Mobility and Energy Management in 5G Ultra-Dense Networks

by

Li Sun

A dissertation proposal submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
December 10, 2022

Keywords: ultra-dense network, 5G communication, mobility management, information compression, energy efficiency

Approved by

Richard Chapman, Chair, Associate Professor of Computer Science and Software Engineering
Xiao Qin, Professor of Computer Science and Software Engineering
David Umphress, Professor of Computer Science and Software Engineering
Shiwen Mao, Professor of Electrical and Computer Engineering

Abstract

Triggered by the development of 5G technologies, the demand for mobile data has grown tremendously in recent years. It leads to an urgent need to upgrade the current macro-cell-based architecture of network infrastructure from which no substantial amount of future system performance gains could be obtained. Ultra-dense network (UDN) is a promising technique to meet the requirement of explosive data traffic in the 5G era, because of its ability to provide better spectrum efficiency. However, a large number of small base stations (SBSs) or access points (APs), associated with massive MIMO and millimeter wave (mmWave) technologies, make the UDN suffer from severe interference, signaling overhead, and power consumption issues. Therefore, effective mobility and energy management strategies that take account of the architecture of UDN are required to take advantage of 5G technologies fully.

In this dissertation, I propose several novel strategies and methods to improve mobility and energy management in 5G UDNs. Specifically, I address the issue of frequent handover in mmWave UDN with the goal of enhancing time-frequency resource efficiency. By considering the spatial and temporal features of handover, I propose two multi-armed-bandit (MAB) based handover strategies to reduce the handover frequency by exploiting the empirical knowledge distribution of the user's geographical location and the line-of-sight (LOS) blockage. Secondly, to address the signaling overhead issue of centralized downlink precoding in cell-free massive MIMO systems, I propose a novel bandwidth-efficient global zero-forcing precoding strategy associated with a model-based CSI compression method leveraging the physical structure of Rician fading channels. Thirdly, to address the power consumption

and scalability issues in cell-free massive MIMO systems, I propose a multi-agent deep reinforcement learning-based AP activation strategy as a scalable solution to improve energy efficiency in UDNs.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

The emerging 5G communication is characterized by higher data throughput, lower latency and higher reliability, which is realized by some key technologies, such as mmWave (millimeter wave) and massive MIMO (multiple-input-multiple-output). However, the traditional macro-cell based network architecture, in which a macro cell provides low-frequency coverage for miles, becomes inadequate to support the 5G communication, since the high-frequency waves used in 5G communication cannot propagate for such long distance. For this reason, ultra-dense network (UDN) is introduced as an ideal solution to integrate these advanced technologies. An UDN, which is usually in a form of small-cell network, is usually implemented by deploying a large number of small cells providing high-frequency coverage for around 100 yards, within which users are served by the surrounding small base stations (SBSs) that are close to them. Although the high-frequency signal brings the benefit of abundant bandwidth, its short-distance serving range and vulnerability to blockage cause severe service outage at cell edge and render frequent handover, especially in a real-world mobile scenario, which significantly wastes spectrum resource and degrades users' experience. The frequent handover has become a bottleneck for UDNs to be widely used.

To further mitigate frequent handover in UDNs, the cell-free massive MIMO network is introduced as a promising variant of UDN. In a cell-free massive MIMO network, a large number of antennas or access points (APs) are distributed within the network and jointly provide service to UEs. Because of the user-centric nature, the cell-free massive MIMO system significantly reduces the inter-cell interference and is able to provide uniform quality of service to the UEs. However, severe inter-user interference caused by the huge number of APs is the main issue of it. Although the global zero-forcing (GZF) precoding can eliminate

the inter-user interference theoretically, it will render unaffordable signaling overhead due to heavy channel state information exchange, which wastes bandwidth resources and delays communication. How to reduce the overhead in the GZF is a significant but challenging problem.

Another critical issue of UDNs is the high power consumption. To provide wireless service, the AP consumes power to generate radio wave. When the number of APs becomes extremely large, the total power consumption in the network will be considerable. Considering that the number of APs is much larger than that of UEs, wisely turn off idle or under-loaded APs is an effective strategy to reduce the power consumption. However, the straightforward centralized decision-making mode is not suitable for UDNs because of the huge number of APs which would cause explosively growing solution space and severe communication overhead. Therefore, how to make a shrewd distributed AP activation strategy by turning off idle APs is a problem that is worth of being investigated.

To address these issues mentioned above, I investigate the following three problems in this dissertation. Firstly, I propose two novel handover strategies to reduce the handover in mmWave ultra-dense cellular networks. These two strategies exploit the spatial and temporal features of handover and generate BS-selection decisions based on a multi-armed bandit (MAB) model by learning the distribution of UE's moving trajectory and line-of-sight (LOS) link blockage from historical handover events. To support the proposed handover strategies, I develop a signal space partitioning scheme to extract the handover features and design an acceleration technique to improve the proposed SBS-selection algorithms. Secondly, I propose a novel bandwidth-efficient global zero-forcing precoding strategy for cell-free massive MIMO systems, associated with a model-based channel state information (CSI) overhead reduction mechanism. By exploiting the physical structure of Rician fading channels, I propose to decompose a channel matrix into a LOS and a NLOS components, and design tailored methods to compress these two parts, respectively. Thirdly, I investigate the distributed AP activation problem in user-centric cell-free massive MIMO systems. To solve the scalability

issue due to the numerous APs, I propose a multi-agent-deep-reinforcement-learning based approach to solve this problem effectively.

The main contributions of this research can be summarized as follows:

- I propose a novel handover strategy to reduce unnecessary handovers in mmWave UDN by carefully deciding the next SBS a user should be switched to so that the new user-BS connection after the handover can last as long as possible. Without prior knowledge of the user's mobility and environment, the proposed handover strategy extracts the spatial feature of handovers by exploiting the available received signal strength information to explore the empirical distribution of the user's post-handover trajectory and LOS blockage, which is learned online through a multi-armed bandit (MAB) framework. This work is based on my published paper [64].

- Based on the above MAB-based handover mechanism framework, I future explore the temporal feature of handovers based on the proposed spatial feature. The newly added dimension of the handover feature provides an effective way to further distinguish different handover events. By incorporating the spatial-temporal feature of handovers, I propose an advanced handover strategy and a Lin-UCB-based BS selection algorithm to further improve the handover decisions. This work is based on my published paper [62].

- By exploiting the physical structure of Rician fading channels, I propose a novel model-based CSI compression mechanism for centralized downlink precoding in cell-free massive MIMO systems, which decomposes a channel matrix into a line-of-sight (LOS) and a non-line-of-sight (NLOS) components, and then compresses them using a model-based method and a singular-value-decomposition (SVD)-based method, respectively. I also present two optimization-based algorithms to obtain the phase information of the LOS component of the channel, which is then used by the proposed channel matrix decomposition. This work is based on my published paper [63].

- I propose a distributed solution to solve the access point activation (APA) problem in cell-free massive MIMO networks to reduce power consumption. I leverage the user-centric characteristic and design a multi-agent deep reinforcement learning (MADRL) algorithm by which each AP independently decides whether it needs to be switched on or off. For comparison, I also design a centralized approach by which a centralized controller decides to switch on/off for all APs.

The rest of this dissertation is organized as follows. In Chapter 2, I present the state of the art of the existing research. The following two chapters: Chapter 3 and Chapter 4, are the background of my dissertation. In Chapter 3, I introduce part I of the background of this dissertation, which refers to multi-armed bandit based optimal handover policies for mmWave cellular networks. In Chapter 4, I introduce part II of the background of this dissertation, which refers to the bandwidth-efficient precoding in cell-free massive MIMO networks with Rician fading channels. Moreover, in Chapter 5, I propose a multi-agent deep reinforcement learning framework to solve the access point activation problem in cell-free massive MIMO networks. In Chapter 6, I conclude my research and point out future work.

Chapter 2

Related Work

This research work is related to four fields: (1) handover strategies for mmWave cellular networks, (2) machine-learning-based methods for handover in wireless communication, (3) overhead compression methods for wireless communication, and (4) BS/AP activation strategies in UDNs.

2.1 Handover Strategies for mmWave Networks

Although the research on handover management in mmWave band is still in the initial stage, there have been many fundamental works that pave the path to my research. Some methods are designed for single connectivity implementation, in which a UE is served by a single base station through beamforming from the base station, and needs a handover when the current connection is lost [84, 4, 76, 23]. In particular, [84] explored the physical characteristic of 60GHz outdoor mmWave picocells. In this paper, extensive measurements and system-level simulations on the propagation range, attenuation, and sensitivity to blockage, demonstrated the feasibility and benefits of 60GHz outdoor mmWave picocells. [4] formulated a user association problem in the mmWave network as a mixed integer linear programming model. To solve this NP-hard problem, the authors proposed a novel distributed algorithm based on Lagrangian duality theory and subgradient methods. [23] took the reallocation cost of potential handovers and the channel variability of the mmWave channel into account, and investigated an optimal and fair cell selection policy. [76] considered a joint association and relaying problem with load balancing in mmWave networks and formulated this problem as a stochastic optimization problem. To efficiently solve this problem, the authors

proposed a distributed auction algorithm where the UEs and relays are treated as independent agents and act asynchronously to achieve optimal UE-relay association solutions. The single connectivity implementation suffers from long handover delay since the initial access of the new beam requires expensive signaling and long training time [7]. Some works improve the handover delay by employing statistical predictive models such as finite-state Markov chain [24, 18] and Markov decision process (MDP) [43, 79] to predict the possibility of an outage in the next time slot based on the current channel state. Moreover, [50] introduced a linear-regression-based direction of pass detection algorithm to reduce handover delay. In addition, the content caching technique has been utilized to lower the handover failure rate and smooth handover [53, 58, 59]. Specifically, to solve the long handover delays and connection latency, [53] developed a caching-based mmWave network to precache video contents for serving users in handover and reducing connection and retrieval delays. The authors modeled it as a dynamic programming problem and proposed a cell-by-cell decomposition method to solve it. [58] analyzed the impact of caching on the number of handovers and the average handover failure. Based on the analysis, [59] investigated a cache-enabled mobility management problem that is formulated as a dynamic matching game between UEs and BSs. A distributed algorithm was also presented to obtain a dynamically stable handover mechanism.

Another strategy of handover management is multi-connectivity, which maintains beamforming from multiple base stations to a user, so that the user is still under cover if its LOS to one base station is lost [54, 22, 51, 82]. In particular, to address the issue of vulnerable mmWave signals which are susceptible to blockage, [51] designed a dual connectivity protocol for a mmWave cellular network to allow mobile UE to maintain two connections to 4G and 5G BSs at the same time. The authors also developed an uplink control signaling system to enable rapid path switching when any one link is lost and measurement-based channel models to capture the physical characteristics of mmWave signals. It is demonstrated that the proposed method outperforms conventional handover mechanisms. [22] pointed out

that multi-connectivity is a robust way to enhance mmWave communication but it has a challenging issue of directional tracking due to the highly directional beams and fast varying channels. To overcome this issue, the authors proposed a measurement system based on measuring sounding signals and received signal strength to make centralized handover and scheduling decisions. [82] proposed a received signal strength (RSS) prediction-based multi-connectivity handover scheme to address the frequent unnecessary handover problem in UDNs. This scheme is associated with a new handover triggering mechanism and makes BS selection decisions based on RSS prediction. Since multi-connectivity has low efficiency in beam utilization and suffers from multi-fold user capacity loss, it is not considered in this research.

2.2 Machine Learning Based Methods for Handover

Machine learning provides another promising tool to improve handover decisions. In particular, the authors in [66] introduced a reinforcement-learning (RL) based handover policy to reduce the number of handovers in HetNet. In [33], the authors utilized RL to predict user's mobility and applied proactive handover to improve the throughput. However, their method needs the user's velocity and location information obtained via a dedicated tracking device. In [36], the authors considered a communication system consisting of users and unmanned aerial vehicles (UAVs), and proposed a user association algorithm based on RL to reduce redundant handovers.

Moreover, the authors in [45] introduced a partially blind handover scheme that uses an embedded XGBoost classifier to predict the success rate of handover. In [3], the authors employed deep learning (DL) framework to predict upcoming failure events and implement proactive handover based on historical beamforming vectors. However, they did not describe how it works in multi-user scenarios. In [48], the authors built a convolutional neural network to predict the signal power that will be received in a short time. But their solution relies on a costly camera device, which is not scalable in practice.

As an integration of DL and RL, deep reinforcement learning (DRL) is also utilized to reduce handover frequency in wireless communication. In [73], the authors proposed an asynchronous multi-user DRL scheme with a deep neural network (DNN) as a handover controller to reduce handover frequency. Similarly, the authors in [27] proposed a handover scheme based on a deep Q-network. Different from [73], [27] utilized the historical received uplink SINR on APs to characterize the UE's state and leveraged the convolutional neural network and the recurrent neural network to extract UE's features.

2.3 Overhead compression methods for wireless communication

Interference management is rather challenging in the cell-free massive MIMO system due to the numerous APs. As an effort to eliminate the inter-user interference in downlink transmission, LZF precoding and its variants [9, 28] are widely applied, in which each AP conducts ZF precoding independently for its served UEs. However, these precoding techniques can only partially mitigate the interference since the precoding is conducted locally at each AP without any collaboration with other APs and consequently the precoding decision is "blindly" in some way. In addition, it has another assumption that the number of antennas at each AP is larger than the number of mutually orthogonal pilots, which limits the application of LZF in the systems where the APs are equipped with few antennas. Although the GZF is able to further mitigate the interference theoretically, it suffers from huge overhead caused by CSI exchange. Due to the existence of a huge number of antennas, the CSI feedback is too large to be transmitted without compression. How to efficiently reduce the CSI feedback is of great importance to improving the practicality and performance of wireless networks [55, 80].

As an early CSI compression method, the quantization-based compression has been widely utilized in C-RANs and MIMO networks [32, 31, 70]. Because of its low complexity, it is suitable for remote radio heads with limited computation ability. But it is too simple to be adequate for more complex application scenarios. In recent years, compressing sensing

(CS) has been applied to reduce the CSI feedback in FDD massive MIMO systems [34, 15, 60]. It exploits the spatial correlation of CSI and represents the sparse signal by a few elements through random projections [60]. In particular, [15] proposed a channel feedback reduction technique based on the theory of CS. This technique not only substantially reduces feedback overhead but also guarantees CSI recovery for BS. Kuo et.al. [34] considered CS-based channel feedback reduction techniques in a massive MIMO system associated with two adaptive feedback protocols to efficiently configure the feedback content. Considering the spatial correlation and channel conditions, Sim et.al. [60] proposed a new reconstruction algorithm for CS. A new codebook for the compressed channel quantization without any assumption of no other-cell interference was also developed. The two major issues of the CS-related methods are their complex recovery computation at BS and the limitations and requirements of the sparsifying basis. In order to address these issues, the PCA (principle component analysis)-based schemes are proposed [21, 81, 29]. Specifically, Ge et al. [21] proposed a PCA-based feedback reduction scheme by exploiting the spatial correlation characteristics of a massive MIMO channel model, by which a UE is able to compress high-dimensional CSI into a low-dimensional one. Zhang et al. [81] proposed to generate a compression matrix over a long-term period while feeding back the compressed CSI to the BS in a short-term one. The authors also investigated the information distortion of the proposed scheme and derived a closed-form expression for the distortion. Also considering spatially correlated fading channels, Joung et al. [29] further analyzed the compression feedback error, bit error rate, and the spectral efficiency of the PCA-based compression used in massive MIMO systems. The PCA-based compression matrix, being signal-dependent and a statistical basis, is a trade-off between the DCT basis and the KLT basis [21, 81]. It only requires the UE and the BS to have the same spatial correlation matrix in a long-term period. Another advantage of PCA-based compression schemes is the simplified decoding computation at BS, which only needs to conduct matrix operation on the received low-dimension CSI, instead of solving the underdetermined linear system as in CS.

Machine learning (ML) is another powerful tool to solve the CSI feedback-reduction problem [25]. The advantage of ML over CS lies in that, it requires no sparsifying basis or random projection and has a simpler CSI recovery procedure [74]. This motivates a growing research effort in this field [74, 37, 72, 39, 78, 38]. Specifically, convolutional neural networks (CNNs) have been introduced as the encoder and the decoder, which conducts random projection and inverse transformation from codewords to original channels, respectively [74, 37, 38, 78]. Moreover, the recurrent neural network (RNN) is utilized to extract interframe correlation [72], and redesign the feature compression and decompression modules [39]. However, the ML-based overhead-reduction method requires massive training data sets and nontrivial hyperparameter tuning.

2.4 AP activation strategy in ultra-dense networks

How to improve the energy efficiency in UDNs by turning off redundant BSs/APs is a hot topic in the past decade. Many researchers have made remarkable contributions to this problem and its variants. More details can be found in [75, 26, 19, 57]. Recently, some papers have focused on the AP activation (APA) strategy in cell-free massive MIMO networks. In particular, the performance of the APA technique in cell-free massive MIMO networks powered by adaptively switched on/off APs is analyzed in [30]. Moreover, various APA strategies are proposed in [16, 20, 44, 42, 71], which are designed for different scenarios. In particular, Garcia-Morales et al. [20] focused on a cell-free massive MIMO network with mmWave band in which the inhomogeneous nature of spatial traffic distribution is considered and investigated energy efficient AP sleep-mode techniques. The authors analyzed and compared different AP switch strategies based on the use of goodness-of-fit (GoF) tests. Femenias et al. [16] not only analyzed and compared different AP switch strategies in cell-free massive MIMO systems but also considered line-of-sight (LOS) and non-line-of-sight (NLOS) links between APs and UEs, the different antenna array architectures, and specific power consumption models. Mishra and Vijayakumar considered a special cell-free massive

MIMO with the architecture of radio stripe and proposed a random AP switch strategy which is suitable for this kind of cell-free massive MIMO to improve the energy efficiency [44]. Wang et al. proposed a network decomposition method to divide a cell-free network into subnetworks in order to reduce signaling overhead and system complexity, which allows the APs not serving any UEs to be in sleep mode [71]. A rate-constrained network decomposition algorithm is developed to solve this problem. Moreover, to solve the AP activation problem in a cell-free MIMO network with the purpose to improve energy efficiency, Mendoza et al. proposed a deep-reinforcement-learning-based method to find the optimal set of APs which need to be turned on [42]. Xu et al. also took BS activation decision into their resource allocation strategy to minimize power consumption in Cloud Radio Access Networks (CRANs). They presented a centralized deep-reinforcement-learning-based framework to find a power-efficient resource allocation solution [77].

2.5 Discussion

Upon the plentiful research results in the related fields, the limitations of the existing works and the contributions of this dissertation are summarized as follows:

- Few of the existing handover strategies explicitly considers the impact of distributions of user's mobility and LOS blockage on handover frequency in mmWave cellular networks. Most research in the literature uses throughput [54, 24, 51, 33, 66], or delay [43, 18, 79, 50], or failure rate [53, 58, 59, 45, 3] as the evaluation criteria for handover policy. Rather than these metrics, I focus on the unobstructed time for a LOS link which more directly reflects the quality of a handover decision due to the directivity of mmWave communication. The estimation of the unobstructed LOS time requires certain knowledge of the user's post-handover trajectory and LOS blockage, whose acquisition has not been studied in the literature.

- Most of the existing handover solutions have specific requirements, for example, the prior knowledge on the distribution of channel state [43] and user's mobility [24, 79, 45, 73], or auxiliary devices to obtain user's movement information [14, 33, 48]. However, such requirements cannot always be satisfied in practice. Hence, I propose two novel handover strategies which leverage the available RSS information to extract the user's spatial and temporal features to guide the handover decision without any pre-knowledge of the user's exact mobility information. I also propose an online learning framework based on the MAB process with low computational complexity. This online learning has a simple structure that requires no offline training phase or hyperparameter tuning, hence is easy to implement.

- In most of the related work on feedback reduction, the wireless channels are modeled to be Rayleigh fading channels which do not fit the characteristics of UDNs. Most of the current CSI feedback compression methods only consider the statistical characteristics of the channel matrix. In contrast, the proposed GZF precoding strategy is customized for Rician fading channels in UDNs and fully exploits the physical structure of Rician fading channels to efficiently reduce the information loss in CSI compression and therefore is more reliable.

- Unlike many existing methods of CSI compression that rely on sparsefying-basis to make the channel vector sparse and need time-consuming iterative algorithms to recover CSI, the proposed CSI compression strategy is built on simple SVD and only involves basic matrix operation. Moreover, this method does not need off-line training or hyperparameter tuning which is necessary for ML-based methods and hence is easy to implement.

- The current research on how to turn off BSs/APs in UDNs mainly focuses on small-cell networks, in which each UE is served by at most one BS/AP, and the existing AP activation strategies are not tailored for cell-free massive MIMO systems. Moreover,

the widely-used centralized decision is not suitable for UDNs when the number of APs is huge. To address these issues, I propose a multi-agent deep reinforcement learning approach to solve the AP activation problem in cell-free massive MIMO systems.

In summary, I propose two novel online-learning-based contextual handover mechanisms, which can learn the empirical distribution of the user's post-handover trajectory and LOS blockage, and use the learning outcome to reduce unnecessary handovers in ultra-dense mmWave cellular networks. Depending on the availability of information, two different MAB formulations are proposed for the learning, one focused on features of handover events in the space domain, and the other on features in both the space and the time domains. Two effective BS-selection algorithms are developed for these two mechanisms, respectively. Moreover, in order to address the issue caused by the high-dimension feature of handover in a complex scenario, a novel acceleration technique is presented to increase the efficiency of the algorithms. Note that, none of these proposed mechanisms requires the knowledge of the user's exact (or fine-grained) mobility information. In addition, to address the issue of severe overhead generated in cell-free massive MIMO systems, I propose an efficient CSI compression method that exploits the physical structure of Rician fading channel. This method does not rely on any statistical characteristics of the channel matrix, or sparsefying-basis, and is easy to implement. To handle the scalability issue in solving the AP activation problem in UDNs, I propose a multi-agent deep reinforcement learning approach that is tailored for cell-free massive MIMO systems.

Chapter 3

Background: Part I

3.1 Motivation

As a supporting technology of 5G cellular communication, millimeter wave (mmWave) brings a remarkable benefit in bandwidth due to its ultra-high frequency (30 $\sim$ 300 GHz). This characteristic makes the mmWave communication highly relies on the line-of-sight (LOS) path between the base station (BS) and the user equipment (UE). However, when the UE is moving, the LOS path is easily blocked by obstacles, such as buildings, advertising boards, and tree-tops. Unlike the microwave, the millimeter wave lacks the ability to penetrate through or circumvent the obstacle because of its too-short wavelength. That results in a sudden drop of signal strength when the LOS path is blocked and a subsequent handover, a process switching the UE to another BS to maintain the communication. That is why the handover frequency in mmWave cellular networks is much higher than that in 4G networks [69]. Frequent handover not only wastes spectrum resources but also impairs users' experience. How to efficiently mitigate the frequent handover in mmWave cellular networks is an important but challenging problem.

Although the handovers are frequent in mmWave cellular networks, 61% of them are unnecessary. In other words, a large part of handovers could be avoided if better handover decisions for the UEs could be taken. Therefore, a wise handover strategy is essential to reduce unnecessary handovers for improving the spectrum efficiency and the user experience. A straightforward strategy is to choose the BS that provides the strongest signal strength. However, it may lead to a "short-sighted" handover decision. In particular, the newly built connection between the UE and the chosen BS may be blocked again in a few seconds because of the UE's trajectory and the location of obstacles. A wise strategy should consider not

only the instantaneous signal strength but also be aware of which BSs are able to provide connections that can last for a long time.

In recent years, more and more researchers pay attention to reducing unnecessary handover in mmWave cellular networks by applying Markov theory [43, 13], machine learning [65, 36, 73, 27], and other methods [22, 82]. Although these works have made remarkable contributions to solving this problem, they still have limitations and could be improved. Some handover strategies have special requirements, such as an antenna array equipped at UE and exhaustive direction search [22]. Some strategies have the assumption of pre-knowledge about the distribution of channel state [43] or the mobility of UE [36]. Some others need time-consuming offline training [27]. These requirements and assumptions prohibit the wide usage of these handover strategies in UDNs. Therefore, an advanced handover strategy requires less assumption and pre-knowledge is more suitable in practice. I notice that the contextual relationship between LOS link, user's movement, and obstacle location could be leveraged to direct good handover decisions. In particular, the unobstructed time for a LOS link is determined by the user's moving trajectory and the distribution of the obstacles after the handover. Therefore, the awareness of this information is of benefit to making a wise handover decision. Unfortunately, this knowledge usually is not available when a handover event happens. A straightforward way to solve this problem is to predict the user's post-handover trajectory, but that requires the physical location information of the user, which is not always available in reality.

In this chapter, I propose two handover strategies to address this issue by exploiting the empirical distribution of the user's post-handover trajectory and LOS link blockage. These strategies are based on an instinct observation: if a strategy applied on a handover is optimal, it is quite likely to be also optimal to other handovers of the same feature. The feature of a handover reflects the user's state and the environment where the handover event happens, which is related to the user's post-handover trajectory and the distribution of blockage. A feature is treated as a label according to which a specific policy is applied on

the handovers of that feature. To differentiate the handovers of different features, I propose a novel partitioning scheme to extract the handover feature in spatial and temporal domains by leveraging the available signal strength information. Two handover strategies are then designed based on the extracted feature to direct handover decisions based on the empirical knowledge accumulated from past handover events. These strategies are easy to implement and require no exact location information of users.

In these strategies, the empirical knowledge is learned online through a multi-armed bandit (MAB) framework, in which the goal is to maximize the expected unobstructed time for the post-handover connection between BS and UE. Specifically, the centralized controller maintains an individual MAB process for each spatial feature of handover represented by a block that corresponds to a specific physical area, which will be explained in Section 3.4.1. When a handover event happens, the controller extracts its spatial (and temporal) feature and retrieves the MAB model corresponding to the feature. The MAB, which maintains the empirical knowledge then chooses a BS among all candidate BSs as the handover decision. Being switched to this BS, the UE will receive an instantaneous reward, which will be interpreted in Section IV. This reward will then be used by the MAB model to update its accumulated knowledge to direct future handover decisions of the same feature. Compared with the existing handover strategies including UEs' trajectory prediction, one advantage of the proposed strategies is that they do not require the exact location information of UE. Instead, they are able to infer the location and mobility information of UE by exploiting the available received signal strength (RSS). It is demonstrated through simulation that, although the inferred information is coarse-grained, it is beneficial to make better handover decisions than some existing counterparts.

This work is mostly related to the SMART scheme [66], which also uses the MAB framework to direct handover decisions in mmWave cellular networks. The difference between my work and the SMART scheme is that my work differentiates handover events by their spatial and temporal features, and applies specific MAB models to deal with different

handovers according to their features, while the SMART does not consider any individual characteristics of handover and uses a single MAB model to make all handover decisions. Moreover, although some works consider a similar context in handover management, they require auxiliary devices to collect user's speed information [24], or aim to avoid exhaustively beam-searching to reduce handover delay [50], which is different from my work.

3.2 System Model

Let's consider a cellular network $\mathcal{G}$ consisting of a set of mmWave small-cell base stations (SBSs), denoted as $\mathcal{L}$, and a set of single-antenna user equipments (UEs), denoted as $\mathcal{K}$. These SBSs and UEs are randomly distributed in the network. A centralized controller (CC) takes charge of handover in this network through the control plane. This UDN can be illustrated in Fig. 3.1.

