based on the trained DNN, the features are input from the test dataset to predict AOA and AOD, which consists of complete new unseen samples for the trained DNN. The absolute error is calculated between the predicted and real values and show the results of the probability density functions (PDF) of prediction errors. From Fig. 3.8, the AOA prediction absolute error (blue line) keeps in around $\pm 2°$. It is defined that if the error is over $7°$, it will be out of connection from BS. Then there is a calculation of the number of positions in testing dataset, which are out of connection from BS. There are 40 out of 7034 positions out of connection with BS (the percentage is about (0.5%)). Further, for AOD prediction absolute error (red line),It can be seen that the error percentge is around 7/7034 (0.1%). However, despite these minor variations, the proposed DNN method is able to achieve accurate predictions for both AOA and AOD with an error below $7°$.
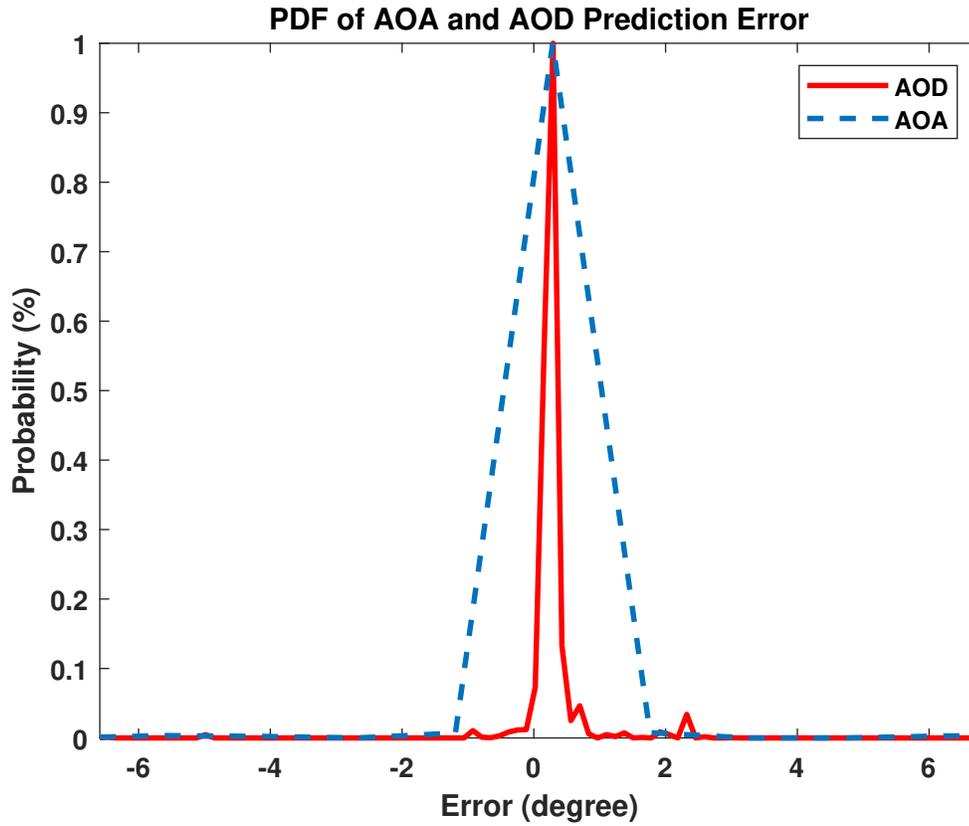
Figure 3.8: PDF of AOA/AOD predicted error.

3.4.2 Prediction Performance with imperfect testing data

Next, the performance of the proposed DNN with imperfect testing dataset is explored, which is more related to the practical scenario. By adding errors into features, specifically AOA and AOD of the test dataset for AOA and AOD prediction. The errors follow a truncated normal distribution and the upper and lower bounds are $\pm 10°$, $\pm 7°$, and $\pm 2°$. In experience, if the AOA and AOD prediction errors are over $\pm 10°$, the UE will be out of connection with the BS. When the bound is between $\pm 10°$ and $\pm 7°$, the UE has connection with BS but poor signal quality. The signal quality will be better with the error reducing. When the bound is between $\pm 7°$ and $\pm 2°$, the UE will have the prosperity communication experience with the BS. The performance of the proposed algorithm under imperfect testing dataset is shown in Table III, in addition to the case with no errors. From Table III, it can be seen that the proposed system can still maintain a very low prediction error when the added errors are smaller than the threshold. This clearly shows the robustness of the system and the advantage of utilizing machine learning, more specifically DNN, than
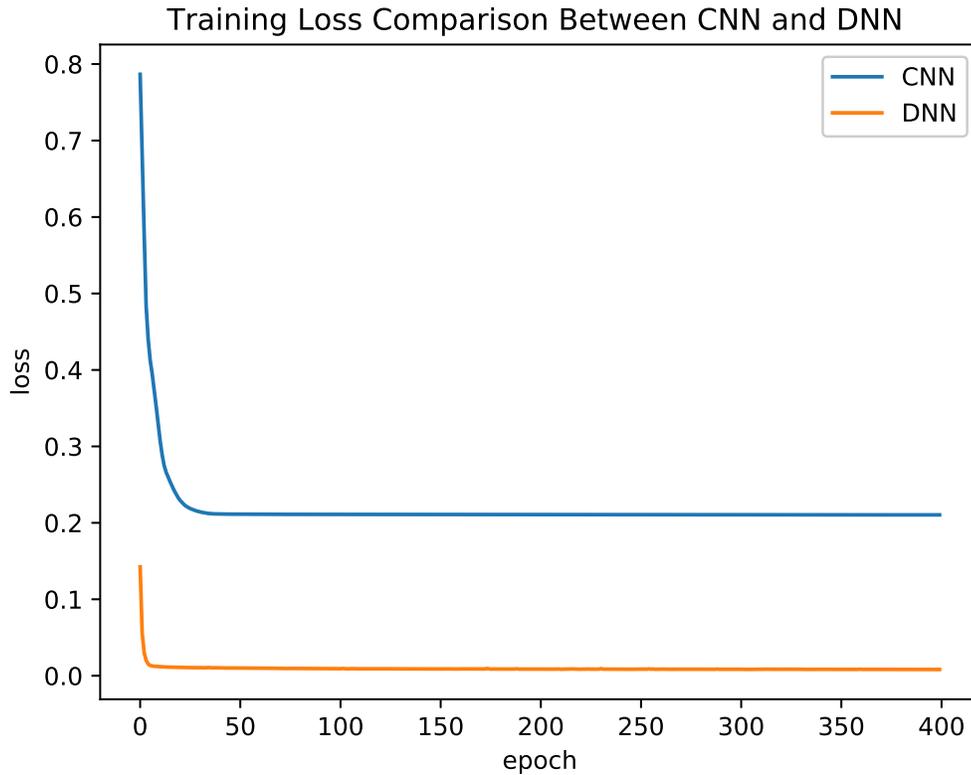
Figure 3.9: Training loss comparison between CNN and DNN for NLOS beam tracking

other methods. Some errors are added into the dataset to make it more particular. The results shows that the predication algorithm can maintain the accuracy when the added errors are smaller than $\pm 7°$.

Further, in order to make a comparison with other method, the performance of a CNN-based prediction of AOA and AOD is evaluated. The results are shown in Fig. 3.9. As it can be seen, the loss generated by the CNN is much higher than that of the DNN and even by increasing the number of epochs, it can be seen that it does not improve. The reason for this is that architecture of the CNN is not good for this problem and CNN is more suitable for receiving and processing pixel data. However, the dataset is consisted of numeral numbers.

3.4.3   AOA/AOD Prediction with UE Trajectory Prediction

In this simulation, the UE locations predicted by DWA are generated for 30 times with a random starting point, as stated in Section 3.3.3. The location information is input as
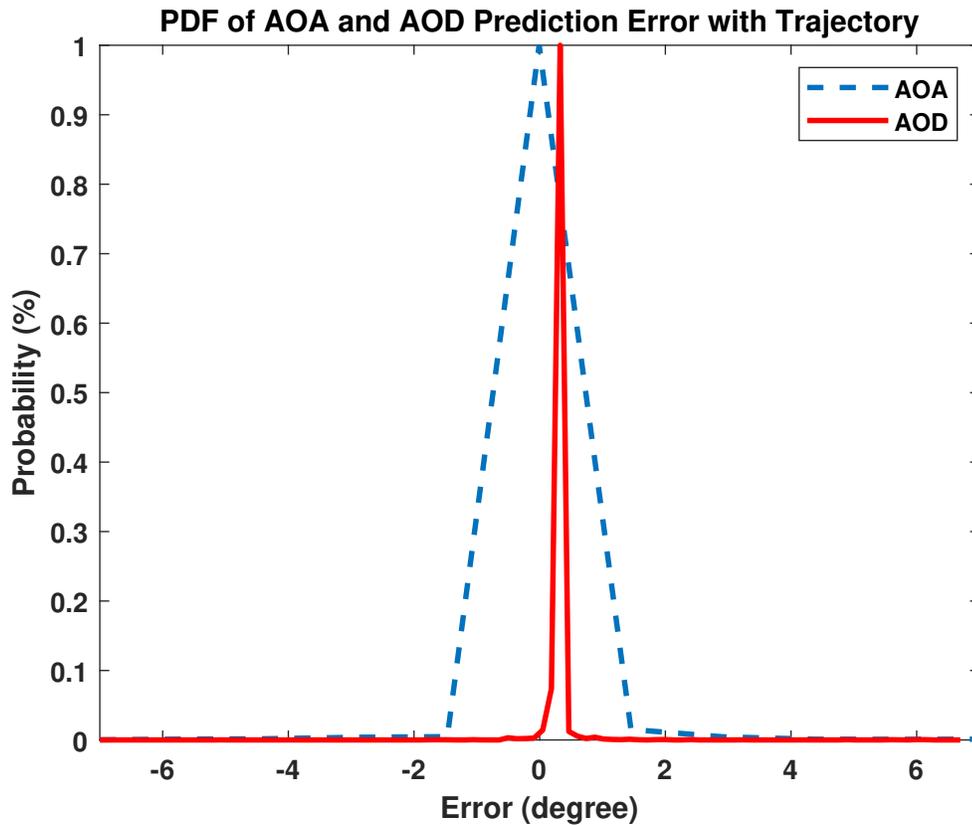
Figure 3.10: PDF of AOA/AOD predicted error with trajectory prediction.

the only feature into the trained DNN for AOA and AOD prediction. From Fig. 3.10, AOA and AOD prediction error with trajectory prediction are located around $\pm 2°$, which maintain closed level with AOA and AOD predicted error. As mentioned in Section 3.3.3, the dynamic window approach is a reactive collision avoidance algorithm. The BS could locate the UE position and be aware of the surrounding environment, such as buildings, streets, and road. In this case, based on collision avoidance principle of the UE, the dynamic window approach can predict the UE's route from its starting point to its destination. Therefore, the BS generates the location information of the predicted UE route. And based on the location information as the input of the DNN, the trained DNN can predict the AOA/AOD of the UE on each position of its route. From this experiment, it shows that the proposed system has the ability to predict the UEs' AOA and AOD with totally unknown locations. The reason why the performance results of AOA/AOD prediction with UE trajectory prediction are better than the performance of only DNN based method shown in Fig. 3.8. is that when only location information generated by DWA is input, the trained DNN can be more easily focused. The reason behind is that when training the DNN, the

Table 3.3: Prediction Performance in the Present of Errors

Scenarios	AOD	AOD
No Error	7 (0.1%)	40 (0.5%)
2°	10 (0.14%)	62 (0.8%)
7°	35 (0.5%)	81 (1.15%)
10°	3353 (47%)	3550 (50%)

weight of location information account for the majority. Thus, the accuracy of the prediction results rises. On the contrary, when predicting AOA and AOD via DNN without DWA (the performance is shown in Fig. 3.8), in addition to the location information, the inputs in the neural network also include received power and the number of clusters from raw data obtained by K-means. In this case, other factors may affect the result more even through the location information is accurate. Therefore, the prediction results of Fig. 3.10 is slightly better than that of Fig. 3.8.

3.4.4 Computation Complexity

The comparison of computation complexity between traditional method such as exhaustive search and the proposed method is made in this section. The exhaustive search browses all possible AOA/AOD and chooses the best result. However, when compared the predicted outcome of AOA/AOD with the particular value to generate the prediction error, the particular value is already the best received power, which is as same as the exhaustive search result for each position in the simulation area. It is considered the online deployment part of the method and searching part of the exhaustive search method. More specifically, only the prediction procedure for each position of the proposed method is under consideration rather that training part and for exhaustive search the channel estimation procedure is ignored. The equation which is adapted to calculate the computation complexity of DNN is:

$$F = \sum_{i=2}^{n} \eta_i \times \eta_{i-1} + \sum_{i=1}^{n} \eta_i, \tag{3.14}$$

where η is the number of nodes in each layers and i is the layer. In this case, according to Table I (it can be found in the Appendix of this letter), there is one input layer with 1028 nodes, four hidden layers with 512, 256, 128, 128, and 64 nodes, and one output layer with 4 nodes. It is assumed that the number of additions and multiplications in the DNN have the same computation burden in our case. Therefore, the computation complexity of

the proposed method is the computation between each layers and the computation of activation function (Relu in the method) in each node, which is 6.99×10^5. For the exhaustive search, for each position, both azimuth and elevation angle have to be considered. When calculating the prediction error, the precision is one degree. Therefore, there are 360 different angles to be considered for both azimuth and elevation. In this case, based on the simplest Bubble Sort computation complexity calculation, the computation complexity for exhaustive search is 1.67×10^9, which is around 42,000 times larger than the proposed method.

3.5 Conclusion

A deep learning enabled method to prediction the AOA and AOD in NLOS channel for mmWave communication is proposed in this chapter. Firstly, the simulation model with NLOS scenario and channel model of AOA/AOD are built to generate the dataset for DNN training. The neural network is trained with some channel features, such as received power, location. Results indicate that the absolute error, calculated between the real and the predicted are quite low, validating the effectiveness of the proposed solution. Further, some errors with truncated normal distribution is added in the beam angle to evaluate the robustness of the proposed system. When the error is below a given threshold of $7°$, the system still has good performance. Finally, the UE trajectory with DWA and generate location input is predicted. Further, input it into the trained DNN to evaluate the performance of trajectory prediction. The error in this case is close to the original location information from data generation.

Chapter 4

A Novel Handover Scheme using RL

4.1 Introduction

As mentioned in previous chapter, to overcome the drawbacks of mmWave and make full use of it, beamforming and dense small cell base stations (SCBSs) architecture [192, 193] play a major role in mmWave communication. Especially, cellular networks with dense SCBSs could improve the efficient propagation of mmWave while beamforming offers a potential solution for mmWave to avoid the blockage.

However, with the increase of SCBS in the cellular network and the propagation becoming directional, there is a great challenge for the handover (HO) in mmWave cellular network [86]. Specifically, with the SCBS increasing, the inter-cell handover becomes more frequent, leading to higher HO rates. The user equipments (UEs) need to switch from one SCBS (or one beam) to another while moving to maintain the communication quality [150]. In particular, HO mechanisms affect not only the quality of service (QoS) on UE side but also the network performance [3]. Since there is a limitation of the resource in BS, growing HO rates usually bring some problems to the network, such as increasing the HO failures rate and higher signalling overheads, which reduces the system performance [86]. Further, since most beamforming techniques in particular are directional, the HO event also occurs when UEs move from one beamforming covering region to another. In this case, the intra-cell HO also grows significantly compared with the traditional network structure. According to [194, 195], the average handover interval could be lower than 0.75 seconds in the typical mmWave cellular network scenarios and approximately 61% of handovers are unnecessary. Therefore, how to improve the HO efficiency in mmWave cellular network is a key issue to be resolved.

In the traditional communication network, to reduce the redundant of handover, the

3rd Generation Partnership Project (3GPP) [196] defines that handover is triggered when the Reference Signal Received Power (RSRP) of current serving BS is lower than the threshold and RSRP of targeting BS is stronger than the current serving BS. However, this method is not adapted to the mmWave cellular network, resulting in the frequent HO problem and increase the HO overhand [197]. Therefore, an optimized handover mechanism is to establish crucially. With the environment of mmWave scenario becoming complex, plenty of optimization problems are nonlinear, making the traditional mathematical tool less efficient to solve the problems. In this case, one of the widely-used AI algorithms, reinforcement learning (RL), could be designed for a smart handover mechanism in mmWave cellular network, via the interactions with the network environment. However, only in this way, could it not meet the quality of service (QoS) of the mmWave network, since RL method focuses on the handover trigger decisions. Further, with the number of UEs and SCBSs increasing, the resource allocation becomes difficult. In other words, resource allocation should be optimized in conjunction with handover decisions. RL method typically estimates and evaluate the UEs' action through the interaction with the environment, which takes a long time to coverage. In this case, after the RL algorithm makes the handover trigger decision, the optimization theory is implemented to manage the resource allocation, target BS and beam selection in each SCBS, which not only improves the overall system performance, including total throughput and delay but also guarantees the QoS of each UE.

This chapter proposes a novel handover scheme called the optimization-based MAPPO (O-MAPPO) method to help UEs make the optimal handover decision regarding targeting beam and BS and improve the overall system performance, including increasing total system throughput and reducing total system delay. Further, with the assistant of the proposed method, the demand of individual UE, in terms of QoS is met. From the numerical results, It is demonstrated that the proposed method achieves better performance with the comparisons of other typical RL algorithms, such as Deep Deterministic Policy Gradient (DDPG) and Deep Q-learning (DPQ). The main contributions of this chapter are as follows:

1. The O-MAPPO method consists of two parts. An intelligent handover trigger condition scheme based on RL algorithm called MAPPO is implemented in the mmWave cellular network to assist each UE in making the best handover trigger decision. With the help of this method, the reliability of handover in the network is improved, including the reduction of HO rate and HO failures.

2. An optimal handover decision scheme based on optimization theory is designed to manage the resources in each SCBS, such as bandwidth allocation and target beam

and BS selection, which optimal the overall system throughput and delay while ensure the individual UE meeting the demand of QoS. Further, the information generated by the optimal handover decision scheme is used as the observation and state of the MAPPO algorithm, making the handover decision more promising.

3. Further, a handover penalty mechanism is applied to reduce the HO rate while avoiding unnecessary handover. In this case, the system is optimized in the perspective of energy efficiency.

The rest of the chapter is organized as follows. A system model is proposed in Section 4.2. The basic framework of O-MAPPO is stated in Section 4.3. The design of intelligent handover trigger condition scheme based on the RL algorithm called MAPPO for handover decision of UEs is discussed in Section 4.4. The design of optimal handover decision to management the resources in BSs and improve the system performance and guarantee the QoS of each UE is proposed in Section 4.5. Simulations results and analysis are given in Section Section 4.6. Finally, in Section 4.7 concludes the chapter.

4.2 System Model

4.2.1 Network Topology

The network topology is shown in Fig. 4.1. The mmWave cellular network is presented, consisting of one macro base station (MBS) and M small cell base station (SCBS) with N beams in each BS. The set of BSs is denoted as $M = \{0, 1, 2, ..., M\}$, in which 0 represents the index of MBS while $\{1, 2, ..., M\}$ is the index of SCBS. It is assumed that each BS has the same number of beams and the set of beams in each BS is denoted as $N = \{0, 1, 2, ..., N\}$. Further, the set of UEs is defined as $I = \{0, 1, 2, ..., I\}$. Each UE is served by either the MBS or one SCBS with only one beam. UEs are located at random positions within the coverage of MBS at the initial stage. The UE mobility model is random walk [198].

