CITY UNIVERSITY OF HONG KONG
香港城市大學

Optimal Placement of Access Points for Infrastructure-based Wireless Communication Networks

無線通信網絡的最優接入點部署方案研究

Submitted to
Department of Electrical Engineering
電機工程學系
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
哲學博士學位

by

ZHANG Yue
張鉞

February 2021
二零二一年二月

Abstract

The widespread popularity of wireless smart devices and proliferation of multimedia services have posed unprecedented challenges to the designers of wireless communication networks. To meet the fast growing data traffic demand, a great number of wireless access points (APs) should be deployed, which, nevertheless, would also cause severe interference if they are not properly placed. Therefore, it is of paramount importance to study how to optimize the AP placement of infrastructure-based wireless communication networks.

The difficulty of optimizing network topology mainly lies in the time-varying nature of users' positions. To address this issue, most of the existing works assumed that either the positions of users are fixed, or the spatial distributions of users are known. The problems were then solved with deterministic optimization techniques. In practice, however, it is challenging to characterize the distribution of users. Instead, sample positions of mobile users that are constantly measured by APs and stored in the central server are abundant. Therefore, stochastic optimization approaches, which rely upon a collection of samples to produce statistical estimates of the random variables, could be more suitable for solving network topology optimization problems.

This thesis is devoted to the AP placement optimization of infrastructure-based wireless communication networks, including distributed antenna systems (DASs), millimeter wave (mmWave) cellular networks and IEEE 802.11 networks. It begins from the base station (BS) antenna layout optimization of a multi-user DAS. The objective is formulated as the uplink ergodic sum capacity averaged over the distribution of users. A computationally efficient algorithm is proposed to solve it based on stochastic gradient descent. In order to calculate the gradient of the ergodic sum capacity, an accurate closed-form approximation of

ergodic sum capacity is also derived. Simulation results corroborate that the proposed algorithm converges fast and significantly outperforms the existing representative BS antenna placement schemes.

The stochastic optimization framework is further extended to optimize the BS placement in mmWave cellular networks. The aim is to minimize the long-term outage probability, which is related to the BS placement and user association. The BS placement problem is formulated as maximizing the average number of physically accessible BSs of each user under an inaccessible probability constraint. As both the objective and constraint functions are in the form of expectation, a novel algorithm based on cooperative stochastic approximation (CSA) is proposed to effectively determine the optimal placement of BSs. For user association, a load-balancing user association scheme is proposed, which can achieve similar outage performance to the Hungarian-algorithm-based optimal user association scheme, but with much less running time. Combined with the proposed user association scheme, the proposed BS placement scheme can significantly improve the network outage probability in the long term, especially when the aggregation degree of users is large.

Finally, the optimal placement and coverage of APs for IEEE 802.11 networks are investigated under the stochastic optimization framework. The objective is to maximize the average network throughput under an outage probability constraint. A CSA-based algorithm is developed to solve the problem. Simulation results demonstrate that the proposed AP placement algorithm can quickly converge, and gains over the representative grid installation AP placement scheme increase as the outage constraint becomes loose or the users' spatial distribution becomes more clustered. The methodology in this thesis offers important insights into practical network topology design for infrastructure-based wireless communication networks.

Acknowledgements

Every journey has an end. At the moment of finishing this dissertation, I realize that I am so lucky to have met and received help from many people in the past five years. In this acknowledgement, I would like to deliver my sincere appreciation to them.

First of all, I would like to express my deepest gratitude to my supervisor, Prof. Lin Dai, for her continuous support and patient guidance during my Ph.D. study. I will cherish the memory of our numerous face-to-face meetings and discussions, from which I have learnt a lot of means for developing systematic thinking and academic writing skills. Prof. Dai is a sharp thinker and could always help me find a way out when I was having difficulty with some research problems. Her enthusiasm for academia and rigorous attitude towards research have deeply influenced me during these years and will inspire me to pursue excellence in my future academic life. It is a great honor for me to have been one of her students.