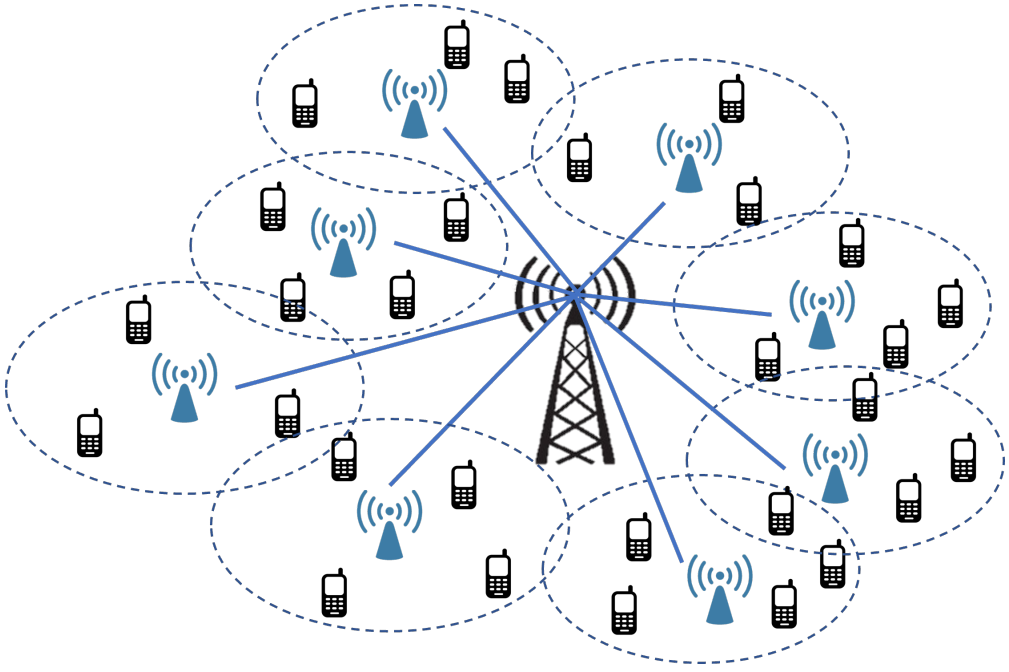


Figure 3.1: Ultra-dense network

3.2.1 Propagation Model

In this chapter, I describe the channel of the mmWave band by the 3GPP standard probabilistic LOS model. The statistic path loss model is formulated as [66, 2]

$$PL(d)[dB] = \alpha + 10\beta \log_{10}(d) + \xi, \xi \sim \mathbb{N}(0, \sigma^2), \tag{3.1}$$

where d is the distance (m) between the SBS and the UE, α and β are the least square fittings of floating intercept and slope respectively over the measured distances, and ξ represents a lognormal shadowing with variance σ^2. It is assumed that a perfect beamforming technique is applied and therefore the inter-user interference could be ignored. The signal-to-noise ratio (SNR) of the signal received by the UE $k \in \mathcal{K}$ from the SBS $l \in \mathcal{L}$ is modeled as [66]

$$SNR_k^l = \frac{P_l \times G \times PL(d)^{-1}}{P_n}, \tag{3.2}$$

where P_l is the transmit power of SBS l, P_n is the noise power and G is the antenna gain. The antenna gain in mmWave communication highly depends on the direction of beams formed by the SBS. Since I assume that SBS is equipped with directional antennas with a sectorized gain pattern, the antenna gain G is actually a function of the angle of departure ω from the SBS to the UE. According to [61], this function can be represented by

$$G(\omega) = \begin{cases} G_{\max}, & \text{if } |\omega| \leq \omega_{\mathrm{s}} \\ G_{\min}, & \text{otherwise,} \end{cases} \tag{3.3}$$

where $G_{\max}$ is the main lobe gain, $G_{\min}$ is the side lobe gain, and ω_{s} is the main lobe width of the SBS. Since I assume that the perfect beam tracking technique can be used to maintain mmWave link [66], the UE could always be in the main lobe and have main lobe gain as long as its LOS path to the SBS is not blocked.

I assume that an SBS is able to serve at most $U_{\max}$ UEs simultaneously, and all served UEs equally share its bandwidth. The downlink transmission rate from the SBS l to the UE k can be calculated as follows:

$$h_k^l = \frac{B_{\mathrm{w}}}{U_l} \log_2(1 + SNR_n^l), \tag{3.4}$$

where B_{w} is the bandwidth of SBS and U_l is the number of UEs simultaneously served by the SBS l.

3.2.2 Blockage and Mobility Model

Due to the vulnerability of mmWave signal to obstacles, the transmission rate of a LOS link would quickly drop to zero immediately when the link is blocked. In the simulation of this chapter, a user is characterized by a random moving speed and direction, while an obstacle is modeled as an object with a fixed radius and location. A link is blocked whenever there is an obstacle to which the distance from the link is less than its radius. Given a LOS link and a set of randomly distributed obstacles, the link is blocked whenever there is an obstacle to which the distance from the link is less than its radius. This modeling could be used to represent any fixed obstacle, such as a tree top, an advertising board or a building. Note that there is no assumption on the blockage and mobility models of the proposed mechanisms and their analysis. In other words, they are general enough to work under any blockage and mobility models that may appear in practical applications. To evaluate the performance of the proposed handover mechanisms, in the simulations I assume a randomly distributed circular obstacle model (i.e., an obstacle is modeled as a circle whose location is randomly distributed) and a random waypoint mobility model. The main reason why I select this model is that it is easy to simulate but still general enough. Even though I assume a unified radius for all obstacles in the simulations, this assumption does not undermine the generality of the obstacle model in the sense that the blockage time caused by an arbitrary

obstacle to an arbitrary user in the model is still a random variable. This is because this blockage time depends on not only the size of the obstacle but also the distance between the obstacle and the user, the user's moving direction, and speed, which are randomly distributed in the model. Therefore, this model has already been able to capture the fundamental effects of heterogeneous blockage time that could have been caused by a more complicated obstacle model. Moreover, in this chapter, I mainly consider the static obstacle and mobile-user scenario. The more challenging mobile-obstacle scenario is out of the scope of this chapter and will be considered in future work.

Moreover, no communication through non-LOS (NLOS) is considered in this chapter. Although it is feasible to use NLOS for communication when there is no LOS, according to the commercial tests from Qualcomm [12], Samsung [56], and NI [6], simply switching the beam to an NLOS component of the mmWave channel when the LOS is blocked [40, 67, 83] may not always be a good solution to this handover problem. Switching beams to NLOS may either lead to a transmission rate that is orders of magnitude lower than the LOS rate if the transmission power is not increased or cause a huge spike in power consumption if one wishes to retain a comparable transmission rate because of the huge difference in the path loss between LOS and NLOS. It is obvious that, rather than simply switching to the NLOS, a more sophisticated handover to another base station or a relay that has a new LOS with the user, may avoid the above weaknesses and thus generate a more desirable solution. I do not consider the effect of reflected signals by obstacles in this work.

3.2.3 Handover Trigger Condition

To guarantee the quality of service, the handover trigger condition for a UE k associated with SBS l is described as

$$SNR_k^l < SNR_{\min} - h_{\text{ys}}, \tag{3.5}$$

where $SNR_{\min}$ is the minimum SNR required for a certain service level, and h_{ys} is a hysteresis parameter for avoiding frequent handover. Although how to select a proper value for h_{ys} is

an interesting issue, it is not the key point of this research. For simplicity, I set h_{ys} to be zero. Note that any specific value of h_{ys} does not influence the proposed handover strategies.

3.3 Online Learning of Contextual Handover Mechanisms

Among the six handover events defined by the 3GPP technical specification, I focus on the BS selection for Event A2 (i.e., a handover will be triggered whenever the received signal strength goes below a pre-defined threshold [1]), since handover triggered by A2 is common but challenging in mmWave band. The general framework of the proposed MAB-based online learning of contextual handover mechanisms is illustrated in Fig. 3.2 and elaborated in the following.

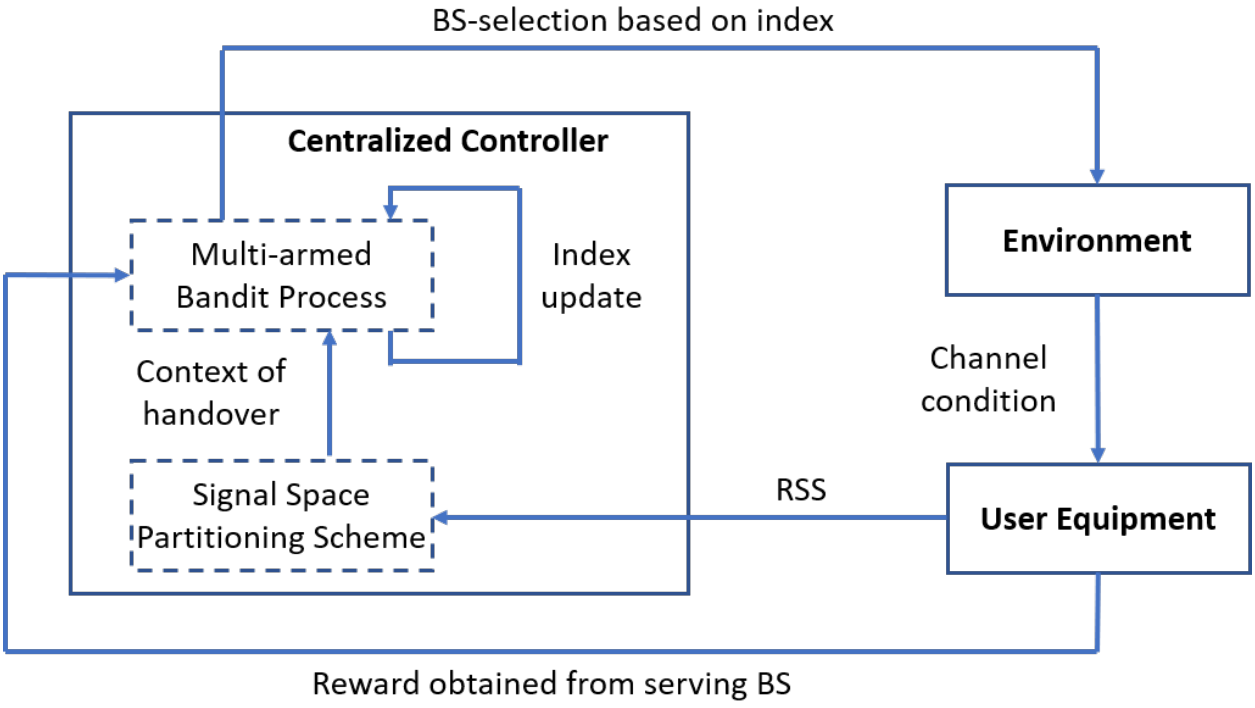


Figure 3.2: Framework of online learning of contextual handover mechanisms

3.4 Spatial Contextual Handover Mechanism (SCH)

3.4.1 Signal Space Partitioning Scheme

In my strategy, the BS that is able to provide the longest unobstructed time tends to be chosen as the handover decision. The unobstructed time of the LOS link is determined by the user's post-handover trajectory and the location of the blockage, which has a close relationship with the location of the UE. Therefore, the location of the UE should be taken as a spatial feature of handover. However, it is impossible to obtain the exact geographical location of UE without any auxiliary positioning device. To solve this problem, I propose a *signal space partitioning scheme*, which exploits the RSS information at UE as a label to differentiate handovers happening at different locations.

This scheme comes from the observation that the collection of RSSs from surrounding SBSs could be used as a signature to identify the user's location. In particular, it is typical that multiple small cells overlap in a UDN and a UE would receive multiple signals from multiple SBSs. Since these SBSs different distances from it, the signals transmitted from them have different strengths at the UE. For example, a UE k at any location is likely to receive from multiple surrounding SBSs, which form an available SBS set $\mathcal{M}_k$ for the UE k. Each SBS in $\mathcal{M}_k$ has an SNR and all SNRs constitute a signal vector for the UE k, denoted as $\mathbf{v}_k$. Each entry $v_z \in \mathbf{v}_k$ is a quantized version of the SNR received from the SBS denoted by the zth entry. The quantization is defined as follows. Choose J quantization thresholds $e_0, \ldots, e_{J-1}$, where J is a parameter and $e_{j_1} < e_{j_2}, 0 \leq j_1 < j_2 \leq J-1$, then define the quantized SNR as

$$v_z = \begin{cases} J, & \text{if } SNR_z \geq e_{J-1}, \\ j, & \text{if } e_{j-1} \leq SNR_z < e_j, 1 \leq j < J-1, \\ 0, & \text{if } SNR_z < e_0. \end{cases} \tag{3.6}$$

In this way, any geographic area can be mapped to a specific instance of a signal vector. This idea could be illustrated in Fig. 3.3.

In Fig. 3.3, there are three SBSs, A, B, and C, whose small cells overlap. Within each cell, a UE is able to receive a signal from the corresponding SBS and the RSS is above the minimum required threshold. Here I consider a binary quantization criterion, i.e., use a single quantization threshold e_0. There are 9 UEs distributed in the network. If a UE is in the small cell of an SBS and there is no blockage on the LOS link between the UE and that SBS, the UE's quantized SNR corresponding to the SBS is 1, otherwise 0. In this setting, all UEs' signal vectors are listed in Table 3.1. It is easy to see that UEs at different locations receive different signal vectors, hence have different spatial features. In particular, note that UE 8 does not have the same signal vector as UE 4, even though both of them reside in the overlap between the small cells of SBS A and SBS B. This disparity in the received signal vector arises from the blockage of the LOS between SBS B and UE 8 caused by the obstacle. Instead, UE 8 has the same signal vector as UE 1, also due to the blockage of the obstacle. As a result, UE 1 and UE 8 are considered to have the same spatial feature in my model.

Table 3.1: Signal vectors

UE	Quantized SNR			Signal Vector
	A	B	C	
1	1	0	0	1 0 0
2	0	1	0	0 1 0
3	0	0	1	0 0 1
4	1	1	0	1 1 0
5	1	0	1	1 0 1
6	0	1	1	0 1 1
7	1	1	1	1 1 1
8	1	0	0	1 0 0
9	0	0	0	0 0 0

Whenever a handover is triggered, the CC collects the UE's instantaneous RSS and identifies its signal vector by utilizing the proposed signal space partitioning scheme, which indicates where the handover event happens. According to Eq. (3.1), the RSS is related to the propagation distance between BS and UE. Therefore, the partition in the signal space

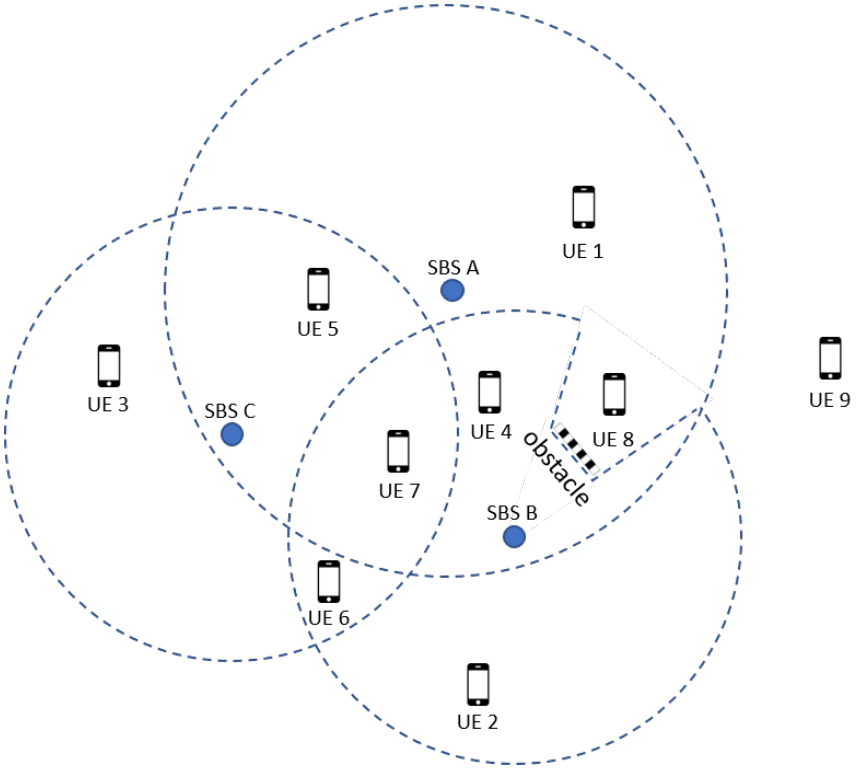


Figure 3.3: Illustration of signal space partition

could be mapped to the geographic space and each signal vector is corresponding to a unique geographic area, called *block*. This scheme is to identify handovers with the same spatial feature by assigning them the same block id. The quantization thresholds play a key role in the partitioning scheme since it determines the amount and the size of the blocks, referred to as the granularity of the partition.

3.4.2 Spatial Contextual BS-Selection based on Empirical Knowledge of Post-handover Trajectory

In the SCH mechanism, given M blocks (each corresponding to a unique signal vector), the CC maintains M independent MAB models, each serving a block by making BS-selection decisions for handover events happening in the block. Each MAB model has multiple arms corresponding to the candidate SBSs in the block which the MAB model serves. Each arm is associated with an expected reward of choosing the corresponding SBS as the handover

decision, which is accumulated in the past handovers. Choosing an arm means selecting the corresponding SBS as the decision for the handover event and the UE will be switched to that SBS. The newly obtained reward after playing that arm will be used to update the expected reward of that arm.

3.5 UCB-Based BS-Selection Algorithm for SCH

With the context of spatial feature extracted by the proposed partitioning scheme, the CC is able to retrieve the MAB model corresponding to the block where the handover event happens and make a BS-selection decision. In this subsection, I will introduce the BS-selection algorithm for this spatial contextual handover strategy.

3.5.1 Multi-Armed Bandit Model

In a block g_i, a UE has a set of candidate SBSs to choose from, denoted as $\mathcal{L}_i$. Let SNR_i^l be the SNR received by a UE from SBS l in block g_i, then $\mathcal{L}_i$ is defined as

$$\mathcal{L}_i = \{l \mid SNR_i^l \geq SNR_{\min}, l \in \mathcal{L}\}. \tag{3.7}$$

After the CC chooses a SBS $l \in \mathcal{L}_i$ for a handover in the block g_i in trial t at time τ, the UE will be served by the SBS l until it needs another handover at time τ'. Then the UE get an instantaneous reward associated with the SBS l in block g_i, denoted as $r_{i,l}^t = \tau' - \tau$. Since τ' is determined by user's moving direction, speed, and the location of blockage, the reward $r_{i,l}^t$ could be treated as an i.i.d. random variable. The BS selection in block g_i is formulated by a MAB model [41], denoted as $\mathcal{M}_i = \{\mathcal{L}_i, \mu_{i,l}^t\}$, where $l \in \mathcal{L}_i$, and $\mu_{i,l}^t$ is the expected reward of the SBS l in block g_i in trial t.

Suppose $a_{i,t}$ as the SBS truly selected by the CC following a certain handover strategy in the block g_i at trial t. I define the reward as the difference between the actually accumulated

reward and the optimal one, which could be formulated as

$$R_{i,T} = \max_{l \in \mathcal{L}_i} \mathbb{E}\left[\sum_{t=1}^{T} r_{i,l}^{t}\right] - \mathbb{E}\left[\sum_{t=1}^{T} r_{i,a_{i,t}}^{t}\right]. \tag{3.8}$$

With the model $\mathcal{M}_i$, the handover decision problem in block g_i with the aim to choose the SBS which brings the longest unobstructed LOS connection time is transformed to find the optimal policy for the corresponding MAB problem that minimizes the regret.

3.5.2 Estimation of Expected Reward

Due to the lack of full knowledge about the distribution of each SBS's reward, it can only be estimated based on historical observations [66]. Denote $T_{i,k}$ and $\bar{r}_{i,l}(T_i^l)$, as the number of times that the SBS l has been chosen and the sample mean of reward of the SBS l in block g_i, respectively. Given an instantaneous reward $r_{i,l}^t$, $T_{i,l}$ and $\bar{r}_{i,l}^t$ can be updated as

$$\bar{r}_{i,l}(T_{i,l}+1) = \frac{T_{i,l} \times \bar{r}_{i,l}(T_{i,l}) + r_{i,l}^t}{T_{i,l}+1}, \tag{3.9}$$

$$T_{i,l} := T_{i,l} + 1. \tag{3.10}$$

The initial values of $T_{i,l}$ and $\bar{r}_{i,l}(0)$ are set to be 0. The sample mean value $\bar{r}_{i,l}(T_{i,l})$ are used as the approximation of the expected reward of the SBS l in block g_i. Since an instantaneous reward is defined as the time difference between two adjacent handovers, it is determined by the distribution of blockage and the user's mobility, which include the moving direction and the speed. I consider the speed of UE as a factor that contributes to the randomness of the reward. In this subsection, the speed of UE is modeled as a random variable following a certain distribution, such as the Gaussian distribution.

3.5.3 Exploration and Exploitation

How to deal with the trade-off between exploration and exploitation is a key problem in reinforcement learning. In this subsection, I utilize the widely-used UCB policy proposed by [5] to handle this trade-off since it can achieve logarithmic regret with low computation complexity [66]. According to the UCB, the selected SBS should be

$$k^* = \arg\max_{k \in \mathcal{L}_i} \left(\bar{r}_{i,k}(T_{i,k}) + \sqrt{\frac{2 \ln F_i}{T_{i,k}}} \right). \tag{3.11}$$

In Eq. (3.11), $\bar{r}_{i,l}(T_{i,l}) + \sqrt{\frac{2 \ln F_i}{T_{i,l}}}$ indicates the index of the SBS l in block g_i, where F_i denotes the total number of handovers happened in the block. The first and the second item act as the exploitation part and the exploration part, respectively.

3.5.4 Dynamic Block Set Construction

In the SCH, the handover decision is based on the block information in the network. However, there is no assumption that there is any pre-knowledge about the blocks that could be used at the initial stage of the algorithm. Hence, the block set $\mathcal{B} =$ initially. When a handover event happens, the CC calculates the UE's signal vector. If the signal vector already exists, then the id of the corresponding block is retrieved. If not, the corresponding block will be assigned a new id and then added to $\mathcal{B}$. Meanwhile, a new MAB model for the new block is created. In this way, the block set is dynamically maintained.

3.5.5 Acceleration Technique

When a post-handover connection, built between the UE k and the SBS l in the block g_i, is blocked, a reward $r_{i,l}$ would be reported and only $\bar{r}_{i,l}$ would be updated (time- and trial- related subscripts are omitted). However, since the UE's post-handover trajectory is instantiated at this moment, it is able to update some other SBSs' rewards on this trajectory simultaneously, by using the so-called *virtual update*. Specifically, in the previous handover,

if the CC switched the UE k to the SBS l in block g_i, the CC was also aware of the set of SBSs which were not selected, denoted as $\bar{\mathcal{L}}_{i,l} = \mathcal{L}_i \backslash \{l\}$, and pretended to build a virtual LOS link between the UE k and each $l' \in \bar{\mathcal{L}}_{i,l}$. During the UE's post-handover movement, in addition to checking the handover trigger condition on the true LOS link, the CC kept checking that on each virtual LOS link. If the virtual LOS path between the UE k and the SBS l' was blocked, the observed reward $r_{i,l'}$ was calculated and used to update the sample mean $\bar{r}_{i,l'}$, although the corresponding handover event did not truly occur. By this virtual update, any trajectory of UE can be used to update multiple sample means and the efficiency of the algorithm can be improved significantly.

The UCB-based BS-selection algorithm for the the SCH is summarized in Algorithm 1.

Algorithm 1 UCB-based BS-selection algorithm for the SCH

Input: Cellular network $\mathcal{G}$ which consists of a set $\mathcal{L}$ of SBSs and a set of obstacles
$\mathcal{B} = \emptyset$;
while Event A2 handover trigger condition is met for a UE k **do**
    Record the current time τ;
    Identify the block g_i where UE n resides, associated with the available SBS set $\mathcal{L}_i \subseteq \mathcal{L}$;
    if $g_i \notin \mathcal{B}$ **then**
        $T_{l,k} \leftarrow 0$;
        $\bar{r}_{i,l}(0) \leftarrow 0$;
        $F_i \leftarrow 0$;
        $\mathcal{B} \leftarrow \mathcal{B} \cup g_i$;
    end if
    $a_i = \arg\max_{l \in \mathcal{L}_i} \left(\bar{r}_{i,l}(T_{i,l}) + \sqrt{\frac{2 \ln F_i}{T_{i,l}}} \right)$;
    Switch the UE k to the SBS a_i;
    Observe the reward $r_{i,a_i} = \tau' - \tau$ when the next handover occurs for UE k at time τ';
    $\bar{r}_{i,a_i}(T_{i,l}+1) \leftarrow \frac{T_{i,l} \times \bar{r}_{i,a_i}(T_{i,l}) + r_{i,a_i}}{T_{i,l}+1}$;
    $T_{i,l} \leftarrow T_{i,l} + 1$;
    $F_i \leftarrow F_i + 1$;
    Update $\bar{r}_{i,l'}(T_{i,l'}+1)$, $T_{i,l'}$ and F_i for $l' \in \bar{\mathcal{B}}_{i,a_i}$ in the same way, if the virtual reward $r_{i,l'}$ is obtained.
end while

3.6 Space-Time Contextual Handover Mechanism (STCH)

3.6.1 Temporal Feature Extraction

The UE's mobility includes not only its instantaneous location but also the moving direction, which refers to the location change. This location change can not be reflected by a single block id. Hence, I propose to use a sequence of block ids, i.e., a *block concatenation*, as a label to identify the UE's moving direction. In particular, the CC maintains a block concatenation for each UE, which contains the ids of the blocks that the UE has passed in chronological order. This block concatenation that reflects the location change over a period of time could be used as a coarse-grained moving direction of UE.

The handover events shown in Fig. 3.4 illustrate how the block concatenation is used to represent the UE's moving direction. Suppose the UEs 1, 2, and 3 need handover in the block a. Each of them has its own moving direction, which is indicated by the corresponding arrow. It is assumed that all these three UEs are able to receive signals from the SBS A, B, and C. According to the SCH strategy, these three UEs have the same spatial feature since they are in the same block, and they are treated homogeneously by the MAB model corresponding to the block a. Therefore, they will be switched to the same SBS, suppose to be the SBS B. Obviously, considering the individual moving directions of these UEs, the SBS B is not the optimal solution for all of them. In particular, except the UE 2, UE 1 and UE 3 are moving away from the SBS B. So the SBS B is not optimal for them. Due to their moving directions, the SBS C and the SBS A can provide longer LOS connection times for them, respectively. That requires further distinguishing the UEs by considering their moving directions besides their locations. By using the concatenation of historical block Ids, I can classify those UEs according to their pre-handover trajectories, and hence offer an opportunity to better tailor the handover decisions for them. For example, being aware that the UEs 1, 2, and 3 come from the blocks d, b, and c, I can switch them to the SBS C, B, and A, respectively, and that leads to better handover decisions.

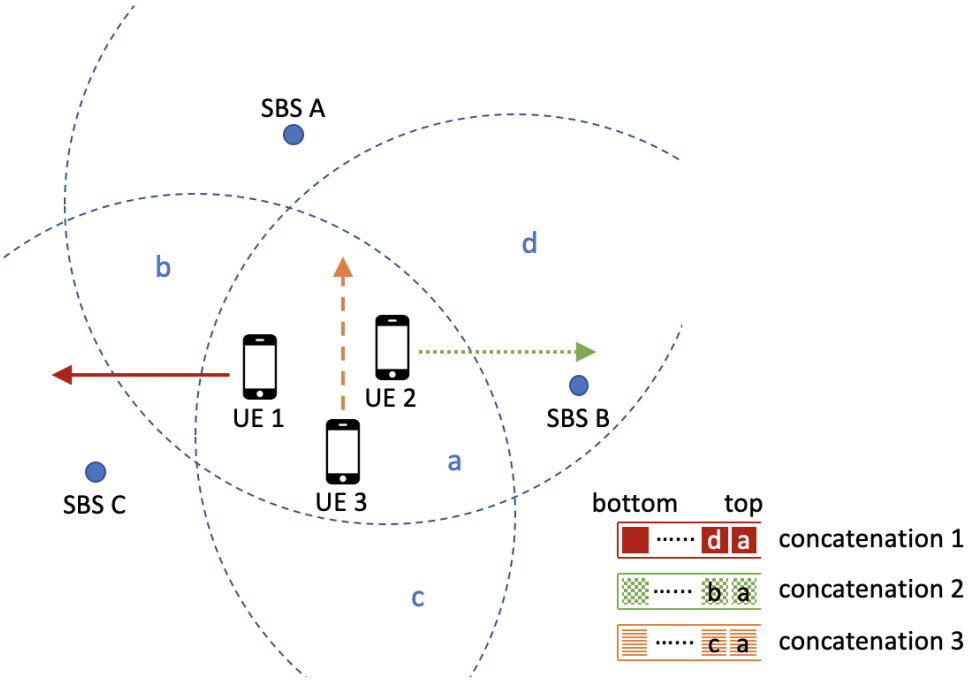


Figure 3.4: Differentiation of UEs' moving directions

It is worth noting that, although this block-concatenation-based estimation of UE's direction is coarse-grained, it provides valuable information to further differentiate UEs with the same spatial feature. Obviously, with finer partition granularity of the signal space and longer block concatenation, the concatenation-based temporal feature is more differentiable. However, maintaining the entire block-traversing history for every UE is expensive, I use the one-step-look-back construct, which is the simplest implementation of the block concatenation. In particular, given a UE k in the block g_k, denote $\mathbf{g}_k = (g_k^-, g_k)$ as the block concatenation of UE, where g_k^- is the immediate preceding block to g_k. Using this one-step-look-back construct not only avoids the "curse of dimensionality" caused by long concatenation but also saves storage and reduces the space complexity of the system.