The channel information of UEs is periodically measured. When UE moves, HO trigger conditions are learnt by RL when either the current SINR cannot meet the demand of UE's service or UE moves to overlapping area. Further, there are two handover cases in the network: inter-cell handover and intra-cell handover. Inter-cell handover occurs among the different BSs. Especially when UEs move to the overlapping area and the current SINR is lower than the threshold. Intra-cell handover triggers when UEs moves
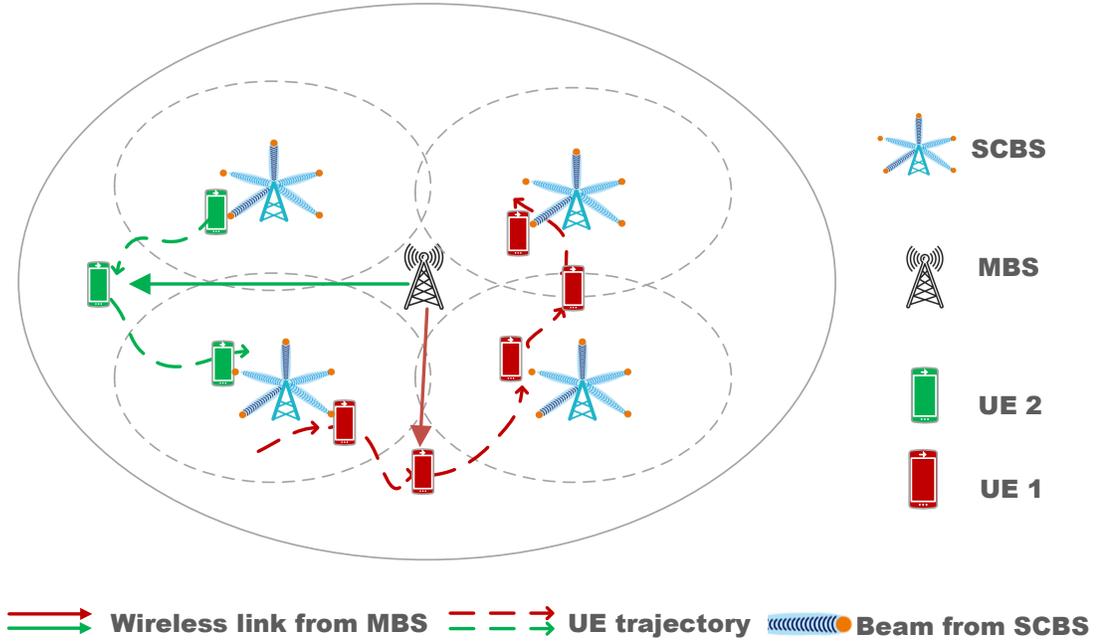
Figure 4.1: UEs and BS Distribution.

within the same SCBS, but the current serving SINR cannot meet the demand. When RL decides the handover trigger conditions, the channel information is used to optimize the decisions of beam and BS selection and bandwidth allocation. The UE can either switch to another BS or maintain a different beam in the current BS.

4.2.2 Channel Model

The channel models between BS and UE are presented in this subsection. First, the channel model of MBS is introduced. It is considered that there is an omnidirectional antenna applied in the MBS to assurance the signal coverage. The path loss (in dB) model of the MBS is [4]

$$PL(d)[dB] = \alpha_M + 10\kappa_M \log_{10}(d) + \psi + \xi, \qquad (4.1)$$

where d is the distance in meters, κ_M is the path loss exponent representing the slope of the best linear fit to the propagation measurement in the mmWave band, α_M is the path loss factor, ψ is random small-scale fading, and ξ is the random lognoraml shadowing.

On the UE i side, it is defined that d_i^0 is the distance between UE and the MBS while p_i^0 denotes the transmission power from MBS to UE, which satisfies $\sum_i^I p_i^0 = P_M$. Since there is co-channel interference in the mmWave band due to the shared bandwidth, the

SINR received by UE from MBS is:

$$SINR_i = \frac{PL^{-1}p_i^0}{\beta_i + N_M \omega_i^0},\tag{4.2}$$

where β_i is the co-channel interference [1], N_M is the noise power spectral density of MBS, and ω_i^0 represents the bandwidth allocated to the UE from MBS.

Second, the channel model of mmWave SCBS is presented. In practice, there are two kinds of channels among different SCBS in mmWave band: LOS and NLOS channel [37]. It is considered that there is a probabilistic LOS-NLOS channel model defined in 3GPP standard [3], which means there are two different channels (LOS and NLOS) for UE in SCBS and the channel can change with its probability. It is defined that v_i^m is the probability of LOS channel adopted from SCBS ($m \in M, m \neq 0$) to UE ($i \in I$). According to [199], where there is an estimation method for LOS channel probability with the building density in the simulation area, the LOS probability from the SBS and UE is:

$$v_i^m = \exp(-\frac{2D_B X_B d_i^m}{\pi}), m \neq 0,\tag{4.3}$$

where D_B is the building density, X_B is the expectation length of the buildings, and d_i^m is the distance from UE to SCBS. In this case, according to [3], the path loss model of SCBS is:

$$pl(d)[dB] = \alpha_S + 10\kappa_S \log_{10}(d),\tag{4.4}$$

where d is the distance in meters, α_S and κ_S is the same as that in equation (1), which is path loss factor and exponential decay factor, respectively[2]. The random small-scale fading (ψ) and random lognormal shadowing (ξ) are ignored since the LOS-NLOS probability mode has already considered.

Assuming that the directional antennas are equipped on all SCBSs to support beamforming and beam tracking in mmWave system, while there is an omnidirection antenna on UE side in order to calculate the antenna gain on the SCBS side. In this case, according to [3], the antenna gain is:

$$g(\phi) = \begin{cases} g_{max}, & |\phi| < \dfrac{\phi_S}{2} \\ g_{min}, & otherwise, \end{cases}\tag{4.5}$$

---

[1] The interference is the sum power received on the UE side from MBS nearby small cell base station.

[2] α_S and κ_S have different values in LOS and NLOS cases.

where ϕ is the angle between UE and BS, and ϕ_S is the width of the antenna main lobe. In our case, there is perfect beam tracking performed, which means the UE is always served by main lobe to obtain the maximum antenna gain.

Since the interference among SCBSs can be ignored in mmWave system, the signal to noise ratio (SNR) is calculated as [3]

$$SNR_i^m = \frac{g_{max}pl^{-1}p_i^m}{N_S}, m \neq 0 \in M, \tag{4.6}$$

where p_i^m is the transmission power between UE and SCBS, satisfying $\sum_i^I p_i^m = P_S$, and N_S is the noise power spectral density among SCBSs.

4.3   Framework of O-MAPPO Handover Scheme

This section proposes the O-MAPPO framework, which contains two parts: intelligent handover trigger condition and optimal handover decision. Specifically, MAPPO algorithm is used to learn the HO trigger condition in intelligent handover trigger condition part. After MAPPO makes the trigger decision, the SINR between UE and BS are calculated based on the channel model and sent to the optimal handover decision part. In this part, the beams and BSs selection as well as bandwidth allocation are optimized and evaluated, with which the throughput and delay of all UEs are calculated. The calculation results are then passed to the MAPPO as the observation and state to evaluate the handover trigger decision according to the reward function. The basic structure of O-MAPPO framework is shown in Fig. 4.2.

In more details, there are two handover trigger scenarios applied in the proposed method: handover triggers either in the SCBSs overlapping are, or the serving SINR can not meet the demand. When UEs are moving, MAPPO needs to decide the handover trigger conditions. When the handover trigger occurs, the MAPPO algorithm searches the nearest three target beams in current BS or another BSs with the shortest distance, which can provide the highest SINR to UE and then send them to the optimal handover decision scheme to make the beams and BSs selection. Further, the optimal handover decision scheme allocates the bandwidth based on the package length of different UEs. It calculates the overall system throughput and delay with the resource allocation information. Meanwhile, during the allocation and calculation, there is a threshold of throughput and delay for individual UE to guarantee the QoS. The allocation and calculation results are then feedback to the MAPPO algorithm as states and observations. According to the
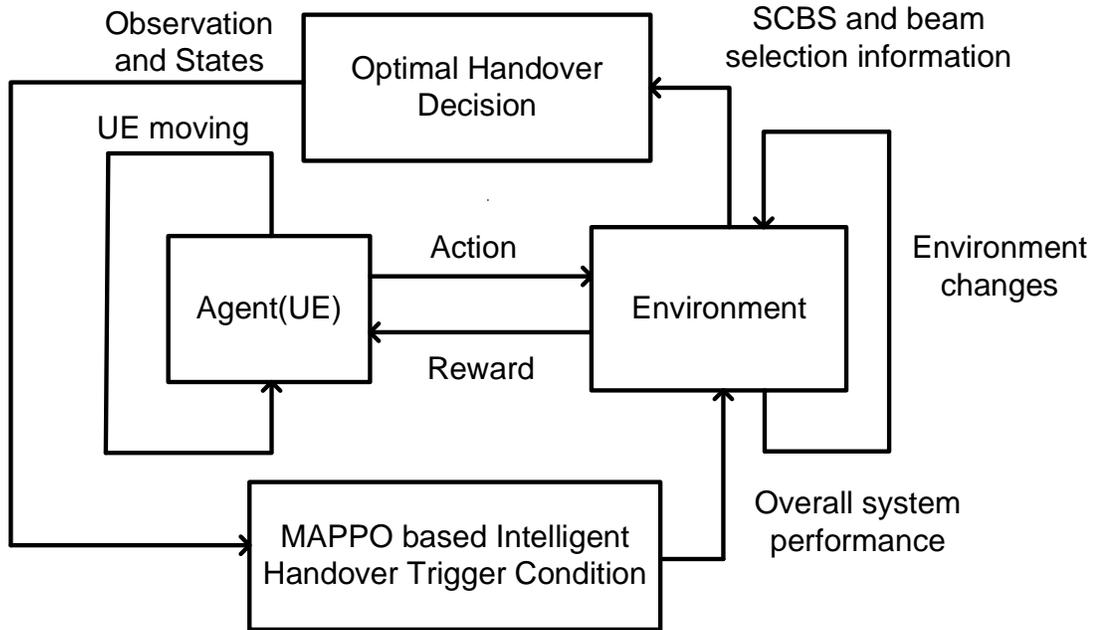
Figure 4.2: O-MAPPO framework.

reward function in the MAPPO algorithm, each handover trigger decision will be either rewarded or punished. In this way, all UEs learn how to make the best handover trigger decision, which improves the overall system performance and makes the QoS of individual UE promising.

4.4 MAPPO based Intelligent Handover Trigger Condition Design

This section is based on the design of intelligent handover trigger condition with MAPPO algorithm to learn the handover trigger condition. The proposed MAPPO is a centralized training with MBS while decentralized execution with BSs the UE connecting framework [150]. The centralized critic and decentralized policy are learnt by the MBS for each UE with the proposed algorithm. Each UE updates its policy based on recent learning results from MBS periodically. Since the UEs in the mmWave system are interactive, the problem is modeled as a fully cooperative multi-agent task with reinforcement learning. This problem can be described as $\Gamma =< \mathscr{S}, \mathscr{A}, \mathscr{P}, \mathscr{R}, \mathscr{O}, \mathscr{N}, \gamma >$. $\mathscr{S}$ is the state space

while $\mathscr{A}$ is the shared action space for each agent. $o_i = \mathscr{O}(s;i)$ is the local observation for agent i at global state s. $\mathscr{P}(s'|s,A)$ represents the transition probability, while $\mathscr{R}$ is the shared reward function. γ is the discount factor, which is $\Sigma_t \gamma^t \mathscr{R}(s^t, a^t)$.

Action

The action of each UE in the system contains handover trigger or not at each time step t. To guarantee the QoS of the UE, MAPPO generates three candidate BSs with the shortest distance and calculate the SINR of UEs. At time step t, the action of UE i is expressed as:

$$a_t^i = \{0, 1\}, \tag{4.7}$$

where 1 represents trigger. Since all the UEs in the system are required to consider, A is denoted as the action space of all UEs, which is defined as:

$$A_t = (a_t^1, a_t^2, a_t^3, ..., a_t^i) \tag{4.8}$$

The reason why the action set of all the UEs' is generated is that there is a trade-off between the single UE reward and the overall system reward. The maximum reward of the single UE usually is not optimal in terms of the overall system reward. The specific statement to solve the trade-off is presented by the end of this sub-section.

State and Observation

The state and observation of each UE are based on beams and BSs selection, handover delay, bandwidth allocation, and system overall throughput, which is evaluated and calculated by the optimal handover decision scheme after the action are taken.

The current serving beam n in its BS m of the UE i is chosen at the previous time step $t-1$. At the start of each time step, the public information is sent by MBS to each UE. Specifically, for each BS $m \in M$, the total number of served UEs is defined as $I_t^m = \Sigma_{i \in I} n_{t-1}^i = m$. Therefore, the public information at time step t is $\mathbf{I}_t = (n_t^0, n_t^1, ..., n_t^m)$. At the beginning of each time t, the optimal handover decision scheme calculates the handover delay, bandwidth allocation, and overall system throughput based on UEs' actions taken at last time step $t-1$. In this case, for each UE, the observation can be denoted as:

$$s_t^i = (d_{t-1}^{i,m}, r_{t-1}, b_{t-1}^{i,m}, \mathbf{I}_t), \tag{4.9}$$

where $d_{t-1}^{i,m}$ is the handover delay of each UE at previous time step $t-1$, $b_{t-1}^{i,m}$ is the band-

width allocation of each UE at previous time step, and r_{t-1} is the overall system throughput at previous time step. Therefore, the global state as the ensemble of observations of all UEs can be defined as:

$$S_t = (s_t^1, s_t^1, ..., s_t^I) \in \mathscr{S}, \tag{4.10}$$

where $\mathscr{S}$ is the state space.

Reward

The reward of the the algorithm is divided in two parts: overall system throughput and delay evaluation and handover rate (HOR). Firstly, since the switch decision leads to the changes of throughput and delay, it is important to evaluate the handover trigger decision based on that. Therefore, the system performance reward is defined as:

$$\mathscr{R}^i = \begin{cases} 10\delta, & Rt > Rt_1 \wedge Dt < Dt_1 \\ \delta, & Rt > Rt_2 \wedge Dt < Dt_2 \\ -\delta, & Others, \end{cases} \tag{4.11}$$

where Rt_1 and Dt_1 are higher bound of system total throughput and delay, while Rt_2 and Dt_2 are lower bound.

Second, the HO penalty is defined, which is to avoid the unnecessary handover trigger decisions:

$$P_{HO}^i(s_t^u, a_t^u) = \varepsilon \not\Vdash \left\{ b_t^i \neq b_{t-1}^i \right\}, \tag{4.12}$$

where $\varepsilon \geq 0$ is the weighting factor. Therefore, the local reward of UE in time step t is expressed as:

$$\mathscr{R}^i = \begin{cases} 10\delta P_{HO}^i(s_t^u, a_t^u) & Rt > Rt_1 \wedge Dt < Dt_1 \\ \delta & Rt > Rt_2 \wedge Dt < Dt_2 \\ -\delta P_{HO}^i(s_t^u, a_t^u) & Others. \end{cases} \tag{4.13}$$

Since the problem is a multi-agent problem, it is modeled as a fully cooperative multi-agent task, where the total reward of UE is

$$\mathscr{R}(S_t, A_t) = \sum_{i=1}^{i} \mathscr{R}^i(s_t^i, A_t). \tag{4.14}$$

The total reward $\mathscr{R}(S_t, A_t)$ can guild agents to balance the trade-off between SINR trigger condition and HOR with the adjusting weighting factor ε.

Q-value and Policy

The state-action value function Q^π, the state value function V^π, and the advantage function A^π are defined as follow:

$$Q^\pi(s_t, a_t) = E_{s_{t+1}, a_{t+1}, \dots}[R_t | s_t, a_t],$$
$$V^\pi(s_t) = E_{a_t, s_{t+1}, \dots}[R_t | s_t], \qquad (4.15)$$
$$A^\pi(s_t, a_t) = Q^\pi(s_t, a_t) - V^\pi(s_t),$$

where π is the joint policy. The parameters ω for the critic $V_\omega(s_t)$ is updated by minimizing the loss

$$J(\omega) = \hat{\mathbb{E}}[(V_\omega(s_t) - y_t)],$$
$$y_t = \mathscr{R}_t + \gamma V_\omega(s_{t+1}), \qquad (4.16)$$

where $\mathscr{R}_t$ is the reward in time t, and V_ω is the target state-value function [200].

According to [150], the independent proximal policy optimization (IPPO) is one of the RL methods that implement the PPO algorithm on each UE independently, where each UE learns the actor and critic on its own. However, this method cannot approach the true overall state value since the state and action information is updated locally on the UE side. In this case, there is no global state information, and jointly action information shared on the UE side, which makes the advantage function of IPPO less accurate. In addition, the lack of joint actions makes it more difficult for the UE to learn about cooperation policies and assess the influence of UE action on the reward.

Therefore, the multi-agent proximal policy optimization (MAPPO) algorithm is proposed, a centralized training with a decentralized execution framework to improve the performance of the IPPO. In this case, global information is implemented for training the decentralized policies of each UE. More specifically, the global information is supposed to collect in MBS, and the learning procedure is also processed in the MBS.

The decentralised actors and centralised critics framework are implemented since the joint advantage function has strong relevance with the policy gradients. In this case, with the global information such as UE action a_t and UE state s_t available, the centralized critic evaluates the joint value function (Q or V) in the training process. At the same time, decentralized actors estimate based on UE's observations locally. When the training process finishes, global information is no longer required, which means the UEs can implement the actions in the decentralized actors. The basic MAPPO structure is shown in Fig. 4.3, in which there is a neural network in each actor and critic.

The state-value function $V_\omega^i(s_t)$ is estimated in the centralized critics with the critic
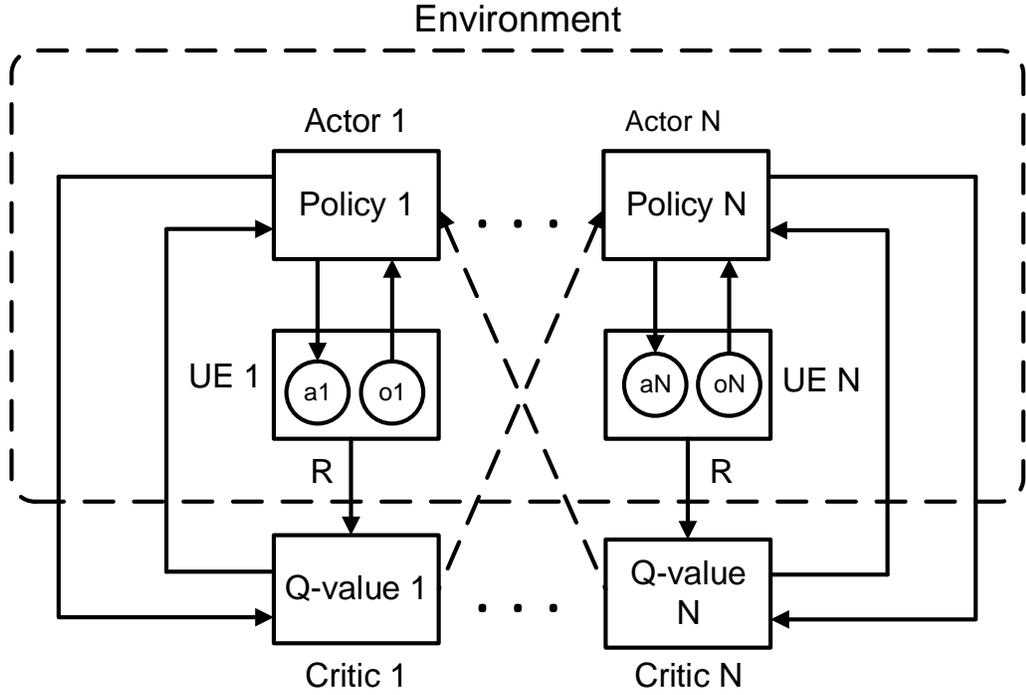
Figure 4.3: MPPO structure.

parameters ω^i of the UE. Since the expectation is replaced by sample averages in RL, the policy is updated with the gradient:

$$\Delta\theta^i = \nabla^i_\theta \hat{\mathbb{E}}_t \left\{ f(\mathscr{R}_t(\theta^i), A^i(s_t, a_t)) \right\}, \tag{4.17}$$

where $A^i(s_t, a_t)$ is the estimation of joint advantage function, which is calculated by generalized advantage estimation (GAE) [201] with the state-value function $V^i_\omega(s_t)$.