Second, I would like to offer my thanks to my co-supervisor, Prof. Eric Wong, for his constant encouragement and valuable suggestions. His optimism always shined through for me. I have learnt a lot from him about how to manage my emotions. Special thanks should be given to Prof. Stephen Boyd for introducing to me the stochastic optimization tools, which greatly contributed to the completion of this thesis. I also highly appreciate the generous support from Prof. Albert Sung, Prof. Taejoon Kim, Dr. Bo Bai, and Prof. Xijun Wang. I have benefited a lot from their insightful comments and valuable discussions.

Many thanks also go to my groupmates Yayu Gao, Junyuan Wang, Wen Zhan, Pei Liu, Yitong Li, Zhiyang Liu, Xinghua Sun, and Nian

Peng. They were always helpful when I encountered problems or difficulties in my research. I am also grateful to my officemates Shansuo Liang, Lei Liu, Chulong Liang, Hong Wang, Yaru Fu, Salwa Mostafa, Jiajun Chen, Miaomiao Dong, Rongbin Zhang, Guojun Xiong, Wei Zhang, Hao Wang, Zhanglei Shi, Chao Wang, and many others. I will always remember the happy time we spent together.

Last but not least, I would never have been able to finish my Ph.D. study without the unconditional love, support, and inspiration from my family: my mom, my dad, and my grandma. I thank them from the bottom of my heart.

Contents

List of Figures

Abbreviations

AO	Asymptotically-Orthogonal
AP	Access Point
B&B	Branch-and-Bound
BS	Base Station
BSS	Basic Service Set
CS	Candidate Site
CSA	Cooperative Stochastic Approximation
CST	Carrier Sense Threshold
DAS	Distributed Antenna System
DCF	Distributed Coordination Function
DR	Doubly-Regular
i.i.d.	Independent and Identically Distributed
LoS	Line-of-Sight
MIMO	Multiple-Input-Multiple-Output
mmWave	Millimeter Wave
PDF	Probability Density Function
RP	Reference Point

RSSI	Received-Signal-Strength-Indicator
RTS/CTS	Request-to-Send/Clear-to-Send
SA	Stochastic Approximation
SAA	Sample-Average Approximation
SGD	Stochastic Gradient Descent
SINR	Signal-to-Interference-plus-Noise Ratio
SNR	Signal-to-Noise Ratio

Notations

$\circ$	Hadamard product
$\mathbb{E}[\cdot]$	Expectation operator
$\lvert\mathcal{X}\rvert$	Cardinality of a set $\mathcal{X}$
$\lVert\mathbf{x}\rVert$	Euclidean norm of a vector $\mathbf{x}$
$\mathbf{X}'$	Transpose of a matrix $\mathbf{X}$
$\mathbf{X}^\dagger$	Conjugate transpose of a matrix $\mathbf{X}$
$\det(\mathbf{X})$	Determinant of a matrix $\mathbf{X}$
$\mathbf{I}_N$	An $N \times N$ identity matrix
$\mathbb{C}$	Set of complex numbers
$\mathbb{R}$	Set of real numbers
$\mathbb{R}_+$	Set of positive real numbers
M	Total number of BS antennas in a DAS
N	Total number of user antennas in a DAS
N_c	Number of antennas at each BS antenna cluster in a DAS
N_u	Number of antennas at each user in a DAS
L	Number of APs
$\bar{L}_{\text{ac}}$	Average number of accessible BSs of each user in an mmWave