3.6.2 Space-Time Contextual BS-Selection Based on Empirical Knowledge of Post-handover Trajectory

In the STCH strategy, the CC maintains a space-time contextual MAB model for each block to make BS-selection decisions for handover events happening in the block. In particular, for any block g_i, I define the set of its preceding blocks $\mathcal{B}_i = \{g_k^- | g_k = g_i, \forall \tilde{\mathbf{g}}_k = (g_k^-, g_k), \forall k \in \mathcal{K}\}$ from all block concatenations of all UEs, where $\mathcal{K}$ is the set of UEs. The contextual MAB process corresponding to block g_i maintains a coefficient vector $\boldsymbol{\theta}_{i,l}$ for each candidate (or arm) SBS $l \in \mathcal{L}_i$, where $\mathcal{L}_i$ is the set of all available SBSs a UE may handover to in block g_i. The elements in $\boldsymbol{\theta}_{i,k}$, each corresponding to a unique preceding block $g_j \in \mathcal{B}_i$, represent the expected rewards a UE in block g_i can receive by handing over to the SBS k, considering that the UE has come to g_i from various (coarse-grained) directions, respectively (i.e., one element per direction). As described in the previous section, here each coarse-grained direction is represented by a unique preceding block in $\mathcal{B}_i$.

Let $\mathbf{x}$ denote the context which reflects the temporal feature $\tilde{\mathbf{g}}_k = (g_k^-, g_k)$ of a handover event happening on the UE k in block g_i. Given $\mathbf{x}$ and $\boldsymbol{\theta}_{i,l}$ for $\forall l \in \mathcal{L}_i$, the MAB model will calculate the expected reward for handing over to the SBS l as $\bar{r}_{i,l} = \theta_{i,l}^j$, where $\theta_{i,l}^j$ is the element in $\boldsymbol{\theta}_{i,l}$ corresponding to preceding block g_j. The actual handover decision is made according to a specific criterion, e.g., choosing the SBS, say a, with the largest expected reward. Once a decision is made, the UE will be switched to and served by the SBS a from then on. When the LOS connection between the UE and the SBS a is lost, e.g., the propagation distance is too long or there is a blockage, an actual reward $r_{i,a}$ that represents the empirical connection time will be calculated. Subsequently, $\mathbf{x}$, a, and $r_{i,a}$ will be used to update the expected rewards in $\boldsymbol{\theta}_{i,a}$, as will be described in details in Section 3.7.

3.7 LinUCB-Based BS-Selection Algorithm for STCH

In this section, I formulate this decision as a contextual MAB problem and propose a LinUCB-based BS-selection algorithm that considers the time-space feature of handover for STCH.

3.7.1 Contextual Multi-Armed Bandit Model

Given the handovers with the same spatial feature, the contextual MAB model identifies the temporal feature of each handover and applies a tailored policy. The BS-selection algorithm for the STCH is based on the LinUCB algorithm which has been widely utilized in many industry fields [35, 11]. When a handover event happens in trial t ($t = 1, 2, 3, \ldots, T$), the algorithm firstly identifies the block g_i where it occurs and finds out the candidate SBS set $\mathcal{L}_i$. The context (i.e., temporal feature of handover) $\mathbf{x}_{i,l}^t$ with dimension of d_i, which is associated with the SBS $l \in \mathcal{L}_i$, is then extracted, and the details will be discussed in Section 3.7.2. According to [35], the expected reward received from the SBS l at trail t is linear in its d_i-dimension context $\mathbf{x}_{i,l}^t$ with unknown coefficient vector $\boldsymbol{\theta}_{i,k}$ and is shown as

$$\bar{r}_{i,l}^t = \mathbb{E}[r_{i,l}^t | \mathbf{x}_{i,l}^t] = {\mathbf{x}_{i,l}^t}^\top \boldsymbol{\theta}_{i,l}. \tag{3.12}$$

The d_i-dimension coefficient vectors $\boldsymbol{\theta}_{i,l}$ are updated with the accumulated observations and used to guide future handover decisions accompanied with the context $\mathbf{x}_{i,l}^t$. Denote $\mathbf{D}_{i,k}$ as a $m \times d_i$ matrix which consists of m contexts observed in the past for SBS l in block g_i, and $\mathbf{c}_{i,l}$ as a m-dimension vector which indicates the rewards of the m observations. By applying ridge regression to $\mathbf{D}_{i,l}$ and $\mathbf{c}_{i,l}$, the optimal $\boldsymbol{\theta}_{i,l}^*$ is estimated as

$$\hat{\boldsymbol{\theta}}_{i,l} = (\mathbf{D}_{i,l}^\top \mathbf{D}_{i,l} + \mathbf{I}_{d_i})^{-1} \mathbf{D}_{i,l}^\top \mathbf{c}_{i,l}, \tag{3.13}$$

where $\mathbf{I}_{d_i}$ is a d_i-dimension identity matrix. The algorithm chooses SBS in each trial t in block g_i as

$$a_{i,t} = \arg\max_{l \in \mathcal{L}_i} \left({\mathbf{x}_{i,l}^t}^\top \hat{\boldsymbol{\theta}}_{i,l} + \eta \sqrt{{\mathbf{x}_{i,l}^t}^\top \mathbf{A}_{i,l}^{-1} \mathbf{x}_{i,l}^t} \right), \tag{3.14}$$

where $\mathbf{A}_{i,l} = \mathbf{D}_{i,l}^\top \mathbf{D}_{i,l} + \mathbf{I}_{d_i}$, and $\eta > 0$ is the hyperparameter. Specifically, ${\mathbf{x}_{i,l}^t}^\top \hat{\boldsymbol{\theta}}_{i,l}$ is the predicted reward for the SBS l, while $\sqrt{{\mathbf{x}_{i,l}^t}^\top \mathbf{A}_{i,l}^{-1} \mathbf{x}_{i,l}^t}$ indicates the standard deviation of reward. Given previous observations, the algorithm chooses the SBS $a_{i,t}$ with the optimal expected reward according to $\mathbf{x}_{i,a_{i,t}}^t$. When the reward $r_{i,a_{i,t}}^t$ is obtained, the new observation $(\mathbf{x}_{i,a_{i,t}}^t, a_{i,t}, r_{t,a_{i,t}}^t)$ is used to update the BS-selection policy.

3.7.2 Context Construction

In the STCH strategy, once there is a handover triggered, the CC firstly identifies its spatial feature, i.e., the block where it occurs, and then extracts its temporal feature based on the UE's block concatenation. These features are then used to construct the context of the proposed contextual MAB model for the STCH. Specifically, given the preceding blocks $\mathcal{B}_i$ of a block g_i and $d_i = |\mathcal{B}_i|$, $i = 1, \ldots, M$, I define a context associated with the SBS $l \in \mathcal{L}_i$ for any handover event raised in block g_i in trial t as a d_i-dimension 0-1 vector, denoted by $\mathbf{x}_{i,l}^t = (x_{i1}^t, \ldots, x_{i,d_i}^t)$. Each element $x_{ij}^t \in \mathbf{x}_{i,l}^t$ corresponds to a proceeding block $g_{ij} \in \mathcal{B}_i$, where g_{ij} denotes the jth preceding block in $\mathcal{B}_i$. Note that I consider that all SBS $l \in \mathcal{L}_i$ share the same context construction. The binary value of $x_{ij}^t \in \mathbf{x}_{i,l}^t$ depends on whether the g_{ij} matches the temporal feature of the handover, say $\tilde{\mathbf{g}}_k = (g_k^-, g_k)$. In particular, $x_{ij}^t = 1$ if $g_{ij} = g_k^-$, and $x_{ij}^t = 0$ otherwise, where $j = 1, \ldots, d_i$.

From the very beginning, the CC has no knowledge about the block set in the STCH and needs to dynamically construct it, which is similar to the block set construction in the SCH, except for the trigger condition. In particular, in the STCH, the block set is constructed all the time during UE's movement and the CC keeps updating UE's block concatenation, while in the SCH, the block is constructed only when a handover is triggered. Due to the lack of prior knowledge in the initial stage, I set $\mathcal{B}_i = 0, \forall g_i \in \mathcal{B}$. When a handover event

happens in the block g_i, which is associated with a preceding block g_k^-, the CC adds it into $\mathcal{B}_i$ if $g_k^- \notin \mathcal{B}_i$. Accordingly, the dimension of the temporal feature of any handover events raised in the block g_i increases by 1, as well as the dimensions of related coefficients.

3.7.3 Acceleration Technique

The acceleration technique introduced in Section 3.5.5 can also be applied to improve the efficiency of the LinUCB-based BS-selection algorithm for the STCH. Specifically, when switching a UE k to the chosen SBS, suppose to be SBS l, in block g_i in trial t with context $\mathbf{x}_{i,l}^t$, the CC also build a virtual connection from the UE to each SBS in $\bar{\mathcal{L}}_{i,l}$. In the UE's post-handover trajectory, if a virtual connection between the UE to a SBS $l' \in \bar{\mathcal{L}}_{i,l}$, is blocked, the instantaneous virtual reward $r_{i,l'}^t$ is obtained. Then the associated virtual observation $(\mathbf{x}_{i,l'}^t, l', r_{i,l'}^t)$ is used to update the coefficients for the SBS l' in the LinUCB model of block g_i. In this way, not only the coefficients for the SBS l, but also those for some SBS $l' \in \bar{\mathcal{L}}_{i,l}$ are able to be updated by a single instance of UE's post-handover trajectory. As mentioned in Section 3.5.5, the virtual update can fully exploit true experience to increase the efficiency of the BS-selection algorithm.

The LinUCB-based BS-selection algorithm for STCH is summarized in Algorithm 2.

3.8 Complexity Analysis

To analyze the computation complexity of the proposed algorithm, I divide the total cost of the system into the following three parts.

(1) Model-related cost. The CC maintains a MAB model for each block. For an UCB-based BS-selection algorithm, the space complexity is $\mathbf{O}(L_{\mathrm{b}})$, where L_{b} is the average number of candidate SBSs in a block. It is because each arm requires two storage units to maintain its expected reward and its chosen times, respectively. Moreover, the time complexity is also $\mathbf{O}(L_{\mathrm{b}})$, since the CC needs to compute and search for the biggest index among the arms which refer to the candidate SBSs. For the same reason, in

the LinUCB-based BS-selection algorithm, the time complexity is also $\mathbf{O}(L_{\mathrm{b}})$. Due to the time-space context, the STCH needs extra storage units to keep the temporal feature and the space complexity is $\mathbf{O}(d_{\mathrm{b}}L_{\mathrm{b}})$, where d_{b} is the average dimension of the temporal feature of a handover in a block. For the whole system, there are M MAB models, where M is the number of the blocks as well as the signal vectors identified according to the signal space partitioning scheme. Theoretically, the number of blocks is huge. Suppose there are L SBSs in the network and J quantization thresholds, there exist totally $(J+1)^L$ different signal vectors in the signal space. However, because the

Algorithm 2 LinUCB-based BS-selection algorithm for STCH

Input: Cellular network $\mathcal{G}$ which consists of a set $\mathcal{L}$ of SBSs and a set of obstacles, $\eta \in \mathbb{R}_+$
$\mathcal{B} = \emptyset$;
Keep tracking each UE k and updating its block concatenation $\tilde{\mathbf{g}}_k = (g_k^-, g_k)$;;
if $g_k \notin \mathcal{B}$ **then**
    $\mathcal{B}_k = \emptyset$, $d_k = 0$;
    $\mathbf{A}_{k,l} = \emptyset$, $\mathbf{b}_{k,l} = \emptyset$;
    $\mathcal{B} \leftarrow \mathcal{B} \cup g_k$;
end if
while Event A2 handover trigger condition is met for UE k **do**
    Identify the block g_i where UE k resides, associated with the available SBS set $\mathcal{L}_i$ and the corresponding trial number t;
    Retrieve $\tilde{\mathbf{g}}_k = (g_k^-, g_i)$ for UE k from the memory;
    if $g_k^- \notin \mathcal{B}_i$ **then**
        $\mathbf{A}_{i,l} \leftarrow [\mathbf{A}_{i,l}|v_0]$, where v_0 is a d_i-dimension zero vector;
        $\mathbf{A}_{i,l}^\top \leftarrow [\mathbf{A}_{i,l}^\top|v_1]$, where v_1 is a (d_i+1)-dimension vector, $v_1 = [0, \ldots, 0, 1]$;
        $\mathbf{b}_{i,l} \leftarrow [\mathbf{b}_{i,l}|0]$;
        $d_i \leftarrow d_i + 1$;
        $\mathcal{B}_i \leftarrow \mathcal{B}_i \cup g_k^-$;
    end if
    Observe context $\mathbf{x}_{i,l}^t$ from $\tilde{\mathbf{g}}_k$ for all SBS $l \in \mathcal{L}_i$;
    $\hat{\boldsymbol{\theta}}_{i,l} \leftarrow \mathbf{A}_{i,l}^{-1}\mathbf{b}_{i,l}$;
    $q_{i,l}^t \leftarrow {\mathbf{x}_{i,l}^t}^\top \hat{\boldsymbol{\theta}}_{i,l} + \eta\sqrt{{\mathbf{x}_{i,l}^t}^\top \mathbf{A}_{i,l}^{-1}\mathbf{x}_{i,l}^t}$;
    Choose SBS $a_{i,t} = \arg\max_{l \in \mathcal{L}_i} q_{i,l}^t$ and observe a reward $r_{i,l}^t$;
    $\mathbf{A}_{i,a_{i,t}} \leftarrow \mathbf{A}_{i,a_{i,t}} + \mathbf{x}_{i,a_{i,t}}^t {\mathbf{x}_{i,l_{i,t}}^t}^\top$;
    $\mathbf{b}_{i,a_{i,t}} \leftarrow \mathbf{b}_{i,a_{i,t}} + r_{i,l}^t \mathbf{x}_{i,a_{i,t}}^t$;
    Update $\mathbf{A}_{i,l'}$ and $\mathbf{b}_{i,l'}$ for $l' \in \bar{\mathcal{L}}_{i,a_{i,t}}$ in the same way, if the virtual reward $r_{i,l'}^t$ is obtained.
end while

propagation range of the mmWave signal is limited, the number of available SBSs for a UE is much smaller than that of the whole SBSs. It means that the signal space is sparse and the vast majority of the signal vectors will not show up in practice. Therefore, the number of MAB models $M \ll (J+1)^L$. Hence, the total MAB related space complexity is $\mathbf{O}(ML_{\mathrm{b}})$ for the SCH, and $\mathbf{O}(Md_{\mathrm{b}}L_{\mathrm{b}})$ for the STCH.

(2) Block-set-related cost. Since the CC keeps the mapping between each block and its associated signal vector, the space complexity to maintain the signal vector and the blocks is $\mathbf{O}(M)$. When a handover is triggered, the CC needs to retrieve the id of the block where the handover event occurs according to a given signal vector. This search contributes to the time complexity of $\mathbf{O}(M)$.

(3) UE-related cost. In order to support the BS-selection algorithm, the CC needs to keep the reward information for all UEs, which causes the space complexity of $\mathbf{O}(K)$, where K is the average number of all UEs in the network over time. If the CC applies the acceleration technique, it also needs to keep checking the virtual LOS links, and the extra space complexity of $\mathbf{O}(KL_{\mathrm{b}})$ is required to keep the virtual rewards for all UEs. Especially, for the STCH, the CC also needs to keep a block concatenation for each UE, which requires additional space complexity of $\mathbf{O}(ZK)$, where Z is the size of the block concatenation. When the one-step-back-look construct is used, L could be reduced to 2.

To sum up, the total costs as well as the computation complexity of the proposed schemes could be summarized in Table 3.2.

3.9 Numerical Experiments

In this section, I evaluate the performance of the proposed handover strategies in various scenarios on two important metrics, i.e., the average number of handovers per UE (ANH) and the average lasting time per each LOS connection (ALT). These two metrics directly

Table 3.2: Complexity Analysis

	Time Complexity	Space Complexity
SCH	$\mathbf{O}(M + L_{\mathrm{b}})$	$\mathbf{O}(ML_{\mathrm{b}} + K)$
SCH with acceleration	$\mathbf{O}(M + L_{\mathrm{b}})$	$\mathbf{O}(ML_{\mathrm{b}} + KL_{\mathrm{b}})$
STCH	$\mathbf{O}(M + L_{\mathrm{b}})$	$\mathbf{O}(Md_{\mathrm{b}}L_{\mathrm{b}} + KZ)$
STCH with acceleration	$\mathbf{O}(M + L_{\mathrm{b}})$	$\mathbf{O}(Md_{\mathrm{b}}L_{\mathrm{b}} + K(Z + L_{\mathrm{b}}))$

reflect the quality of the handover decision. I compare the proposed two handover strategies with two counterparts: the Rate-first handover (RFH) and the SMART [66]. In the RFH, the CC chooses the BS that provides the maximum transmission rate as the handover decision, while the SMART proposes to use a single MAB model to make the handover decision for all handover events in the network.

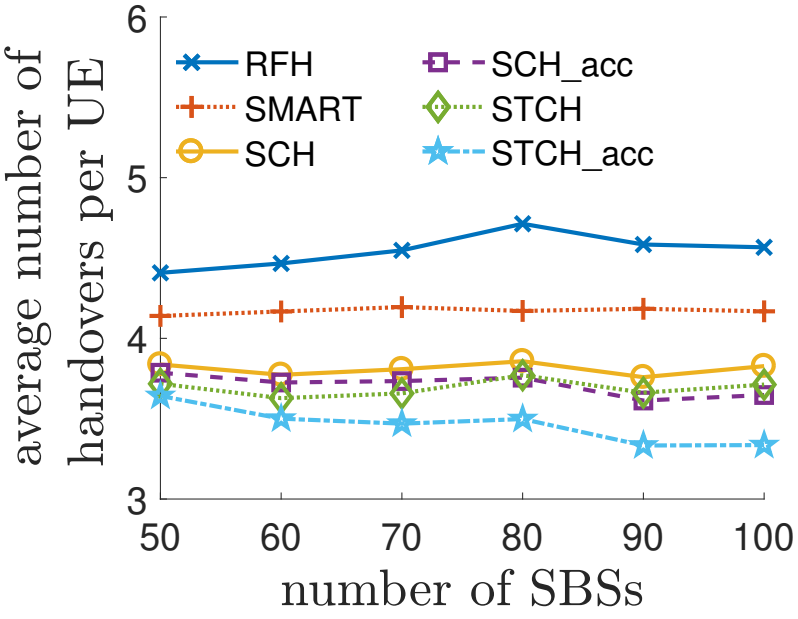
3.9.1 Experiment Settings

I consider a cellular network $\mathcal{G}$ which covers a 100(m)×100(m) square region and consists of a certain number of mmWave-band SBSs (100 by default) and single-antenna UEs. The transmit power of SBS is set to 30 dBm, and the noise power is -57 dBm. Similar to [2], I set the parameters α and β in Eq. (3.1) as 61.4 and 2, respectively, corresponding to a carrier frequency of 28 GHz. The channel gain G_{max} of main lobe is 18 dB as in [61]. The bandwidth of SBS is set as 1000 MHz. I assume that a certain number of identical obstacles (20 by default) with a radius of 1(m) are randomly distributed in the network. The minimum required transmission rate $h_{\min}$ is set to be 1000 Mbps. The number of UEs entering into the network per time slot has a Poisson distribution with parameter λ. For a new coming UE, its initial position is uniformly distributed at the border of the network and its moving orientation is also uniformly distributed. The UE's moving velocity is supposed to be random following $\mathbb{N}(\mu_{\mathrm{v}}, \sigma_{\mathrm{v}}^2)$, where μ_{v} is the mean value and σ_{v} is the standard deviation. I set $\mu_{\mathrm{v}} = 1$ and $\sigma_{\mathrm{v}} = 0.1$ by default. Any UE's experience is used to update the accumulated reward of SBS until it moves out of the network region. Furthermore, the hyperparameter η in Eq. (3.14) is empirically set to be 1 because that leads to the best results in my experiments.

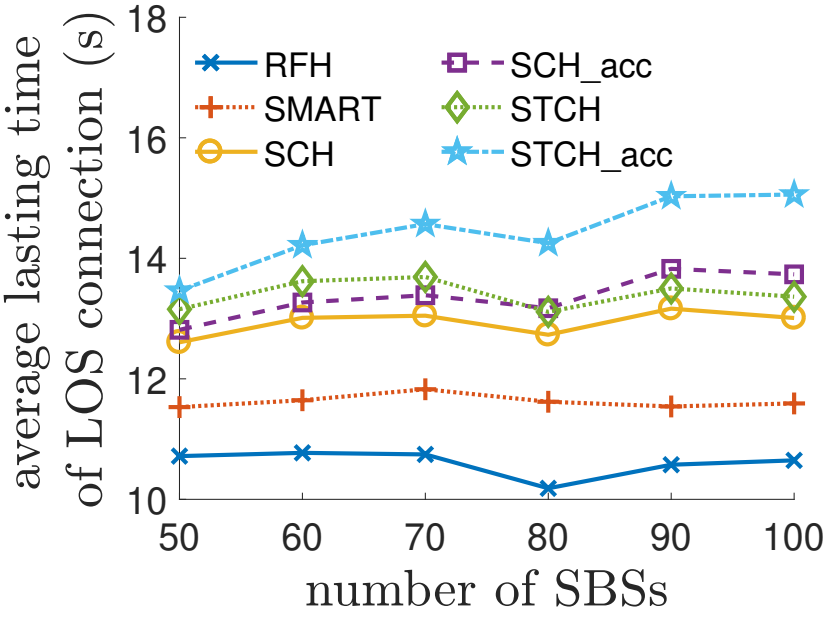
Moreover, the UE's arrival rate and the threshold of SNR are set to be 3 (/iteration) and 20 (dB), respectively by default, if not specified.

3.9.2 Number of SBSs

In this experiment, I compare the performances of the considered handover strategies, i.e., RFH, SMART, SCH, SCH with acceleration (SCH_acc), STCH and STCH with acceleration (STCH_acc), with different numbers of SBSs on the metrics of ANH and ALT. I consider six instances with different numbers of SBS: 50, 60, 70, 80, 90, and 100, and run 10000 iterations (time unit) for each. The results are shown in Fig. 3.5. It can be found that the proposed contextual handover strategies perform better than the other two, i.e., the RFH and the SMART. Compared with the SMART, which obviously outperforms the RFH, the SCH improves the ANH and the ALT by up to 10.0% and 14.0%, while the STCH improves even better, by up to 12.4% and 17.0%, respectively when given 90 SBSs. It is demonstrated that better handover decisions could be obtained by considering the temporal feature than by only taking account of the spatial feature. It is worth noting that, when the acceleration technique is applied, the SCH_acc improves the ANH and the ALT by 13.6% and 19.8%, while the STCH_acc improves by 20.3% and 30.2%, respectively compared with the SMART. This shows the advantages of the acceleration technique. It could also be found that, when the number of SBSs grows, there is no significant improvement in the performances of the SCH and the STCH. It is because when there are more SBSs, there will be more blocks identified and more MAB models maintained by the CC. Therefore, with a certain number of given iterations, the average number of training samples allocated to each MAB model is smaller and hence easily causes models under-trained. Fortunately, this issue could be mitigated by applying the proposed acceleration technique which increases the utility efficiency of each training sample and achieves good performance even with limited iterations.



(a) Performance on the average number of handovers per UE



(b) Performance on the average lasting time per LOS connection

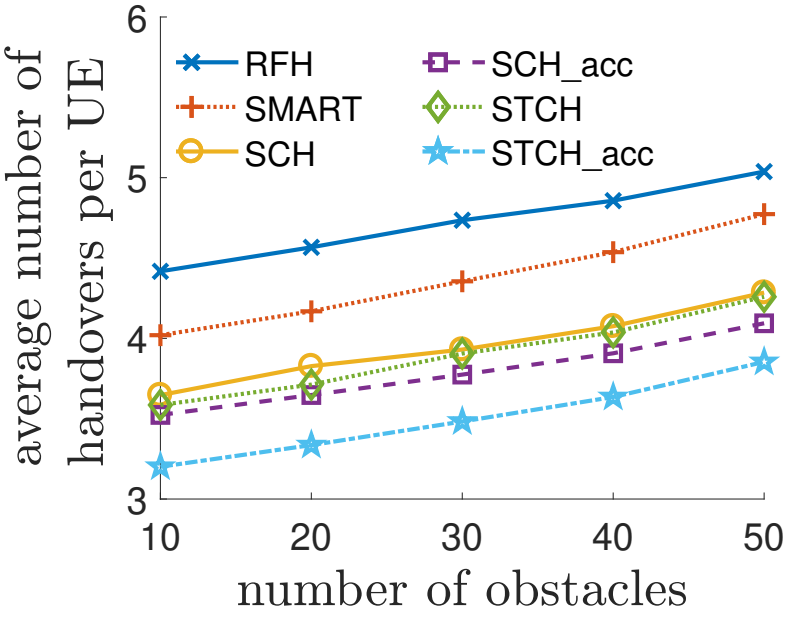
Figure 3.5: Comparison of performances with different numbers of SBSs

3.9.3 Density of Obstacles

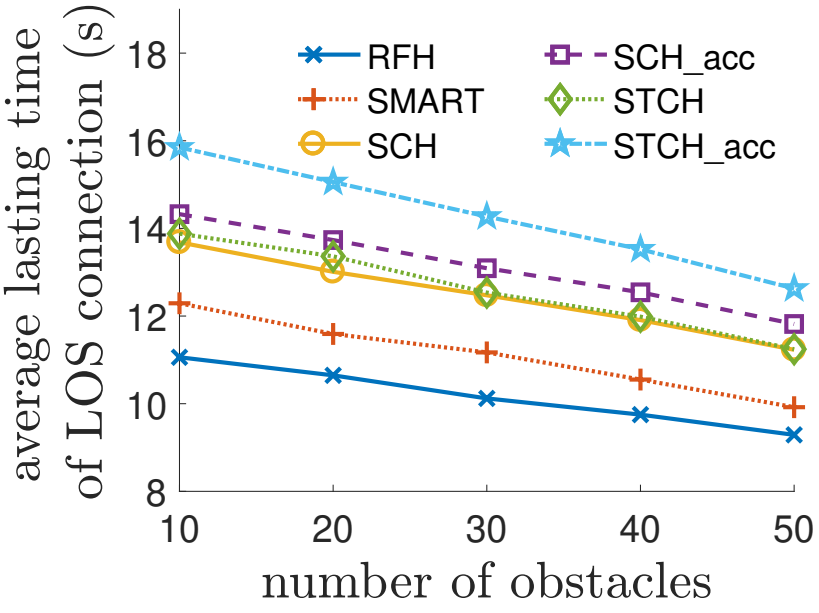
In this simulation, I investigate the impact of the density of obstacles on the performances of the proposed handover strategies. Fig. 3.6 illustrates ANH and ALT obtained under these considered strategies with various numbers of obstacles, from 10 to 50, after 10000 iterations. With a higher density of obstacles, more handovers are raised, and the LOS connection time decrease, no matter under which strategy. This trend is in accord with the intuition that, there are usually more handover events in a complex environment with many obstacles than in a simple environment with few obstacles. Furthermore, the proposed handover strategies, i.e., SCH, SCH_acc, STCH and STCH_acc, observably outperform the other two counterparts, i.e., RFH and SMART, no matter in which scenario.

3.9.4 UE's Arrival Rate

In this simulation, I compare the performances of the proposed handover strategies with different arrival rates of UE. The arrival rate of UE reflects the crowdedness of the network. I consider five values for λ: 1, 2, 3, 4, and 5 to simulate different practical scenarios in terms of the crowdedness, and run 10000 iterations for each. The results are displayed in Fig. 3.7. As shown, in a crowded scenario (i.e., when λ approaches 5), the performances of RFH and SMART decrease, while all the proposed strategies still have good performances, which are even better than those in a less crowded scenario. The reason is that, in a crowded network, there are so many users providing sufficient training samples to make the MAB models well-trained. While in a less crowded situation, the MAB models may get under-trained due to a lack of enough training samples. This is the reason why two out of the four proposed strategies, i.e., SCH and STCH, perform even worse than the SMART when $\lambda = 1$. However, if well trained in a crowded scenario, their performances could be improved significantly.