According to [150], there is a credit assignment problem, since it is not clear how a specific UE's action contribute to the reward. In order to solve it, the counterfactual baseline method proposed in [202] are used. In our case, a centralized critic $Q_{\omega^i}(s_t, a_t)$ is proposed to evaluate the action-value function. The joint quantities is denoted to UE as $-i$. Therefore, the advantage function for each UE is calculated by comparing the Q-value estimated by the critic for the executed action a^i_t to a counterfactual baseline that marginalizes out a^i_t, maintaining the actions of other UEs same:

$$A^i(s_t, a_t) = \hat{Q}^i(s_t, a_t) - b(s_t, a_t^{-i}), \tag{4.18}$$

where $b(s_t, a_t^{-i})$ is the counterfactual baseline, which can be defined as

$$b(s_t, a_t^{-i}) = \sum_{a^i} \pi_{old}^i(a^i \mid z_t^i) Q_\omega^i(s_t, (a_t^{-i}, a_i)), \tag{4.19}$$

where π_{old}^i is the initial guess of the optimal policy and $\hat{Q}^i$ is the estimation of $Q^{\pi_{old}}$, which is calculated by the temporal-difference (TD) [203]. Although each $\hat{Q}^i$ is calculated by separated critics, the joint action-value function $Q^{\pi_{old}}(s_t, a_t)$ is same.

Algorithm 2: MAPPO procedure.

Initiate critic Q_{ω^i} and actor π^i with θ^i, $\forall i$.
Initiate the initial policies π_{old}^i and target critic $Q_{\hat{\omega}^i}$.
initiate state.
for iteration=1,2,...,T **do**
    initiate state.
    for an episode **do**
        Executes action for each agent.
        Get reward and the new state.
    end
    Get the movement of each UE.
    Calculate the $\hat{Q}^i(s_t, a_t)$.
    Calculate all UEs' action space $A^i(s_t^i, a_t)$.
    Store the data $\{z_t^i, Q^i(s_t, a_t), A^i(s_t^i, a_t)\}$ into database D.
    for k=1,2,3,...,K **do**
        Shuffle and relabel the data.
        for j=0,1,2,...,$\frac{T}{B}-1$ **do**
            Select groups of data D_j:
            Calculate new action space.
            for l=1,2,...,L **do**
                Calculate gradient ascent $\Delta\theta^i$.
                Use minibatch Adam [186] to apply gradient ascent θ^i.
                Calculate gradient ascent $\Delta\omega^i$.
                Use minibatch Adam [186] to apply gradient ascent ω^i.
            end
        end
    end
    Update θ^i and ω^i for each UE.
    Clear database D.
end

Further, the discrete action space $\mathscr{A}$ is considered. After the state-action (s_t, a_t) is input into the critic, the scalar $Q_{\omega^i}(s_t, a_t)$ is obtained. However, $|\mathscr{A}|$ times evaluation is necessary when computing the counterfactual baseline, which makes the time-consuming

when the action space is getting larger. In this case, the critic structure is used [202]. The input of the neural network of the critic of UE is $Q_{\omega^i}(s_t, a_t)$ while the output is the state-action values of the UE. In order to distinguish whether the specific UE's action is marginalized, the critic structure requires there must be a critic for each UE. The procedure of MAPPO is presented in the Algorithm 1.

4.5 Optimal Beam Selection and Bandwidth Allocation for Handover UEs

After the handover trigger decision is made, the related channel information is sent to optimal handover decision scheme to optimize bandwidth allocation as well as beams and BSs selection, with the aim of improving the performance of handover delay as well as the overall system throughput while guaranteeing QoS of each UE. Further, the results of calculation and allocation will then feedback to the MAPPO algorithm to evaluate the handover trigger decision through reward function. From 3GPP [204], the handover delay in mmWave system is defined as:

$$D = T_R + T_I + T_T, \tag{4.20}$$

where T_R is the Radio Resource Control procedure delay, which is decided by the mmWave system; T_I is the handover interruption time which includes target cell searching time, target cell tracking and acquiring time, and interruption uncertainty time, which is also decided by the system. Thus, the handover transmission time T_T between UE and BS is the key to optimize the handover delay, which is expressed as:

$$T_T^i = \frac{PL_i}{R_i}, \tag{4.21}$$

where PL is the package length and R is the throughput of the system and it is defined by Shannon Formula:

$$R_i = B_i \log(1 + SINR_i), \tag{4.22}$$

where B is the bandwidth taken by the UE.

Based on the system model, the optimal handover decision scheme is denoted to improve the system performance in terms of throughput and delay while guarantee the QoS

of each UE, which can be denoted as:

$$\min \sum_{i \in I} \sum_{n \in N} \sum_{m \in M} x^i_{m,n} \times Delay_i, \qquad (4.23)$$

$$\text{s.t.} \quad \sum_{i \in I} \sum_{n \in N} x^i_{m,n} \times B_i \leq B_0 \quad MHz, \forall m \in M, \qquad (4.24)$$

$$\sum_{m \in M} \sum_{n \in N} x^i_{m,n} \times R_i \geq R_0 \quad Mbps, \forall i \in I, \qquad (4.25)$$

$$\sum_{m \in M} \sum_{n \in N} x^i_{m,n} = 1, \forall i \in I, \qquad (4.26)$$

$$Delay_i \leq D_0 \quad ms, \quad R_i \geq R_{i0} \quad Mbps \quad \forall i \in I, \qquad (4.27)$$

where $x^i_{m,n} \in (1,0)$ describes the UE connection statues. When $x^i_{m,n} = 1$, it means UE i connects with the beam n in BS m; on the contrary, $x^i_{m,n} = 0$. Since the maximum band-width of each SCBS is fixed, (24) is the constrain showing that the maximum bandwidth for each SCBS, which can provide to the UE. (25) is the constrain, which aims to optimize the total throughput of all UEs. A lower bound of the throughput is set, which formulates the minimum throughput all the UEs can gain from the system. In addition, each UE can only connect with one beam in one BS, which is constrained by (26). The (27) is the minimum delay and throughput that each UE must reach.

From the proposed optimal handover decision scheme ((23) - (27)), as it can be seen, since $x^i_{m,n} \in (0,1)$ and the equation (22) is a nonlinear function, the optimization function is a zero-one mix integer nonlinear problem and there are two unknown parameters ($x^i_{m,n}$ and B_i) to figure out. In this case, the problem is divided into three parts. Firstly, the Se-quential Quadratic Programming (SQP) algorithm [205] is utilized to solve the nonlinear part. In this case, the integer $x^i_{m,n}$ is relaxed as a continuous variable, which ranges from 0 to 1. Secondly, after the continuous $x^i_{m,n}$ is obtained, the tight relaxation algorithm [206] is used to transfer continuous variable into integer variable. Thirdly, after solving the integer problem, the rest of the optimization function becomes a linear problem: the function of B_i. It is solved with a linear algorithm to obtain the most suitable bandwidth B_i for each UE. With the bandwidth allocation of each UE, the throughput can be calculated; thereby, the delay of different UEs can be obtained.

According to SQP algorithm, the Lagrangian function of the optimization function is:

$$L = F + \alpha h_1 + \beta h_2 + \gamma h_3, \qquad (4.28)$$

where

$$F = \sum_{i \in I} \sum_{n \in N} \sum_{m \in M} x^i_{m,n} \times \frac{PL}{\log(1 + SINR_i)} \tag{4.29}$$

$$h_1 = \sum_{i \in I} \sum_{n \in N} x^i_{m,n} \times B_i \tag{4.30}$$

$$h_2 = \sum_{m \in M} \sum_{n \in N} x^i_{m,n} \times B_i \log(1 + SINR_i) \tag{4.31}$$

$$h_3 = \sum_{m \in M} \sum_{n \in N} x^i_{m,n} \tag{4.32}$$

Here, the *PL* is the package length and *SINR* is channel state information, which means that the two variable is known in the system. In this case, the optimization function is a nonlinear problem as the function of bandwidth (B_i). Then the first order approximation of the gradient of the Lagrangian function is figured out as:

$$\nabla L = \begin{bmatrix} \frac{dL}{dx} \\ \frac{dL}{d\alpha} \\ \frac{dL}{d\beta} \\ \frac{dL}{d\gamma} \end{bmatrix} = \begin{bmatrix} \nabla F + \alpha \nabla h_1 + \beta \nabla h_2 + \gamma \nabla h_3 \\ h_1 \\ h_2 \\ h_3 \end{bmatrix}. \tag{4.33}$$

Then, the second order approximation of the gradient of the Lagrangian function is:

$$\nabla^2 L = \begin{bmatrix} \nabla^2_x L & \nabla h_1 & \nabla h_2 & \nabla h_3 \\ \nabla h_1 & 0 & 0 & 0 \\ \nabla h_2 & 0 & 0 & 0 \\ \nabla h_3 & 0 & 0 & 0 \end{bmatrix}. \tag{4.34}$$

It is defined that $p = \frac{\nabla^2 L}{\nabla L} = \frac{\nabla^2 L(p)}{\nabla L(p)}$. In this case, the nonlinear optimization function can be simplified as:

$$\min(p)F_k(x) + \nabla F_k^T p + \frac{1}{2} p^T \nabla^2_x L_k p \tag{4.35}$$

$$\text{s.t.} \quad \nabla h_1 p + h_1 \le B_0 \quad MHz, \forall m \in M \tag{4.36}$$

$$\nabla h_2 p + h_2 \ge R_0 \quad Mbps, \forall i \in I \tag{4.37}$$

$$\nabla h_3 p + h_3 = 1, \forall i \in I \tag{4.38}$$

$$Delay_i \le D_0 \quad ms, \quad R_i \ge R_{i0} \quad Mbps \quad \forall i \in I. \tag{4.39}$$

Therefore, the nonlinear function is approximated in the linear function and the continuous $x_{m,n}^i$ can be obtained.

The tight relaxation is utilized to transfer continuous $x_{m,n}^i$ into integer. According to [207], the method used is implicit enumeration method. As it shown in Fig. 4.4, the procedure searching the possible $x_{m,n}^i$ is:

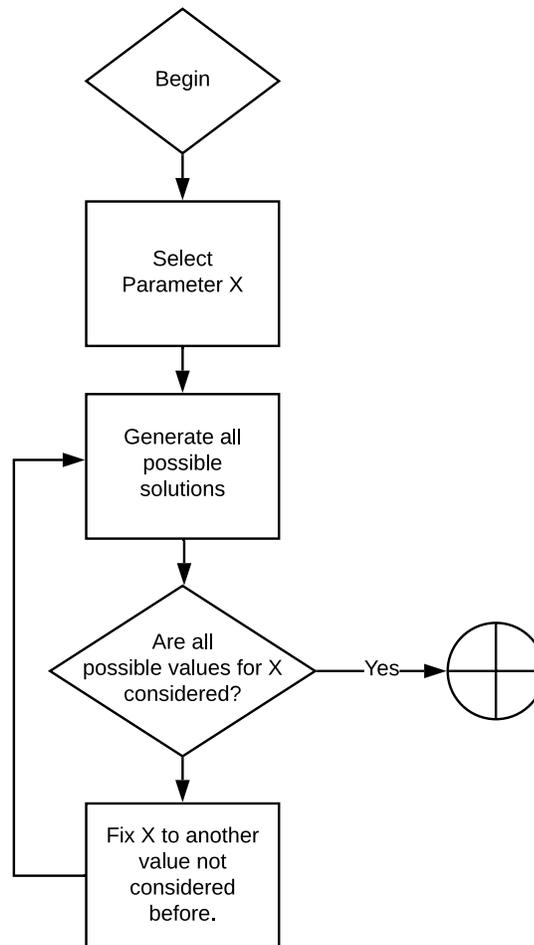Figure 4.4: Zero-One tight relaxation procedure flow diagram.

1. Since a set of variable $x_{m,n}^i$ can be obtained with the help one the SQP algorithm for each UE, the size of which is $m \times n$, we select the top three largest $x_{m,n}^i$ to one of the permissible integer values.

2. Resolve the problem in the remaining variable.

3. fix one of three $x_{m,n}^i$ to another permissible value.

4. repeat (2) and (3) until all possible for $x_{m,n}^i$ are considered.

This algorithm is a basic search that implies a general state of search in which all possible solutions are considered either explicitly or implicitly. Finally, the zero-one nonlinear integer problem is approximated into linear problem, which is easily to be solved.

4.6 Results and Discussion

In this section, the simulation setups are presented and then some numerical results with discussions and analysis are shown.

4.6.1 Simulation Setup

A two-tier heterogeneous mmWave cellular network is proposed, consisting of one microwave MBS, M_s mmWave SCBSs with the number of UE I, and for each SCBS, there are total N beams. Specifically, we donate $M_s = 6$, $N = 8$ and $I = 10$ as default. Cartesian coordinates describe the location of BS and SCBS. It is assume that there is an effective propagation coverage with 200 meters radius. The MBS is located in origin $(0,0)$, and the rest of the six SCBSs are evenly distributed in the considered area. Further, in each SCBS, eight directional beams are equally distributed with $45°$. The coverage of each SCBS has an overlapping area with its neighbour. In this case, a handover event occurs either when the current SINR of UEs is lower than the threshold or when the UEs are in an overlapping area for inter-cell handover. On the other hand, intra-cell handover occurs when UEs move from one beam area to another. Further, the UEs that the SCBS cannot cover are served by the MBS, which usually occurs on edge of the effective propagation area. The log-normal shadow fading of MBS ξ has zero mean and 3dB standard deviation, and the small-scale fading in linear value from $10^{\frac{\psi}{10}}$ follows an exponential distribution with unit mean [150]. The other channel parameters is summarized in Table I. The initial location of UEs are randomly distributed, and the movement of them obeys the random walk method [198] with the velocity 2 m/s.

For the hyper-parameters of MAPPO, the Root Mean Squared Propagation (RMSprop) optimizer with the learning rate $lr = 5 \times 10^{-4}$ is applied. One hidden layer with 64 units using Rectified Linear Unit (ReLU) activation function for the policies and critics is considered. The minibatch size to be 1, discount factor to be 0.9, and clip loss value 0.2 are set.

Table 4.1: Channel parameters of mmWave cellular network [3–5]

Parameters	Value
Bandwidth of SCBSs	100 MHz
Bandwidth of MBS	20 MHz
Pathloss parameters of LOS	$\alpha_S = 70, \kappa_S = 2$
Pathloss parameters of NLOS	$\alpha_S = 70, \kappa_S = 2$
Pathloss parameters of MBS channel	$\alpha_M = 70, \kappa_M = 2$
MmWave noise power density	-163 dBm/Hz
Microwave noise power density	-174 dBm/Hz
The maximum antenna gain	10 dB
Interval of each timestep	100 ms
Interruption time	100 ms
Building density	$1 \times 10^{-4}/m^2$
Expected length of buildings	25 m
Transmission power of MBS	46 dBm
Transmission power of SCBS	30 dBm

4.6.2 Results and Discussion

In the following sub-section, to evaluate the system performance in terms of system total throughput and delay, some simulations are implemented with different reward conditions.

Simulation I: The number of UE (I) is 10 and the configured upper bounds and lower bound of throughput (Rt_1), delay (Dt_1), (Rt_2), and delay (Dt_2) can be found in Table I, which means if the throughput or delay cannot reach the lower bound, the training episodes will be done in that round and the training will start again.

First, the training loss is shown in Fig. 4.5. As can be seen, the training loss reduces with the increasing training episodes, and it eventually converges to 0.1, which means the proposed method's good performance in this simulation and all the numerical results are promising.

Second, the probability density function (PDF) and distribution of system total throughput are shown in Fig. 4.6. During the training process, most of the system total throughput is distributed from 1800*Mbps* to 2300*Mbps*, and the episodes, of which throughput is lower than 1800*Mbps*, is 19.5%. Furthermore, from the system throughput distribution curve, the lower throughput occurs in the early stage of the training process. In this stage, the reward is not large enough, which implies that the learning process of UEs is bad. With the training continuing, the reward becomes larger and larger, and the system achieves good control of the SCBSs and beams allocation for each UE. Therefore, the total system
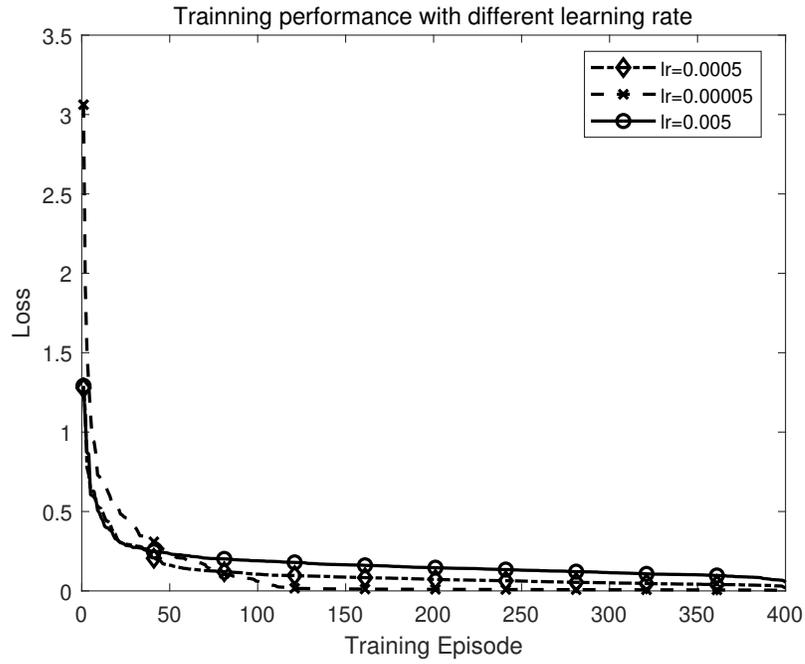
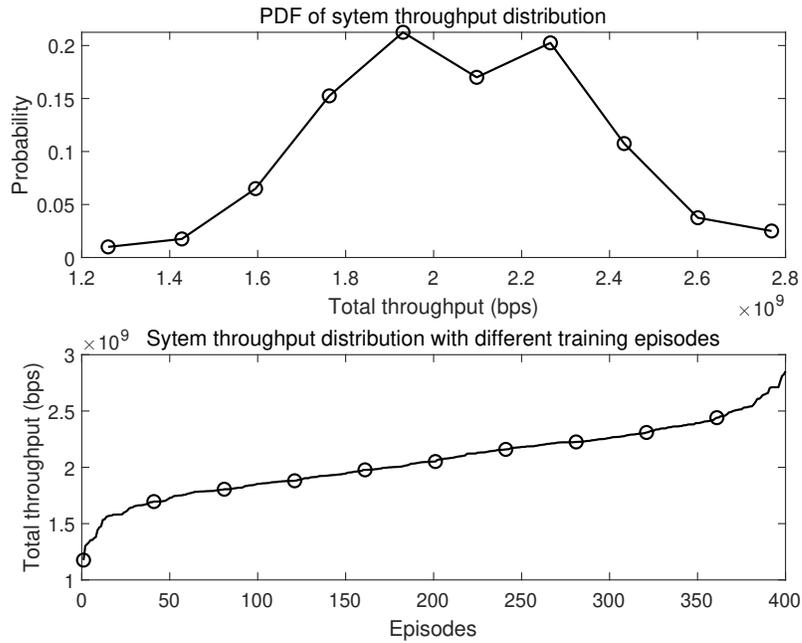Figure 4.5: Training loss in Simulation I.