	cellular network
K	Number of users
$K_{\max}$	Maximum number of users that can be served by one BS in an mmWave cellular network
K_{out}	Number of users outside the coverage areas of all the APs in an IEEE 802.11 network
$\mathcal{L}$	Set of APs
$\mathcal{L}_k$	Set of BSs that include user k in their coverage areas in an mmWave cellular network
$\mathcal{K}$	Set of users
$\mathcal{K}_i^*$	Set of users who associate with AP i in an IEEE 802.11 network
$\mathbf{r}_l^{\text{AP}}$	Position of AP l
$\mathbf{r}_l^{\text{AP}*}$	Optimized position of AP l
$\mathbf{r}_k^{\text{user}}$	Position of user k
$\mathbf{G}$	Channel gain matrix between users and BS antennas in a DAS
$\mathbf{H}$	Small-scale fading matrix between users and BS antennas in a DAS
$\mathbf{\Gamma}$	Large-scale fading matrix between users and BS antennas in a DAS
$\gamma_{l,k}$	Large-scale fading coefficient between BS antenna cluster l and user k in a DAS
$d_{l,k}$	Distance between AP l and user k
α	Path-loss factor
β	Ratio of the number of user antennas to the number of BS antennas at each cluster in a DAS

ρ_0	Transmit SNR of each user in a DAS
C	Uplink normalized ergodic sum capacity of a DAS
$\mathcal{C}$	Asymptotic normalized ergodic sum capacity of a DAS
$\tilde{\mathcal{C}}^{(0)}$	Zero-order approximation of the asymptotic normalized ergodic sum capacity $\mathcal{C}$
$\tilde{\mathcal{C}}^{(1)}$	First-order approximation of the asymptotic normalized ergodic sum capacity $\mathcal{C}$
$\mathcal{C}_{DR}$	DR approximation of the asymptotic normalized ergodic sum capacity $\mathcal{C}$
$\mathcal{C}_{AO}$	AO approximation of the asymptotic normalized ergodic sum capacity $\mathcal{C}$
ϵ	Euler-Mascheroni constant
Ω	Ratio of the total number of BS antennas to the total number of user antennas in a DAS
$\mathbb{W}_0(\cdot)$	Principal branch of Lambert W function
$\Pi_{\mathcal{X}}(\mathbf{x})$	Euclidean projection of a vector $\mathbf{x}$ on a set $\mathcal{X}$
$\mathcal{A}$	Area where APs are placed
T	Total number of iterations of each algorithm
μ_δ	Aggregation factor of users' distributions
E_{bui}	Side length of each block in an mmWave cellular netowrk
W_{str}	Street width in an mmWave cellular netowrk
R_{B}	Coverage radius of each BS in an mmWave cellular network
R_{A}	Coverage radius of each AP in an IEEE 802.11 network
R_{A}^*	Optimized coverage radius of each AP in an IEEE 802.11 network
$I(\cdot)$	Indicator function

$\mathcal{I}_{l,k}^{\text{out}}$	Indicator of whether user k is inside the coverage range of BS l in an mmWave cellular network
$\mathcal{I}_{l,k}^{\text{block}}$	Indicator of whether there is a LoS path between BS l and user k in an mmWave cellular network
$\mathcal{I}_{l,k}^{\text{associate}}$	Indicator of whether user k associates with BS l in an mmWave cellular network
$S(\cdot)$	Sigmoid function
ξ	Steepness parameter of the sigmoid function
P_{iac}	Inaccessible probability of an mmWave cellular network
P_{iac}^{*}	Maximum average inaccessible probability of an mmWave cellular network
P_{outB}	Instantaneous outage probability of an mmWave cellular network
$\bar{P}_{\text{outA}}$	Maximum allowable outage probability of an IEEE 802.11 network
$\mathcal{G}_{i,\mathcal{S}}$	Group of users who associate with AP i and can be heard by the APs in the set $\mathcal{S}$ in an IEEE 802.11 network
$k^{(i,\mathcal{S})}$	Number of users in Group $\mathcal{G}_{i,\mathcal{S}}$
$\hat{\lambda}$	Total network throughput of an IEEE 802.11 network
$\lambda^{(i,\mathcal{S})}$	Throughput of a user from Group $\mathcal{G}_{i,\mathcal{S}}$
$\varepsilon^{(i,\mathcal{S})}$	Probability of sensing the channels of BSS set $\mathcal{S}$ idle by users in Group $\mathcal{G}_{i,\mathcal{S}}$
ϑ	Initial backoff window size of each user in an IEEE 802.11 network
ϱ	Cutoff phase of each user in an IEEE 802.11 network
τ_T	Successful transmission time in an IEEE 802.11 network
τ_F	Collision time in an IEEE 802.11 network
$p^{(j)}$	Limiting probability of successful transmission of head-of-line packets