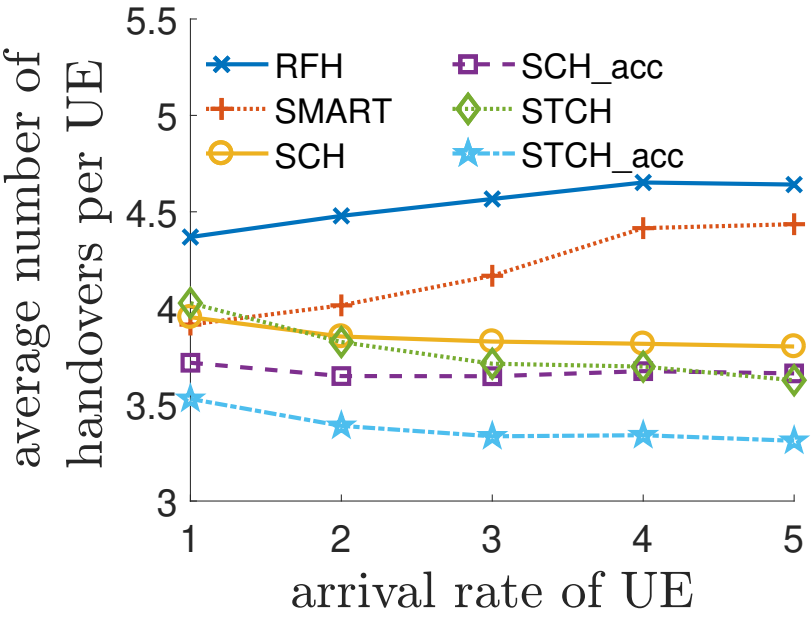


(a) Performance on the average number of handovers per UE

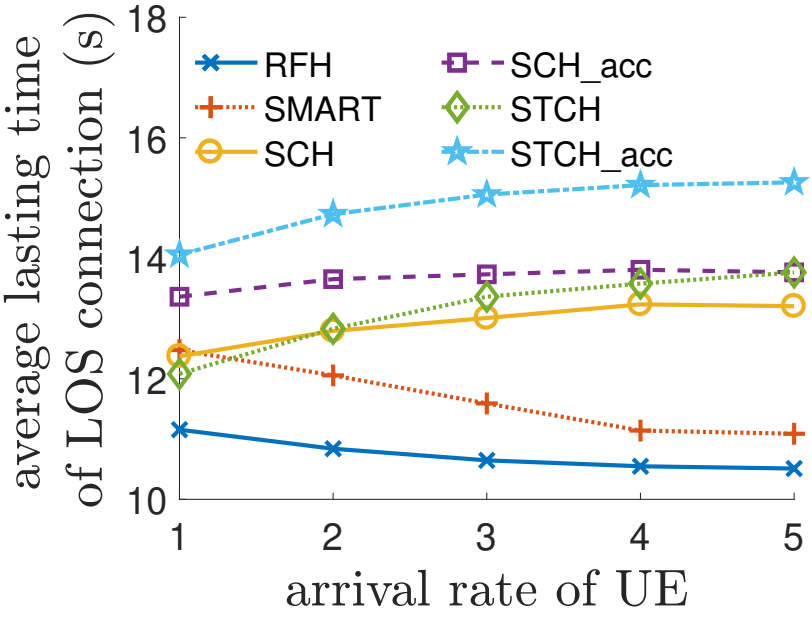


(b) Performance on the average lasting time per LOS connection

Figure 3.6: Comparison of performances with different numbers of obstacles



(a) Performance on the average number of handovers per UE



(b) Performance on the average lasting time per LOS connection

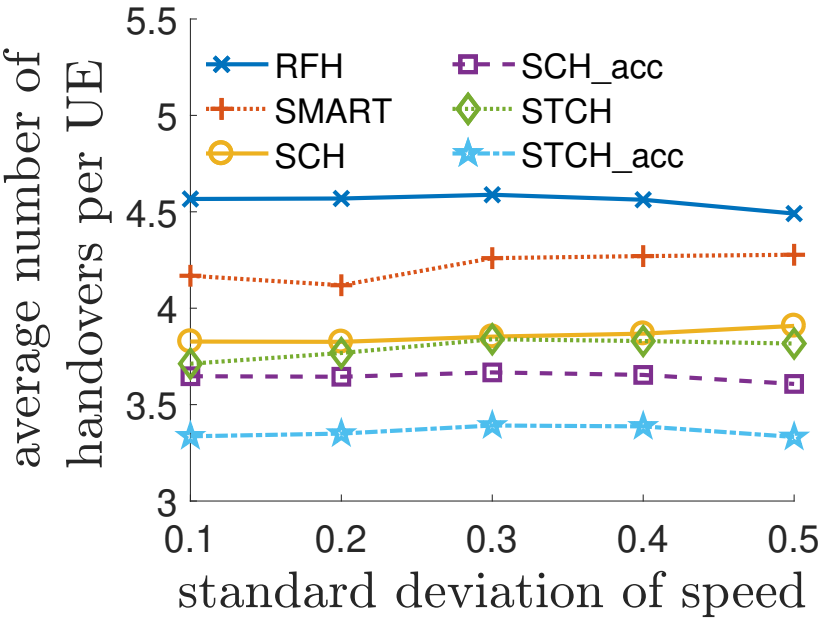
Figure 3.7: Comparison of performances with different arrival rates

3.9.5 Variance of UE's Speed

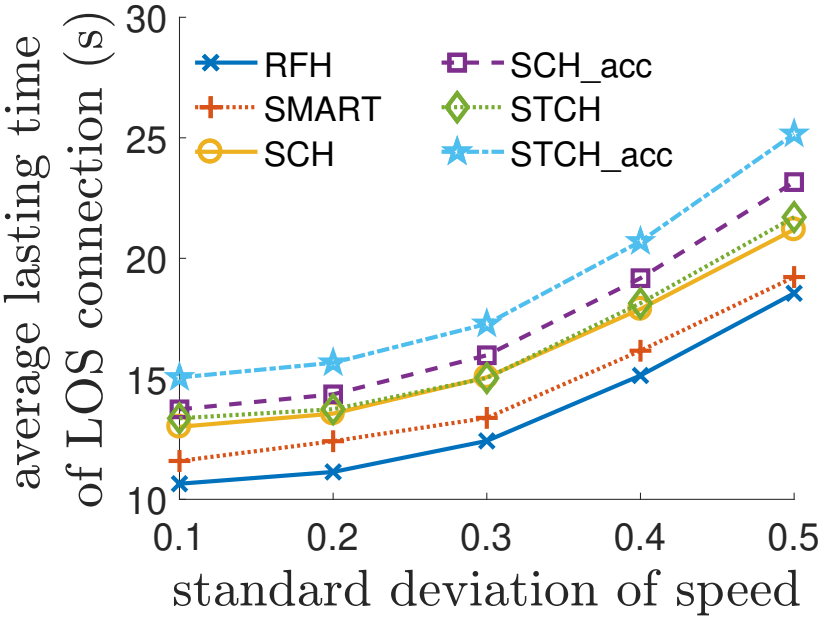
Since the instantaneous reward obtained by a UE is impacted by its moving speed, as described in Section 3.5.2, it is worth investigating the influence of the variance of UE's speed on the performances of the proposed handover strategies. In this part, I consider the random moving speed of UEs following Gaussian distribution, denoted as $\mathbb{N}(1, \sigma_{\mathrm{v}}^2)$. As the variance of UE's speed is indicated by σ_{v}, I choose five values for it, from 0.1 to 0.5, and run 10000 iterations for each instance. The simulation results are shown in Fig. 3.8. As expected, the performances of the proposed strategies on the ANH and the ALT outperform the two benchmarks. Note that, the ANHs of all considered strategies generally keep the same, no matter with which variance of speed. The reason lies in that, given the distribution of speed, the impact of the variance of reward caused by random speeds of UE on the performance of the ANH could be averaged out over time. This result demonstrates the stability of the proposed handover strategies in the scenario with UEs whose speeds are various.

3.9.6 Granularity of Partition

As described in Section 3.4.1, the quantization thresholds determine the granularity of the partition of the signal space as well as the network into blocks. To figure out the impact of the granularity of partition on the performance of the proposed handover strategies, I try four values for J: 1, 2, 3, and 4, and run 100000 iterations for each instance. The results are displayed in Fig. 3.9. The simulation results show that, for any proposed strategy, the performances with different quantization thresholds are comparable. In other words, although the performances under fine granularity (i.e., $J = 4$) of partitioning are better than those under coarse granularity (i.e., $J = 1$), their difference is ignorable. However, the increment of the required storage with fine granularity is much larger than that with coarse granularity. I show the number of identified blocks under these scenarios with different values of J in Fig. 3.10. As shown, when the value of J grows, the number of identified blocks increases exponentially. That means the CC has to spend much higher storage cost



(a) Performance on the average number of handovers per UE



(b) Performance on the average lasting time per LOS connection

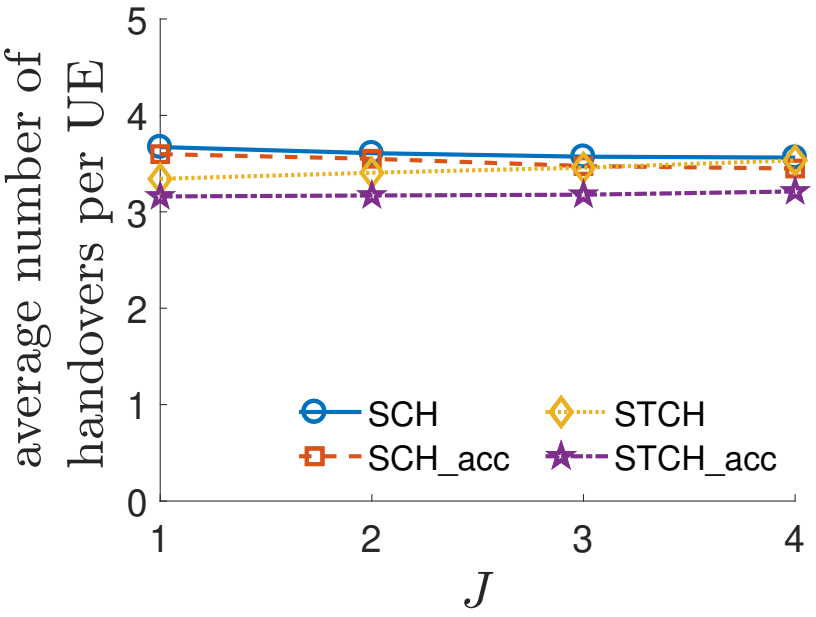
Figure 3.8: Comparison of performances with different variances of UE's speed

to support the fine granularity of partitioning. Obviously, the output is not proportional to the input when adopting fine granularity of partitioning. At least in this simulation, the coarse granularity of partitioning, i.e., $J = 1$, is good enough to achieve good performance while requiring the lowest computation cost.

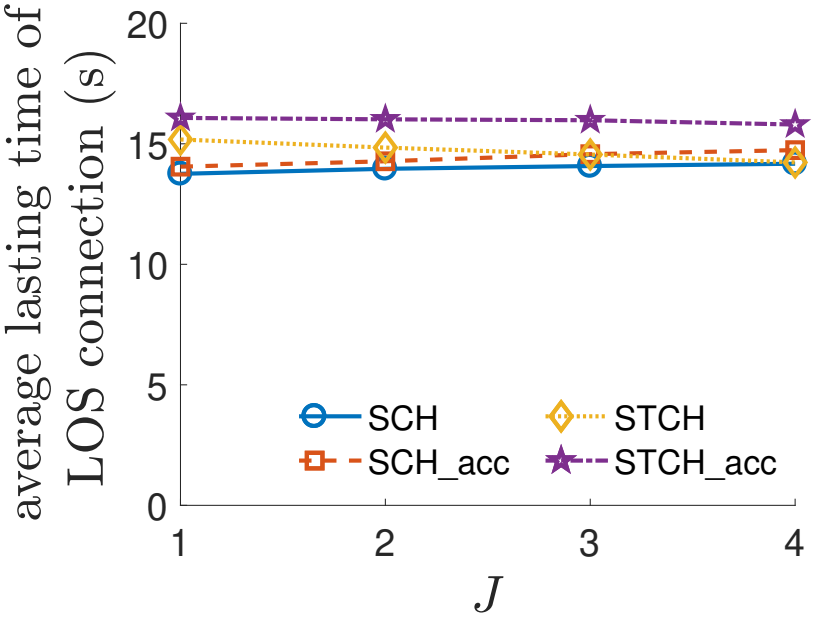
3.9.7 Regulation of UE's Movement

Unlike in the simulations mentioned above, in many real-world scenarios, especially in urban areas, the UE's movement is highly restricted. Specifically, since the UEs normally move along with sidewalks, the UE's position is actually limited within the area of a sidewalk but not the whole network, and the UE's moving direction is along with the sidewalk. In this simulation, I investigate the impact of the regulation of UE's movement on the performances of the proposed handover strategies. Following the grid-based scenario according to [68], I divide the whole network into 400 identical square areas each with the size of 5(m) $\times$ 5(m), as shown in Fig. 3.11. Note that these square areas have nothing to do with the blocks mentioned before. In this grid-based network, I set four sidewalks which are indicated by grey areas within which all UEs move. In particular, the UEs enter the network area from one end of a sidewalk and their moving directions are indicated by the corresponding yellow arrows. I deploy 80 SBSs in the network, represented by blue circles, and each of them locates at the center of a small area along the sidewalks. This deployment simulates the sidewalks with street lamps equipped with SBSs. Besides, 20 obstacles are randomly distributed on the sidewalks, represented by red triangles.

In order to express the variance of UE's moving direction, I introduce a parameter $\gamma \in [0.1, 0.5]$ to describe the homogeneity of UEs' movement within the sidewalks. In particular, if $\gamma = 0.1$, ten percent of UEs in the same sidewalk move in the same direction while ninety percents move oppositely; if $\gamma = 0.5$, half of the UEs in the same sidewalk move oppositely. Besides, I choose different values for $SNR_{\min}$: 18, 19, and 20, and run 10000 iterations for each instance with distinct combination of γ and $SNR_{\min}$. The simulation



(a) Performance on the average number of handovers per UE



(b) Performance on the average lasting time per LOS connection

Figure 3.9: Comparison of performances with different numbers of quantizing thresholds

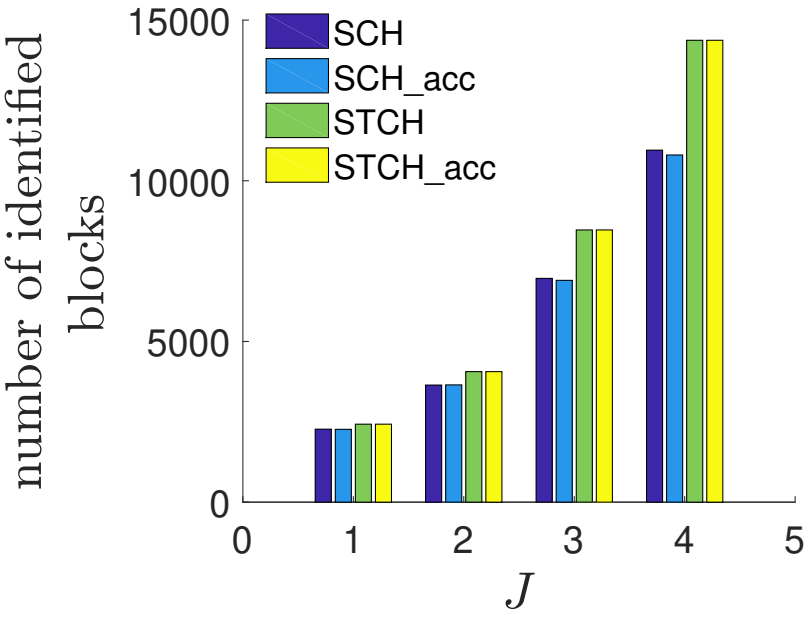


Figure 3.10: Increase of the number of identified blocks

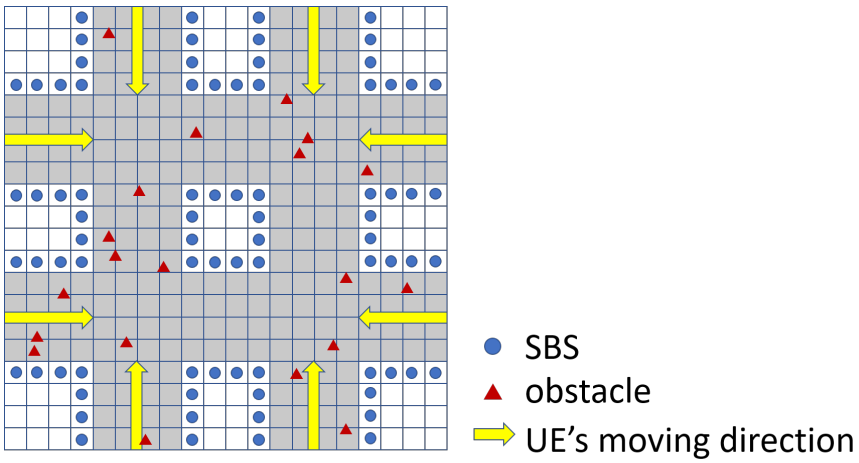


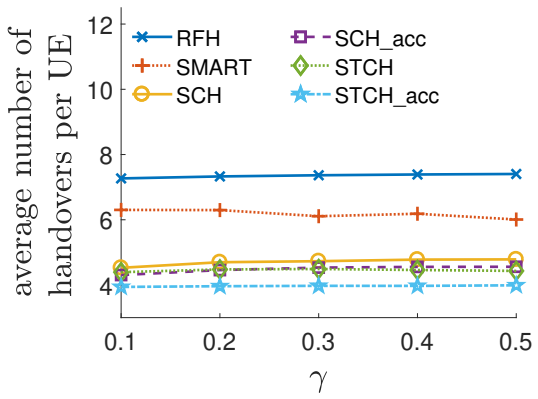
Figure 3.11: Scenario with sidewalks

results are shown in Fig. 3.12. Not surprisingly, the proposed contextual handover strategies are superior to the other two, no matter in the scenario with regular movement or the one with the irregular movement of UE. More importantly, the performance of the proposed strategies does not change significantly when UEs' movement becomes irregular. That means the proposed strategies have stable performances in real-world scenarios.

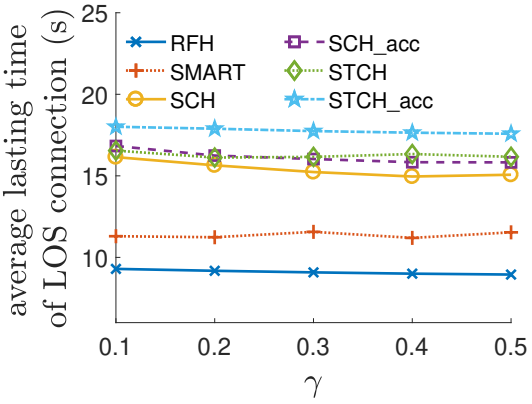
Moreover, I am also interested in how the performances of the proposed handover strategies change with different requirements of service level, i.e., the values of $SNR_{\min}$. For the the ANH, by comparing Fig. 3.12(a), 3.12(c) and 3.12(e), I found that the performances of the proposed strategies when $SNR_{\min} = 18$ is better than those when $SNR_{\min} = 20$. It is because, with a lower SNR requirement, the candidate SBSs for a handover event are more than those with a higher requirement. Generally, it is able to make better handover decisions given more candidate SBSs. The same conclusion could be made on the ALT by comparing Fig. 3.12(b), 3.12(d) and 3.12(f). In addition, the performance difference among the proposed strategies becomes smaller when $SNR_{\min}$ increases. Specifically, when $SNR_{\min} = 18$, the performance differences between the best strategy, i.e., STCH_acc, and the worst one, i.e., SCH, is 16.6% and 16.7% on the ANH and the ALT, respectively. However, when $SNR_{\min}$ increases to 20, these two differences significantly reduce to 5.0% and 4.3%, respectively. That means the advantage of temporal context could be fully exploited with a low requirement of service level.

3.10 Summary

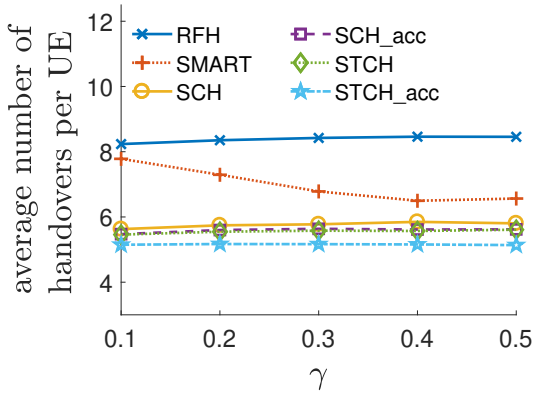
In this chapter, I investigate the significant benefits of exploiting the spatial and temporal features of handover in making better handover decisions in ultra-dense mmWave cellular networks. In particular, with the purpose to choose the optimal SBS that is able to provide the longest LOS connection time to a UE that needs handover, I formulate the handover decision-making as a MAB problem associated with the context of the handover. To extract the spatial and temporal features of handover, I propose a signal space partition scheme to



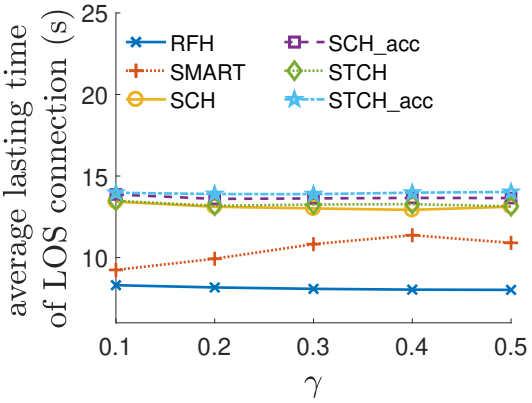
(a) Performance on the average number of handovers per UE ($SNR_{\min} = 18$)



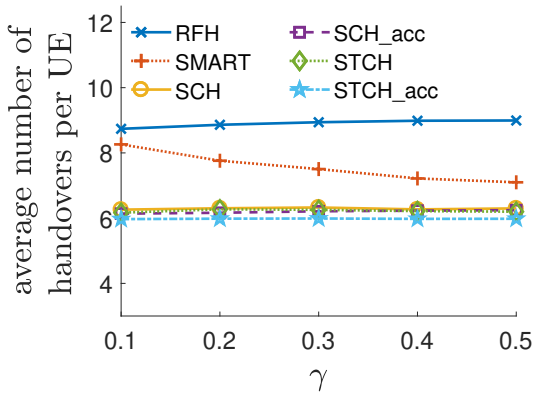
(b) Performance on the average lasting time per LOS connection ($SNR_{\min} = 18$)



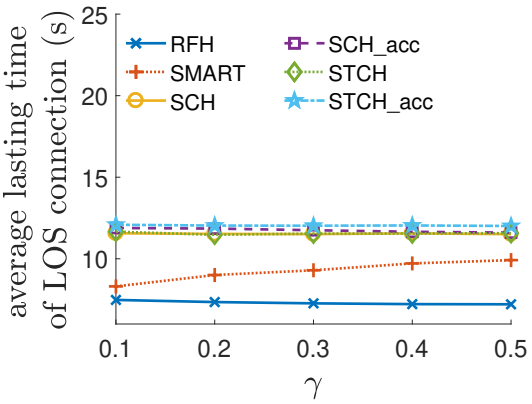
(c) Performance on the average number of handovers per UE ($SNR_{\min} = 19$)



(d) Performance on the average lasting time per LOS connection ($SNR_{\min} = 19$)



(e) Performance on the average number of handovers per UE ($SNR_{\min} = 20$)



(f) Performance on the average lasting time per LOS connection ($SNR_{\min} = 20$)

Figure 3.12: Comparison of performances on different moving directions of UE

indicate the location where the handover event happens, and a derived block concatenation structure to figure out the UE's moving trajectory. Based on these, an efficient BS-selection algorithm and a LinUCB-based BS-selection algorithm are designed for the spatial context-based MAB model and the temporal context-based MAB model, respectively. Moreover, I also propose an acceleration technique to improve the learning efficiency of the MAB model.

Chapter 4

Background: Part II

4.1 Motivation

In a cell-free massive MIMO system, a large number of distributed antennas or access points (APs) jointly serve the user equipments (UEs) in the network. It has become a promising solution in 5G and 6G communication due to its ability to reduce inter-cell interference and mitigate frequent handover by providing uniform quality of service (QoS) to UEs. However, one of the main issues of the cell-free massive MIMO system is the inter-user interference caused by the massive distributed APs. To address this issue, local zero-forcing (LZF) precoding [9, 28] and global zero-forcing (GZF) precoding [46, 47] are introduced into the cell-free massive MIMO domain. In the LZF, an AP equipped with multiple antennas conducts zero-forcing (ZF) precoding for downlink transmission to the UEs it serves. Due to the lack of collaboration with other APs, the LZF only eliminates the local inter-user interference but not the interference that comes from other APs. In the GZF, a centralized processing unit (CPU) collects the channel state information (CSI) from all APs and conducts the downlink precoding for them. Since the CPU has full knowledge of CSI in the network, it is able to make a better precoding decision than the LZF. However, this benefit comes with the cost of huge overhead due to the heavy CSI exchange between the CPU and the APs, which wastes bandwidth resources and aggravates transmission delay. The overhead issue has become a bottleneck that prohibits the GZF to be widely applied in ultra-dense networks (UDNs).

Overhead compression is an effective way to address this issue. Although many solutions have been developed to compress the CSI feedback in massive MIMO, they are not tailored for cell-free massive MIMO systems. On the one hand, most of them are built for conventional

massive MIMO systems considering the channels to be Rayleigh fading due to a large amount of reflection and scattering. On the other hand, these methods are statistic-based and focus on exploring the statistical characteristics of the channel matrix. However, in a cell-free massive MIMO system, channels are more suitable to be modeled as Rician fading because the APs are close to the UEs and little obstacles exist on the propagation path [49]. This propagation characteristic makes the channels in cell-free massive MIMO have remarkably different physical structures from the channels in conventional massive MIMO. Therefore, a model-based CSI compression method that considers and fully exploits this unique physical structure of channels is able to reduce the overhead in cell-free massive MIMO more efficiently than the statistic-based methods but is overlooked in the literature.

Unlike Rayleigh fading channels, a Rician fading channel consists of two components: a line-of-sight (LOS) component and a non-line-of-sight (NLOS) component. Compared with the NLOS component that is caused by multi-path reflection and scattering, the LOS component is more deterministic because it relies on the direct propagation path between the transmitter and the receiver, and contributes the majority of signal power since it suffers less energy loss. By leveraging this characteristic, a straightforward idea is to apply different overhead-reduction operations on these two parts separately, rather than directly compressing the original channel matrix, in order to shrink the overhead. In particular, the main component can be calculated based on the information on the direct propagation path, and the residual component can be compressed using a matrix compression method. Therefore, there are two key problems that need to be solved: (1) how to decompose a Rician fading channel matrix into two components: the main component and a residual component, respectively; and (2) how to compress these two components effectively.

In this chapter, I propose a novel bandwidth-efficient GZF precoding strategy in order to suppress interference in downlink transmission for cell-free massive MIMO systems, associated with a model-based CSI overhead-reduction mechanism customized for Rician fading channels. Specifically, the AP exploits side information of direct path, i.e., the signal phase

and the angle of arrival (AoA), to decompose a Rician fading channel into a LOS component and an NLOS component, and then conducts model-based compression and singular-value-decomposition (SVD) on them, respectively.

4.2 System Model

I consider a cell-free massive MIMO system consisting of one CPU, L APs equipped with N antennas and K single-antenna UEs. Denote $\mathcal{L}$, and $\mathcal{K}$ as the set of APs and that of UEs, respectively. These APs jointly serve the UEs and communicate with the CPU through fronthaul links with limited capacity. Throughout this chapter, I use upper and lower case boldfaced letters to describe matrix and vector, respectively. Moreover, I denote the transpose of matrix by $(\cdot)^{\mathrm{T}}$, the Hermitian transpose of matrix by $(\cdot)^{\mathrm{H}}$, and the l_2-norm of vector by $\|\cdot\|$.

4.2.1 Propagation Model

The channel matrix of the AP $l \in \mathcal{L}$ is denoted by $\mathbf{G}_l = [\mathbf{g}_{l1}, \mathbf{g}_{l2}, \ldots, \mathbf{g}_{lK}] \in \mathbb{C}^{N\times K}$. Each column $\mathbf{g}_{lk} \in \mathbb{C}^{N\times 1}$ denotes the channel vector between the AP l and the UE $k \in \mathcal{K}$. Each element of $\mathbf{g}_{lk}$ is modeled as a Rician fading channel, which is a combination of a dominant LOS component and a NLOS component following Rayleigh distribution. The channel vector $\mathbf{g}_{lk}$ is represented as

$$\mathbf{g}_{lk} = \sqrt{\beta_{lk}} \left(\sqrt{\frac{K_{lk}}{K_{lk}+1}} \bar{\mathbf{h}}_{lk} + \sqrt{\frac{1}{K_{lk}+1}} \check{\mathbf{h}}_{lk} \right), \tag{4.1}$$

where β_{lk} is the large-scale fading coefficient and K_{lk} is the K-factor that represents the ratio of signal power in the dominant component over the scattered power on the corresponding channel. $\bar{\mathbf{h}}_{lk}$ and $\check{\mathbf{h}}_{lk}$ denote the LOS and the NLOS components, respectively.