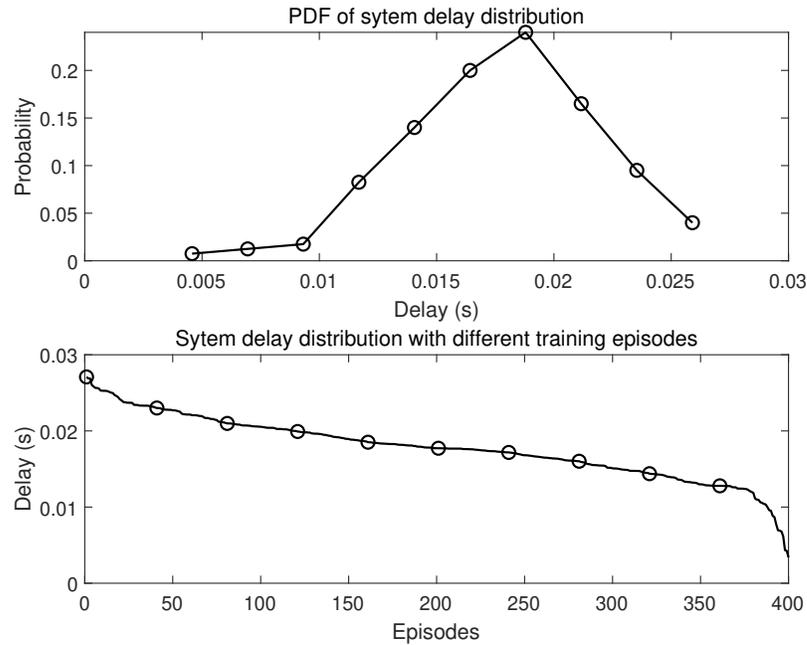Figure 4.6: System throughput performance in Simulation I.

Figure 4.7: System delay performance in Simulation I.

throughput increases.

Third, the PDF and distribution of system total delay are shown in Fig. 4.7. Alone with the training progress, the system total delay mainly distributes from $0.013s$ to $0.023s$. There are 0.16% of episodes, of which the delay is higher than lower reward bound Dt_2, and there are 53% of episodes, of which the delay is lower than the upper reward bound Dt_1. Further, the total delay reduces with the training episode increasing, which indicates that UEs achieving a better learning ability. The performance in terms of total system delay is better than total system throughput. The reason is that two conditions (high throughput and low delay) must trigger at the same time to obtain a good reward. Moreover, the upper trigger condition for the system delay is easier than system throughput.

Simulation II: In this simulation, there is a comparison between the proposed algorithm with other typical algorithms, such as Deep Deterministic Policy Gradient (DDPG), Deep Q Learning (DQN), exhaustive search and random connection in terms of system total delay and throughput performance, respectively. The lower bound and upper bound of reward are as same as that in Simulation I, and all of the algorithms are compared when the number of UE is 10 ($I = 10$).

From Fig. 4.8 and Fig. 4.9, it can be seen that the proposed method has the best performance, especially when the total system throughput is compared. The proposed method has 10% improvement over the DDPG and 25% over DQN. For the total system
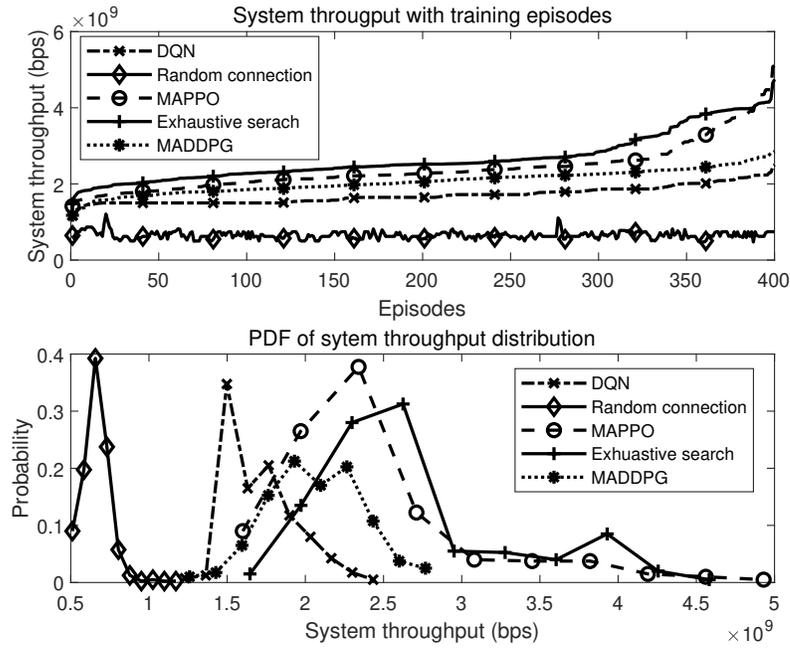
Figure 4.8: System throughput performance with different algorithms $UE = 10$ in Simulation II.
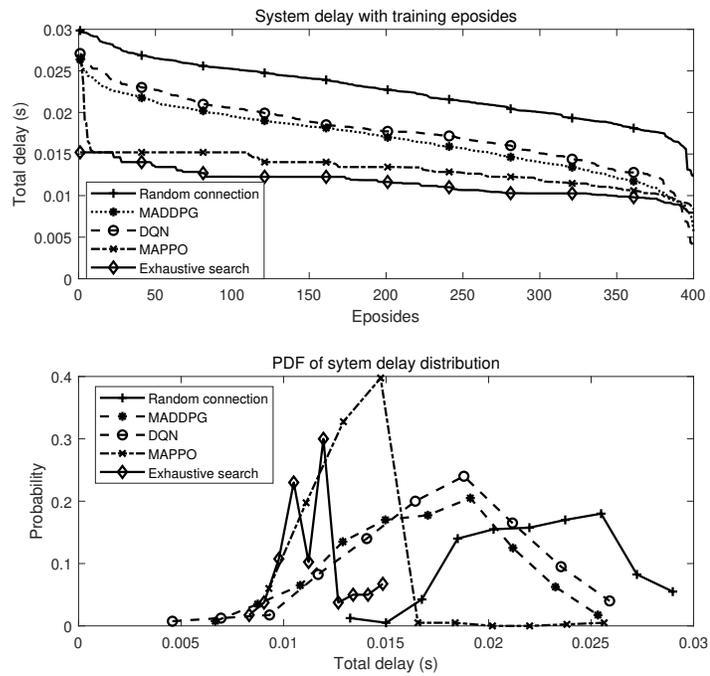
Figure 4.9: System delay performance with different algorithms $UE = 10$ in Simulation II.

delay performance, there is no obvious improvement against DDPG, but the method can achieve the minimum total delay with training processing.
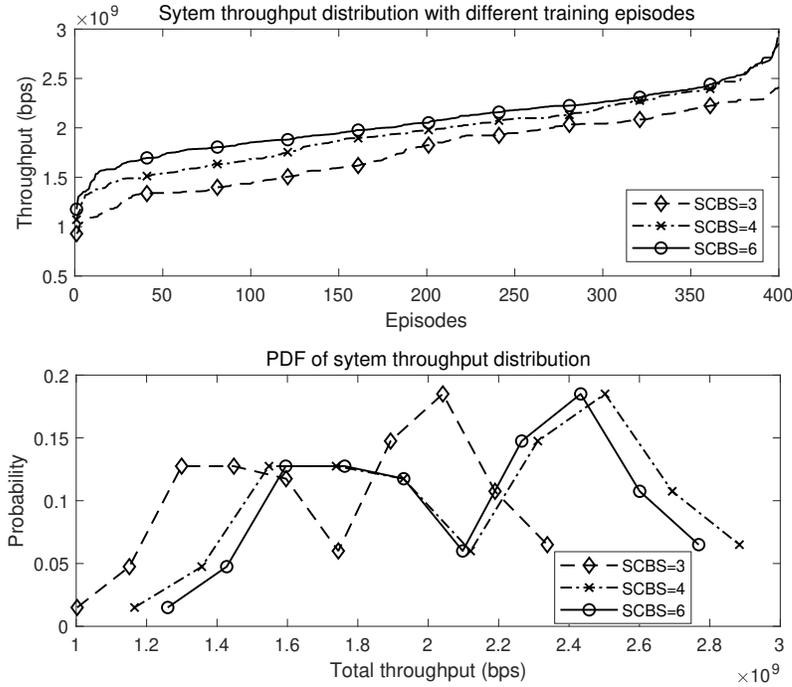
Figure 4.10: System delay performance with different SCBSs in Simulation III.

Simulation III: This section compares the system performance with different SCBSs to explore the system threshold of SCBS allocation when the number of UEs is 10. The number of SCBSs ranges from 3 to 7 in this simulation. However, when SCBSs are over six, the system throughput and delay performance are almost the same. Therefore, the performance of default SCBSs setting ($M_s = 6$) is shown in Fig. 4.10. The system total throughput performance reaches its maximum value when the number of SCBSs is 6, which has a small gap compared with that when $M_s = 4$. Further, the performance of $M_s = 3$ is the worst. When it comes to the total system delay performance, shown in Fig. 4.11, the more SCBSs lead the better performance until the system reaches its threshold. The reason is that since the MBS effective signal coverage is fixed, when the number of SCBSs is reduced, the radius of each SCBSs increases, making the propagation distance of mmWave in each cell longer. MmWave attenuates with propagation distance increases, which leads to the SINR of each UE reducing in high probability when connecting with SCBS. Thus, the system performance worsens with the number of SCBSs reducing. Meanwhile, when the number of SCBSs increases, the handover rate must be
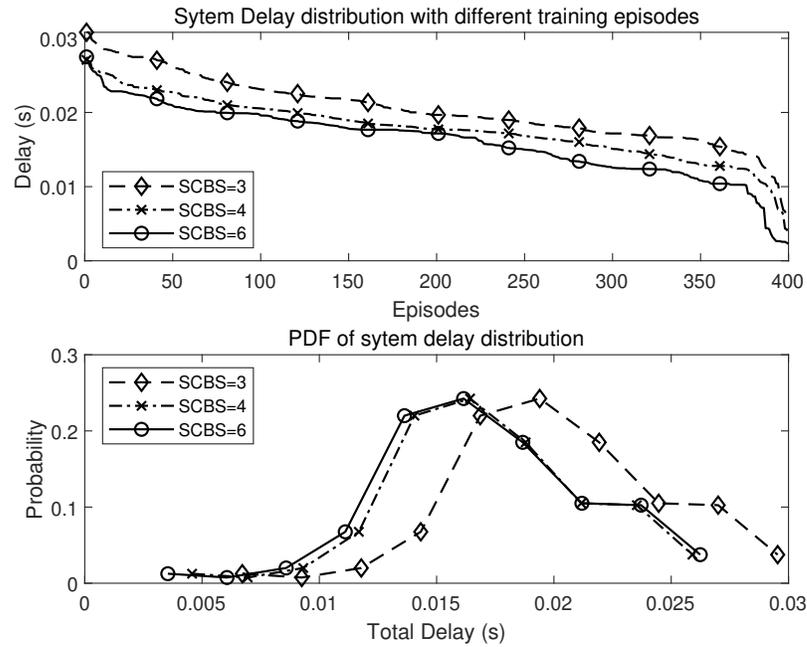
Figure 4.11: System delay performance with different SCBSs in Simulation III.

larger. Since each cell's radius is smaller, the overlapping area in the region grows, which means more handover events will occur.

4.7 Conclusion

An optimization-theory-based on one of the RL methods called O-MAPPO is proposed in this chapter to optimize the total system delay and throughput. Specifically, the RL algorithm called MAPPO is applied to improve the handover trigger decision. After the handover triggers, the related channel information is sent to the optimal handover decision scheme to optimize the beams and BSs selection, bandwidth allocation, maximize the overall system throughput and delay. Further, to avoid unnecessary HO events and reduce the HO rate, HO penalty strategy is implemented to improve the efficiency of the system. Simulation results demonstrate that with the training processing, The proposed method could achieve good performance in terms of total system throughput and delay compared with some typical RL algorithms, such as MADDPG and DQN.

Chapter 5

Intelligent Beam Management

5.1 Introduction

Despite the performance gains that beamforming and dense network provide in the mmWave cellular network, there are still some challenges existing. For example, the distribution of UEs and traffic could changes rapidly within a short time period [208]. Further, wireless networks of 5G and beyond are expected to serve UEs with diversified and tight QoS requirements [209]. With such network dynamics, beam management in the cellular network becomes more challenging. In addition, in current sub 6 GHz band system, the beam control procedures, including initial access, beam adjustment, and beam recovery, are performed with omidirectional signals, and beamforming then can be implemented only after the physical link is established. However, in mmWave band, there is high probability to generate a mismatch with the omidirectional signal between the relatively short range at which a cell can be detected or the control signals can be received. Further, directional beams can significantly delay the access procedures and make the performance more sensitive to the beam alignment [93].In this case, a new beam management scheme fast beam discovery, high transmission rate, and low energy consumption is essential to be explored.

To meet the demands mentioned above, one of the widely-used AI algorithms, RL, can be used to design a smart beam management mechanism in the mmWave cellular network for mobile UEs. On one hand, compared with the typical mathematical tool, RL algorithm is more efficient to solve the problem in the dynamic and complex environment, since the action decisions will be optimized as the environment changes. On the other hand, RL algorithm can make the smart action decisions of beam measurement and beam determination procedures in beam management, which leads to the optimization of system performance, such as overall system throughput. Specifically, the different beams in

the SCBS is intelligently managed to serve the different UEs, which could provide the maximum transmission rate to the UEs.

This chapter proposes an RL-based beam management scheme on the SCBSs side to maximize the overall system throughput in mmWave dense cellular network. Specifically, an RL method named multi-agent deep deterministic policy gradient (MADDPG) is applied for the multi-agent scenario to optimize the beam determination procedure. Some directional beams with different beam sectors are used to track the UEs and provide the signal link with strongest signal to interference and noise ratio (SINR). From the numerical results, it is demonstrated that the proposed method achieves better performance with the comparisons of other typical beam management method and other RL algorithms.

5.2   System Model

5.2.1   MmWave Cellular Network Topology

A two-tier heterogeneous network is considered, which consists of one macro base station (MBS) and M small cell base stations (SCBS) with N beams in each SCBS. The set of BSs is denoted as $M = \{0, 1, 2, ..., M\}$, in which 0 represents the index of MBS and the index of SCBSs starts from 1 to M. Further, the set of UEs and beams are denoted as $I = \{0, 1, 2, ..., I\}$ and $N = \{0, 1, 2, ..., N\}$, respectively. The SCBSs are distributed uniformly within ranges of the coverage of MBS while UEs are deployed randomly. SCBSs have priority to serve the UE within their coverage. Each UE could only be served by one SCBS with one beam, however, different beams could serve multiple UEs. Further, the transmittable directions of beams in each SCBS the are in the sectors of $\lambda = 360°/\alpha$, where α is the width of beam in degree [162]. The distribution of BSs, UEs and beams is shown in Fig. 5.1. The mobility of UE is based on random walk model [210]. SCBSs serve the UE in its range and MBS takes care of mobility and control signals when UE is moving out of the range of any SCBS. Handover triggers when UE moving to the overlapping areas among SCBS. If UE keeps moving in the small cell, the current serving beam will track it until it moves out the range.

5.2.2   Channel Model

The system achieves in a synchronous equal-length time-step, the time slot of which is $t = 0, 1, 2, ...$ and the length of time-step is Δt [197]. A centralized training is processed in MBS with global information and decentralized execution applied in SCBSs. It is con-
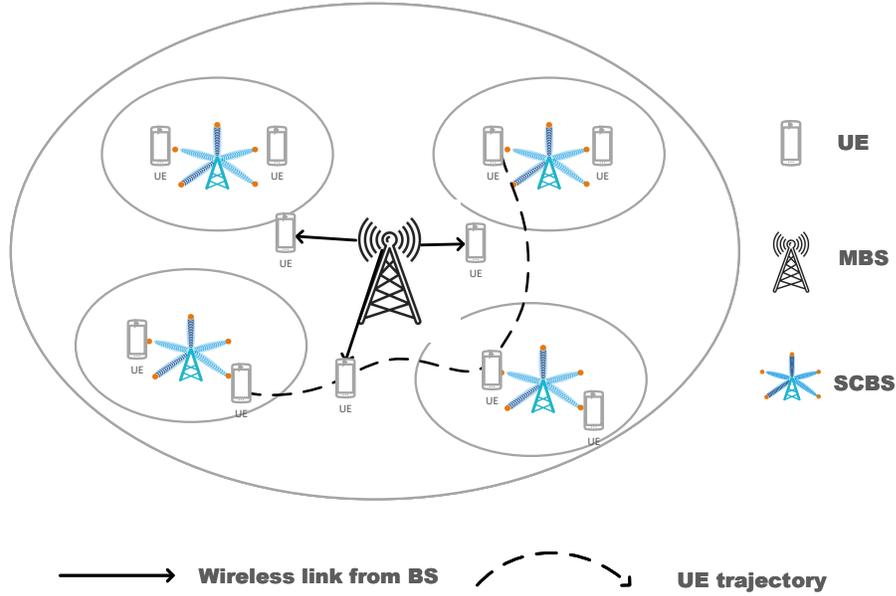
Figure 5.1: UEs and BS Distribution.

sidered that there is an omi-directional antenna deployed in the MBS to assure the signal coverage. An alpha plus beta model is employed for the path loss between transmitter and receiver, which is given by [211]

$$PL(d)[dB] = \alpha_M + 10\kappa_M \log_{10}(d) + \psi + \xi, \tag{5.1}$$

where d is the distance in meters, κ_M is the path loss exponent representing the slope of the best linear fit to the propagation measurement in the mmWave band, α_M is the path loss factor, ψ is random small-scale fading, and ξ is the random lognoraml shadowing.

The distance between each UE and MBS is d_i^0 and the transmitted power of MBS i is p_i^0, which meets the satisfaction of $\sum_i^I p_i^0 = P_M$, where P_M is the total transmitted power of one SCBS. Further, due to the shared bandwidth in MBS, there is a co-channel interference β_i [1]. Therefore, the SINR of each UE is denoted as:

$$SINR_i = \frac{p_i^0/PL}{\beta_i + N_M \omega_i^0}, \tag{5.2}$$

where PL is in linear, N_M is the noise power spectral density of MBS, and ω_i^0 represents the bandwidth allocated to the UE from MBS.