	of BSS j given that the channel of BSS j is idle in an IEEE 802.11 network
$\tilde{\boldsymbol{\omega}}_{\mathcal{C}}$	Gradient of the asymptotic normalized ergodic sum capacity $\mathcal{C}$ in terms of the positions of BS antenna clusters in a DAS
$\tilde{\boldsymbol{\omega}}_{\bar{L}_{\text{ac}}}$	Gradient of the average number of accessible BSs of each user $\bar{L}_{\text{ac}}$ in terms of the positions of BSs in an mmWave cellular network
$\tilde{\boldsymbol{\omega}}_{P_{\text{iac}}}$	Gradient of the inaccessible probability P_{iac} in terms of the positions of BSs in an mmWave cellular network
$\tilde{\boldsymbol{\omega}}_{\text{B}}^{\text{cover}}$	Gradient of the coverage component in terms of the positions of BSs in an mmWave cellular network
$\tilde{\boldsymbol{\omega}}_{\text{B}}^{\text{block}}$	Gradient of the blockage component in terms of the positions of BSs in an mmWave cellular network
$\tilde{\boldsymbol{\omega}}_{\hat{\lambda}}$	Gradient of the total network throughput $\hat{\lambda}$ in terms of the positions and coverage radius of APs in an IEEE 802.11 network
$\tilde{\boldsymbol{\omega}}_{K_{\text{out}}}$	Gradient of the number of users outside the coverage areas of all the APs K_{out} in terms of the positions and coverage radius of APs in an IEEE 802.11 network
η	Scalar step size of SA-based algorithms
$\boldsymbol{\eta}$	Vectorized step size of SA-based algorithms
η^{cover}	Step size corresponding to $\tilde{\boldsymbol{\omega}}_{\text{B}}^{\text{cover}}$ of the CSA-based optimal BS placement scheme (Algorithm 2)
η^{block}	Step size corresponding to $\tilde{\boldsymbol{\omega}}_{\text{B}}^{\text{block}}$ of the CSA-based optimal BS placement scheme (Algorithm 2)

Chapter 1

Introduction

1.1 Infrastructure-based Wireless Communication Networks

1.1.1 Wireless Communication: The Technology Shaping Our World

"Wireless is freedom. It's about being unleashed from the telephone cord and having the ability to be virtually anywhere when you want to be."

—— *Martin Cooper*

This quote reveals how wireless communication has profoundly impacted our world. With handheld communication devices set free from wires and the number of wireless access points (APs) growing exponentially, individuals can easily contact each other from anywhere in the world. This "wireless freedom" not only promotes worldwide interchange of information, but also greatly boosts people's productivity. In 2020, with the spread of pandemic COVID-19 across more than 200 countries, many employees have had to work from home and collaborate remotely with their teammates in different places. When many companies realized that such working mode could save workplace maintenance cost and offer employees great flexibility, they were willing to continue this policy for a substantially long time, or even permanently [1]. The flexible and productive remote-work setup could not have been imaginable without the well-developed wireless communication infrastructure.

Besides providing ubiquitous wireless access, the rapid development of wireless communication technology has also significantly increased the transmission data rate, which in turn has bred many earthshaking applications. For instance, video live streaming and video-sharing social networking services, such as Twitch and TikTok, are becoming more and more popular nowadays. These applications have revolutionized the ways people access information and entertain themselves. Without a doubt, our way of living has been dramatically changed by wireless communication technology and its related applications.