The value of β_{lk} depends on the distance d_{lk} between the AP l and the UE k, and is modeled as [2]

$$\beta_{lk} = \alpha + 10\rho \log_{10}(d_{lk}) + \xi[dB], \xi \sim \mathbb{N}(0, \sigma_{\mathrm{S}}^2), \tag{4.2}$$

where α and ρ are the least square fittings of floating intercept and slope respectively over the measured distance, and ξ represents a lognormal shadowing with variance σ_{S}^2. Since β_{lk} does not change frequently, I assume that it is known at both CPU and AP. The large-scale fading coefficients for the LOS and the NLOS components are computed as $\bar{\beta}_{lk} = \frac{K_{lk}}{K_{lk}+1}\beta_{lk}$ and $\check{\beta}_{lk} = \frac{1}{K_{lk}+1}\beta_{lk}$, respectively.

The LOS components $\bar{\mathbf{h}}_{lk}$ are modeled as

$$\bar{\mathbf{h}}_{lk} = \left[e^{j\varphi_{lk}^{(1)}}, e^{j\varphi_{lk}^{(2)}}, \ldots, e^{j\varphi_{lk}^{(N)}}\right]^{\mathrm{T}}, \tag{4.3}$$

where $\varphi_{lk}^{(n)}, 1 \leq n \leq N$ is the instantaneous phase of the signal received by the nth antanna at the AP l from the UE k. The elements in the NLOS components $\check{\mathbf{h}}_{lk}$ are modeled as i.i.d. random variables drawn from $\mathbb{CN}(\mathbf{0}, \mathbf{R}_{lk})$, where $\mathbf{R}_{lk}$ describes the spatial correlation of the NLOS components.

4.2.2 User-Centric Association

The user-centric association is an inherent characteristic of a cell-free massive MIMO system, in which each UE is only served by surrounding APs, rather than all APs in the network. To implement the user-centric association, it is assumed that the AP l serves the UE k when their propagation distance is smaller than a predefined threshold $\bar{d}$. I denote the set of APs that serve the UE k as $\mathcal{L}(k)$ and the set of UEs that are served by the AP l as $\mathcal{K}(l)$. Note that other criteria such as the signal strength can also be used to determine the serving range of the APs, and this would not affect the design and performance of the precoding strategy proposed below.

4.2.3 Communication Process

The transmission between APs and UEs is supposed to use in TDD mode and synchronization among all APs is assumed. Specifically, each coherence block consists of three stages: uplink training, CSI uploading, and downlink transmission.

(1) Uplink Training

In the uplink training stage, all UEs simultaneously send their assigned pilot signals to APs. Upon receiving the copies of these pilots, each AP utilizes a well-designed channel estimation method, such as MMSE estimation, to estimate the channels between itself and the UEs. In this chapter, I assume that the AP $l \in \mathcal{L}$ somehow is able to have perfect CSI, i.e., the channel matrix $\mathbf{G}_l$, and focus on the GZF precoding strategy involved in the subsequent stages.

(2) CSI Uploading

After channel estimation, the AP l compresses its local channel matrix $\mathbf{G}_l$ and uploads the compressed CSI to the CPU through fronthaul link. Then the CPU recovers the original channel matrix $\mathbf{G}_l$ based on the compressed CSI and gets the estimation $\tilde{\mathbf{G}}_l$. The details will be discussed in Section 4.3.

(3) Downlink Transmission

Based on the recovered channel matrix $\tilde{\mathbf{G}} = \left[\tilde{\mathbf{G}}_1, \tilde{\mathbf{G}}_2, \ldots, \tilde{\mathbf{G}}_L\right]^{\mathrm{T}} \in \mathbb{C}^{LN \times K}$, the CPU computes the GZF precoding matrix for downlink transmission. In global predocing, the whole cell-free massive MIMO system can be treated as a virtual AP with LN antennas. For the UE $k \in \mathcal{K}$, the normalized precoding vector $\mathbf{w}_k \in \mathbb{C}^{LN \times 1}$ is calculated as

$$\mathbf{w}_k = \frac{\mathbf{f}_k}{\|\mathbf{f}_k\|_F^2}, \tag{4.4}$$

where $\mathbf{F} = [\mathbf{f}_1, \ldots, \mathbf{f}_K] = \tilde{\mathbf{G}} / \left(\tilde{\mathbf{G}}^{\mathrm{H}} \tilde{\mathbf{G}} \right)$. The collection of the precoding vectors for all UEs is $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_K] = [\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_L]^{\mathrm{T}}$, where $\mathbf{W}_l \in \mathbb{C}^{N \times K}$ is the precoding matrix for the AP $l \in \mathcal{L}$. The CPU sends back the precoding matrix $\mathbf{W}_l, l \in \mathcal{L}$ to the AP l and then the APl calculates the transmitted signal as

$$\mathbf{x}_l = \sum_{k \in \mathcal{K}(l)} \sqrt{\eta_k} \mathbf{w}_{lk} q_k, \tag{4.5}$$

where $\mathbf{w}_{lk} \in \mathbb{C}^{N \times 1}$ is the kth column of $\mathbf{W}_l$, η_k denotes the power allocated to the UE k, and q_k is the data intended to the UE k, $\mathbb{E}\{|q_k|^2\} = 1$.

The received signal at the UE k is formulated as

$$y_k = \sum_{l \in \mathcal{L}(k)} \mathbf{g}_{lk}^{\mathrm{H}} \sqrt{\eta_k} \mathbf{w}_{lk} q_k + \sum_{\substack{t \in \mathcal{K} \\ t \neq k}} \sum_{l \in \mathcal{L}(t)} \mathbf{g}_{lk}^{\mathrm{H}} \sqrt{\eta_t} \mathbf{w}_{lt} q_t + n_k, \tag{4.6}$$

where the first item is the desired signal, the second stands for the interference, and $n_k \sim \mathbb{CN}(0, \sigma_{\mathrm{N}}^2)$ is the additive white Gaussian noise at the UE k.

4.2.4 Performance Metrics

In this section, the GZF precoding strategy is evaluated based on two key performance metrics: spectral efficiency and upload overhead.

(1) Spectral Efficiency

Based on Eq. (4.6), the SINR at the UE k is calculated as

$$SINR_k^{\mathrm{G}} = \frac{\left| \sum_{l \in \mathcal{L}(k)} \mathbf{g}_{lk}^{\mathrm{H}} \sqrt{\eta_k} \mathbf{w}_{lk} \right|^2}{\sum_{\substack{t \in \mathcal{K} \\ t \neq k}} \left| \sum_{l \in \mathcal{L}(t)} \mathbf{g}_{lk}^{\mathrm{H}} \sqrt{\eta_t} \mathbf{w}_{lt} \right|^2 + \sigma_{\mathrm{N}}^2}. \tag{4.7}$$

The achievable data rate for the UE k is calculated as

$$r_k^{\mathrm{G}} = \log_2\left(1 + SINR_k^{\mathrm{G}}\right). \tag{4.8}$$

(2) Upload Overhead

In the GZF precoding, all APs upload their local channel matrices to the CPU through fronthaul links. If each AP uploads its full channel matrix, the total overhead is $2L \times N \times K$ (consider each element in a channel matrix as a complex number containing a real part and an imaginary part). Considering the large amount of APs in the network, the consequent overhead is too huge to be affordable and causes unacceptable transmission delays. Therefore, the upload overhead should be considered an important metric to evaluate the performance of a precoding strategy.

4.3 Bandwidth-Efficient Global ZF Precoding Strategy

In order to improve the efficiency of GZF precoding in cell-free massive MIMO systems, I propose a bandwidth-efficient precoding strategy, by exploiting the physical structure of Rician fading channels, to address the upload overhead challenge. This strategy consists of two parts: the AP-side operation and the CPU-side operation, where the CSI is compressed and recovered, respectively. The details are discussed as follows.

4.3.1 AP-Side Operation

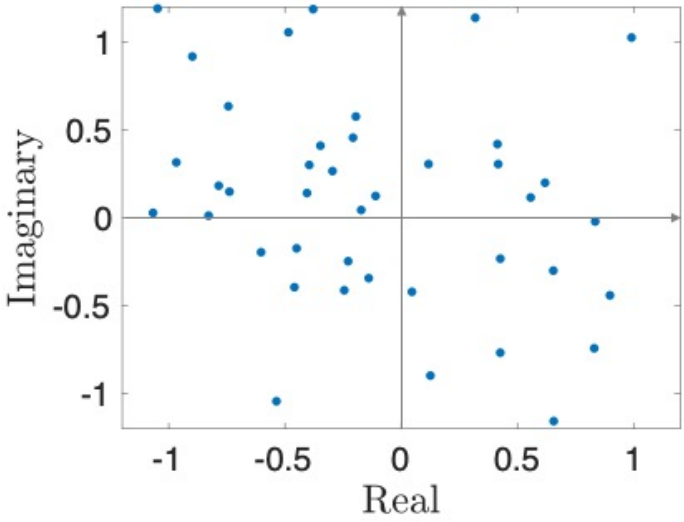
On the AP side, the AP compresses its channel matrix before uploading it to the CPU. Unlike the conventional CSI compression methods that conduct compression directly on the original channel matrix, in the proposed strategy, the AP first decomposes the original channel matrix into the main matrix and a residual matrix, and then compresses them respectively.

(1) Rician Fading Channel Decomposition

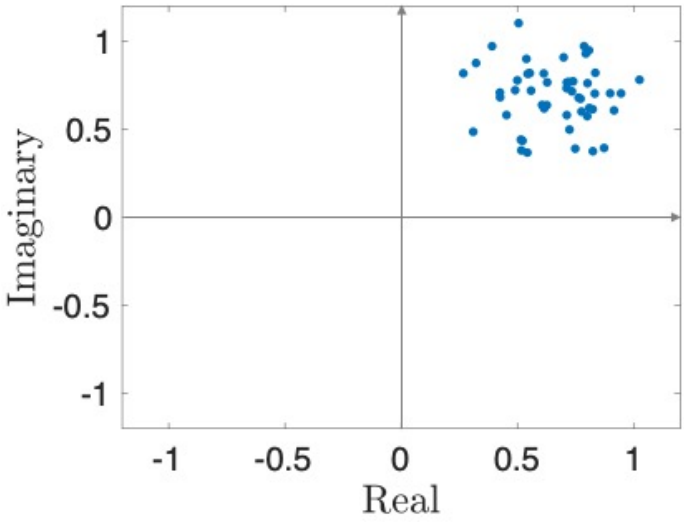
Different from a Rayleigh fading channel, a Rician fading channel exhibits higher directivity as a result of the LOS component that contributes the most of the signal power. To illustrate this directivity, I collect 50 samples of the channel gain for a Rayleigh fading channel and a Rician fading channel between two arbitrary pairs of transmitting and receiving antenna, respectively, and display them in Fig. 4.1. As shown in Fig. 4.1(a), all sample points of the Rayleigh fading channel are uniformly distributed in the whole signal space; while for the Rician fading channel, the sample points are constricted within a limited area, shown in Fig. 4.2(b). This characteristic makes it able to decompose a Rician fading channel matrix into the main and a residual matrix, which correspond to the LOS and the NLOS components, respectively. According to Eq. (4.1), the channel between the nth antenna of the AP l and the UE k can be decomposed as follows:

$$g_{lk}^{(n)} = \sqrt{\bar{\beta}_{lk}}\bar{h}_{lk}^{(n)} + \sqrt{\check{\beta}_{lk}}\check{h}_{lk}^{(n)}, \tag{4.9}$$

where $\bar{h}_{lk}^{(n)}$ is the deterministic LOS component that relies on the phase of the received signal at the receiving antenna, and $\check{h}_{lk}^{(n)}$ denotes the NLOS component that is modeled as an i.i.d. random variable. The relationship between these two components is described in Fig. 4.2(a). Given a Rician channel gain $g_{lk}^{(n)}$, if the LOS component $\bar{h}_{lk}^{(n)}$ is known, the NLOS component $\check{h}_{lk}^{(n)}$ can be easily calculated. In this way, the sample points of the Rician fading channel shown in Fig. 4.1(b) can be decomposed and the corresponding LOS component and NLOS components are shown in Fig. 4.2(b). Compared with directly compressing the original channel matrix $\mathbf{G}_l$, the information loss caused by compressing the residual matrix brings a smaller impact on CSI recovery since the NLOS component contributes only the minority of the signal power, i.e., its magnitude is rather small. The prerequisite for this implementation is to extract the LOS component from a Rician fading channel, as will be discussed in the following.

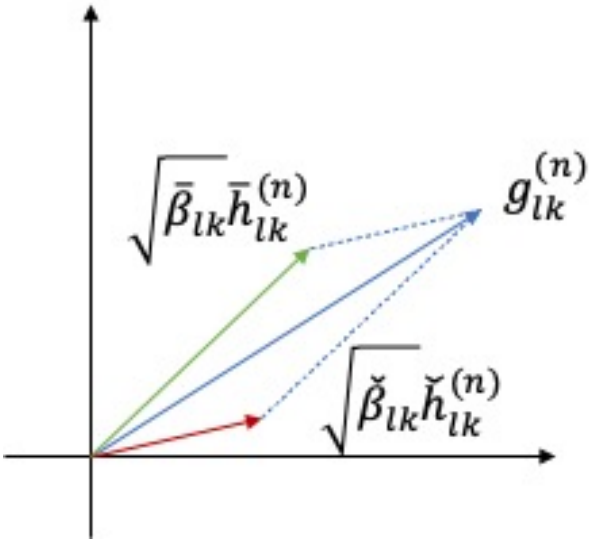


(a) Sample points of Rayleigh fading channel

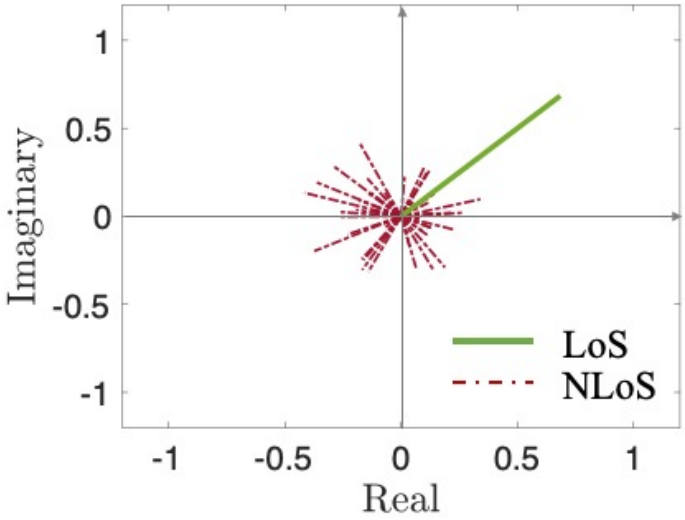


(b) Sample points of Rician fading channel

Figure 4.1: Spacial correlations of different fading types



(a) Decomposition of Rician fading channel



(b) Decomposition of the sample points in Fig. 4.1(b)

Figure 4.2: Rician fading channel decomposition

(2) LOS Components Extraction

Supposing that the antenna spacing of the horizontal uniform linear array (ULA) equipped on AP is $d_A = 1/2\lambda$, where λ is the wavelength. For any AP in the network, considering that the UEs are in the far-filed of the AP, the signal that reaches the antenna array at the AP from a UE can be viewed as a plane wave from a generic azimuth angle [8]. Denote the distance between the closest antenna (suppose to be the 1st antenna) of the AP l and the UE k as d, the LOS components $\bar{\mathbf{h}}_{lk}$ described in Eq. (4.3) can be further modeled as

$$\begin{aligned}\bar{\mathbf{h}}_{lk} &= \left[e^{j\varphi_{lk}^{(1)}}, \ldots, e^{j\varphi_{lk}^{(N)}}\right]^{\mathrm{T}} \\ &= \Big[e^{j\left(2\pi\frac{d}{\lambda}+\varphi_{lk}^{(0)}\right)}, e^{j\left(2\pi\frac{d+d_A\sin(\theta_{lk})}{\lambda}+\varphi_{lk}^{(0)}\right)}, \ldots, \\ &\quad e^{j\left(2\pi\frac{d+(N-1)d_A\sin(\theta_{lk})}{\lambda}+\varphi_{lk}^{(0)}\right)}\Big]^{\mathrm{T}},\end{aligned} \tag{4.10}$$

where θ_{lk} is the angle of arrival (AoA) from the UE k to the AP l and $\varphi_{lk}^{(0)}$ is the initial phase. If the AoA θ_{lk} and the signal phase $\varphi_{lk}^{(1)} = 2\pi\frac{d}{\lambda} + \varphi_{lk}^{(0)}$ at the 1st antenna are known, the phases at all antennas can be aligned and Eq. (4.10) can be simplified as [80]:

$$\bar{\mathbf{h}}_{lk} = e^{j\varphi_{lk}^{(1)}}\left[1, e^{j\pi\sin(\theta_{lk})}, \ldots, e^{j(N-1)\pi\sin(\theta_{lk})}\right]^{\mathrm{T}}. \tag{4.11}$$

In other words, the LOS components $\bar{\mathbf{h}}_{lk}$ becomes a function of θ_{lk} and $\varphi_{lk}^{(1)}$. In the rest of the paper, I use φ_{lk} instead by omitting the superscript of $\varphi_{lk}^{(1)}$. As illustrated in Fig. 4.3, the phase shifts between any two adjacent antennas are identical to be $\pi\sin(\theta_{lk})$. This property makes it enable to infer each element in $\bar{\mathbf{h}}_{lk}$ based on the given φ_{lk} and θ_{lk}. In this way, the LOS components $\bar{\mathbf{h}}_{lk}$ can be extracted from the observed channel vector $\mathbf{g}_{lk}$ and then the NLOS components $\check{\mathbf{h}}_{lk}$ is calculated as

$$\check{\mathbf{h}}_{lk} = \frac{1}{\sqrt{\check{\beta}_{lk}}}\left(\mathbf{g}_{lk} - \sqrt{\bar{\beta}_{lk}}\bar{\mathbf{h}}_{lk}\right) \tag{4.12}$$

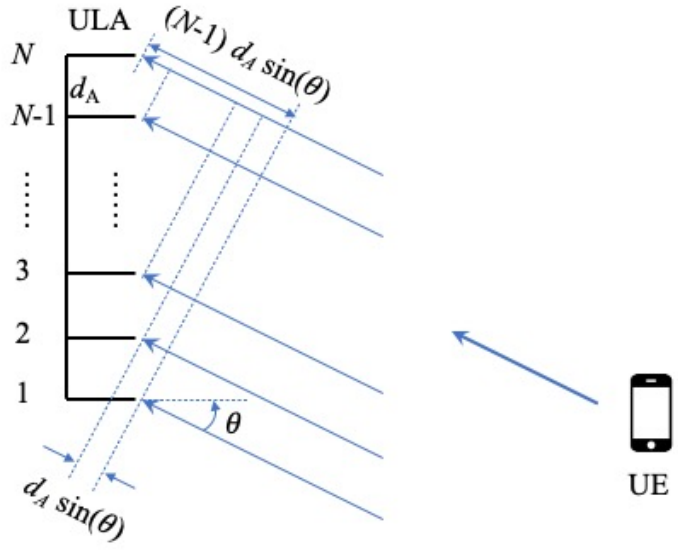


Figure 4.3: LOS propagation

(3) Side Information Calculation

Since the LOS components are determined by the signal phase and the AoA, they will change with these two key factors with the UE's movement. These changes are identical for all antennas at the same AP [49], and can be tracked as follows.

The first step is to estimate the AoA θ_{lk}, for $l \in \mathcal{L}$, $k \in \mathcal{K}$. There have been many efficient algorithms to estimate the AoA on ULA with omni-directional antenna elements, such as MUSIC and ESPRIT. In this section, I use MUSIC to estimate θ_{lk}. The details are omitted.

The next step is to calculate the signal phase φ_{lk}. I design the following two algorithms for different scenarios where the antenna array at AP is large or small, respectively.

Algorithm I: When the antenna array is large, I firstly calculate the mean value of channel vector $\mathbf{g}_{lk}$ as follow

$$\frac{1}{N}\sum_{1\le n\le N} g_{lk}^{(n)} = \frac{1}{N}\sqrt{\bar{\beta}_{lk}}\sum_{1\le n\le N}\bar{h}_{lk}^{(n)} + \frac{1}{N}\sqrt{\check{\beta}_{lk}}\sum_{1\le n\le N}\check{h}_{lk}^{(n)} \tag{4.13}$$

As a common assumption, the number of the antennas is assumed to be large enough, then $\frac{1}{N}\sum \check{h}_{lk}^{(n)} \to 0$, since $\check{\mathbf{h}}_{lk} \sim \mathbb{CN}(\mathbf{0}, \mathbf{R}_{lk})$. Hence,

$$\frac{1}{N}\sum_{1\le n\le N} g_{lk}^{(n)} \approx \frac{1}{N}\sqrt{\bar{\beta}_{lk}}\sum_{1\le n\le N} \bar{h}_{lk}^{(n)}. \tag{4.14}$$

According to Eq. (4.11),

$$\sum_{1\le n\le N} \bar{h}_{lk}^{(n)} = e^{j\varphi_{lk}}\left(1 + e^{j\pi\sin(\theta_{lk})} + \cdots + e^{j(N-1)\pi\sin(\theta_{lk})}\right). \tag{4.15}$$

Substituting Eq. (4.15) into Eq. (4.14), I get

$$\frac{1}{N}\sum_{1\le n\le N} g_{lk}^{(n)} \approx \frac{1}{N}\sqrt{\bar{\beta}_{lk}}e^{j\varphi_{lk}}\left(1 + \cdots + e^{j(N-1)\pi\sin(\theta_{lk})}\right). \tag{4.16}$$

As $\mathbf{g}_{lk}$ is known and θ_{lk} has already been calculated, there is only one variable in Eq. (4.16), which then can be rewritten as

$$e^{j\varphi_{lk}} = a + bi, \tag{4.17}$$

where a and b are the calculation results. According to Euler's formula and Taylor's series for trigonometric functions,

$$\begin{aligned} e^{j\varphi_{lk}} &= \cos(\varphi_{lk}) + i\sin(\varphi_{lk}) \\ &= 1 - \frac{\varphi_{lk}^2}{2} + o(\varphi_{lk}^2) + i\left(\varphi_{lk} - \frac{\varphi_{lk}^3}{3!} + o(\varphi_{lk}^3)\right) \\ &\approx 1 - \frac{\varphi_{lk}^2}{2} + i\left(\varphi_{lk} - \frac{\varphi_{lk}^3}{3!}\right). \end{aligned} \tag{4.18}$$

Combining Eq. (4.17) and Eq. (4.18), the signal phase φ_{lk} is calculated by solving the following optimization problem

$$\text{minimize}_{\varphi_{lk}}\left(1 - \varphi_{lk}^2/2 - a\right)^2 + \left(\varphi_{lk} - \varphi_{lk}^3/6 - b\right)^2. \tag{4.19}$$

Algorithm II: When the antenna array is small, I treat the LOS component as a function of φ_{lk} according to Eq. (4.11), denoted by $\bar{\mathbf{h}}_{lk}(\varphi_{lk})$. Then the optimal φ_{lk} can be estimated by solving the following MMSE problem

$$\text{minimize}_{\varphi_{lk}} \ \|\sqrt{\bar{\beta}_{lk}}\bar{\mathbf{h}}_{lk}(\varphi_{lk}) - \mathbf{g}_{lk}\|^2. \tag{4.20}$$

The above two optimization problems can be easily solved. The collection of the calculated side information, i.e., $\boldsymbol{\varphi}_l = [\varphi_{l1}, \ldots, \varphi_{l,K}]$ and $\boldsymbol{\theta}_l = [\theta_{l1}, \ldots, \theta_{l,K}]$, is the compressed CSI of the LOS components $\bar{\mathbf{H}}_l = \left[\bar{\mathbf{h}}_{l1}, \ldots, \bar{\mathbf{h}}_{l,K}\right]$.

(4) Compression of NLOS Components

Given the obtained NLOS components, denoted as $\check{\mathbf{H}}_l = \left[\check{\mathbf{h}}_{l1}, \check{\mathbf{h}}_{l2}, \ldots, \check{\mathbf{h}}_{l,K}\right]$, according to Eq. (4.12), the AP l applies the SVD technique to compress it. Specifically, for the given channel matrix $\check{\mathbf{H}}_l$ of the AP l, its SVD is calculated as

$$\check{\mathbf{H}}_l = \mathbf{U}_l\mathbf{S}_l\mathbf{V}_l^{\mathrm{H}}, \tag{4.21}$$

where $\mathbf{U}_l = [\mathbf{u}_{l1}, \ldots, \mathbf{u}_{l,R}] \in \mathbb{C}^{N\times R}$ and $\mathbf{V}_l = [\mathbf{v}_{l1}, \ldots, \mathbf{v}_{l,R}] \in \mathbb{C}^{K\times R}$ are semi-unitary matrices that contain the left-singular vectors and the right-singular vectors of $\check{\mathbf{H}}_l$, respectively, with $R = \text{rank}(\check{\mathbf{H}}_l)$, and $\mathbf{S}_l = \text{diag}\,(s_{l1}, \ldots, s_{l,R}) \in \mathbb{C}^{R\times R}$ is a diagonal matrix whose diagonal elements $s_{li} > 0$ $(1 \le i \le R)$ are the singular values of $\check{\mathbf{H}}_l$ and sorted in descending order. Since the top largest singular values contain the most information, the rest can be discarded for the purpose of compression. That being said, the AP l is able to keep the most significant information of $\check{\mathbf{H}}_l$ by choosing the first $M(M \le R)$ singular values $\tilde{\mathbf{S}}_l = \text{diag}\,(s_{l1}, \ldots, s_{l,M})$ and the corresponding semi-unitary matrices, i.e., $\tilde{\mathbf{U}}_l = [\mathbf{u}_{l1}, \ldots, \mathbf{u}_{l,M}]$ and $\tilde{\mathbf{V}}_l = [\mathbf{v}_{l1}, \ldots, \mathbf{v}_{l,M}]$, to be the compressed CSI of the NLOS component. Together with the compressed CSI of $\bar{\mathbf{H}}_l$, the total volume of the uploaded data is $2M(N + K + 1) + 2K$. Compared with uploading $\mathbf{G}_l$ with the volume of $2 \times N \times K$, the proposed compression

method is able to save the bandwidth of the fronthaul links when $M \leq \lfloor \frac{(N-1)K}{N+K+1} \rfloor$, with a compression ratio of $\frac{M(N+K+1)+K}{NK}$.

4.3.2 CPU-Side Operation

On the CPU side, the CPU receives the compressed CSI from each AP and then recovers the original local channel matrix. After that, the CPU implements the global ZF precoding and power allocation, and then sends the result back to all APs.

(1) CSI Recovery

Upon receiving the compressed CSI uploaded from all APs, the CPU recovers the whole channel matrix $\mathbf{G}$ in the following steps.

Step 1: $\check{\mathbf{H}}_l$ recovery. Based on $\tilde{\mathbf{S}}_l$, $\tilde{\mathbf{U}}_l$ and $\tilde{\mathbf{V}}_l$, $\check{\mathbf{H}}_l$ can be recovered as

$$\tilde{\mathbf{H}}_l = \tilde{\mathbf{U}}_l \tilde{\mathbf{S}}_l \tilde{\mathbf{V}}_l^{\mathrm{H}}. \tag{4.22}$$

Step 2: $\bar{\mathbf{H}}_l$ recovery. Based on φ_{lk} and θ_{lk}, $\bar{\mathbf{h}}_l$ can be calculated according to Eq. (4.11). Then the collection of $\bar{\mathbf{h}}_{lk}$, $k \in \mathcal{K}$, is $\bar{\mathbf{H}}_l = \left[\bar{\mathbf{h}}_{l1}, \ldots, \bar{\mathbf{h}}_{l,K}\right]$.

Step 3: $\mathbf{G}_l$ recovery. Upon obtaining $\bar{\mathbf{h}}_{lk}$ and $\tilde{\mathbf{h}}_{lk}$ from $\bar{\mathbf{H}}_l$ and $\tilde{\mathbf{H}}_l$, respectively, the channel vector $\mathbf{g}_{lk}$ can be recovered as

$$\tilde{\mathbf{g}}_{lk} = \sqrt{\bar{\beta}_{lk}} \bar{\mathbf{h}}_{lk} + \sqrt{\check{\beta}_{lk}} \tilde{\mathbf{h}}_{lk}. \tag{4.23}$$

And $\tilde{\mathbf{G}}_l = [\tilde{\mathbf{g}}_{l1}, \ldots, \tilde{\mathbf{g}}_{l,K}]$ is used as the recovery of $\mathbf{G}_l$.

Step 4: $\mathbf{G}$ recovery. Collecting all $\mathbf{G}_l$, $l \in \mathcal{L}$, I get the recovery of $\mathbf{G}$ as $\tilde{\mathbf{G}} = \left[\tilde{\mathbf{G}}_1, \ldots, \tilde{\mathbf{G}}_L\right]^{\mathrm{T}}$.

(2) GZF Precoding

After obtaining $\tilde{\mathbf{G}}$, the CPU computes the precoding matrix $\mathbf{W}$ according to Eq. (4.4).