---

[1]The interference is the sum power received on the UE side from MBS nearby small cell base station.

Before the channel mode of SCBSs is defined, the line-of-sight (LOS) and non-line-of-sight (NLOS) mode need to be defined first. Since the penetrability of mmWave is poor, which leads to severe fading even outages of mmWave link in SCBSs. When UEs are served by NLOS link, there is a high signal attenuation, which makes the received power significantly bad. Thus, the SINR of UEs is greatly affected. In order to fit the real environment, the statistical mode is considered to denote the probability of the LOS, NLOS and outage [212], which is

$$v_i^m = \exp(-\frac{2D_B X_B d_i^m}{\pi}), m \neq 0, \tag{5.3}$$

where D_B is the building density, X_B is the expectation length of the buildings, and d_i^m is the distance from UE to SCBS.

Furthermore, the channel model of SCBSs can also be defined with the alpha plus beta model and the path loss of which is denoted as

$$pl(d)[dB] = \alpha_S + 10\kappa_S \log_{10}(d), \tag{5.4}$$

where d is the distance in meters, α_S and κ_S is same as that in Eqt. (1), which is path loss factor and exponential decay factor, respectively[2]. The random small-scale fading (ψ) and random lognormal shadowing (ξ) are ignored since the LOS-NLOS probability mode has already considered.

Assuming there is an omnidirection antenna on UE side and directional antenna on SCBSs side, an ideal sectorized antenna gain model is applied to approximate the real-world antenna pattern [193], which can obtain the important features of mmWave antenna. Therefore, the antenna gain is denoted as

$$g(\phi) = \begin{cases} g_{max}, & |\phi| < \dfrac{\phi_S}{2} \\ g_{min}, & otherwise \end{cases} \tag{5.5}$$

where ϕ is the angle between UE and BS, and ϕ_S is the width of the antenna main lobe. In our case, it is assume that the perfect beam alignment performed, which means the UE is always served by main lobe to obtain the maximum antenna gain. To make the procedure of beamforming more effective and energy efficiency, the hybrid precoding technique is applied to obtain the beamforming gain.

With the path loss model and antenna gain model, the SNR of UEs served by SCBSs

---

[2] α_S and κ_S have different values in LOS and NLOS cases.

is

$$SNR_i^m = \frac{g_{max}(p_i^m/PL)}{N_S}, m \neq 0 \in M,$$ (5.6)

where PL is in linear, p_i^m is the transmission power between UE and SCBS, satisfying $\sum_i^I p_i^m = P_S$, and N_S is the noise power spectral density among SCBSs. According to [3,5], the interference could be ignored.

5.3 Methodology

In this section, the beam management scheme is proposed based on MADDPG algorithm to improve over all system throughput.

5.3.1 Problem Formulation

In this section, the beam management problem is formulated as an optimization problem to improve the system performance in terms of maximizing the over all system throughput, meanwhile the QoS of UE in this cases, such as transmission delay and achievable date rate is guaranteed by constraints in optimization functions. From 3GPP [213], the transmission delay of each UE on the uplink is defined as

$$D = T_{package} + T_{transmission},$$ (5.7)

where $T_{package}$ is the package transmission time and $T_{transmission}$ signal transmission time. Since signal transmission time is closely related to distance and transmission speed, in the mmWave cellular network, compared with the transmission speed is the speed of light and the range of SCBSs is fixed and the radius of them is less than 200 meters. Therefore, $T_{transmission}$ is neglectable. In this case, the delay of each UE is

$$D = \frac{P}{R},$$ (5.8)

where P is the package length and R is the achievable data rate of each UE, which can be calculated according to Shannon Formula

$$R = B\log(1 + SINR),$$ (5.9)

where B is the available bandwidth of each beam.

Therefore, the optimization function to maximize the overall system throughput is

$$\max \sum_{i \in I} \sum_{n \in N} \sum_{m \in M} x^i_{m,n} \times R_i, \tag{5.10}$$

$$\text{s.t.} \quad \sum_{i \in I} \sum_{n \in N} x^i_{m,n} \times D_i \leq D_{total} \quad ms, \forall m \in M, \tag{5.11}$$

$$\sum_{m \in M} \sum_{n \in N} x^i_{m,n} = 1, \forall i \in I, \tag{5.12}$$

$$D_i \leq D_0 \quad ms, \quad R_i \geq R_{i0} \quad Mbps \quad \forall i \in I, \tag{5.13}$$

where $x^i_{m,n} \in [0,1]$ describes the UE connection statues, in which $x^i_{m,n} = 1$ represents UE connects with beam n in BS m. Constraint (11) is constrain that the overall system delay must be lower than D_{total}. Constraint (12) is the constrain that each UE can only be served by one beam and BS. Constraint (13) is the constrain to guarantee the basic QoS of each UE in terms of achievable data rate and transmission delay. The optimization function is a zero-one mix integer nonlinear problem, which can not directly be solved by traditional methods. However, with the priori information, e.g., BS and beam connection status, this problem can be transferred as linear problem and solved by regular methods. To obtain the priori information, more efficient and available methods should be designed and developed.

5.3.2 Markov Game for Multi-agents

The proposed beam management problem is modeled as cooperation Markov game, which is a discrete-time Markov decision process (MDP) [214]. A Markov game for I agents is modeled as $(\mathscr{S}, \mathscr{A}, \mathscr{R}, \mathscr{P}, \gamma)$, which represents state space, action space, reward space, transition probability, and reward discount factor. At time-step Δt, the agent observes the current state and make an action based on a certain policy, which leads to a new state with the transition probability. Then the algorithm will give a reward or penalty with discount factor to the action based on the comparison between current state and new state. The details of Markov game is defined as

Agent: Each SCBS is regarded as the agent in the proposed model. Compared with UEs as the agent, the beams in base station can obtain more information from the environment, such as UE location and the status of other beams within or without same SCBS, which make beam management more efficiency. Further, since the proposed method is an online algorithm, the computational ability of SCBSs is more powerful than UE devices.

Action: the action of SCBSs is the direction that each beam can change t, which is

$\lambda = 360°/\alpha$. In other words, there is total λ different azimuth that each beam can select. It makes there are the number of λ discrete actions in the action space. Thus, the action of each beam can be denoted as

$$a_t^i = (0, ..., \lambda), \tag{5.14}$$

and the action space for all agent is

$$A_t = (a_t^1, a_t^1, ..., a_t^N) \in \mathscr{A}. \tag{5.15}$$

State and Observations: the state in the model is based on the results of optimization function in the last section and the observations of each agent are SINR, the amount of transmission delay and achievable data rate the beam can provide to the serving UEs. The total number of served UE can be expressed as $I_t^m = \sum_{i \in I} n_{t-1}^i = m$, therefore, the public information at time step Δt is $\mathbf{I}_t = (n_t^0, n_t^1, ..., n_t^m)$. At the beginning of each time t, the SINR, achievable data rate, and delay are calculated based on the actions agents taken at previous time $t-1$, therefore the observation in the model can be defined as

$$s_t^i = (d_{t-1}^{i,m}, r_{t-1}, \mathbf{I}_t), \tag{5.16}$$

where $d_{t-1}^{i,m}$ is the transmission delay of beam N can provide to the serving UE while r_{t-1} is the achievable data rate, which is sent to the optimization function mentioned before to maximize the overall system throughput. Therefore, for each agent, the global state can be denoted as

$$S_t = (s_t^1, s_t^1, ..., s_t^N) \in \mathscr{S}, \tag{5.17}$$

where $\mathscr{S}$ is the state space.

Reward: the reward of the model is based on the overall system throughput. After generating the global state of all agents, the performance is evaluated in terms of the amount of achievable date each beam can provide to the serving UEs, which then can be used for calculating the system overall throughput. In this case, the reward function of the model is

$$\mathscr{R}^n = \begin{cases} \beta_0, & R_t > R_{t1} \\ \beta_1, & R_{t2} < R_t < R_{t1} \\ -\beta_2, & Others \end{cases} \tag{5.18}$$

where R_{t1} is the highly expected achievable data rate the agent can provide while R_{t2} is the lower bound achievable data rate the agent must provide. Correspondingly, the reward β_0 and β_1 is to evaluate the related actions. If the decisions made by agent fail to provide

the achievable data rate R_{t2}, the actions is penalized.

Further, the transition probability and reward discount factor are defined $P(s_n^{\Delta t}|s_n^{\Delta(t-1)}, a_1^{\Delta(t-1)}...a_n^{\Delta(t-1)})$ and $\Sigma_{n=0}^N \gamma^n \mathscr{R}_n^{\Delta t}$, respectively [161].

5.3.3 MADDPG Scheme Design

The performance of the traditional RL methods, such as Q-Learning and policy gradient are not satisfactory in multi-agent RL environment. The reasons is that the policy of each agent is changing with the training progressing, leading to the environment becoming non-stationary from the perspective of any individual agent. Hence, the environment cannot be explained by the changes in the agents' own policy. Moreover, policy gradient methods typically exhibit high variance when coordination of multiple agents is required, which makes instability of adversarial training methods [215]. This chapter proposes an RL method called MADDPG to solve the optimization function mention above. All RL methods including MADDPG are based on model free algorithm, which does not depend on the complete knowledge and statistical information in the system. With some common information, such as BS location, the initial UE location and channel information, MADDPG can solve the complicated problems and thus optimize system overall throughput [216].

Since MADDPG is a multi-agent extension of DDPG which is a variant of DPG [217]. Let us start with DPG and the gradient of objective $J(\theta) = \mathbb{E}_{s \to p^\mu}[\mathscr{R}(s,a)]$ can be denoted as

$$\Delta_\theta J(\theta) = \mathbb{E}_{s \to D}[\Delta_\theta \mu_\theta(a|s)\Delta_a Q^\mu(s,a)|_{a=\mu_\theta(s)}], \tag{5.19}$$

where μ_θ is the deterministic policy, Q^μ is the Q-value in critic and μ is the policy.

However, the updated network $Q(s,a|\theta^Q)$ cannot be directly used for calculating the target value, therefore DDPG applies a copy of actor and critic network, which denotes as $Q'(s,a)$ and μ_θ' to compute the target value. Then, the DDPG can be extended to the multi-agent scenario. Specifically, for a game with L agents, denoting $\mu_\theta = \{\mu_\theta 1,...,\mu_\theta L\}$ as the set of all deterministic policies for all agents, which can be simplified as $\theta = \{\theta_1,...,\theta_L\}$. The gradient of the expected return for each agent is

$$\Delta_{\theta i}J(\mu\theta_i) = \mathbb{E}_{\mathbf{o},\mathbf{a} \to D}$$
$$[\Delta_{\theta i}\mu_{\theta i}(a_i|s_i)\Delta_{ai}Q^\mu(\mathbf{o},a_1,...,a_L)|_{a_i=\mu_{\theta i}(s_i)}], \tag{5.20}$$

where $Q^\mu(\mathbf{o},a_1,...,a_L)$ is the centralized action-value function with the inputs of all the agents' action $\mathbf{a}$ and the observation $\mathbf{o}$.

The critic network is update by minimizing the loss function:

$$Loss(\theta_i) = \mathbb{E}_{\mathbf{o},\mathbf{o}',a,r}[(Q_i^{\mu}(\mathbf{0},a_1,...,a_L) - y)^2], \tag{5.21}$$

where

$$y = r_i + \gamma Q_i^{\mu'}(\mathbf{o}',a_1',...,a_L')|_{a_{i'}=\mu'(o_{i'})}. \tag{5.22}$$

The training algorithm of MADDPG can be found in Algorithm 1.

Algorithm 3: MADDPG procudure.

---

for episode = 1 to M **do**

    Initialize a random process $\mathcal{N}$ for action exploration

    Receive initial state S

    for t=1 to max-episode-length **do**

        for each agent i **do**

            Select action a_i w.r.t the current policy and exploration

            Execute actions A and observe reward $\mathcal{R}$ and new state s'

            Store $(S,A,\mathcal{R},S')$ in replay buffer $\mathcal{D}$

            $x \leftarrow x'$

        end

        for agent $i = 1$ to N **do**

            Sample a random minibatch of X samples $(S^j,A^j,\mathcal{R}^j,S'^j)$ from $\mathcal{D}$

            Set $y = r_i + \gamma Q_i^{\mu'}(\mathbf{o}',a_1',...,a_L')|_{a_{i'}=\mu'(o_{i'})}$

            Update critic by minimizing loss function $Loss(\theta_i$

            Update actor using the sampled policy gradient $\Delta_{\theta_i}J(\mu\theta_i)$

        end

        Update the target network parameters for each agent:

          $\theta_i` \leftarrow \tau\theta_i + (1-\tau)\theta_i'$

    end

end

5.4 Numerical Results and Discussions

In this section, the simulation setup is first introduced and then some numerical results with discussions and analysis are presented.

Table 5.1: Channel parameters of mmWave cellular network.

Parameters	Value
Bandwidth of SCBSs	100 MHz
Bandwidth of MBS	20 MHz
Pathloss parameters of LOS	$\alpha_S = 70, \kappa_S = 2$
Pathloss parameters of NLOS	$\alpha_S = 70, \kappa_S = 2$
Pathloss parameters of MBS channel	$\alpha_M = 70, \kappa_M = 2$
MmWave noise power density	-163 dBm/Hz
Microwave noise power density	-174 dBm/Hz
The maximum antenna gain	10 dB
Interval of each timestep	100 ms
Interruption time	100 ms
Building density	$1 \times 10^{-4}/m^2$
Expected length of buildings	25 m
Transmission power of MBS	46 dBm
Transmission power of SCBS	30 dBm

5.4.1 Simulation Settings

There is a two-tier heterogeneous mmWave cellular network consisting of one microwave MBS, M_s mmWave SCBSs with the number of UEs I, and for each SCBS, there are total N beams. Specifically, we donate $M_s = 6$, $N = 8$ and $I = 5$ as default. In each SCBS, there are $\lambda = 360°/\alpha$ in-directional beams and we donate $\alpha = 45°$. It is assume that there is an effective propagation coverage with 200 meters radius. The MBS is located at the center, and the rest of the six SCBSs are evenly distributed in the considered area. The log-normal shadow fading of MBS ξ has zero mean and 3dB standard deviation, and the small-scale fading in linear value from $10^{\frac{\psi}{10}}$ follows an exponential distribution with unit mean [150]. The other channel parameters is summarized in Table I. The initial location of UEs are randomly distributed, and the movement of them obeys the random walk method [198] with the velocity $2\ m/s$.

The simulation results are demonstrated by comparing the proposed work with following benchmark algorithms.

1. Traditional Beam management scheme: from [218], the beam sweeping procedure is carried out with the exhaustive search, i.e., there is a predefined codebook of di-
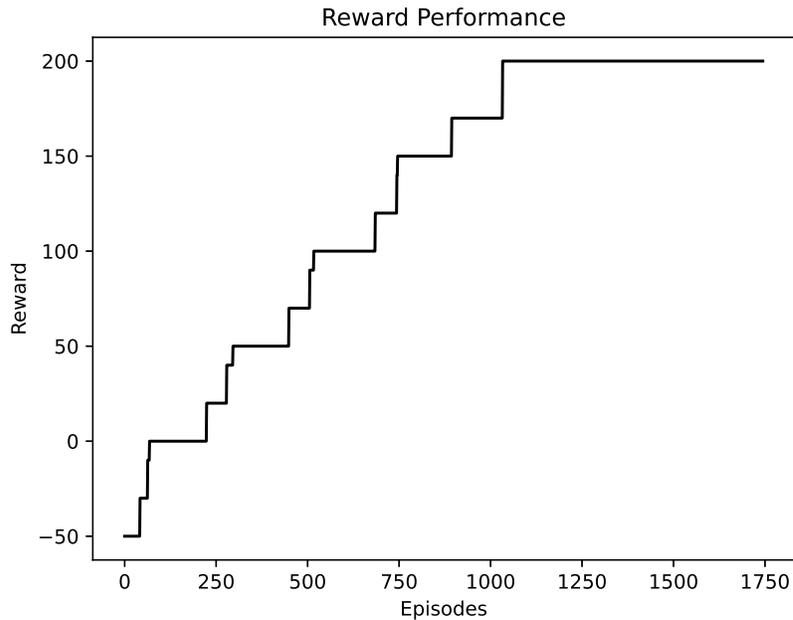
Figure 5.2: Reward performance of MADDPG.

rections which covers the whole angular space for both UEs and BSs. It is applied sequentially to transmit/receive synchronization and reference signals. Further, in the beam determination procedure, the mobile terminal selects the beam which provides the maximum SINR. Therefore, on the SCBSs side, there is a design that the beams sweep the UEs in their signal coverage and serve the UEs which beams could provide the highest SINR to.

2. Deep RL (DRL): [162] adopts Deep Q-learning to optimize the beam management and improve the radio resources allocation efficiency. Further, in [160], Long Short Term Memory (LSTM) based DRL is used for resource block allocation. However, DRL has bad performance when it is in the multi-agent scenario. Therefore, the Asynchronous Actor-Critic Agent (A3C) is applied, which generates the advantages of DQN and Policy Gradients. A3C uses multiple agents with each agent having its own network parameters and a copy of the environment. This agents interact with their respective environments Asynchronously, learning with each interaction.

For the hyper-parameters of MADDPG and A3C, the same parameters: learning rate in actor is 1×10^{-4}, learning rate in critic is 1×10^{-3}, epsilon greedy is 0.1, discount factor is 0.95 are applied. The policies and critics are parameterized by adopting the two-layer Rectified Linear Unit (ReLU) with 64 units per layer and Adam optimizer.
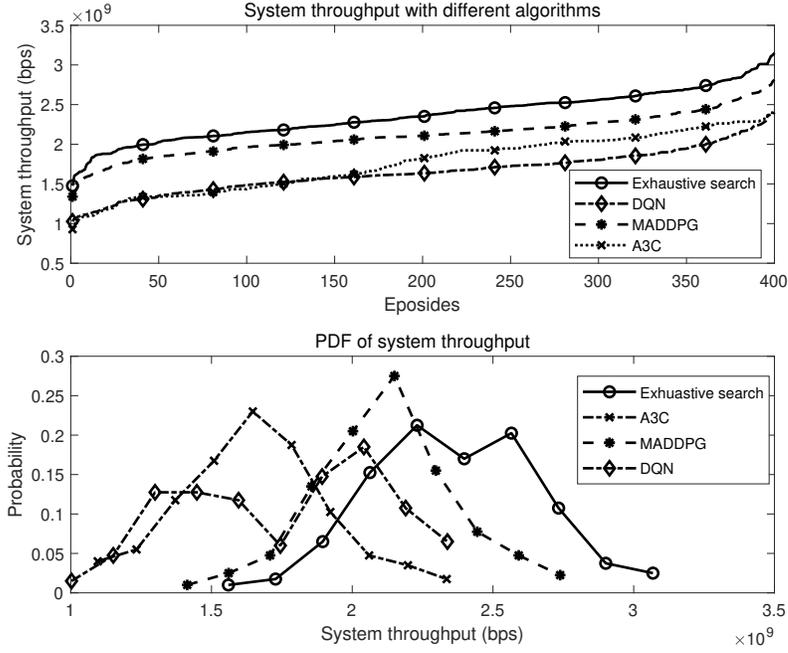
Figure 5.3: Overall throughput comparison with different schemes.