(3) Power Allocation

Given the estimated channel matrix $\tilde{\mathbf{G}}$ and the precoding matrix $\mathbf{W}$, the CPU sets the transmit power for each UE. The optimal transmit power allocation can be determined by solving the following optimization problem

$$\begin{aligned} &\text{maximize}_{\boldsymbol{\eta}} \sum_{k\in\mathcal{K}} \log_2\left(1+SINR_k^{\mathrm{G}}\right) \\ &\text{s.t. Eq.(4.7)}, \forall k \in \mathcal{K}, \\ &\sum_{k\in\mathcal{K}(l)} \eta_k \leq P_l, \forall l \in \mathcal{L}, \\ &\eta_k \geq 0, \forall k \in \mathcal{K}, \end{aligned} \tag{4.24}$$

where $\boldsymbol{\eta} = [\eta_1, \ldots, \eta_K]$ and P_l is the maximum power that the AP l can provide. The power constraints require that the total transmitted power from an AP to its served UEs can not exceed its power capacity. The above power allocation problem is modeled as non-linear programming and could be solved by using the interior-point algorithm offered by the Matlab toolbox.

4.4 Simulations and Analysis

This section is to evaluate the performances of the proposed GZF precoding strategy in various scenarios. I compare it with three counterparts: GZF precoding with fully uploading (GZF-FU), GZF precoding with preliminary SVD (GZF-PSVD), and LZF precoding.

The communication process in both the GZF-FU and the GZF-PSVD strategies are similar to that in my strategy, except for the following difference. In the GZF-FU strategy, each AP uploads its original local channel matrix; and in the GZF-PSVD strategy, the SVD is conducted directly on the original local channel matrix without channel decomposition as in my strategy.

In the LZF precoding strategy, all APs implement downlink precoding independently based on their own local channel matrices without any information exchanging with the CPU or other APs. Specifically, in the LZF precoding, the AP $l \in \mathcal{L}$ computes its normalized precoding vector as

$$\mathbf{w}_{lk} = \frac{\mathbf{f}_{lk}}{\|\mathbf{f}_{lk}\|_F^2}, \tag{4.25}$$

where $\mathbf{F}_l = [\mathbf{f}_{l1}, \ldots, \mathbf{f}_{lK}] = \mathbf{G}_l / \left(\mathbf{G}_l^{\mathrm{H}} \mathbf{G}_l\right)$. The AP $l \in \mathcal{L}$ independently allocates its power to its served UEs by solving the following optimization problem

$$\begin{aligned} \text{maximize}_{\boldsymbol{\eta}_l} \quad & \sum_{k \in \mathcal{K}(l)} \log_2\left(1 + SINR_{lk}^{\mathrm{L}}\right) \\ \text{s.t.} \quad & SINR_{lk}^{\mathrm{L}} = \left|\mathbf{g}_{lk}^{\mathrm{H}} \sqrt{\eta_{lk}} \mathbf{w}_{lk}\right|^2 / \sigma_{\mathrm{N}}^2, \forall k \in \mathcal{K}(l), \\ & \sum_{k \in \mathcal{K}(l)} \eta_{lk} \leq P_l, \\ & \eta_{lk} \geq 0, \forall k \in \mathcal{K}(l), \end{aligned} \tag{4.26}$$

where $\mathbf{g}_{lk} \in \mathbb{C}^{N \times 1}$ indicates the channel vector between the AP l and the UE k, and η_{lt} denotes the power allocated to the UE k by the AP l. The optimal solution of the above problem is denoted as $\boldsymbol{\eta}_l^* = [\eta_{l1}^*, \ldots, \eta_{lK}^*]$. When all APs conduct the downlink transmission with their optimal power allocations, the received signal at the UE k is

$$y_k = \sum_{l \in \mathcal{L}(k)} \mathbf{g}_{lk}^{\mathrm{H}} \sqrt{\eta_{lk}^*} \mathbf{w}_{lk} q_k + \sum_{\substack{t \in \mathcal{K} \\ t \neq k}} \sum_{l \in \mathcal{L}(t)} \mathbf{g}_{lk}^{\mathrm{H}} \sqrt{\eta_{lt}^*} \mathbf{w}_{lt} q_t + n_k. \tag{4.27}$$

The SINR at the UE k is calculated as

$$SINR_k^{\mathrm{L}} = \frac{\left|\sum_{l \in \mathcal{L}(k)} \mathbf{g}_{lk}^{\mathrm{H}} \sqrt{\eta_{lk}^*} \mathbf{w}_{lk}\right|^2}{\sum_{\substack{t \in \mathcal{K} \\ t \neq k}} \left|\sum_{l \in \mathcal{L}(t)} \mathbf{g}_{lk}^{\mathrm{H}} \sqrt{\eta_{lt}^*} \mathbf{w}_{lt}\right|^2 + \sigma_{\mathrm{N}}^2}, \tag{4.28}$$

and the achievable sum data rate of all UEs is $\sum_{k \in \mathcal{K}} \log_2\left(1 + SINR_k^{\mathrm{L}}\right)$.

4.4.1 Simulation Setup

In the following simulations, I consider a cell-free massive MIMO system consisting of 25 APs (i.e., $L = 25$), each of which is equipped with 20 antennas (i.e., $N = 20$), and 16 single-antenna UEs (i.e., $K = 16$). The APs and the UEs are uniformly distributed in a 100(m) $\times$ 100(m) square area. The UEs are moving in the area with random directions at the speed of 1(m)/iteration. In all simulations, I consider 10 iterations. The AP's serving range is set to be 50(m) (i.e., $\bar{d} = 50$) and the maximum transmit power is set to be 1. The noise power is $\sigma_{\mathrm{N}}^2 = -92$ dBm [10]. For Eq. (4.2), I consider the conventional 1900 MHz frequency band and set $\alpha = -30.18$, $\rho = -2.6$ and $\sigma_{\mathrm{S}} = 4$ [49]. The K-factor in Eq. (4.1) is modeled as [49]

$$K_{lk} = 10^{1.3-0.003d_{lk}}[dB], \tag{4.29}$$

where d_{lk} is the distance between the AP l and the UE k. The spatial correlation of the NLOS components $\mathbf{R}_{lk}$ is calculated according to [49], and the details are omitted in this chapter. I estimate the φ_{lk} and θ_{lk}, $l \in \mathcal{L}$ and $k \in \mathcal{K}$ according to Sec. 4.3.1, and implement the simulations as follows.

4.4.2 Compression Ratio

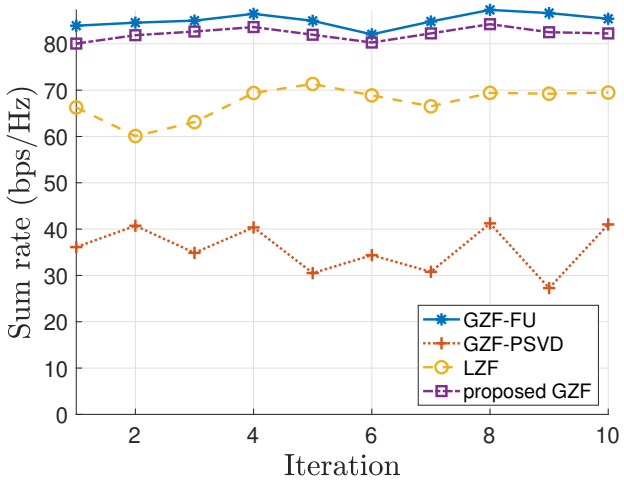
Since the number of singular values chosen is a key factor in SVD which determines the compression ratio and impacts the CSI recovery accuracy, I evaluate these strategies under different values of M in this simulation. The results of the spectral efficiencies and the overhead are shown in Fig. 4.4 and Fig. 4.5, respectively.

In Fig. 4.4, the spectral efficiencies of the considered strategies are compared when $M =$ 1, 2 and 3. It is illustrated that the GZF-FU precoding outperforms all other strategies in all scenarios since the CPU has full CSI knowledge and is able to make a better precoding decision to suppress interference. In the LZF strategy, since each AP makes its own precoding decision and has no collaboration with other APs, the interference suppression is not very

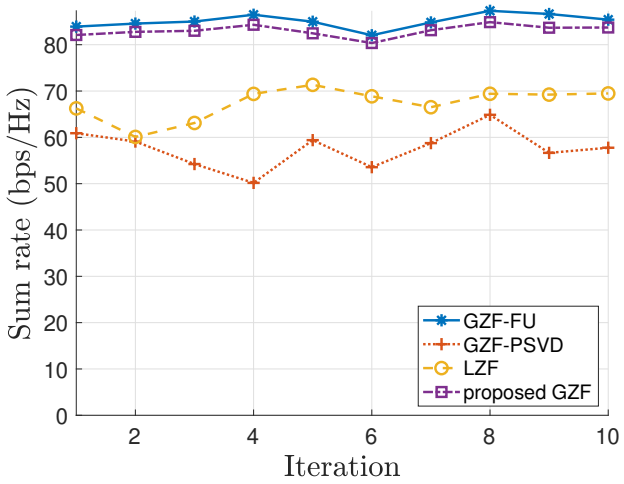
good. Although the GZF-PSVD is also a global precoding strategy, like the GZF-FU, its performance is even worse than the LZF's performance when $M = 1$, as shown in Fig. 4.4(a). The reason is that when only one singular value is used, very limited information is contained in the compressed CSI, which leads to low CSI recovery accuracy. It is worth noting that, compared with the GZF-PSVD, the proposed GZF precoding strategy significantly increases the spectral efficiency by 134.4% on average. Besides, my strategy is slightly sub-optimal, with an average gap of only 3.5% relative to the GZF-FU strategy. It is also observed in Fig. 4.5 that my strategy drastically cuts the overhead by 83.4% compared with the GZF-FU. That being said, the proposed precoding strategy can provide slightly suboptimal performance with much less overhead and hence can reach outstanding bandwidth efficiency.

Moreover, the spectral efficiencies of both the GZF-PSVD strategy and the proposed GZF strategy increase with the number of chosen singular values and the increasing rate of the former is larger than that of the latter. The reason is that, in the GZF-PSVD, the more singular values are chosen, the more information is kept in the compressed CSI. Therefore, the CPU can recover the channel matrix with higher accuracy and consequently make a better precoding decision. However, in my strategy, since the compressed CSI only contains the NLOS components that contribute little signal power, choosing more singular values does not bring obvious performance improvement as in the GZF-PSVD.

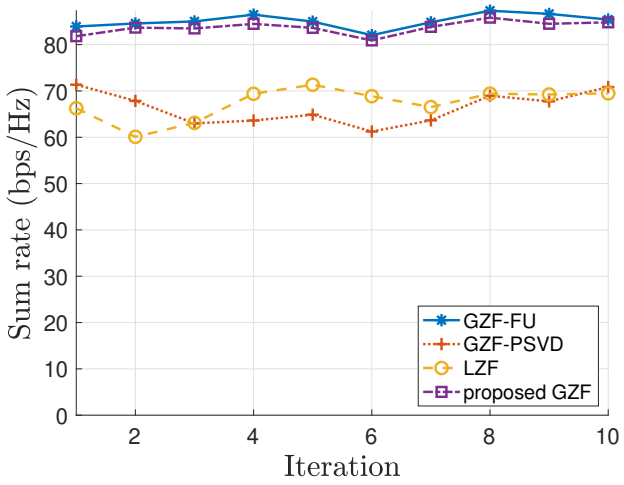
But there is no free lunch. For the GZF-PSVD, its performance is improved at the cost of significantly increased overhead. As displayed in Fig. 4.5, to achieve comparable performance with the LZF, the GZF-PSVD needs 3 singular values (i.e., $M = 3$). That needs three-fold overhead when $M = 1$. In contrast, the performance of my strategy when $M = 1$ is even higher than that of the GZF-PSVD when $M = 3$, but with much less overhead. This indicates that my strategy is able to achieve a sub-optimal performance with substantially low overhead.



(a) $M = 1$



(b) $M = 2$



(c) $M = 3$

Figure 4.4: Comparison of spectral efficiency with different values of M

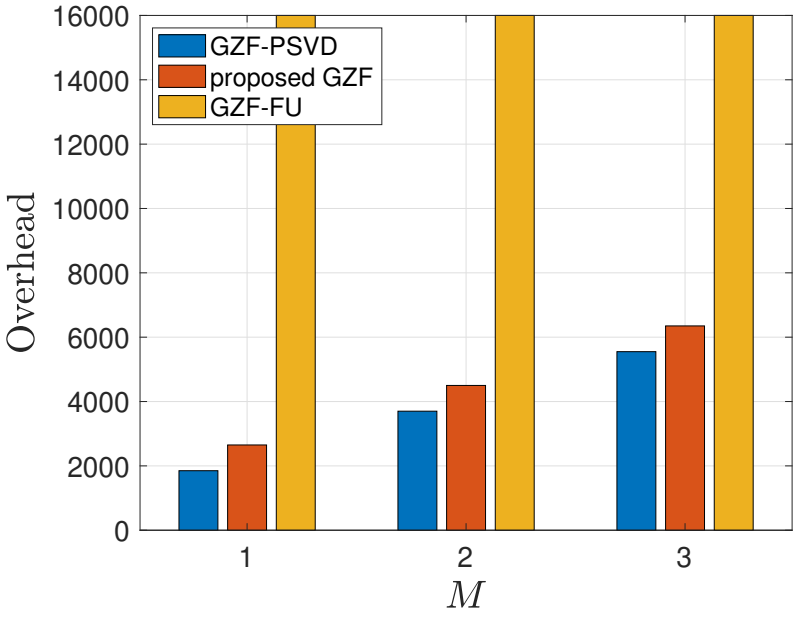


Figure 4.5: Comparison of overhead with different values of M

4.4.3 K-Factor

The K-factor indicates the degree of dominance of the LOS component over the NLOS component. A larger K-factor means that the LOS component contributes more power to channel gain. Especially, if the K-factor equals 0, there is no direct path between the transmitter and the receiver, meaning that the Rician fading channel decays to a Rayleigh fading channel. In this simulation, I evaluate the considered strategies in the scenario of Rayleigh fading channels. The simulation results when $M = 1$ are shown in Fig. 4.6. It is demonstrated that the performance of the proposed precoding strategy is identical to that of the GZF-PSVD strategy. It is because there is no LOS component in this case. That means, in Rayleigh fading channels, my strategy degrades to the GZF-PSVD whose performance is the worst. Therefore, this strategy is not suitable for Rayleigh fading channels.

4.4.4 Frequency Band

In this simulation, I evaluate the impact of the frequency band on the performance of the proposed precoding strategy. Different from the setting in Sec. 4.4.2, I choose another set of

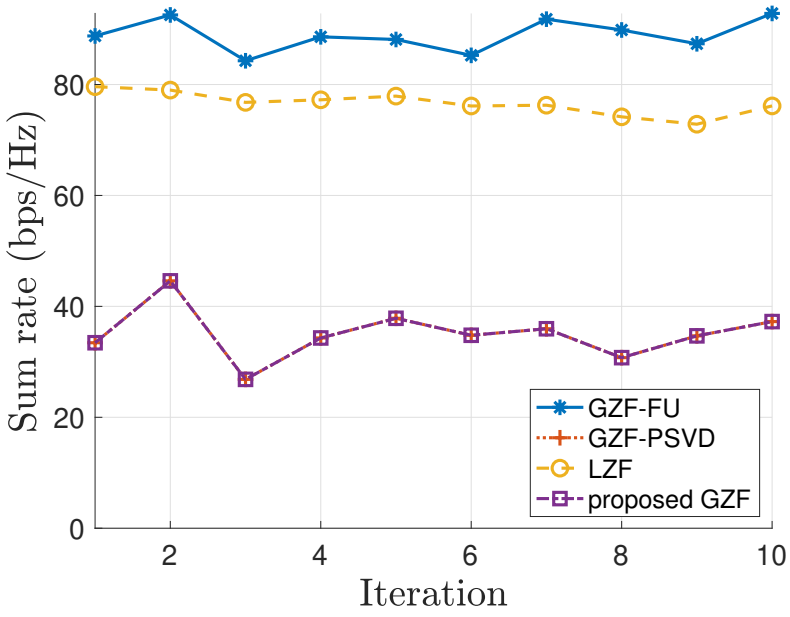


Figure 4.6: Comparison of spectral efficiency when the K-factor = 0

path loss parameters, i.e., $\alpha = -61.4$, $\rho = -2$ and $\sigma_S = 5.8$ for mmWave 28 GHz frequency band [2], and set $M = 1$. The simulation results are shown in Fig. 4.7. Comparing Fig. 4.4(a) with Fig. 4.7, it could be found that, in the conventional band, my strategy achieves 96.5% of the spectral efficiency of the GZF-FU strategy on average; and in the mmWave band, this proportion is 96.7%, showing that the proposed strategy has similar performances in these two frequency bands and performs well in both scenarios. This indicates the great flexibility and reliability of my strategy when being applied to various application scenarios with different frequency bands.

4.5 Summary

In this chapter, I propose a bandwidth-efficient GZF precoding strategy for cell-free massive MIMO systems considering Rician fading channels. In order to reduce the CSI overhead on fronthaul links, I propose to decompose the channel matrix into two components by exploiting the physical structure of Rician fading channels and then design two corresponding efficient compression methods. Moreover, two effective algorithms are developed to calculate

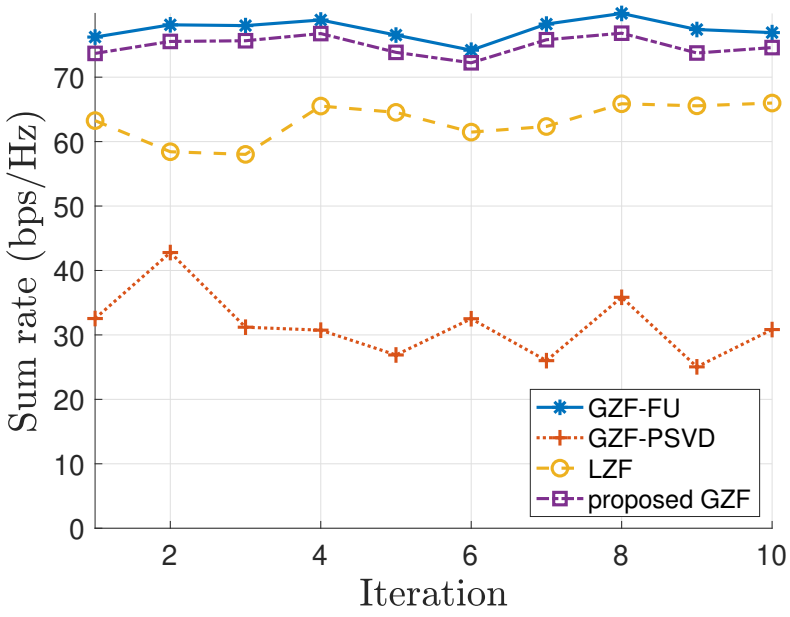


Figure 4.7: Comparison of spectral efficiency in 28 GHz frequency band

the necessary information that is required for the proposed CSI method. Through well-designed simulations, I demonstrate that the proposed GZF precoding strategy has a good ability to significantly reduce the upload overhead while achieving comparable performance to the traditional GZF without CSI compression.

Chapter 5

Multi-Agent Deep Reinforcement Learning for Access Point Activation Strategy in Cell-Free Massive MIMO Networks

5.1 Introduction

As an advanced ultra-dense small-cell network, the cell-free massive MIMO network has been introduced and attracted much attention in recent years due to its advantages of low inter-cell interference and uniform quality of service. In a cell-free network, a large number of antennas or access points (APs) are distributed within the network and jointly serve user equipments (UEs). This architecture makes it able to avoid handover and provide seamless wireless services.

However, the cell-free massive MIMO network still suffers from two major issues. One is the huge power consumption. To provide wireless services, an AP consists of multiple components with different functionalities, such as amplifiers, filters, mixers, and synthesizers. Typically, the power consumption at AP is composed of the power expenditure on the RF transmissions, the circuit power expenditure on RF chain, and the power expenditure on site-specific and architecture-specific factors [52]. When the network scale becomes large, i.e., as there are a large number of APs deployed in the network, huge power will be consumed network-wide. A practical approach is to turn off some APs to reduce unnecessary power consumption, especially when the demand traffic is not heavy. This approach is referred to as the AP sleep mode strategy or activation strategy [75]. However, blindly turning off the APs might also affect the quality of the service. Therefore, the AP activation decision should be carefully made to balance the trade-off between the service level and power consumption, especially in large-scale cell-free massive MIMO network implementation.

The other one is the scalability issue. In a traditional cell-free massive MIMO network, there is a centralized controller that takes charge of the radio-related decisions for all APs, such as precoding, power control, and switching on/off. When the number of APs increases, the computation time needed for making those decisions at the controller will become less affordable due to its limited computational capacity. Besides, to make reasonable decisions, the controller needs to have a comprehensive knowledge of the whole network, referring to the resources (e.g., bandwidth) and environment (e.g., channel state information (CSI)). This will lead to considerable signaling overhead that wastes bandwidth resources. To address this scalability issue, a promising solution is to apply a distributed control, in which each AP makes the decision by itself and no centralized controller exists. Since each AP only needs the local environment knowledge and undertakes lightweight computation, the scalability issue in the cell-free massive MIMO networks is expected to be effectively solved.

How to improve the energy efficiency in UDNs by turning off redundant BSs/APs is a hot topic in the past decade. Many researchers have made remarkable contributions to this problem and its variants. More details can be found in [75, 26, 19, 57]. Recently, some papers have focused on the AP activation (APA) strategy in cell-free massive MIMO networks. In particular, the performance of the APA technique in cell-free massive MIMO networks is analyzed in [30]. Moreover, various APA strategies are proposed in [16, 20, 44, 42, 71], which are designed for different scenarios, such as in mmWave domain [20], with homogeneously and heterogeneously distributed users [16, 20], and for radio stripe-based deployment [44]. The widely used approaches to solve the APA problem include random strategy [44], optimization theory [16, 20, 71], and machine learning [77, 42]. However, all these papers assume that the APA decision is made by a centralized controller and, therefore, none of the approaches can address the scalability issue. The authors in [17] proposed a distributed solution to solve the APA problem in mmWave cellular networks, but the approach they proposed is for small-cell networks where each UE is served by at most one BS and is not applicable for cell-free networks.

Considering the inherent relationship between distributed control and multi-agent decision-making, I fill this gap by proposing a multi-agent deep reinforcement learning (MADRL) approach to solve the APA problem in cell-free massive MIMO networks with static and dynamic user demand. In particular, Ileverage the user-centric characteristic of the cell-free massive MIMO network to design a DQN-based MADRL approach. In the proposed approach, each AP acts as an independent agent that makes the switching on/off decision-based on its local knowledge of the network. By comparing it with a centralized deep reinforcement learning (CDRL) on the reduced power consumption, I demonstrate that the MADRL can efficiently solve the APA problem and it outperforms the CDRL.

The rest of this paper is organized as follows. In Section 5.2, I formulate the APA problem in a user-centric cell-free massive MIMO network as an optimization problem. In Section 5.3, a MADRL approach is proposed to solve the APA problem, and in Section 5.4, a CDRL is presented. The performance of the proposed approach is evaluated in Section 5.5. Finally, I conclude this research in Section 5.6.

5.2 Problem Formulation

Consider a cell-free massive MIMO cellular network $\mathcal{G}$, composited of a set of access points (APs) $\mathcal{M} = \{m_1, m_2, \ldots, m_M\}$ and a set of user equipments (UEs) $\mathcal{K} = \{k_1, k_2, \ldots, k_K\}$, which is shown in Fig. 5.1. The APs, each of which is equipped with N antennas, as well as the single-antenna UEs, are independently distributed in the network. The APs jointly provide downlink service to the UEs with a predefined minimum requirement of the received signal-interference-noise-ratio (SINR) for the UEs. All APs are connected to a centralized processing unit (CPU) via fronthual. The CPU takes charge of the data exchange and processing for wireless communication in the network.

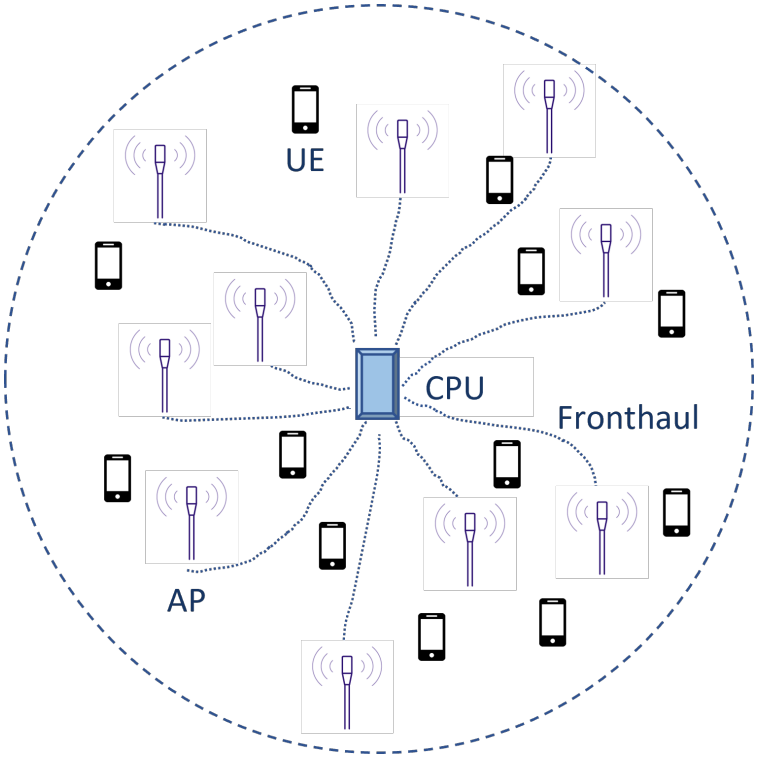


Figure 5.1: Cell-free massive MIMO network

5.2.1 Propagation Model

In this research, the channels between the APs and the UEs are modeled as Rayleigh fading channels. Note that the method proposed in this paper can also be applied to other channel fading models, such as Rician fading. For the AP $m \in \mathcal{M}$ and the UE $k \in \mathcal{K}$, the channel vector $\mathbf{g}_{mk} \in \mathbb{C}^{N \times 1}$ between them is modeled as

$$\mathbf{g}_{mk} = \sqrt{\beta_{mk}}\mathbf{h}_{mk}, m \in \mathcal{M}, k \in \mathcal{K}, \tag{5.1}$$

where β_{mk} denotes the large-scale fading coefficient and $\mathbf{h}_{mk} \sim \mathbb{CN}(\mathbf{0}, \mathbf{I}_N)$ represents the small-scale fading. The estimation of the CSI is outside of the scope of this paper. I assume that the APs have the full knowledge of the CSI and apply the maximum ratio transmission (MRT) precoding for downlink transmission. In particular, given the channel vector $\mathbf{g}_{mk}$ for the AP m and the UE k, the corresponding precoding vector $\mathbf{w}_{mk} = \mathbf{g}_{mk}^*$. The transmitted

signal $\mathbf{x}_m \in \mathbb{C}^{N\times 1}$ at the AP m is calculated as

$$\mathbf{x}_m = \sum_{k\in\mathcal{K}(m)} \sqrt{p_{mk}}\mathbf{w}_{mk}q_k, \tag{5.2}$$

where $\mathcal{K}(m)$ denotes the subset of UEs that are served by the AP m, p_{mk} is the power allocated to the UE k by the AP m, and q_k represents the data intended to the UE k, $\mathbb{E}\{|q_k|^2\} = 1$. Denoting the subset of APs that serve the UE k as $\mathcal{M}(k)$, the received signal at the UE k is calculated as

$$y_k = \sum_{m\in\mathcal{M}(k)} \mathbf{g}_{mk}^* \sqrt{p_{mk}}\mathbf{w}_{mk}q_k + \sum_{\substack{t\in\mathcal{K}\\ t\neq k}} \sum_{m\in\mathcal{M}(t)} \mathbf{g}_{mk}^* \sqrt{p_{mt}}\mathbf{w}_{mt}q_t + n_k, \tag{5.3}$$

where the first item is the desired signal, the second stands for the interference, and $n_k \in \mathbb{CN}(0, \sigma_N^2)$ denotes the additive white Gaussian noise at the UE k.

Based on the above setting, the received SINR at the UE k can be calculated as

$$SINR_k = \frac{\left|\sum_{m\in\mathcal{M}(k)} \mathbf{g}_{mk}^* \sqrt{p_{mk}}\mathbf{w}_{mk}\right|^2}{\sum_{\substack{t\in\mathcal{K}\\ t\neq k}} \left|\sum_{m\in\mathcal{M}(t)} \mathbf{g}_{mk}^* \sqrt{p_{mt}}\mathbf{w}_{mt}\right|^2 + \sigma_N^2}, \tag{5.4}$$

where σ_N^2 is the noise power.

5.2.2 User-Centric Association

In a typical cell-free massive MIMO network, a UE is assumed to be served by all APs in the network. However, this assumption is not realistic in a real-world network. On the one hand, different APs have different distances to a UE. The APs that are near a UE provide stronger signals than those that are far away. On the other hand, an AP in a UDN has limited transmitting power. The APs that are far away from the UE have little impact on it. Therefore, in this research, I consider a user-centric cell-free massive MIMO network, in which each UE is only served by nearby APs. In particular, for the UE $k \in \mathcal{K}$,

$\mathcal{M}(k) = \{m | d_{mk} \leq \bar{d}, m \in \mathcal{M}\}$, where d_{mk} is the distance between the UE k and the AP m, and $\bar{d}$ is a predefined threshold.

5.2.3 AP Activation Problem

In order to provide seamless and reliable services, it is typical that, in a cell-free massive MIMO network, there are much more APs than UEs, i.e., $M \gg K$. However, when the number of UEs is not large or the demand traffic is not heavy, e.g., at midnight, some APs may be redundant. In other words, only a subset of APs could fulfill the total demand in the network. In this case, for the purpose of energy saving, I can turn off the redundant APs if the remaining activated APs have the capacity to serve all UEs. The problem is how to find the maximum subset of APs which can be turned off while maintaining the minimum requirement of wireless service. Here, I consider two scenarios: static and dynamic. In the static scenario, all UEs are active, i.e., each UE has a demand to be met. In particular, let x_m denote whether the AP m is active or not, i.e., $x_m = 1$ if the AP m is active and $x_m = 0$ otherwise. Supposing the AP m consume energy e_m when it is active, this APA problem (P1) is formulated as follows.

$$\mathbf{P1:} \min \sum_{m \in \mathcal{M}} e_m x_m \tag{5.5a}$$

$$s.t.\ r_k \geq \bar{r}_k, k \in \mathcal{K}, \tag{5.5b}$$

$$r_k = \frac{\left| \sum_{m \in \mathcal{M}(k)} \mathbf{g}_{mk}^* \sqrt{p_{mk}} \mathbf{w}_{mk} x_m \right|^2}{\sum_{\substack{t \in \mathcal{K} \\ t \neq k}} \left| \sum_{m \in \mathcal{M}(t)} \mathbf{g}_{mk}^* \sqrt{p_{mt}} \mathbf{w}_{mt} x_m \right|^2 + \sigma_{\mathrm{N}}^2}, k \in \mathcal{K}, \tag{5.5c}$$

$$p_{mk} = \frac{P_m}{|\mathcal{K}(m)|}, m \in \mathcal{M}, \tag{5.5d}$$

$$x_m \in \{0, 1\}, m \in \mathcal{M}, \tag{5.5e}$$

where P_m is the total transmit power of the AP m and $\bar{r}_k$ is the minimum requirement of received SINR of the UE k. In the above model, the objective function (5a) is to minimize

the number of activated APs. The constraints (5b) and (5c) introduce the requirement of the received SINR for each UE. The constraints (5d) indicate the power allocation for the APs. According to the user-centric association, I assume that each AP equally allocates its transmit power to the UEs that are served by it.

In the dynamic scenario, whether a UE is active, i.e., raises a demand, or not is an i.i.d. random variable at each time slot. Specifically, let $z_k(t)$ denote the status of the UE k at time $t \in \mathcal{T}$, i.e., $z_k(t) = 1$ if the UE k is active, while $z_k(t) = 0$ otherwise. Moreover, I use $x_m(t)$ and $r_k(t)$ to denote the status of the AP m and the SINR of the UE k at time t. The formulation of this problem (P2) in the dynamic scenario is

$$\textbf{P2:} \min \sum_{t\in\mathcal{T}} \sum_{m\in\mathcal{M}} e_m x_m(t) \tag{5.6a}$$

$$s.t.\ r_k(t) \geq \bar{r}_k z_k(t), k \in \mathcal{K}, t \in \mathcal{T}, \tag{5.6b}$$

$$r_k(t) = \frac{\left|\sum_{m\in\mathcal{M}(k)} \mathbf{g}_{mk}^* \sqrt{p_{mk}} \mathbf{w}_{mk} x_m(t)\right|^2 z_k(t)}{\sum_{\substack{t\in\mathcal{K} \\ t\neq k}} \left|\sum_{m\in\mathcal{M}(t)} \mathbf{g}_{mk}^* \sqrt{p_{mt}} \mathbf{w}_{mt} x_m(t)\right|^2 + \sigma_{\mathrm{N}}^2}, \tag{5.6c}$$

$$k \in \mathcal{K}, t \in \mathcal{T},$$

$$p_{mk} = \frac{P_m}{|\mathcal{K}(m)|}, m \in \mathcal{M}, \tag{5.6d}$$

$$x_m(t) \in \{0, 1\}, m \in \mathcal{M}, t \in \mathcal{T}. \tag{5.6e}$$

It is obvious that problem P1 is NP-hard and P2 is dynamic which is even more complex. Since it is difficult to apply optimization-based methods to solve them, I design two efficient deep reinforcement learning (DRL) approaches to solve them, which are described in the next section. Another reason to use DRL-based approaches is that, compared with traditional optimization-based methods, they are more energy-efficient. Using traditional optimization-based methods, we need to calculate the solution frequently in each time slot because of the dynamic environment and system status. The power spent on frequent computation is considerable in large scale of time. In contrast, the DRL-based approaches could be

trained online or offline and the well-trained models could be used for inference based on simple computation which costs less power. Therefore, the energy overhead in DRL-based approaches is less than that of traditional optimization-based methods.

5.3 Multi-Agent Deep Reinforcement Learning Approach

In this section, I propose a MADRL approach to solve the APA problem. In this approach, each AP acts as an independent agent and interacts with the unknown environment to gain experience that is then used to direct its future actions, to reduce the total power consumption in the network while satisfying the required service level. The MADRL is illustrated in Fig. 5.2 and explained in detail in the following.

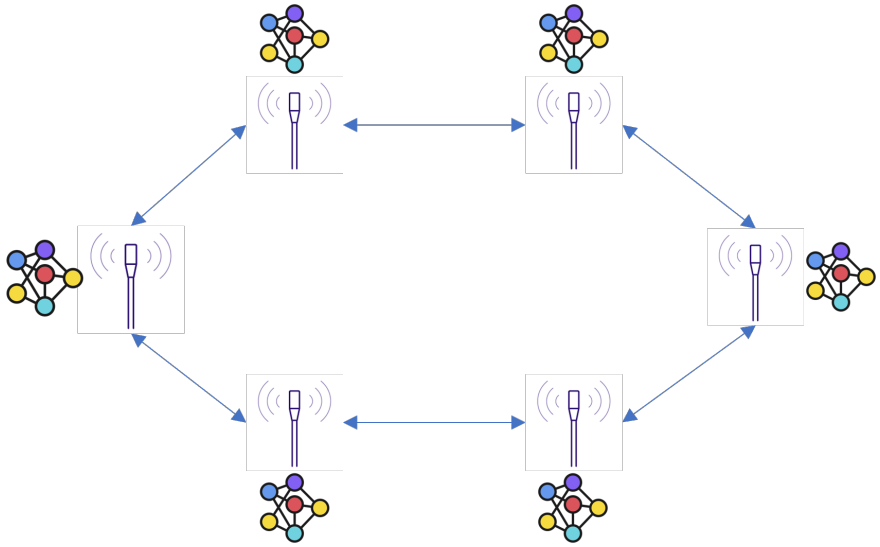


Figure 5.2: Multi-agent deep reinforcement learning

5.3.1 Preliminaries

In my model, each AP is treated as an independent agent to make its own decision, i.e., switch itself on or off, according to its policy given the observation of the environment. In particular, given the instantaneous state s_{mt} at time t, the AP m takes an action a_{mt} based on its policy π_m, which maps a specific state to a corresponding action. After acting

the action, the AP m gets a reward R_{mt} and observes a new state $s_{m,t+1}$, and then repeats this process. The goal of any AP m is to find the optimal policy π_m^* that maximizes its accumulated reward

$$G_{mt} = \sum_{\tau=0}^{\infty} \gamma^{\tau} R_{m,t+\tau}, \tag{5.7}$$

where $\gamma \in [0, 1]$ is a discount factor that scales down the rewards more and more after each step.

5.3.2 Deep Q-Network Framework

My MADRL approach is based on the deep Q-network (DQN) framework, which is described as follows.

(1) State

The state of the AP m at time t is defined as the collection of its neighbors' activation status, i.e., $s_{mt} = \{x_{nt} | n \in \mathcal{N}_m\}$, where $\mathcal{N}_m$ denotes the AP m's neighbor APs. The neighbor of the AP m is defined as a set of APs that is closed to it, i.e., $\mathcal{N}_m = \{n | d_{mn} \leq \bar{d}_N, n \in \mathcal{M}\}$, where d_{mn} is the Euclidean distance between the AP m and the AP n, and $\bar{d}_N$ is a predefined threshold, which determines the neighborhood range of an AP. Although each AP independently determines to be on or off, I allow it to exchange simple information with its neighbors. In particular, an AP can query its neighbor APs' activation status through the connection between them, e.g., wired or wireless fronthaul. Since this information exchange is local, this assumption does not impair the scalability of the solution.

(2) Action

As mentioned above, any AP $m \in \mathcal{M}$ independently decides whether it will be switched on or not. When it observes the instantaneous state s_{mt}, it chooses an action $a_{mt} \in \{0, 1\}$. After the AP m's neighbor APs choose their actions, the AP m's state will change from s_{mt} to $s_{m,t+1}$. Note that $a_{m,t+1}$ is not necessarily different with a_{mt}.

(3) Reward

When all agents take their actions given their respective states at time t, they will receive a reward R_t from the environment. Since my goal is to minimize the number of the power consumption of the network, I define the reward as the sum of the power consumed by the activated APs. Moreover, because of the constraints (5b), I cannot apply RL to solve this problem directly. In order to do so, I need to eliminate the constraints (5b) first. Similar to the Lagrange multiplier method, I transform the constraints as a penalty item and incorporate them into the objective function. In particular, I formulate the reward as

$$R_t = \begin{cases} -\sum_{m \in \mathcal{M}} e_m x_{mt}, & \text{if} \quad r_{kt} \geq \bar{r}_k, \forall k \in \mathcal{K} \\ -B, & \text{otherwise,} \end{cases} \tag{5.8}$$

where r_{kt} is the SINR perceived by the UE k at time t and B is a big number. Note that, in my MADRL model, all APs receive identical rewards.

(4) Policy Update

Each agent's policy is a function in which the input is an observed state and the output is the action mapped to the state. To obtain the AP m's policy π_m, I need to know its Q-function

$$Q_\pi(s_{mt}, a_{mt}) = \mathbb{E}_\pi\{G_{mt} | s_{mt}, a_{mt}\}, \tag{5.9}$$

which shows how good a certain action is, given a state, for an agent following a policy. Q-table is a simple method to estimate the value of $Q(s, a)$ when the state space is discrete and limited. But when the state space is huge, it is computationally intractable to build such a table. In order to address the issue of the "curve of dimension", the DQN uses a neural network, called Q-network, rather than a Q-table, to estimate the Q-function. The Q-network, denoted as Q_{eval}, takes a state as the input and generates an action. Besides the Q-network, I introduce another target-network that has the same structure as the Q-network,

denoted as Q_{target}. Q_{eval} and Q_{target} are parameterized by $\theta_{m,eval}$ and $\theta_{m,target}$, respectively. I calculate

$$q_{target} = R_t + \gamma \max_{a'} Q_{target}(s_{m,t+1}, a'; \theta_{m,target}), \tag{5.10}$$

which is used to update the neural networks. The purpose of introducing the target-network is to make q_{target} stable. Besides, I use the experience replay technique to address the correlation of training data. Each AP maintains a replay buffer D to store the tuples (s, a, r, s') observed recently. When training, the AP m randomly samples a mini-batch of size B_a from D and calculates the loss value as

$$L_{m,eval} = \frac{1}{B_a} \sum_{i=1}^{B_a} \left(q_{target,i} - Q_{eval}(s_i, a_i; \theta_{m,eval})\right)^2. \tag{5.11}$$

The Q_{eval} is updated by performing a gradient descent step on $L_{m,eval}$ as

$$\Delta\theta_{eval} = \alpha L_{eval} \cdot \nabla Q_{eval}(s_t, a_t; \theta_{m,eval}), \tag{5.12}$$

where α is the learning rate. I update the Q_{target} once every C rounds of updating the Q_{eval}, as

$$\theta_{m,target} \leftarrow \theta_{m,eval}. \tag{5.13}$$

The algorithm of the MADRL is summarized in Algorithm 3.

5.4 Centralized Deep Reinforcement Learning Approach

For comparison, I propose a centralized DRL (CDRL) approach that is also based on DQN to solve the APA problem. In the CDRL, the AP activation decision is made by a centralized controller. The controller may locate at a macro BS that takes charge of the control plane of the cell-free massive MIMO network, or at the remote cloud. The controller maintains a DRL model to make the activation decision for all APs in the network. The

Algorithm 3 Multi-Agent DQN algorithm for APA problem

Input: Cellular network $\mathcal{G}$ which consists of an AP set $\mathcal{M}$ and a UE set $\mathcal{K}$
Initialize Q_{eval} with random θ_{eval} for each AP
Initialize Q_{target} with random θ_{target} for each AP
Initialize a replay buffer with size of B_a for each AP
for $episode = 1, \ldots, E$ **do**
    Reset each AP's activation status
    for $t = 1, \ldots, T$ **do**
        for $m = 1, \ldots, M$ **do**
            AP m observes its local state s_{mt} from its neighbors $\mathcal{M}_m$
            With probability ϵ select a random action a_{mt}
            Otherwise, choose $a_{mt} = \arg\max_{a'} Q_{eval}\left(s_{mt}, a'; \theta_{m,eval}\right)$
        end for
        All APs conduct their actions
        for $m = 1, \ldots, M$ **do**
            AP m gets reward R_t and observes the next state $s_{m,t+1}$
            AP m stores tuple $(s_{mt}, a_{mt}, R_t, s_{m,t+1})$ in its replay buffer D
            AP m randomly chooses mini-batch of tuple $(s_{mt}, a_{mt}, R_t, s_{m,t+1})$ from D
            Calculate q_{target} by Eq. (5.10)
            Update $\theta_{m,eval}$ by performing gradient descent step on $L_{m,eval}$
            Every C steps update $\theta_{m,target} \leftarrow \theta_{m,eval}$
        end for
    end for
end for

framework of the CDRL as the same to the MADRL, except for the definition of state and action.

The CDRL is illustrated in Fig. 5.3 . In the CDRL, the controller needs to collect the current state of the whole network. I define the state of the network as the combination of all APs' activation statuses. In particular, the state of the network at time t is denoted as $\mathbf{s}_t = \{x_{1t}, x_{2t}, \ldots, x_{Mt}\}$. At the beginning of each round, each AP uploads its activation status to the controller through the fronthual. I assume the fronthaul is reliable and without any capacity limitation.

When the controller observes the state $\mathbf{s}_t$, it generates an action a_t following the current policy π, i.e., $a_t = \pi(\mathbf{s}_t)$. I define a_t as the id of the AP whose activation status will be flipped over. In particular, suppose the current state is $\mathbf{s}_t = \{x_1, x_2, \ldots, x_M\}$ and the controller chooses the action $a_t \in \{1, \ldots, M\}$, then the next state $\mathbf{s}_{t+1} = \{x_1, x_2, \ldots, 1 - x_{a_t}, \ldots, x_M\}$.

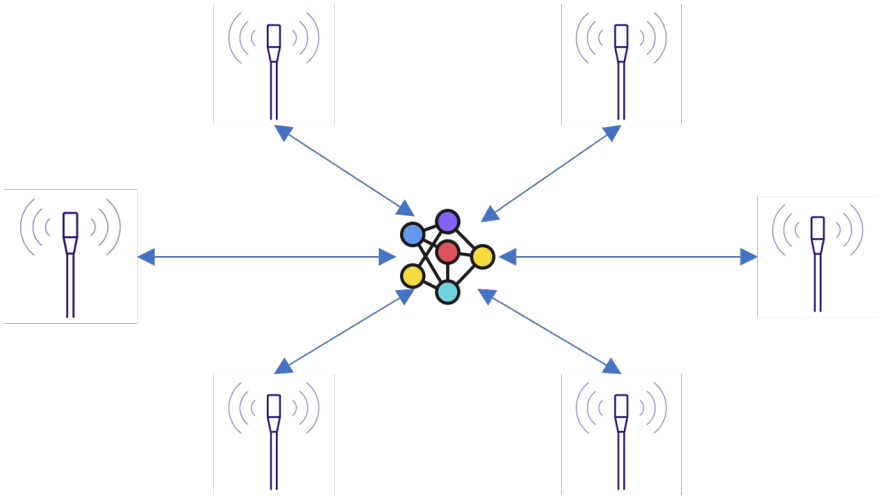


Figure 5.3: Centralized deep reinforcement learning

It is worth noting that, I allow the controller to make no change to the network state. Specifically, if the controller chooses $a_t = M + 1$, it will not change any AP's activation status in this round, i.e., $\mathbf{s}_{t+1} = \mathbf{s}_t$.

The algorithm of the CDRL is summarized in Algorithm 4.

5.5 Performance Evaluation

In the evaluation part, I consider a cell-free massive MIMO cellular network that covers a 100(m) $\times$ 100(m) square region and consists of 50 APs and 10 UEs randomly deployed with uniform distribution in the region. Each AP is equipped with 10 antennas and the UEs are single-antenna. According to the user-centric setting, I set $\bar{d}_N = 20$, which means each UE is served by the APs to which the distance is less than 20(m). Moreover, I set the transmit power of AP as 1w, the minimum requirement of SINR at UE as 0.1, and the noise power $\sigma_N^2 = -92$ dBm. I assume the channels between the APs and the UEs are Rayleigh fading channels and the APs use the Maximum Ratio Transmission (MRT) precoding for the downlink transmission. For the Q-network / target-network, the sizes of the input layer and the output layer are set to be the size of the state space and the action space, respectively.

Algorithm 4 Centralized DQN algorithm for APA problem

Input: Cellular network $\mathcal{G}$ which consists of an AP set $\mathcal{M}$ and a UE set $\mathcal{K}$
Initialize Q_{eval} with random θ_{eval} for the centralized controller
Initialize Q_{target} with random θ_{target} for the centralized controller
Initialize a replay buffer with size of B_a for the centralized controller
for $episode = 1, \ldots, E$ **do**
    Reset each AP's activation status
    for $t = 1, \ldots, T$ **do**
        The centralized controller observes the state s_t
        With probability ϵ select a random action a_t
        Otherwise, choose $a_t = \arg\max_{a'} Q_{eval}(s_t, a'; \theta_{eval})$
        The centralized controller conducts the chosen action
        The centralized controller gets reward R_t and observes the next state s_{t+1}
        The centralized controller stores tuple (s_t, a_t, R_t, s_{t+1}) in its replay buffer D
        The centralized controller randomly chooses mini-batch of tuple (s_t, a_t, R_t, s_{t+1}) from D
        Calculate q_{target} by Eq. (5.10)
        Update θ_{eval} by performing gradient descent step on L_{eval}
        Every C steps update $\theta_{target} \leftarrow \theta_{eval}$
    end for
end for

The other parameters of the neural network are listed in Table 5.1. Moreover, the parameters of the RL are listed in Table 5.2.

Table 5.1: Neural Network Parameters

Notation	Explanation	Value
-	Number of hidden layers	1
-	Size of hidden layer	50
-	Activation function	ReLU
$\|D\|$	Reply buffer size	2000
B_a	Mini-batch size	32
C	Frequency of Q_{target} update	100
-	Optimizer	Adam

5.5.1 Static Scenarios

In this subsection, I evaluate the performances of the proposed DRL-based approaches in static scenarios, where all UEs are active all the time.

Table 5.2: RL Parameters

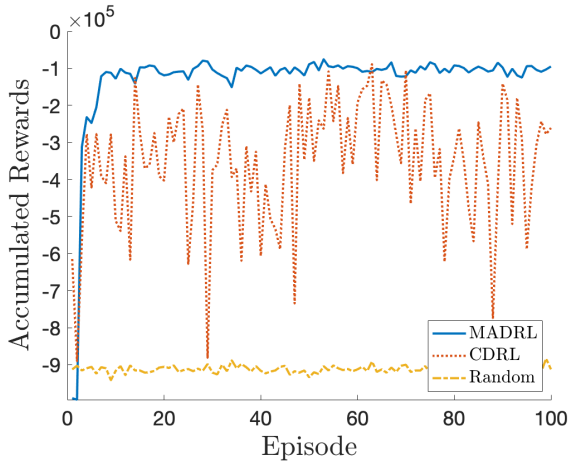
Notation	Explanation	Value
N_{episode}	Number of episodes	100
N_{step}	Number of steps	1000
α	Learning rate	0.01
λ	Discount	0.9
ϵ	Exploration trigger value	0.1

Firstly, I compare the performances of the MADRL and two counterparts, i.e., the CDRL approach and the random approach, on the APA problem. In the random approach, each AP randomly chooses its activation status at each time slot. I consider 50 APs in the network and plot the simulation results in Fig. 5.4. In particular, in Fig. 5.4(a), the accumulated rewards achieved by these three approaches are displayed. To further analyze their performances, I compare the power consumption of these approaches in Fig. 5.4(b). It shows that the random approach renders solutions with the minimum power consumption, while the performances of the MADRL and the CDRL are comparable. However, a solution with low power consumption may be infeasible because of the constraints of service level. To reveal that whether the solutions obtained by these approaches are feasible (can meet the requirement of service level), I also show the number of solutions generated in each episode for each approach, say total penalty, in Fig. 5.4(c). It can be found that, although the solutions obtained by the random approach have the minimum power consumption, most of them cannot meet the requirement of service level. Although the quality of the solutions generated by the CDRL are better, compared with that of the random approach, it exhibits obvious oscillation. In contrast, the MADRL can find much better solutions and shows good convergence. The results demonstrate that, among these two RL approaches, the MADRL outperforms the CDRL. Notably, the MADRL converges quickly and its performance keeps stable over all episodes. In contrast, the CDRL's convergence is not so good and its performance fluctuates over the episodes, even though it reaches some excellent solutions occasionally. The reason why the MADRL is better than the CDRL in solving the APA

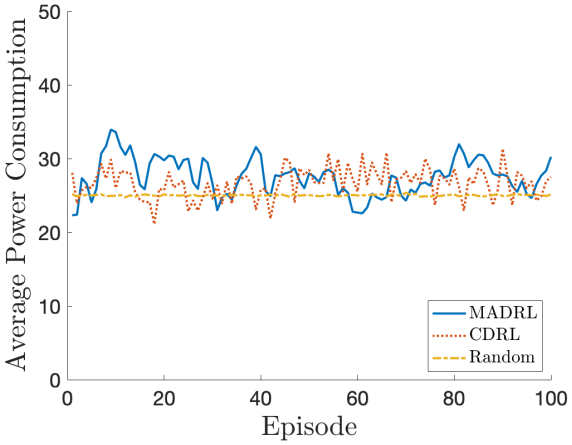
problem lies in that, in the MADRL, each agent has a much smaller state space and action space, which makes it have good convergence. Another reason comes from the user-centric characteristic of the cell-free massive MIMO network. In particular, since each UE is served by the APs that are near it, each AP only needs to collaborate with its neighbor APs, rather than all other APs, to jointly serve a subset of UEs. In other words, those APs that are far from it have little impact on its decision. Therefore, it does not need to be aware of the status of the APs that are far away. That is why an agent in the MADRL only needs a small state space. Moreover, in the MADRL, an agent has a small action space, i.e., $\{0, 1\}$, while in the MADRL, the agent's action space is $\{1, 2, \ldots, M + 1\}$, which is much larger. That also helps the MADRL have good convergence. These results show that, compared with the CDRL, the MADRL not only reduces the communication overhead but also provides better solutions for the APA problem.

Secondly, I investigate the impact of the density of APs on the performance of the MADRL. I consider three numbers of the APs, i.e., 50, 60, and 70, and run the MADRL with these settings respectively. From the results shown in Fig. 5.5, it is demonstrated that, with all these densities, the performances of the MADRL are stable. This shows the ability of the MADRL to provide reliable performance with various densities of APs. Moreover, it is illustrated that the performance of the MADRL with 50 APs is the best. It is because, compared with the other two densities, the state space of the MADRL is smaller when the number of APs is 50. It makes the MADRL converge to better solutions.

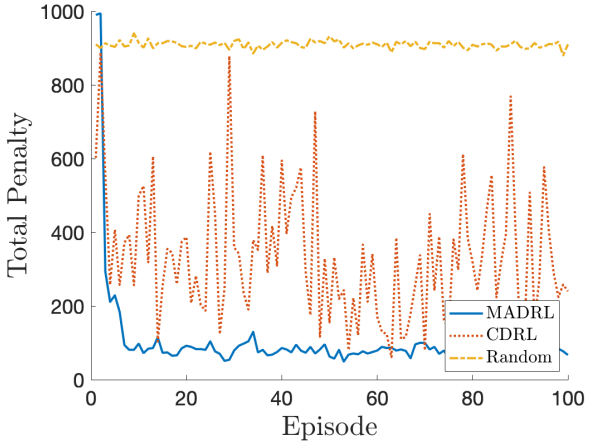
Thirdly, I consider the impact of the neighborhood range on the performance of the MADRL. A larger neighborhood range means that each AP's state space is larger. I choose three values for the neighborhood range and show the results in Fig. 5.6. It is illustrated that the neighborhood range of $\bar{d}_N = 30$ renders the best performance. When the neighborhood range grows larger, for example, to 50 and 70, the performance degrades. The reason is similar to that in the above. With a larger neighborhood range, each agent's state space is larger and more complicated. The larger neighborhood range brings little gain to the



(a) Accumulated rewards of the three approaches



(b) Power consumption of the three approaches



(c) Total penalties of the three approaches

Figure 5.4: Performances of the MADRL, the CDRL, and the random strategies with static demand

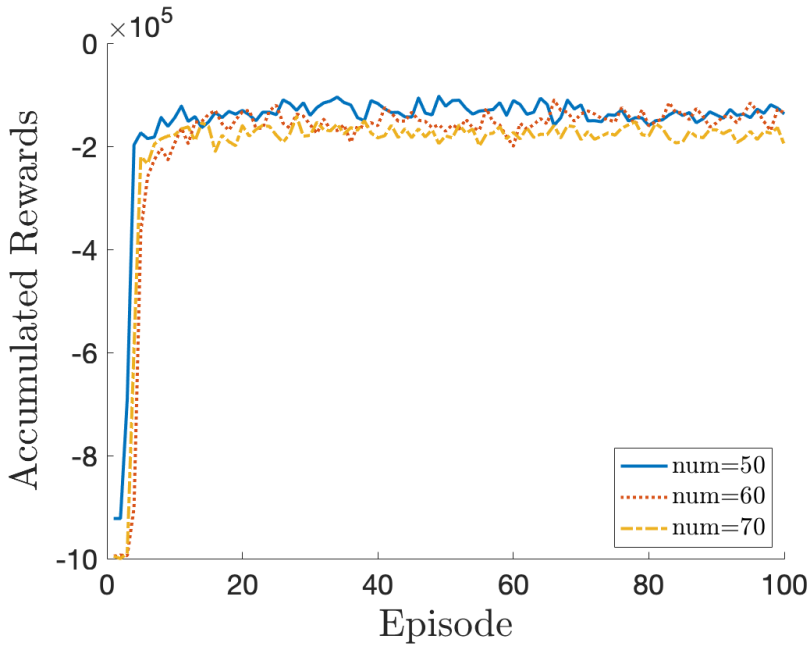


Figure 5.5: Performances of the MADRL with different numbers of APs

learning process but causes more communication overhead. This insight discloses that no matter from the point of view of performance or cost, a large neighborhood range is not a good choice for the APA problem.

5.5.2 Dynamic Scenarios

In this subsection, I evaluate the performances of the proposed DRL-based approaches in the dynamic scenarios, in which whether any UE raises demand is an i.i.d. random variable. In all simulations conducted in this part, I set the number of UEs as 30 and the minimum requirement of SINR at UE as 0.05.