5.4.2 Results and Discussion

To demonstrate the convergence of the proposed method, Fig. 5.2 shows the trend of reward in the training process. As it can be seen, the reward increases with the training episodes growing and finally coverage, which means that the MADDPG algorithm is effective in the simulation as expected. Further, a simulation is conducted to compare the proposed algorithm with benchmark algorithms: a traditional beam selection method and a typical RL algorithm. In Fig. 5.3, as expected, the traditional beam selection method (exhaustive search) obtains the maximum overall throughput in the simulation, the performance of the proposed method (MADDPG) is slightly worse than the exhaustive search, and the performance of another typical RL algorithm (A3C) is the worst. Although the exhaustive search has the best performance, the computational complexity of MADDPG is much lower than that of exhaustive search. Further, the proposed method is based on the exploring and learning from the environment mechanism, which is much more effective than exhaustive search. In addition, when designing the reward function in this simulation, it is denote that when each UE obtain is 3.3×10^8 bps, is will be slightly rewarded and when each UE generate is 3.5×10^8 bps, there is a large reward. In this simulation, there are 5 UEs ($i = 5$), therefore the system will be slightly reward with 1.65×10^9 and get large reward with 1.75×10^9, respectively. From the Probability Density Function (PDF)
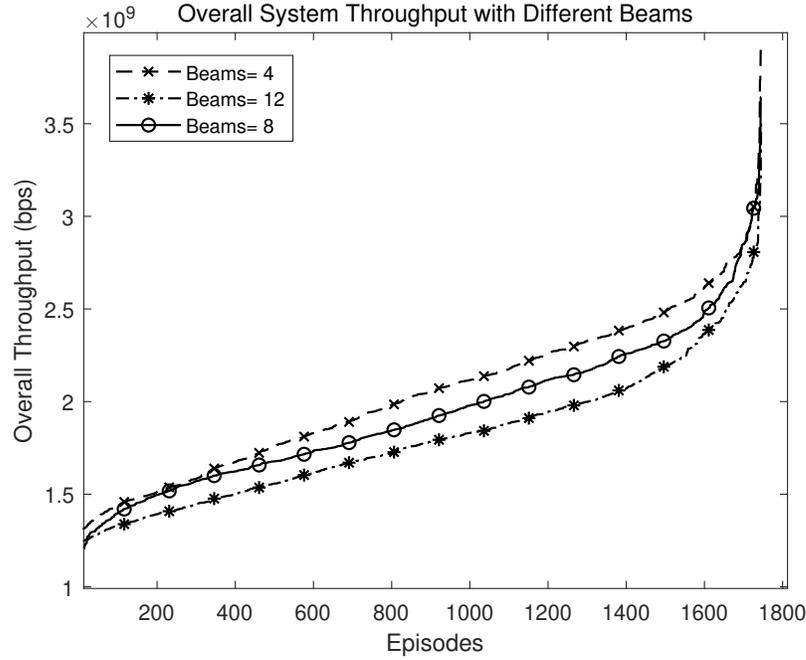
Figure 5.4: Overall throughput comparison with different beams.

of the overall system throughput, it can be seen that the proposed method has larger overall throughput than the that with exhaustive search when the values of overall throughput is over 1.65×10^9 and 1.75×10^9. It demonstrate that the proposed method could guarantee the good overall throughput for the most of time compared with the exhaustive search.

Moreover, next simulation is conducted to describe trend of overall throughput with different available beams in each SCBS. In this simulation, the value of α in $\lambda = 360°/\alpha$ with $90°$, $45°$, and $30°$ is changed, which leads to the available beams grow in each SCBS. In Fig. 5.4, it can be seen, the overall throughput decreases with the available beams. When there are 4 available beams in each SCBS, the beam coverage is the largest, which leads one beam to serve more UEs. Thus, the highest overall throughput is achieved. Further, when there are 12 available beams, the performance of obtaining the overall throughput is the worst, since the resources, i.e. bandwidth, in each SCBS is limited. Hence, there is a trade-off between the resource allocation and the number of available beams in each SCBS.

5.5   Conclusion

A multi-agent reinforcement learning (RL) based beam management scheme is proposed in this chapter, which aims to maximize the overall system throughput in the dense hetero-geneous mmWave network while guaranteed the basic QoS of individual UE. Specifically, an RL algorithm called MADDPG is used on each SCBS with in-directional beams to optimize the beam selection decisions. Numerical results revel the convergence perfor-mance of the MADDPG and the superiority in improving the system overall throughput compared with other typical RL algorithms and the traditional beam selection method.

Chapter 6

Conclusion, Open Issues, and Future Trace

This thesis investigates how ML assists the mmWave cellular network and beyond. Specifically, a DL enabled beam tracking method for NLOS mmWave communication, a novel scheme for handover in mmWave cellular network based on RL and optimization theory, and an RL-based intelligent beam management strategy in mmWave networks are proposed. In this chapter, this thesis's main work and contributions are drawn. In addition, a brief discussion on the open issues and future trends of the ML-enabled mmWave cellular network is proposed.

6.1 Conclusion

For next-generation networks and beyond to keep updating and on par with state-of-the-art intelligent systems, a paradigm change must be considered, which most likely requires the integration of some advanced intelligent solution, such as ML algorithm. Since thousands of parameters must be configured in the next-generation network and beyond, thousands of cells need to be monitored and optimized, and a large amount of data needs to be processed. In this case, it is not efficient and effective to do all these works by humans; instead, the machines will become dominant parts. Therefore, ML will be a promising solution to learn models in a relatively short amount of time and enable an autonomous and intelligent network. As such, ML is expected to play a vital role in the next-generation network and beyond to achieve its full potential.

In this context, the different ML algorithms that assist in optimizing different scenarios in the mmWave cellular network is explored in this thesis. First, the general background

of the mmWave cellular network with its challenges and how ML could benefit the network and provide the potential solutions are proposed in Chapter 1. After that, the specific scenarios is focused on in the mmWave cellular network in terms of NLOS communication, handover management, and beam management. Based on that, a brief introduction of handover management and beam management is proposed in Chapter 2. Further, since there are some different algorithms in the ML and different algorithms are suitable for different occasions, a general background of ML algorithms with advantages and disadvantages of different categories are discussed. Finally, a literature review of ML-based NLOS communication, handover management, and beam management sate-of-the-arts is proposed.

In Chapter 3, a deep learning-enabled method to predict the AOA and AOD in the NLOS channel for mmWave communication. The neural network is trained with some channel features, such as received power, E/AAOA, E/AAOD, location and use deep learning to predict the A/EAOA and A/EAOD. Results indicate that the absolute error calculated between the real and the predicted is quite low, validating the proposed solution. Further, Some errors with truncated normal distribution are added in the beam angle to evaluate the robustness of our system. When the error is below a given threshold, our system still has good performance. Finally, the UE trajectory with DWA and generate location input is predicted. Further, input it into the trained DNN to evaluate the performance of trajectory prediction. The error, in this case, is close to the original location information from data generation. Simulation results show that the prediction errors of the AOA and AOD can be maintained within an acceptable range of $\pm 2°$.

In Chapter 4, an optimization theory based on one of the RL methods called O-MAPPO is proposed to optimize the total system delay and throughput. Specifically, the RL algorithm called MAPPO is applied to improve the handover trigger decision. After the handover triggers, the related channel information is sent to the optimal handover decision scheme to optimize the beams and BSs selection, bandwidth allocation, maximize the overall system throughput and delay. Further, to avoid unnecessary HO events and reduce the HO rate, an HO penalty strategy is implemented to improve the system's efficiency. Simulation results demonstrate that with the training processing, our method could achieve good performance in terms of total system throughput and delay compared with some typical RL algorithms, such as MADDPG and DQN. The numerical results show that the overall system throughput and delay are improved by 10% and 25%, respectively, compared with two typical RL algorithms, Deep Deterministic Policy Gradient (DDPG) and Deep Q-learning (DQL).

In Chapter 5, a multi-agent RL-based beam management scheme aims to maximize the system throughput in the dense heterogeneous mmWave network while guaranteeing the basic QoS of individual UE. Specifically, an RL algorithm called MADDPG is used on each SCBS with in-directional beams to optimize the beam selection decisions. Numerical results reveal the convergence performance of the MADDPG and the superiority in improving the system throughput compared with other typical RL algorithms and the traditional beam selection method.

6.2 Open Issues and Future Traces

Although in this thesis it has been investigated that ML is a promising technology to address the problems for NLOS beam tracking, handover, and beam management, some significant challenges still require to be solved. This section briefly highlights some open issues and future research directions associated with the ML techniques in the mmWave cellular network.

6.2.1 Data Set Availability

To improve the performance of ML, sufficient and quality data for model training is essential. However, ML-based beam management or HO optimization usually requires a user mobility history and location data set. Due to the various data protection regulations [219], it is difficult to generate such a data set. Hence, researchers usually use synthetic data from the network simulations for training the ML model. Further, data uniformity is another big issue to solve, which makes the data set generated not be used across different platforms. Therefore, it is necessary to create quality data sets that could be treated as benchmarks to assess the accuracy of different ML models proposed for handover and beam management optimization.

6.2.2 Privacy and Security

Typically, mobile service operators are responsible for the user data protection, which makes it difficult to release the complete and quality data sets from mobile networks without revealing user identity and protecting their privacy [219]. Further, the security of ML models is another challenge to be considered. For example, DL models could be subject to adversarial attack, in which a fake data set is injected into a training data set, thereby reducing the training accuracy and network performance [220]. Hence, it is necessary to

improve the security of DL models from adversarial attacks. In addition, some privacy-preserving ML algorithms, i.e., federated learning, require developing and employing for 5G cellular network and beyond to improve the security and secrecy for UEs.

6.2.3 Offline and Online Learning

Due to the larger dimension of the 5G cellular network and beyond, there are many parameters for ML to learn to improve the performance of HO and beam management. In this case, to reduce time and space complexity, most network designers usually implement offline ML training. However, there is a high demand for real-time responding and decision-making in the communication system, making offline training ML models less efficient and accurate. In this case, the number of parameters that require to be trained should be reduced by employing the clustering method [149]. Further, some hardware acceleration methods also should be considered to facilitate the ML training process [221]. For some ultra-low-latency scenarios, the online training and offline deployment ML architecture is necessary to be considered, where the model goes through a periodic update and refinement during real-time implementation.

6.2.4 Centralized and Distributed Deployment

ML models could be either centralized or distributed deployment based on different network configurations. Both of these deployments have their advantages and drawbacks. The advantages of distributed deployment are low signaling overhead and less computation. However, there is a big challenge of inaccurate network optimization due to localized or lack of global network information [219]. On the other hand, with the global information for the centralized deployment, it is possible to perform coordinated and collaborative learning, contributing to global network optimization. However, this deployment leads to massive signaling as well as computation overhead due to periodic data collection and end-to-end delay. There is a trade-off between global accuracy and huge overhead when implementing these two deployments. For example, it would be more suitable to apply distributed deployment ML algorithms for HO optimization, with which there is a great improvement when protecting users' privacy, reducing latency as well as communication overhead, and minimizing the energy consumption [222].

Bibliography

[1] S. Geng, J. Kivinen, X. Zhao, and P. Vainikainen, "Millimeter-wave propagation channel characterization for short-range wireless communications," IEEE transactions on vehicular technology, vol. 58, no. 1, pp. 3–13, 2008.

[2] R. Chataut and R. Akl, "Massive mimo systems for 5g and beyond networks—overview, recent trends, challenges, and future research direction," Sensors, vol. 20, no. 10, p. 2753, 2020.

[3] Y. Sun, G. Feng, S. Qin, Y.-C. Liang, and T.-S. P. Yum, "The smart handoff policy for millimeter wave heterogeneous cellular networks," IEEE Transactions on Mobile Computing, vol. 17, no. 6, pp. 1456–1468, 2017.

[4] S. Zang, W. Bao, P. L. Yeoh, B. Vucetic, and Y. Li, "Managing vertical handovers in millimeter wave heterogeneous networks," IEEE Transactions on Communications, vol. 67, no. 2, pp. 1629–1644, 2018.

[5] H. Elshaer, M. N. Kulkarni, F. Boccardi, J. G. Andrews, and M. Dohler, "Downlink and uplink cell association with traditional macrocells and millimeter wave small cells," IEEE Transactions on Wireless Communications, vol. 15, no. 9, pp. 6244–6258, 2016.

[6] X. Wang, L. Kong, F. Kong, F. Qiu, M. Xia, S. Arnon, and G. Chen, "Millimeter wave communication: A comprehensive survey," IEEE Communications Surveys & Tutorials, vol. 20, no. 3, pp. 1616–1653, 2018.

[7] P. V. Klaine, M. A. Imran, O. Onireti, and R. D. Souza, "A survey of machine learning techniques applied to self-organizing cellular networks," IEEE Communications Surveys & Tutorials, vol. 19, no. 4, pp. 2392–2431, 2017.

[8] M. A. Albreem, "5g wireless communication systems: Vision and challenges," in 2015 International Conference on Computer, Communications, and Control Technology (I4CT). IEEE, 2015, pp. 493–497.

[9] M. Shafi, A. F. Molisch, P. J. Smith, T. Haustein, P. Zhu, P. De Silva, F. Tufvesson, A. Benjebbour, and G. Wunder, "5g: A tutorial overview of standards, trials, challenges, deployment, and practice," IEEE journal on selected areas in communications, vol. 35, no. 6, pp. 1201–1221, 2017.

[10] A. Gupta and R. K. Jha, "A survey of 5g network: Architecture and emerging technologies," IEEE access, vol. 3, pp. 1206–1232, 2015.

[11] D. Mi, M. Dianati, L. Zhang, S. Muhaidat, and R. Tafazolli, "Massive mimo performance with imperfect channel reciprocity and channel estimation error," IEEE Transactions on Communications, vol. 65, no. 9, pp. 3734–3749, 2017.

[12] L. Zhang, A. Farhang, G. Feng, and O. Onireti, Radio Access Network Slicing and Virtualization for 5G Vertical Industries. John Wiley & Sons, 2020.

[13] X. Liao, J. Shi, Z. Li, L. Zhang, and B. Xia, "A model-driven deep reinforcement learning heuristic algorithm for resource allocation in ultra-dense cellular networks," IEEE Transactions on Vehicular Technology, vol. 69, no. 1, pp. 983–997, 2019.

[14] Y. Liu, L. Zhang, B. Yang, W. Guo, and M. A. Imran, "Programmable wireless channel for multi-user mimo transmission using meta-surface," in 2019 IEEE Global Communications Conference (GLOBECOM). IEEE, 2019, pp. 1–6.

[15] N. Baldo, L. Giupponi, and J. Mangues-Bafalluy, "Big data empowered self organized networks," in European Wireless 2014; 20th European Wireless Conference. VDE, 2014, pp. 1–8.

[16] L.-C. Wang and S. Rangapillai, "A survey on green 5g cellular networks," in 2012 international conference on signal processing and communications (SPCOM). IEEE, 2012, pp. 1–5.

[17] J. Moysen and L. Giupponi, "From 4g to 5g: Self-organized network management meets machine learning," Computer Communications, vol. 129, pp. 248–268, 2018.

[18] M. E. Morocho-Cayamcela, H. Lee, and W. Lim, "Machine learning for 5g/b5g mobile and wireless communications: Potential, limitations, and future directions," IEEE Access, vol. 7, pp. 137 184–137 206, 2019.

[19] J. Kaur, M. A. Khan, M. Iftikhar, M. Imran, and Q. E. U. Haq, "Machine learning techniques for 5g and beyond," IEEE Access, vol. 9, pp. 23 472–23 488, 2021.

[20] X. Lin, "An overview of 5g advanced evolution in 3gpp release 18," arXiv preprint arXiv:2201.01358, 2022.

[21] O. G. Aliu, A. Imran, M. A. Imran, and B. Evans, "A survey of self organisation in future cellular networks," IEEE Communications Surveys & Tutorials, vol. 15, no. 1, pp. 336–361, 2012.

[22] A. Imran, A. Zoha, and A. Abu-Dayya, "Challenges in 5g: how to empower son with big data for enabling 5g," IEEE network, vol. 28, no. 6, pp. 27–33, 2014.

[23] "5g: A technology vision," 2013, huawei Technologies Co., White paper.

[24] "Looking ahead to 5g," 2014, nokia networks, White paper.

[25] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications (mmwave) for 5g: opportunities and challenges," Wireless networks, vol. 21, no. 8, pp. 2657–2676, 2015.

[26] W. Stallings, "Data and computer communications," Prentice Hall, 2005.

[27] B. Bangerter, S. Talwar, R. Arefi, and K. Stewart, "Networks and devices for the 5g era," IEEE Communications Magazine, vol. 52, no. 2, pp. 90–96, 2014.

[28] F. Khan, Z. Pi, and S. Rajagopal, "Millimeter-wave mobile broadband with large scale spatial processing for 5g mobile communication," in 2012 50th annual allerton conference on communication, control, and computing (Allerton). IEEE, 2012, pp. 1517–1523.

[29] P. Adhikari, "Understanding millimeter wave wireless communication," Loea Corporation, pp. 1–6, 2008.

[30] Z. Qingling and J. Li, "Rain attenuation in millimeter wave ranges," in 2006 7th International Symposium on Antennas, Propagation & EM Theory. IEEE, 2006, pp. 1–4.

[31] Y. Azar, G. N. Wong, K. Wang, R. Mayzus, J. K. Schulz, H. Zhao, F. Gutierrez, D. Hwang, and T. S. Rappaport, "28 ghz propagation measurements for outdoor cellular communications using steerable beam antennas in new york city," in 2013 IEEE international conference on communications (ICC). IEEE, 2013, pp. 5143–5147.

[32] T. Manabe, K. Sato, H. Masuzawa, K. Taira, T. Ihara, Y. Kasashima, and K. Yamaki, "Polarization dependence of multipath propagation and high-speed transmission characteristics of indoor millimeter-wave channel at 60 ghz," IEEE Transactions on Vehicular Technology, vol. 44, no. 2, pp. 268–274, 1995.

[33] G. R. MacCartney and T. S. Rappaport, "73 ghz millimeter wave propagation measurements for outdoor urban mobile and backhaul communications in new york city," in 2014 IEEE International Conference on Communications (ICC). IEEE, 2014, pp. 4862–4867.

[34] S. Sun, G. R. MacCartney, M. K. Samimi, S. Nie, and T. S. Rappaport, "Millimeter wave multi-beam antenna combining for 5g cellular link improvement in new york city," in 2014 IEEE International Conference on Communications (ICC). IEEE, 2014, pp. 5468–5473.

[35] C. R. Anderson and T. S. Rappaport, "In-building wideband partition loss measurements at 2.5 ghz and 60 ghz," CoRR, vol. abs/1701.03415, 2017. [Online]. Available: http://arxiv.org/abs/1701.03415

[36] M. Riza Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," arXiv e-prints, pp. arXiv–1312, 2013.

[37] R. Wang, P. V. Klaine, O. Onireti, Y. Sun, M. A. Imran, and L. Zhang, "Deep learning enabled beam tracking for non-line of sight millimeter wave communications," IEEE Open Journal of the Communications Society, vol. 2, pp. 1710–1720, 2021.

[38] S. Singh, F. Ziliotto, U. Madhow, E. Belding, and M. Rodwell, "Blockage and directivity in 60 ghz wireless personal area networks: From cross-layer model to multihop mac design," IEEE Journal on Selected Areas in Communications, vol. 27, no. 8, pp. 1400–1413, 2009.