Firstly, I consider a dynamic scenario where each UE $k \in \mathcal{K}$ has an individual active probability b_k uniformly distributed in $[0, 1]$ to raise a demand at each time slot. The same as the first simulation in Section 5.5.1, I compare the performances of the MADRL, the CDRL, and the random approach in this scenario. The results are displayed in Fig. 5.7, in which the subfigures 5.7(a), 5.7(b), and 5.7(c) show the comparison on the accumulated reward, the power consumption, and the total penalty, respectively. It is obvious that the

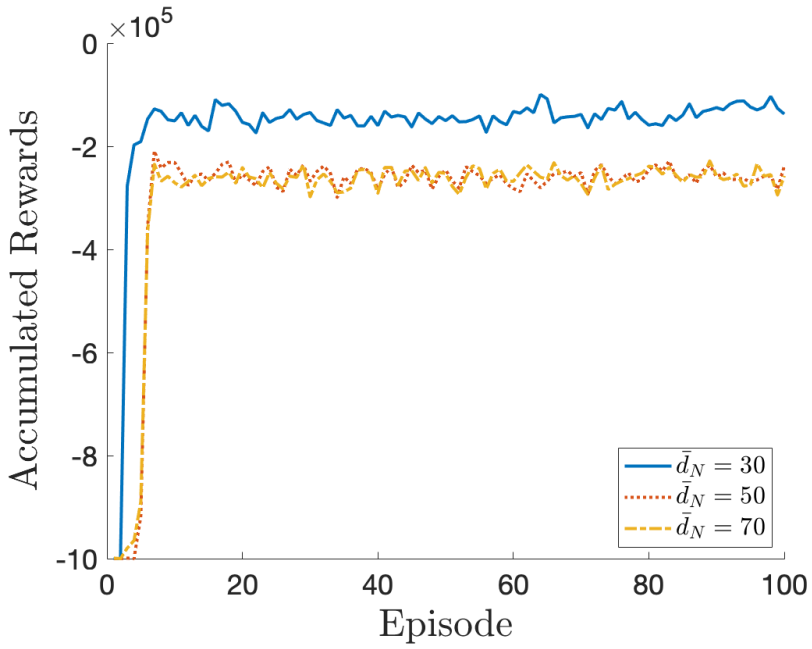
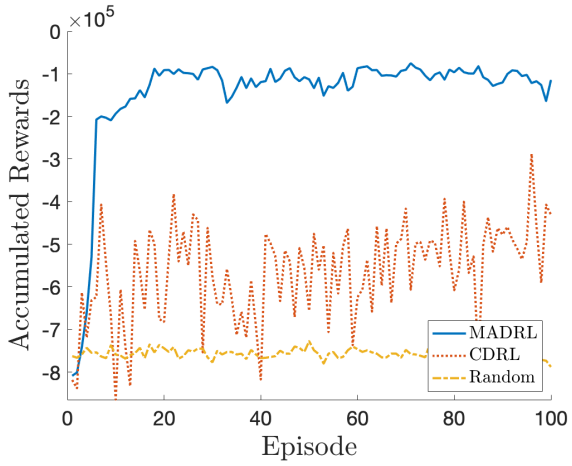


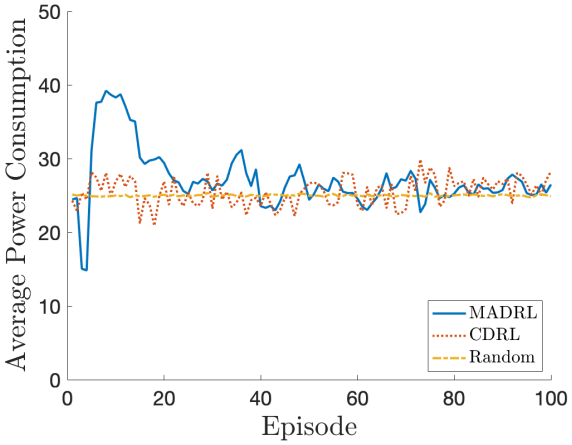
Figure 5.6: Performances of the MADRL with different neighborhood ranges

proposed MADRL still significantly outperforms the other two in both the quality of the solution obtained and the convergence rate, and that the CDRL is insufficient to solve the APA problem in the dynamic scenario due to the huge state and action space, as well as the time-varying demand of UE. This result demonstrates that, compared with the CDRL, the proposed MADRL has an excellent ability to solve the APA problem when user demands are random.

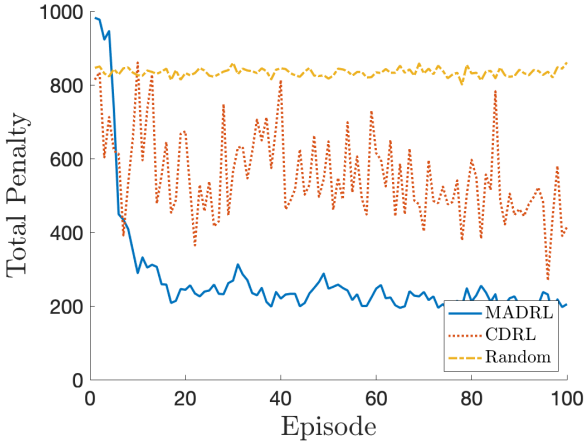
Secondly, I investigate the performance of the MADRL when given different active probabilities of UE. Here I assume that all UEs have an identical value of the active probability b. I choose 4 values for b: 0.25, 0.5, 0.75, and 1, and evaluate the performance of the MADRL in these scenarios, respectively. Especially, $b = 1$ means all UEs are active all the time, which is equivalent to the static scenario. According to the result shown in Fig. 5.8, the MADRL converges to a solution with more active APs when the active probability is high, e.g., $b = 1$, while it tends to choose another solution with less active APs given a low active probability, e.g., $b = 0.25$. It is because when the active probability is low, there are not many UEs raising demand and just a few APs are needed to meet the total demand. When



(a) Accumulated rewards of the three approaches



(b) Power consumption of the three approaches



(c) Total penalties of the three approaches

Figure 5.7: Performances of the MADRL, the CDRL, and the random strategies with dynamic demand

the active probability is higher, more demands are raised and consequently more APs need to be turned on. This result shows that the MADRL can generate effective APA strategies that are suitable for different active probabilities of UE. Moreover, the MADRL shows a good convergence rate in all scenarios. That means the performance of the MADRL is rather stable in various scenarios and it is suitable for general use cases in practice.

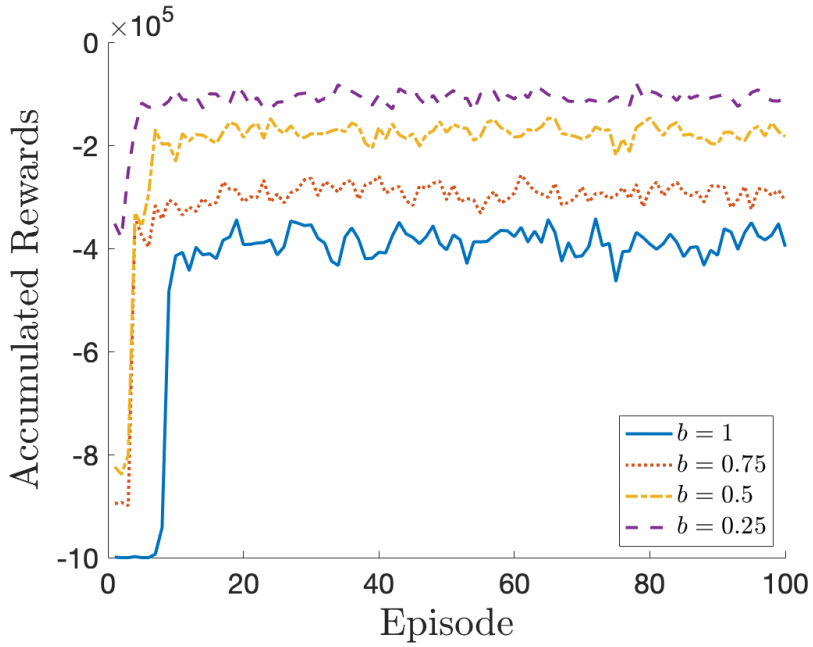


Figure 5.8: Performances of the MADRL with active probabilities of UE

5.6 Summary

In this chapter, I investigate an AP activation problem in a user-centric cell-free massive MIMO network. Considering the static and the dynamic user demand, my goal is to find a strategy to turn off unnecessary APs to reduce power consumption while maintaining the minimum required service level and I model it as an optimization problem. To address the scalability issue in such a UDN, I propose a multi-agent deep reinforcement learning (MADRL) approach and, as a counterpart, a centralized deep reinforcement learning (CDRL) approach. The simulation results demonstrate that, compared with the CDRL, the MADRL is able to find much better solutions and has a much better convergence rate in both static

and dynamic scenarios. More important, the MADRL has great potential to solve the APA problem in cell-free massive MIMO systems in order to reduce power consumption and mitigate the scalability issue.

Chapter 6

Conclusions and Future Work

Because of the large number of APs, mobility and energy management become two critical problems in 5G UDNs. In this dissertation, I propose several novel approaches to solve them effectively. Firstly, I develop two MAB-based handover strategies to reduce the handover in mmWave ultra-dense cellular networks. These two strategies learn the empirical knowledge about the distribution of the user's moving trajectory and LOS link blockage from historical handover events and use it to guide future handover decisions. When a handover event happens, supported by a signal space partitioning scheme, these two strategies extract its spatial and temporal features to retrieve the accumulated knowledge and generate a handover decision by the proposed SBS-selection algorithms. Compared with some counterpart strategies, the proposed handover strategies can significantly reduce the handover frequency. Secondly, I focus on a special UDN: a cell-free massive MIMO system, which is able to further reduce handover. To address the severe inter-user interference issue in cell-free massive MIMO systems, I propose a bandwidth-efficient GZF precoding strategy associated CSI feedback compression mechanism. By leveraging the physical structure of Rician fading channels in cell-free massive MIMO systems, I propose to decompose the channel matrix into LOS and NLOS components, and design two customized methods to compress them, respectively. The simulation result shows that the proposed precoding strategy remarkably reduced the communication overhead with little information loss in CSI compression. Thirdly, being aware of the considerable power consumption UDNs where there are a large number of APs, I propose an AP activation (APA) strategy to wisely turn off idle or under-loaded APs in cell-free massive MIMO systems with the purpose to reduce power consumption. To address the scalability issue when the number of APs grows extremely large, I design a MADRL-based

approach to solve this problem effectively. It is demonstrated through simulation results that the distributed DRL is a strong tool to solve the APA problem in cell-free massive MIMO systems. Compared with the centralized DRL, the multi-agent DRL has better performance and convergence.

My research in this dissertation could be extended in the following directions in the future.

(1) The proposed MAB-based handover strategies can only be applied in scenarios where the obstacles are fixed. In the next step, I will consider how to improve them and make them able to handle mobile obstacles.

(2) In the proposed bandwidth-efficient GZF precoding strategy, it is assumed that the APs are equipped with antenna arrays. When given single-antenna APs, how to solve the same SCI overhead reduction problem is another future work.

(3) In the APA problem considered in Chapter 5, I only consider fixed UEs. The future work is to investigate the performance of MADRL when given mobile UEs and how to improve it.

Acknowledgements

Words cannot express my gratitude to my supervisor Dr. Richard Chapman for his dedicated support and guidance. The journey to this Ph.D. degree is full of challenges and difficulties. When I hit rock bottom in my Ph.D. study and was about to quit, Dr. Chapman helped me out and gave me the power to continue my research. I am also grateful to Dr. Xiao Qin for his kind encouragement to me when I was in the most darkness in my Ph.D. program. Without his generous help and inspiration, I cannot get out of this frustrating dilemma. I would also like to express my gratitude to Dr. Shiwen Mao, who always gives me sincere instructions and suggestions no matter in my research or my future career. I am also thankful to Dr. David Umphress, whose profound knowledge and humorous teaching style deeply impress me from my first day at Auburn University. Additionally, thank Dr. Tao Shu for the guidance and financial support on the work in the background parts.

Moreover, I would like to thank my former lab-mates: Jian Chen, Tian Liu, Xueyang Hu, Hairuo Xu, and Amit Das, with whom I worked together for more than four years. The valuable friendship between us makes my monotonous research life colorful.

Finally, I'm deeply indebted to my wife Jing Hou, my daughter Xiaoran Sun, and my parents, for their strong and incessant love. Without their backup, it is impossible for me to continue my study in the US.

Bibliography

[1] 3GPP. Telecommunication management; network sharing; concepts and requirements. 3GPP TR. 23.251 v15.0.0., Jul. 2018. URL: https://www.etsi.org/deliver/etsi_ts/132100_132199/132130/15.00.00_60/ts_132130v150000p.pdf.

[2] 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*, 32(6):1164–1179, 2014.

[3] 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)*, pages 1055–1059. IEEE, 2018.

[4] G. Athanasiou, P. C. Weeraddana, C. Fischione, and L. Tassiulas. Optimizing client association for load balancing and fairness in millimeter-wave wireless networks. *IEEE/ACM Transactions on Networking*, (3):836–850, 2015.

[5] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2-3):235–256, 2002.

[6] J. Bains. Platform based design accelerates 5g test. In *The Silicon Valley 5G Summit*. National Instruments, 2016. available at http://www.samsung.com/global/business/networks/events/Silicon-Valley-5G-Summit/attachments/S4-NI-Jin-Bains.pdf.

[7] C. N. Barati, S. A. Hosseini, M. Mezzavilla, T. Korakis, S. S. Panwar, S. Rangan, and M. Zorzi. Initial access in millimeter wave cellular systems. *IEEE Transactions on Wireless Communications*, 15(12):7926–7940, 2016.

[8] E. Björnson, J. Hoydis, and L. Sanguinetti. Massive mimo networks: Spectral, energy, and hardware efficiency. *Foundations and Trends in Signal Processing*, 11(3-4):154–655, 2017.

[9] E. Björnson, E. G. Larsson, and M. Debbah. Massive mimo for maximal spectral efficiency: How many users and pilots should be allocated? *IEEE Transactions on Wireless Communications*, 15(2):1293–1308, 2015.

[10] E. Björnson and L. Sanguinetti. Making cell-free massive mimo competitive with mmse processing and centralized implementation. *IEEE Transactions on Wireless Communications*, 19(1):77–90, 2019.

[11] D. Bouneffouf and I. Rish. A survey on practical applications of multi-armed and contextual bandits. *arXiv preprint arXiv:1904.10040*, 2019.

[12] G. Brown and H. Reading. Exploring the potential of mmwave for 5g mobile access. *White paper*, 2016.

[13] C. Chaieb, Z. Mlika, F. Abdelkefi, and W. Ajib. Mobility-aware user association in hetnets with millimeter wave base stations. In *2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC)*, pages 153–157. IEEE, 2018.

[14] I. Colin, A. Thomas, and M. Draief. Parallel contextual bandits in wireless handover optimization. In *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 258–265. IEEE, 2018.

[15] M. E. Eltayeb, T. Y. Al-Naffouri, and H. R. Bahrami. Compressive sensing for feedback reduction in mimo broadcast channels. *IEEE Transactions on communications*, 62(9):3209–3222, 2014.

[16] G. Femenias, N. Lassoued, and F. Riera-Palou. Access point switch on/off strategies for green cell-free massive mimo networking. *IEEE Access*, 8:21788–21803, 2020.

[17] M. Feng, S. Mao, and T. Jiang. Dynamic base station sleep control and rf chain activation for energy-efficient millimeter-wave cellular systems. *IEEE Transactions on Vehicular Technology*, 67(10):9911–9921, 2018.

[18] R. Ford, S. Rangan, E. Mellios, D. Kong, and A. Nix. Markov channel-based performance analysis for millimeter wave mobile networks. In *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6. IEEE, 2017.

[19] P. Gandotra, R. K. Jha, and S. Jain. Green communication in next generation cellular networks: A survey. *IEEE access*, 5:11727–11758, 2017.

[20] J. García-Morales, G. Femenias, and F. Riera-Palou. Energy-efficient access-point sleep-mode techniques for cell-free mmwave massive mimo networks with non-uniform spatial traffic density. *IEEE access*, 8:137587–137605, 2020.

[21] A. Ge, T. Zhang, Z. Hu, and Z. Zeng. Principal component analysis based limited feedback scheme for massive mimo systems. In *2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pages 326–331. IEEE, 2015.

[22] M. Giordani, M. Mezzavilla, S. Rangan, and M. Zorzi. Multi-connectivity in 5g mmwave cellular networks. In *2016 Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net)*, pages 1–7. IEEE, 2016.

[23] S. Goyal, M. Mezzavilla, S. Rangan, S. Panwar, and M. Zorzi. User association in 5g mmwave networks. In *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6. IEEE, 2017.

[24] F. Guidolin, I. Pappalardo, A. Zanella, and M. Zorzi. Context-aware handover policies in hetnets. *IEEE Transactions on Wireless Communications*, 15(3):1895–1906, 2016.

[25] D. Gündüz, P. de Kerret, N. D. Sidiropoulos, D. Gesbert, C. R. Murthy, and M. van der Schaar. Machine learning in the air. *IEEE Journal on Selected Areas in Communications*, 37(10):2184–2199, 2019.

[26] F. Han, S. Zhao, L. Zhang, and J. Wu. Survey of strategies for switching off base stations in heterogeneous networks for greener 5g systems. *IEEE Access*, 4:4959–4973, 2016.

[27] Z. Han, T. Lei, Z. Lu, X. Wen, W. Zheng, and L. Guo. Artificial intelligence-based handoff management for dense wlans: A deep reinforcement learning approach. *IEEE Access*, 7:31688–31701, 2019.

[28] G. Interdonato, M. Karlsson, E. Björnson, and E. G. Larsson. Local partial zero-forcing precoding for cell-free massive mimo. *IEEE Transactions on Wireless Communications*, 2020.

[29] J. Joung, E. Kurniawan, and S. Sun. Channel correlation modeling and its application to massive mimo channel feedback reduction. *IEEE Transactions on Vehicular Technology*, 66(5):3787–3797, 2016.

[30] S. Jung and S.-E. Hong. Performance analysis of access point switch on/off schemes for cell-free mmwave massive mimo udn systems. In *2021 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 644–647. IEEE, 2021.

[31] J. Kang, O. Simeone, J. Kang, and S. Shamai. Fronthaul compression and precoding design for c-rans over ergodic fading channels. *IEEE Transactions on Vehicular Technology*, 65(7):5022–5032, 2015.

[32] J. Kang, O. Simeone, J. Kang, and S. S. Shitz. Joint signal and channel state information compression for the backhaul of uplink network mimo systems. *IEEE Transactions on Wireless Communications*, 13(3):1555–1567, 2014.

[33] 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)*, pages 692–697. IEEE, 2018.

[34] P.-H. Kuo, H. Kung, and P.-A. Ting. Compressive sensing based channel feedback protocols for spatially-correlated massive antenna arrays. In *2012 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 492–497. IEEE, 2012.

[35] L. Li, W. Chu, J. Langford, and R. E. Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 661–670. ACM, 2010.

[36] Q. Li, M. Ding, C. Ma, C. Liu, Z. Lin, and Y.-C. Liang. A reinforcement learning based user association algorithm for uav networks. In *2018 28th International Telecommunication Networks and Applications Conference (ITNAC)*, pages 1–6. IEEE, 2018.

[37] Y. Liao, H. Yao, Y. Hua, and C. Li. Csi feedback based on deep learning for massive mimo systems. *IEEE Access*, 7:86810–86820, 2019.

[38] Z. Liu, L. Zhang, and Z. Ding. Exploiting bi-directional channel reciprocity in deep learning for low rate massive mimo csi feedback. *IEEE Wireless Communications Letters*, 8(3):889–892, 2019.

[39] C. Lu, W. Xu, H. Shen, J. Zhu, and K. Wang. Mimo channel information feedback using deep recurrent network. *IEEE Communications Letters*, 23(1):188–191, 2018.

[40] G. R. MacCartney, H. Yan, S. Sun, and T. S. Rappaport. A flexible wideband millimeter-wave channel sounder with local area and nlos to los transition measurements. In *2017 IEEE International Conference on Communications (ICC)*, pages 1–7. IEEE, 2017.

[41] S. Maghsudi and E. Hossain. Multi-armed bandits with application to 5g small cells. *IEEE Wireless Communications*, 23(3):64–73, 2016.

[42] C. F. Mendoza, S. Schwarz, and M. Rupp. Deep reinforcement learning for dynamic access point activation in cell-free mimo networks. In *WSA 2021; 25th International ITG Workshop on Smart Antennas*, pages 1–6. VDE, 2021.

[43] 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)*, pages 100–105. IEEE, 2016.

[44] A. K. Mishra and P. Vijayakumar. Ap switch on/off strategy in green cell-free massive mimo based radio stripe network. In *2021 5th International Conference on Computer, Communication and Signal Processing (ICCCSP)*, pages 125–130. IEEE, 2021.

[45] F. B. Mismar and B. L. Evans. Partially blind handovers for mmwave new radio aided by sub-6 ghz lte signaling. In *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*, pages 1–5. IEEE, 2018.

[46] E. Nayebi, A. Ashikhmin, T. L. Marzetta, H. Yang, and B. D. Rao. Precoding and power optimization in cell-free massive mimo systems. *IEEE Transactions on Wireless Communications*, 16(7):4445–4459, 2017.

[47] L. D. Nguyen, T. Q. Duong, H. Q. Ngo, and K. Tourki. Energy efficiency in cell-free massive mimo with zero-forcing precoding design. *IEEE Communications Letters*, 21(8):1871–1874, 2017.

[48] T. Nishio, H. Okamoto, K. Nakashima, Y. Koda, K. Yamamoto, M. Morikura, Y. Asai, and R. Miyatake. Proactive received power prediction using machine learning and depth images for mmwave networks. *IEEE Journal on Selected Areas in Communications*, 37(11):2413–2427, 2019.

[49] Ö. Özdogan, E. Björnson, and E. G. Larsson. Massive mimo with spatially correlated rician fading channels. *IEEE Transactions on Communications*, 67(5):3234–3250, 2019.

[50] R. Parada and M. Zorzi. Context-aware handover in mmwave 5g using ue's direction of pass. In *European Wireless 2018; 24th European Wireless Conference*, pages 1–6. VDE, 2018.

[51] M. Polese, M. Giordani, M. Mezzavilla, S. Rangan, and M. Zorzi. Improved handover through dual connectivity in 5g mmwave mobile networks. *IEEE Journal on Selected Areas in Communications*, 35(9):2069–2084, 2017.

[52] K. S. V. Prasad, E. Hossain, and V. K. Bhargava. Energy efficiency in massive mimo-based 5g networks: Opportunities and challenges. *IEEE Wireless Communications*, 24(3):86–94, 2017.

[53] J. Qiao, Y. He, and X. S. Shen. Proactive caching for mobile video streaming in millimeter wave 5g networks. *IEEE Transactions on Wireless Communications*, 15(10):7187–7198, 2016.

[54] S. Rangan, T. S. Rappaport, and E. Erkip. Millimeter-wave cellular wireless networks: Potentials and challenges. *Proceedings of the IEEE*, 102(3):366–385, 2014.

[55] J. C. Roh and B. D. Rao. Efficient feedback methods for mimo channels based on parameterization. *IEEE Transactions on Wireless Communications*, 6(1):282–292, 2007.

[56] W. Roh. 5g: from vision to reality. In *The Silicon Valley 5G Summit*. Samsung Electronics, 2016. available at `http://www.samsung.com/global/business/networks/events/Silicon-Valley-5G-Summit/attachments/S5-Samsung-Wonil-Roh.pdf`.

[57] F. Salahdine, J. Opadere, Q. Liu, T. Han, N. Zhang, and S. Wu. A survey on sleep mode techniques for ultra-dense networks in 5g and beyond. *Computer Networks*, 201:108567, 2021.

[58] O. Semiari, W. Saad, M. Bennis, and B. Maham. Mobility management for heterogeneous networks: Leveraging millimeter wave for seamless handover. In *GLOBECOM 2017-2017 IEEE Global Communications Conference*, pages 1–6. IEEE, 2017.

[59] O. Semiari, W. Saad, M. Bennis, and B. Maham. Caching meets millimeter wave communications for enhanced mobility management in 5g networks. *IEEE Transactions on Wireless Communications*, 17(2):779–793, 2018.

[60] M. S. Sim, J. Park, C.-B. Chae, and R. W. Heath. Compressed channel feedback for correlated massive mimo systems. *Journal of Communications and Networks*, 18(1):95–104, 2016.

[61] S. Singh, M. N. Kulkarni, A. Ghosh, and J. G. Andrews. Tractable model for rate in self-backhauled millimeter wave cellular networks. *IEEE Journal on Selected Areas in Communications*, 33(10):2196–2211, 2015.

[62] L. Sun, J. Hou, and T. Shu. Spatial and temporal contextual multi-armed bandit handovers in ultra-dense mmwave cellular networks. *IEEE Transactions on Mobile Computing*, 20(12):3423–3438, 2020.

[63] L. Sun, J. Hou, and T. Shu. Bandwidth-efficient precoding in cell-free massive mimo networks with rician fading channels. In *2021 18th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, pages 1–9. IEEE, 2021.

[64] L. Sun, J. Hou, and T. Shu. Optimal handover policy for mmwave cellular networks: A multi-armed bandit approach. In *GLOBECOM 2019-2019 IEEE Global Communications Conference*. IEEE, to appear.

[65] 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*, 17(6):1456–1468, 2017.

[66] 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*, (6):1456–1468, 2018.

[67] S. Sur, X. Zhang, P. Ramanathan, and R. Chandra. Beamspy: enabling robust 60 ghz links under blockage. In *13th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 16)*, pages 193–206, 2016.

[68] H. Tabassum, M. Salehi, and E. Hossain. Fundamentals of mobility-aware performance characterization of cellular networks: A tutorial. *IEEE Communications Surveys & Tutorials*, 21(3):2288–2308, 2019.

[69] A. Talukdar, M. Cudak, and A. Ghosh. Handoff rates for millimeterwave 5g systems. In *2014 IEEE 79th Vehicular Technology Conference (VTC Spring)*, pages 1–5. IEEE, 2014.

[70] T. X. Vu, H. D. Nguyen, and T. Q. Quek. Adaptive compression and joint detection for fronthaul uplinks in cloud radio access networks. *IEEE Transactions on Communications*, 63(11):4565–4575, 2015.

[71] J. Wang, L. Dai, L. Yang, and B. Bai. Rate-constrained network decomposition for clustered cell-free networking. In *ICC 2022-IEEE International Conference on Communications*, pages 2549–2554. IEEE, 2022.

[72] T. Wang, C.-K. Wen, S. Jin, and G. Y. Li. Deep learning-based csi feedback approach for time-varying massive mimo channels. *IEEE Wireless Communications Letters*, 8(2):416–419, 2018.

[73] 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*, 5(6):4296–4307, 2018.

[74] C.-K. Wen, W.-T. Shih, and S. Jin. Deep learning for massive mimo csi feedback. *IEEE Wireless Communications Letters*, 7(5):748–751, 2018.

[75] J. Wu, Y. Zhang, M. Zukerman, and E. K.-N. Yung. Energy-efficient base-stations sleep-mode techniques in green cellular networks: A survey. *IEEE communications surveys & tutorials*, 17(2):803–826, 2015.

[76] Y. Xu, H. Shokri-Ghadikolaei, and C. Fischione. Distributed association and relaying with fairness in millimeter wave networks. *IEEE Transactions on Wireless Communications*, 15(12):7955–7970, 2016.

[77] Z. Xu, J. Tang, C. Yin, Y. Wang, G. Xue, J. Wang, and M. C. Gursoy. Recarl: resource allocation in cloud rans with deep reinforcement learning. *IEEE Transactions on Mobile Computing*, 2020.

[78] Q. Yang, M. B. Mashhadi, and D. Gündüz. Deep convolutional compression for massive mimo csi feedback. In *2019 IEEE 29th international workshop on machine learning for signal processing (MLSP)*, pages 1–6. IEEE, 2019.

[79] S. Zang, W. Bao, P. L. Yeoh, H. Chen, Z. Lin, B. Vucetic, and Y. Li. Mobility handover optimization in millimeter wave heterogeneous networks. In *2017 17th International symposium on communications and information technologies (ISCIT)*, pages 1–6. IEEE, 2017.

[80] Q. Zhang, S. Jin, K.-K. Wong, H. Zhu, and M. Matthaiou. Power scaling of uplink massive mimo systems with arbitrary-rank channel means. *IEEE Journal of Selected Topics in Signal Processing*, 8(5):966–981, 2014.

[81] T. Zhang, A. Ge, N. C. Beaulieu, Z. Hu, and J. Loo. A limited feedback scheme for massive mimo systems based on principal component analysis. *EURASIP Journal on Advances in Signal Processing*, 2016(1):1–12, 2016.

[82] F. Zhao, H. Tian, G. Nie, and H. Wu. Received signal strength prediction based multi-connectivity handover scheme for ultra-dense networks. In *2018 24th Asia-Pacific Conference on Communications (APCC)*, pages 233–238. IEEE, 2018.

[83] A. Zhou, X. Zhang, and H. Ma. Beam-forecast: Facilitating mobile 60 ghz networks via model-driven beam steering. In *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*, pages 1–9. IEEE, 2017.

[84] Y. Zhu, Z. Zhang, Z. Marzi, C. Nelson, U. Madhow, B. Y. Zhao, and H. Zheng. Demystifying 60ghz outdoor picocells. In *Proceedings of the 20th annual international conference on Mobile computing and networking*, pages 5–16. ACM, 2014.