[39] S. Collonge, G. Zaharia, and G. E. Zein, "Influence of the human activity on wideband characteristics of the 60 ghz indoor radio channel," IEEE Transactions on Wireless Communications, vol. 3, no. 6, pp. 2396–2406, 2004.

[40] S. H. A. Momo and M. M. Mowla, "Effect of human blockage on an outdoor mmwave channel for 5g communication networks," in 2019 22nd International Conference on Computer and Information Technology (ICCIT). IEEE, 2019, pp. 1–6.

[41] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5g wireless networks: A comprehensive survey," IEEE Communications Surveys & Tutorials, vol. 18, no. 3, pp. 1617–1655, 2016.

[42] H. Holma, A. Toskala, and J. Reunanen, LTE small cell optimization: 3GPP evolution to Release 13. John Wiley & Sons, 2016.

[43] V. Chandrasekhar, J. G. Andrews, and A. Gatherer, "Femtocell networks: a survey," IEEE Communications magazine, vol. 46, no. 9, pp. 59–67, 2008.

[44] C.-X. Wang, F. Haider, X. Gao, X.-H. You, Y. Yang, D. Yuan, H. M. Aggoune, H. Haas, S. Fletcher, and E. Hepsaydir, "Cellular architecture and key technologies for 5g wireless communication networks," IEEE communications magazine, vol. 52, no. 2, pp. 122–130, 2014.

[45] 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, 2012.

[46] K. Sakaguchi, G. K. Tran, H. Shimodaira, S. Nanba, T. Sakurai, K. Takinami, I. Siaud, E. C. Strinati, A. Capone, I. Karls et al., "Millimeter-wave evolution for 5g cellular networks," IEICE Transactions on Communications, vol. 98, no. 3, pp. 388–402, 2015.

[47] F. Haider, C.-X. Wang, H. Haas, D. Yuan, H. Wang, X. Gao, X.-H. You, and E. Hepsaydir, "Spectral efficiency analysis of mobile femtocell based cellular systems," in 2011 IEEE 13th International Conference on Communication Technology. IEEE, 2011, pp. 347–351.

[48] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive mimo for next generation wireless systems," IEEE communications magazine, vol. 52, no. 2, pp. 186–195, 2014.

[49] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive mimo: The next wireless revolution?" IEEE Communications Magazine, vol. 52, no. 9, pp. 56–62, 2014.

[50] E. G. Larsson and L. Van der Perre, "Massive mimo for 5g," 2017.

[51] M. Nguyen, "Massive mimo: A survey of benefits and challenges," ICSES Trans. Comput. Hardw. Electr. Eng, vol. 4, pp. 1–4, 2018.

[52] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," IEEE transactions on wireless communications, vol. 9, no. 11, pp. 3590–3600, 2010.

[53] V. Jungnickel, K. Manolakis, W. Zirwas, B. Panzner, V. Braun, M. Lossow, M. Sternad, R. Apelfröjd, and T. Svensson, "The role of small cells, coordinated multipoint, and massive mimo in 5g," IEEE communications magazine, vol. 52, no. 5, pp. 44–51, 2014.

[54] C.-X. Wang, S. Wu, L. Bai, X. You, J. Wang, and I. Chih-Lin, "Recent advances and future challenges for massive mimo channel measurements and models," Science China Information Sciences, vol. 59, no. 2, pp. 1–16, 2016.

[55] B. Yang, Z. Yu, J. Lan, R. Zhang, J. Zhou, and W. Hong, "Digital beamforming-based massive mimo transceiver for 5g millimeter-wave communications," IEEE Transactions on Microwave Theory and Techniques, vol. 66, no. 7, pp. 3403–3418, 2018.

[56] E. Ali, M. Ismail, R. Nordin, and N. F. Abdulah, "Beamforming techniques for massive mimo systems in 5g: overview, classification, and trends for future research," Frontiers of Information Technology & Electronic Engineering, vol. 18, no. 6, pp. 753–772, 2017.

[57] S. Kutty and D. Sen, "Beamforming for millimeter wave communications: An inclusive survey," IEEE communications surveys & tutorials, vol. 18, no. 2, pp. 949–973, 2015.

[58] W. Roh, J.-Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave Beamforming as An Enabling Technology for 5G Cellular Communications: Theoretical Feasibility and Prototype Results," IEEE communications magazine, vol. 52, no. 2, pp. 106–113, 2014.

[59] W.-C. Liao, T.-H. Chang, W.-K. Ma, and C.-Y. Chi, "Qos-based transmit beamforming in the presence of eavesdroppers: An optimized artificial-noise-aided approach," IEEE Transactions on Signal Processing, vol. 59, no. 3, pp. 1202–1216, 2010.

[60] J. Zhang, X. Yu, and K. B. Letaief, "Hybrid beamforming for 5g and beyond millimeter-wave systems: A holistic view," IEEE Open Journal of the Communications Society, vol. 1, pp. 77–91, 2019.

[61] G. Brown, "Exploring the potential of mmwave for 5g mobile access," Qualcomm White Paper, 2016.

[62] A. A. Momtaz, F. Behnia, R. Amiri, and F. Marvasti, "Nlos identification in range-based source localization: Statistical approach," IEEE Sensors Journal, vol. 18, no. 9, pp. 3745–3751, 2018.

[63] K. Joshi, S. Hong, and S. Katti, "PinPoint: Localizing interfering radios," in 10th USENIX Symposium on Networked Systems Design and Implementation (NSDI 13). Lombard, IL: USENIX Association, Apr. 2013, pp. 241–253. [Online]. Available: https://www.usenix.org/conference/nsdi13/technical-sessions/presentation/joshi

[64] A. Huang, L. Tian, T. Jiang, and J. Zhang, "Nlos identification for wideband mmwave systems at 28 ghz," in 2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring). IEEE, 2019, pp. 1–6.

[65] M. S. Mollel, A. I. Abubakar, M. Ozturk, S. F. Kaijage, M. Kisangiri, S. Hussain, M. A. Imran, and Q. H. Abbasi, "A survey of machine learning applications to handover management in 5g and beyond," IEEE Access, vol. 9, pp. 45 770–45 802, 2021.

[66] Y.-N. R. Li, B. Gao, X. Zhang, and K. Huang, "Beam management in millimeter-wave communications for 5g and beyond," IEEE Access, vol. 8, pp. 13 282–13 293, 2020.

[67] K. Ma, Z. Wang, W. Tian, S. Chen, and L. Hanzo, "Deep learning for beam management: Opportunities, state-of-the-arts and challenges," CoRR, vol. abs/2111.11177, 2021. [Online]. Available: https://arxiv.org/abs/2111.11177

[68] M. Naeem, T. Jamal, J. Diaz-Martinez, S. A. Butt, N. Montesano, M. I. Tariq, E. De-la Hoz-Franco, and E. De-La-Hoz-Valdiris, "Trends and future perspective challenges in big data," in Advances in Intelligent Data Analysis and Applications. Springer, 2022, pp. 309–325.

[69] X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, "Data mining with big data," IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 1, pp. 97–107, 2014.

[70] H. Fourati, R. Maaloul, and L. Chaari, "Self-organizing cellular network approaches applied to 5g networks," in 2019 Global Information Infrastructure and Networking Symposium (GIIS). IEEE, 2019, pp. 1–4.

[71] J. Borras, P. Hatrack, and N. B. Mandayam, "Decision theoretic framework for nlos identification," in VTC'98. 48th IEEE Vehicular Technology Conference. Pathway to Global Wireless Revolution (Cat. No. 98CH36151), vol. 2. IEEE, 1998, pp. 1583–1587.

[72] I. Guvenc, C.-C. Chong, and F. Watanabe, "Nlos identification and mitigation for uwb localization systems," in 2007 IEEE Wireless Communications and Networking Conference. IEEE, 2007, pp. 1571–1576.

[73] S. Marano, 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.

[74] R. Casas, A. Marco, J. Guerrero, and J. Falco, "Robust estimator for non-line-of-sight error mitigation in indoor localization," EURASIP Journal on Advances in Signal Processing, vol. 2006, pp. 1–8, 2006.

[75] J. Khodjaev, Y. Park, and A. S. Malik, "Survey of nlos identification and error mitigation problems in uwb-based positioning algorithms for dense environments," annals of telecommunications-annales des télécommunications, vol. 65, no. 5, pp. 301–311, 2010.

[76] Y. Oguma, T. Nishio, K. Yamamoto, and M. Morikura, "Proactive handover based on human blockage prediction using rgb-d cameras for mmwave communications," IEICE Transactions on Communications, vol. 99, no. 8, pp. 1734–1744, 2016.

[77] T. Wang, S. Wang, and Z.-H. Zhou, "Machine learning for 5g and beyond: From model-based to data-driven mobile wireless networks," China Communications, vol. 16, no. 1, pp. 165–175, 2019.

[78] K. Ma, Z. Wang, W. Tian, S. Chen, and L. Hanzo, "Deep learning for beam management: Opportunities, state-of-the-arts and challenges," arXiv preprint arXiv:2111.11177, 2021.

[79] M. Fallgren, B. Timus et al., "Scenarios, requirements and kpis for 5g mobile and wireless system," METIS deliverable D, vol. 1, no. 1, 2013.

[80] T. Farley, "Mobile telephone history," Telektronikk, vol. 101, no. 3/4, p. 22, 2005.

[81] P. Sharma, "Evolution of mobile wireless communication networks-1g to 5g as well as future prospective of next generation communication network," International Journal of Computer Science and Mobile Computing, vol. 2, no. 8, pp. 47–53, 2013.

[82] S. Shukla, V. Khare, S. Garg, and P. Sharma, "Comparative study of 1g, 2g, 3g and 4g," J. Eng. Comput. Appl. Sci, vol. 2, no. 4, pp. 55–63, 2013.

[83] A. N. Kasim, "A survey mobility management in 5g networks," arXiv preprint arXiv:2006.15598, 2020.

[84] A. Karandikar, N. Akhtar, and M. Mehta, Mobility Management in LTE Heterogeneous Networks. Springer, 2017.

[85] P. Tinkhede and P. Ingole, "Survey of handover decision for next generation," in International Conference on Information Communication and Embedded Systems (ICICES2014). IEEE, 2014, pp. 1–5.

[86] M. Tayyab, X. Gelabert, and R. Jäntti, "A survey on handover management: From lte to nr," IEEE Access, vol. 7, pp. 118 907–118 930, 2019.

[87] X. Gelabert, G. Zhou, and P. Legg, "Mobility performance and suitability of macro cell power-off in lte dense small cell hetnets," in 2013 IEEE 18th International

Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD). IEEE, 2013, pp. 99–103.

[88] M. Tayyab, G. P. Koudouridis, and X. Gelabert, "A simulation study on lte handover and the impact of cell size," in International Conference on Broadband Communications, Networks and Systems. Springer, 2018, pp. 398–408.

[89] K. Kanwal, "Increased energy efficiency in lte networks through reduced early handover," 2017.

[90] A. Ulvan, R. Bestak, and M. Ulvan, "The study of handover procedure in lte-based femtocell network," in WMNC2010. IEEE, 2010, pp. 1–6.

[91] M. Joud, M. García-Lozano, and S. Ruiz, "User specific cell clustering to improve mobility robustness in 5g ultra-dense cellular networks," in 2018 14th Annual Conference on Wireless On-demand Network Systems and Services (WONS). IEEE, 2018, pp. 45–50.

[92] J. Tanveer, A. Haider, R. Ali, and A. Kim, "An overview of reinforcement learning algorithms for handover management in 5g ultra-dense small cell networks," Applied Sciences, vol. 12, no. 1, p. 426, 2022.

[93] M. Giordani, M. Polese, A. Roy, D. Castor, and M. Zorzi, "A tutorial on beam management for 3gpp nr at mmwave frequencies," IEEE Communications Surveys & Tutorials, vol. 21, no. 1, pp. 173–196, 2018.

[94] G. T. 38.802, "Study on new radio access technology, physical layer aspects (release 14)," 2017.

[95] Y. Heng, J. G. Andrews, J. Mo, V. Va, A. Ali, B. L. Ng, and J. C. Zhang, "Six key challenges for beam management in 5.5 g and 6g systems," IEEE Communications Magazine, vol. 59, no. 7, pp. 74–79, 2021.

[96] J. Bang, H. Chung, J. Hong, H. Seo, J. Choi, and S. Kim, "Millimeter-wave communications: Recent developments and challenges of hardware and beam management algorithms," IEEE Communications Magazine, vol. 59, no. 8, pp. 86–92, 2021.

[97] C. Jiang, H. Zhang, Y. Ren, Z. Han, K.-C. Chen, and L. Hanzo, "Machine learning paradigms for next-generation wireless networks," IEEE Wireless Communications, vol. 24, no. 2, pp. 98–105, 2016.

[98] P. Cunningham, M. Cord, and S. J. Delany, "Supervised learning," in Machine learning techniques for multimedia. Springer, 2008, pp. 21–49.

[99] R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms," in Proceedings of the 23rd international conference on Machine learning, 2006, pp. 161–168.

[100] S. Kapoor, D. Grace, and T. Clarke, "A base station selection scheme for handover in a mobility-aware ultra-dense small cell urban vehicular environment," in 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC). IEEE, 2017, pp. 1–5.

[101] H. Zhang, "The optimality of naive bayes," AA, vol. 1, no. 2, p. 3, 2004.

[102] C.-K. Wen, S. Jin, K.-K. Wong, J.-C. Chen, and P. Ting, "Channel estimation for massive mimo using gaussian-mixture bayesian learning," IEEE Transactions on Wireless Communications, vol. 14, no. 3, pp. 1356–1368, 2014.

[103] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," Statistics and computing, vol. 14, no. 3, pp. 199–222, 2004.

[104] V.-s. Feng and S. Y. Chang, "Determination of wireless networks parameters through parallel hierarchical support vector machines," IEEE Transactions on Parallel and Distributed Systems, vol. 23, no. 3, pp. 505–512, 2011.

[105] A. Müller, "Kernel approximations for efficient svms (and other feature extraction methods)," Data Science & Machine Learning Resources, 2012.

[106] J. R. Quinlan, "Induction of decision trees," Machine learning, vol. 1, no. 1, pp. 81–106, 1986.

[107] S. Michaelis and C. Wietfeld, "Comparison of user mobility pattern prediction algorithms to increase handover trigger accuracy," in 2006 IEEE 63rd Vehicular Technology Conference, vol. 2. IEEE, 2006, pp. 952–956.

[108] S. Agatonovic-Kustrin and R. Beresford, "Basic concepts of artificial neural network (ann) modeling and its application in pharmaceutical research," Journal of pharmaceutical and biomedical analysis, vol. 22, no. 5, pp. 717–727, 2000.

[109] J. Schmidhuber, "Deep learning in neural networks: An overview," Neural networks, vol. 61, pp. 85–117, 2015.

[110] I. Goodfellow, Y. Bengio, and A. Courville, <u>Deep learning</u>. MIT press, 2016.

[111] S. J. Pan and Q. Yang, "A survey on transfer learning," <u>IEEE Transactions on knowledge and data engineering</u>, vol. 22, no. 10, pp. 1345–1359, 2009.

[112] H. B. Barlow, "Unsupervised learning," <u>Neural computation</u>, vol. 1, no. 3, pp. 295–311, 1989.

[113] G. E. Hinton, T. J. Sejnowski <u>et al.</u>, <u>Unsupervised learning: foundations of neural computation</u>. MIT press, 1999.

[114] W. H. Press, W. H. Press, B. P. Flannery, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, and W. T. Vetterling, <u>Numerical recipes in Pascal: the art of scientific computing</u>. Cambridge university press, 1989, vol. 1.

[115] A. Ultsch, "Kohonen's self organizing feature maps for exploratory data analysis," <u>Proc. INNC90</u>, pp. 305–308, 1990.

[116] R. S. Sutton and A. G. Barto, <u>Reinforcement learning: An introduction</u>. MIT press, 2018.

[117] C. J. Watkins and P. Dayan, "Q-learning," <u>Machine learning</u>, vol. 8, no. 3-4, pp. 279–292, 1992.

[118] S. Thrun and M. L. Littman, "Reinforcement learning: an introduction," <u>AI Magazine</u>, vol. 21, no. 1, pp. 103–103, 2000.

[119] V. R. Konda and J. N. Tsitsiklis, "Actor-critic algorithms," in <u>Advances in neural information processing systems</u>, 2000, pp. 1008–1014.

[120] A. Maltsev, R. Maslennikov, A. Sevastyanov, A. Lomayev, and A. Khoryaev, "Statistical channel model for 60 ghz wlan systems in conference room environment," in <u>Antennas and Propagation (EuCAP), 2010 Proceedings of the Fourth European Conference on</u>. IEEE, 2010, pp. 1–5.

[121] E. Kurniawan, L. Zhiwei, and S. Sun, "Machine learning-based channel classification and its application to ieee 802.11 ad communications," in <u>GLOBECOM 2017-2017 IEEE Global Communications Conference</u>. IEEE, 2017, pp. 1–6.

[122] J.-S. Choi, W.-H. Lee, J.-H. Lee, J.-H. Lee, and S.-C. Kim, "Deep learning based nlos identification with commodity wlan devices," <u>IEEE Transactions on Vehicular Technology</u>, vol. 67, no. 4, pp. 3295–3303, 2018.

[123] S. Navabi, C. Wang, O. Y. Bursalioglu, and H. Papadopoulos, "Predicting wireless channel features using neural networks," arXiv preprint arXiv:1802.00107, 2018.

[124] M. Giordani, M. Mezzavilla, A. Dhananjay, S. Rangan, and M. Zorzi, "Channel dynamics and snr tracking in millimeter wave cellular systems," in European Wireless 2016; 22th European Wireless Conference; Proceedings of. VDE, 2016, pp. 1–8.

[125] S. Rajagopal, S. Abu-Surra, and M. Malmirchegini, "Channel Feasibility for Outdoor Non-Line-of-Sight MmWave Mobile Communication," in 2012 IEEE vehicular technology conference (VTC Fall). IEEE, 2012, pp. 1–6.

[126] L. Bai, C.-X. Wang, J. Huang, Q. Xu, Y. Yang, G. Goussetis, J. Sun, and W. Zhang, "Predicting Wireless MmWave Massive MIMO Channel Characteristics Using Machine Learning Algorithms," Wireless Communications and Mobile Computing, vol. 2018, 2018.

[127] J.-S. Choi, W.-H. Lee, J.-H. Lee, J.-H. Lee, and S.-C. Kim, "Deep Learning Based NLOS Identification with Commodity WLAN Devices," IEEE Transactions on Vehicular Technology, vol. 67, no. 4, pp. 3295–3303, 2017.

[128] M. Alrabeiah and A. Alkhateeb, "Deep Learning for MmWave Beam and Blockage Prediction Ssing Sub-6GHz Channels," arXiv preprint arXiv:1910.02900, 2019.

[129] J. Gante, G. Falcão, and L. Sousa, "Deep Learning Architectures for Accurate Millimeter Wave Positioning in 5G," Neural Processing Letters, pp. 1–28, 2019.

[130] B. Kim, C. M. Kang, J. Kim, S. H. Lee, C. C. Chung, and J. W. Choi, "Probabilistic Vehicle Trajectory Prediction over Occupancy Grid Pap via Recurrent Neural Network," in 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2017, pp. 399–404.

[131] D. Fox, W. Burgard, and S. Thrun, "The Dynamic Window Approach to Collision Avoidance," IEEE Robotics & Automation Magazine, vol. 4, no. 1, pp. 23–33, 1997.

[132] T. Shimizu, V. Va, G. Bansal, and R. W. Heath, "Millimeter wave v2x communications: Use cases and design considerations of beam management," in 2018 Asia-Pacific Microwave Conference (APMC), 2018, pp. 183–185.

[133] Y. Inoue, Y. Kishiyama, S. Suyama, J. Kepler, M. Cudak, and Y. Okumura, "Field experiments on 5g mmw radio access with beam tracking in small cell environments," in 2015 IEEE Globecom Workshops (GC Wkshps), 2015, pp. 1–6.

[134] T. Kadur, H. Chiang, and G. Fettweis, "Experimental validation of robust beam tracking in a nlos indoor environment," in 2018 25th International Conference on Telecommunications (ICT), 2018, pp. 644–648.

[135] Y. Koda, K. Nakashima, K. Yamamoto, T. Nishio, and M. Morikura, "Cooperative sensing in deep rl-based image-to-decision proactive handover for mmwave networks," in 2020 IEEE 17th Annual Consumer Communications & Networking Conference (CCNC). IEEE, 2020, pp. 1–6.

[136] ——, "Handover management for mmwave networks with proactive performance prediction using camera images and deep reinforcement learning," IEEE Transactions on Cognitive Communications and Networking, vol. 6, no. 2, pp. 802–816, 2020.

[137] A. Klautau, N. González-Prelcic, and R. W. Heath, "Lidar data for deep learning-based mmwave beam-selection," IEEE Wireless Communications Letters, vol. 8, no. 3, pp. 909–912, 2019.

[138] M. Dias, A. Klautau, N. González-Prelcic, and R. W. Heath, "Position and lidar-aided mmwave beam selection using deep learning," in 2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). IEEE, 2019, pp. 1–5.

[139] M. Alrabeiah, J. Booth, A. Hredzak, and A. Alkhateeb, "Viwi vision-aided mmwave beam tracking: Dataset, task, and baseline solutions," arXiv preprint arXiv:2002.02445, 2020.

[140] A. Alkhateeb, I. Beltagy, and S. Alex, "Machine learning for reliable mmwave systems: Blockage prediction and proactive handoff," in 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP). IEEE, 2018, pp. 1055–1059.

[141] L. Sun, J. Hou, and T. Shu, "Optimal handover policy for mmwave cellular networks: A multi-armed bandit approach," in 2019 IEEE Global Communications Conference (GLOBECOM). IEEE, 2019, pp. 1–6.

[142] M. S. Mollel, A. I. Abubakar, M. Ozturk, S. Kaijage, M. Kisangiri, A. Zoha, M. A. Imran, and Q. H. Abbasi, "Intelligent handover decision scheme using double deep reinforcement learning," Physical Communication, vol. 42, p. 101133, 2020.

[143] R. Klus, L. Klus, D. Solomitckii, M. Valkama, and J. Talvitie, "Deep learning based localization and ho optimization in 5g nr networks," in 2020 International Conference on Localization and GNSS (ICL-GNSS). IEEE, 2020, pp. 1–6.

[144] S. Khosravi, H. Shokri-Ghadikolaei, and M. Petrova, "Learning-based handover in mobile millimeter-wave networks," IEEE Transactions on Cognitive Communications and Networking, vol. 7, no. 2, pp. 663–674, 2020.

[145] Y. Koda, K. Yamamoto, T. Nishio, and M. Morikura, "Reinforcement learning based predictive handover for pedestrian-aware mmwave networks," in IEEE INFOCOM 2018-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). IEEE, 2018, pp. 692–697.

[146] F. Guidolin, I. Pappalardo, A. Zanella, and M. Zorzi, "Context-aware handover policies in hetnets," IEEE Transactions on Wireless Communications, vol. 15, no. 3, pp. 1895–1906, 2015.

[147] M. Mezzavilla, S. Goyal, S. Panwar, S. Rangan, and M. Zorzi, "An mdp model for optimal handover decisions in mmwave cellular networks," in 2016 European conference on networks and communications (EuCNC). IEEE, 2016, pp. 100–105.

[148] H. Tabrizi, G. Farhadi, and J. Cioffi, "Dynamic handoff decision in heterogeneous wireless systems: Q-learning approach," in 2012 IEEE international conference on communications (ICC). IEEE, 2012, pp. 3217–3222.

[149] Z. Wang, L. Li, Y. Xu, H. Tian, and S. Cui, "Handover control in wireless systems via asynchronous multiuser deep reinforcement learning," IEEE Internet of Things Journal, vol. 5, no. 6, pp. 4296–4307, 2018.

[150] D. Guo, L. Tang, X. Zhang, and Y.-C. Liang, "Joint optimization of handover control and power allocation based on multi-agent deep reinforcement learning," IEEE Transactions on Vehicular Technology, vol. 69, no. 11, pp. 13 124–13 138, 2020.

[151] Y. Ju, H.-M. Wang, T.-X. Zheng, and Q. Yin, "Secure transmissions in millimeter wave systems," IEEE Transactions on Communications, vol. 65, no. 5, pp. 2114–2127, 2017.

[152] Y. Wang, M. Narasimha, and R. W. Heath Jr, "Mmwave beam prediction with situational awareness: A machine learning approach," arXiv preprint arXiv:1805.08912, 2018.

[153] Y. Liu, X. Fang, M. Xiao, and S. Mumtaz, "Decentralized beam pair selection in multi-beam millimeter-wave networks," IEEE Transactions on Communications, vol. 66, no. 6, pp. 2722–2737, 2018.

[154] Z. Guo and J. Li, "Time-variant beam selection for lens-based millimetre-wave massive mimo systems," Electronics Letters, vol. 53, no. 17, pp. 1226–1228, 2017.

[155] W. Shen, L. Dai, G. Gui, Z. Wang, R. W. Heath, and F. Adachi, "Aod-adaptive subspace codebook for channel feedback in fdd massive mimo systems," in Communications (ICC), 2017 IEEE International Conference on. IEEE, 2017, pp. 1–5.

[156] Y. Long, Z. Chen, J. Fang, and C. Tellambura, "Data-driven-based analog beam selection for hybrid beamforming under mm-wave channels," IEEE Journal of Selected Topics in Signal Processing, vol. 12, no. 2, pp. 340–352, 2018.

[157] A. Klautau, P. Batista, N. Prelcic, Y. Wang, and R. Heath, "5g mimo data for machine learning: Application to beam-selection using deep learning," in Proc. Inf. Theory and Appl. Workshop (ITA), 2016, pp. 1–6.

[158] R. Wang, O. Onireti, L. Zhang, M. A. Imran, G. Ren, J. Qiu, and T. Tian, "Reinforcement learning method for beam management in millimeter-wave networks," in 2019 UK/China Emerging Technologies (UCET). IEEE, 2019, pp. 1–4.

[159] D. C. Araújo and A. L. de Almeida, "Beam management solution using q-learning framework," in 2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP). IEEE, 2019, pp. 594–598.

[160] M. Elsayed and M. Erol-Kantarci, "Radio resource and beam management in 5g mmwave using clustering and deep reinforcement learning," in GLOBECOM 2020-2020 IEEE Global Communications Conference. IEEE, 2020, pp. 1–6.

[161] Y. Zhang, Z. Mou, F. Gao, J. Jiang, R. Ding, and Z. Han, "Uav-enabled secure communications by multi-agent deep reinforcement learning," IEEE Transactions on Vehicular Technology, vol. 69, no. 10, pp. 11 599–11 611, 2020.

[162] T. Zhang, K. Zhu, J. Wang, and Z. Han, "Cost-efficient beam management and resource allocation in millimeter wave backhaul hetnets with hybrid energy supply," IEEE Transactions on Wireless Communications, 2021.

[163] M. Xiao, S. Mumtaz, Y. Huang, L. Dai, Y. Li, M. Matthaiou, G. K. Karagiannidis, E. Björnson, K. Yang, C. 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.

[164] 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.

[165] K. Yu and Y. J. Guo, "Statistical NLOS Identification Based on AOA, TOA, and Signal Strength," IEEE Transactions on Vehicular Technology, vol. 58, no. 1, pp. 274–286, 2009.

[166] S. K. Sharma, T. E. Bogale, L. B. Le, S. Chatzinotas, X. Wang, and B. Ottersten, "Dynamic Spectrum Sharing in 5G Wireless Networks With Full-Duplex Technology: Recent Advances and Research Challenges," IEEE Communications Surveys Tutorials, vol. 20, no. 1, pp. 674–707, 2018.

[167] X. Xue, Y. Wang, X. Wang, and T. E. Bogale, "Joint Source and Relay Precoding in Multiantenna Millimeter-wave Systems," IEEE Transactions on Vehicular Technology, vol. 66, no. 6, pp. 4924–4937, 2017.

[168] M. M. Molu, P. Xiao, M. Khalily, K. Cumanan, L. Zhang, and R. Tafazolli, "Low-complexity and robust hybrid beamforming design for multi-antenna communication systems," IEEE Transactions on Wireless Communications, vol. 17, no. 3, pp. 1445–1459, 2018.

[169] X. Wu, C.-X. Wang, J. Sun, J. Huang, R. Feng, Y. Yang, and X. Ge, "60-GHz Millimeter-wave Channel Measurements and Modeling for Indoor Office Environments," IEEE Transactions on Antennas and Propagation, vol. 65, no. 4, pp. 1912–1924, 2017.

[170] C. Liu, M. Li, S. V. Hanly, P. Whiting, and I. B. Collings, "Millimeter-wave Small Cells: Base Station Discovery, Beam Alignment, and System Design Challenges," IEEE Wireless Communications, vol. 25, no. 4, pp. 40–46, 2018.

[171] T. Mantoro, M. A. Ayu, and M. R. Nugroho, "NLOS and LOS of the 28 GHz Bands Millimeter-wave in 5G Cellular Networks," in 2017 International Conference on Computing, Engineering, and Design (ICCED). IEEE, 2017, pp. 1–5.

[172] J. C. Aviles and A. Kouki, "Position-aided Mm-Wave Beam Training under NLOS Conditions," IEEE Access, vol. 4, pp. 8703–8714, 2016.

[173] O. Abari, H. Hassanieh, M. Rodriguez, and D. Katabi, "Millimeter Wave Communications: From Point-to-Point Links to Agile Network Connections," in Proceedings of the 15th ACM Workshop on Hot Topics in Networks, 2016, pp. 169–175.

[174] Q. C. Li, G. Wu, and T. S. Rappaport, "Channel model for millimeter-wave communications based on geometry statistics," in 2014 IEEE Globecom Workshops (GC Wkshps), 2014, pp. 427–432.

[175] P. Ferrand, M. Amara, S. Valentin, and M. Guillaud, "Trends and Challenges in Wireless Channel Modeling for Evolving Radio Access," IEEE Communications Magazine, vol. 54, no. 7, pp. 93–99, 2016.

[176] M. M. Molu, P. Xiao, M. Khalily, K. Cumanan, L. Zhang, and R. Tafazolli, "Low-Complexity and Robust Hybrid Beamforming Design for Multi-Antenna Communication Systems," IEEE Transactions on Wireless Communications, vol. 17, no. 3, pp. 1445–1459, 2017.

[177] REMCOM, "Remcom," https://www.remcom.com/ Accessed April 4, 2021.

[178] P. Međeđović, M. Veletić, and Blagojević, "Wireless insite software verification via analysis and comparison of simulation and measurement results," in 2012 Proceedings of the 35th International Convention MIPRO, 2012, pp. 776–781.

[179] R. Wang. (2021) "Simulation Data". [Online]. Available: https://drive.google.com/file/d/160G8q2VhSHIEakl-kHV9MynFxbKlaNfF/view?usp=sharing

[180] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," The journal of machine learning research, vol. 15, no. 1, pp. 1929–1958, 2014.

[181] S. Bhanja and A. Das, "Impact of data normalization on deep neural network for time series forecasting," arXiv preprint arXiv:1812.05519, 2018.

[182] Y.-Y. Chen, Y.-H. Lin, C.-C. Kung, M.-H. Chung, I. Yen et al., "Design and Implementation of Cloud Analytics-assisted Smart Power Meters Considering Advanced Artificial Intelligence as Edge Analytics in Demand-side Management for Smart Homes," Sensors, vol. 19, no. 9, p. 2047, 2019.

[183] Y. Li and Y. Yuan, "Convergence Analysis of Two-layer Neural Networks with Relu Activation," in Advances in neural information processing systems, 2017, pp. 597–607.

[184] H. Ye, L. Liang, G. Y. Li, J. Kim, L. Lu, and M. Wu, "Machine Learning for Vehicular Networks," arXiv preprint arXiv:1712.07143, 2017.

[185] I. V. Tetko, D. J. Livingstone, and A. I. Luik, "Neural Network Studies. 1. comparison of Overfitting and Overtraining," Journal of chemical information and computer sciences, vol. 35, no. 5, pp. 826–833, 1995.

[186] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," arXiv preprint arXiv:1412.6980, 2014.

[187] X.-H. Yu and G.-A. Chen, "Efficient Backpropagation Learning Using Optimal Learning Rate and Momentum," Neural Networks, vol. 10, no. 3, pp. 517–527, 1997.

[188] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "Enet: A Deep Neural Network Architecture for Real-time Semantic Segmentation," arXiv preprint arXiv:1606.02147, 2016.

[189] D. Fox, W. Burgard, and S. Thrun, "Controlling Synchro-drive Robots with the Dynamic Window Approach to Collision Avoidance," in Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS'96, vol. 3. IEEE, 1996, pp. 1280–1287.

[190] J.-P. Laumond et al., Robot Motion Planning and Control. Springer, 1998, vol. 229.

[191] E. J. Molinos, Á. Llamazares, and M. Ocaña, "Dynamic Window Based Approaches for Avoiding Obstacles in Moving," Robotics and Autonomous Systems, vol. 118, pp. 112–130, 2019.

[192] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," IEEE assp magazine, vol. 5, no. 2, pp. 4–24, 1988.

[193] T. Bai and R. W. Heath, "Coverage and rate analysis for millimeter-wave cellular networks," IEEE Transactions on Wireless Communications, vol. 14, no. 2, pp. 1100–1114, 2014.

[194] B. Van Quang, R. V. Prasad, and I. Niemegeers, "A survey on handoffs—lessons for 60 ghz based wireless systems," IEEE Communications Surveys & Tutorials, vol. 14, no. 1, pp. 64–86, 2010.

[195] A. Talukdar, M. Cudak, and A. Ghosh, "Handoff rates for millimeterwave 5g systems," in 2014 IEEE 79th Vehicular Technology Conference (VTC Spring). IEEE, 2014, pp. 1–5.

[196] A. Barbuzzi, P. H. Perala, G. Boggia, and K. Pentikousis, "3gpp radio resource control in practice," IEEE Wireless Communications, vol. 19, no. 6, pp. 76–83, 2012.

[197] C. Shen and M. van der Schaar, "A learning approach to frequent handover mitigations in 3gpp mobility protocols," in 2017 IEEE Wireless Communications and Networking Conference (WCNC). IEEE, 2017, pp. 1–6.

[198] F. Spitzer, Principles of random walk. Springer Science & Business Media, 2013, vol. 34.

[199] T. Bai, R. Vaze, and R. W. Heath, "Using random shape theory to model blockage in random cellular networks," in 2012 International Conference on Signal Processing and Communications (SPCOM). IEEE, 2012, pp. 1–5.

[200] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski et al., "Human-level control through deep reinforcement learning," nature, vol. 518, no. 7540, pp. 529–533, 2015.

[201] J. Schulman, P. Moritz, S. Levine, M. Jordan, and P. Abbeel, "High-dimensional continuous control using generalized advantage estimation," arXiv preprint arXiv:1506.02438, 2015.

[202] J. Foerster, G. Farquhar, T. Afouras, N. Nardelli, and S. Whiteson, "Counterfactual multi-agent policy gradients," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, no. 1, 2018.

[203] G. Tesauro et al., "Temporal difference learning and td-gammon," Communications of the ACM, vol. 38, no. 3, pp. 58–68, 1995.

[204] E. U. T. R. Access, "Requirements for support of radio resource management (3gpp ts 36.133 version 12.6. 0 release 12)," ETSI TS, vol. 136, no. 133, p. V11, 2015.

[205] P. T. Boggs and J. W. Tolle, "Sequential quadratic programming," Acta numerica, vol. 4, pp. 1–51, 1995.

[206] I. Kezurer, S. Z. Kovalsky, R. Basri, and Y. Lipman, "Tight relaxation of quadratic matching," in Computer Graphics Forum, vol. 34, no. 5.   Wiley Online Library, 2015, pp. 115–128.

[207] A. B. Jambekar and D. I. Steinberg, "An implicit enumeration algorithm for the all integer programming problem," Computers & Mathematics with Applications, vol. 4, no. 1, pp. 15–31, 1978.

[208] Y. Liu, X. Wang, G. Boudreau, A. B. Sediq, and H. Abou-zeid, "Deep learning based hotspot prediction and beam management for adaptive virtual small cell in 5g networks," IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 4, no. 1, pp. 83–94, 2020.

[209] W. Saad, M. Bennis, and M. Chen, "A vision of 6g wireless systems: Applications, trends, technologies, and open research problems," IEEE network, vol. 34, no. 3, pp. 134–142, 2019.

[210] F. Spitzer, Principles of random walk.   Springer Science & Business Media, 2001, vol. 34.

[211] 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.

[212] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," IEEE journal on selected areas in communications, vol. 32, no. 6, pp. 1164–1179, 2014.

[213] 3GPP, "3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Terminal acoustic characteristics for telephony; Requirements," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 26.131, 12 2014, version 12.3.0. [Online]. Available: https://www.3gpp.org/specifications/releases/68-release-12

[214] M. L. Littman, "Markov games as a framework for multi-agent reinforcement learning," in Machine learning proceedings 1994. Elsevier, 1994, pp. 157–163.

[215] R. Lowe, Y. Wu, A. Tamar, J. Harb, P. Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," arXiv preprint arXiv:1706.02275, 2017.

[216] N. V. Sharma and N. S. Yadav, "Machine learning in wireless communication: A survey," Research Anthology on Artificial Intelligence Applications in Security, pp. 1979–1999, 2021.

[217] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," arXiv preprint arXiv:1509.02971, 2015.

[218] C. Jeong, J. Park, and H. Yu, "Random access in millimeter-wave beamforming cellular networks: issues and approaches," IEEE Communications Magazine, vol. 53, no. 1, pp. 180–185, 2015.

[219] H. Fourati, R. Maaloul, and L. Chaari, "A survey of 5g network systems: challenges and machine learning approaches," International Journal of Machine Learning and Cybernetics, vol. 12, no. 2, pp. 385–431, 2021.

[220] J. Suomalainen, A. Juhola, S. Shahabuddin, A. Mämmelä, and I. Ahmad, "Machine learning threatens 5g security," IEEE Access, vol. 8, pp. 190 822–190 842, 2020.

[221] A. Reuther, P. Michaleas, M. Jones, V. Gadepally, S. Samsi, and J. Kepner, "Survey of machine learning accelerators," in 2020 IEEE high performance extreme computing conference (HPEC). IEEE, 2020, pp. 1–12.

[222] S. Niknam, H. S. Dhillon, and J. H. Reed, "Federated learning for wireless communications: Motivation, opportunities, and challenges," IEEE Communications Magazine, vol. 58, no. 6, pp. 46–51, 2020.