1.1.2 Infrastructure-based Wireless Networks: Concepts and Features

Despite the flexibility and efficiency offered by wireless communications, it is more complicated to establish stable wireless connections between nodes compared with wired options. Generally, multiple wireless nodes are interconnected to form a wireless network and operate under specific protocols for effective information transmission. According to the different network architectures, wireless networks can be categorized into two classes: ad-hoc wireless networks and infrastructure-based wireless networks [2].

An ad-hoc wireless network consists of a group of self-configured and self-organized wireless nodes, as shown in Fig. 1.1a. The transmission between two nodes which are far apart is through multi-hop routing with the aid of intermediate nodes. Ad-hoc networks are highly appealing for many applications, including hazard monitoring and military attack detection thanks to their high flexibility and relatively low deployment cost. However, the distributed control manner and multi-hop routing make ad-hoc networks difficult to support high-data-rate and low-latency applications like multimedia services.

The drawbacks of ad-hoc wireless networks can be overcome by infrastructure-based wireless networks. In an infrastructure-based wireless network, there are two kinds of wireless nodes, namely mobile users and wireless APs. Mobile users have time-varying positions due to their mobility, while wireless APs are facilities that have fixed locations. Furthermore, user equipment is usually powered by batteries with limited capacity for portability. Yet wireless APs could have plentiful power

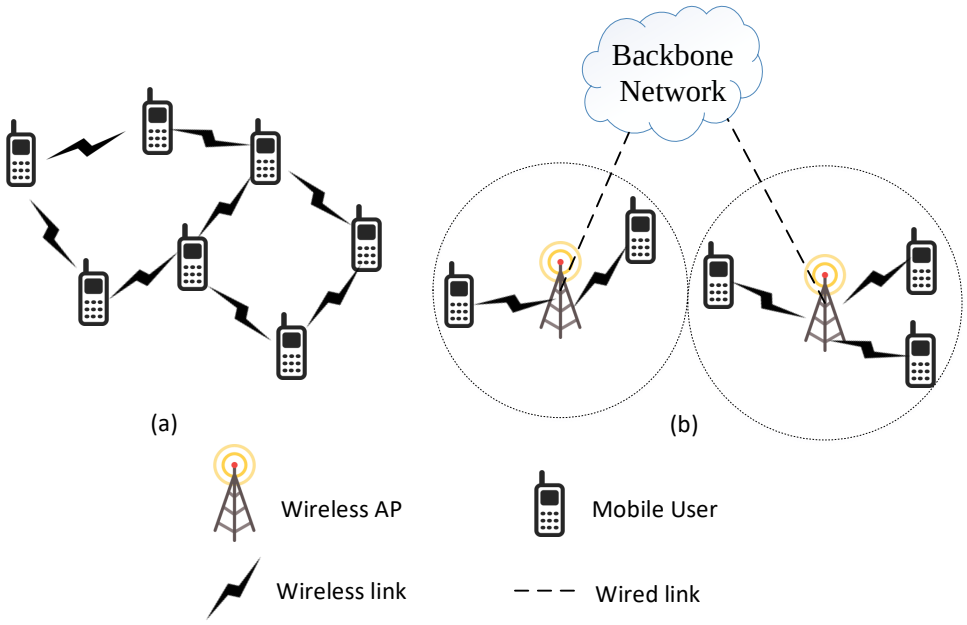

FIGURE 1.1: Graphic illustration of (a) an ad-hoc wireless network and (b) an infrastructure-based wireless network.

supply and are equipped with powerful central processing units, making them capable of providing high-data-rate services for mobile users.

Fig. 1.1b presents a graphic illustration of an infrastructure-based wireless network. As shown in the figure, the radio region of interest is divided into sub-regions according to the coverage range of each AP. The area covered by one AP is called a cell. Each mobile user communicates with the closest AP, and all APs are cabled to the backbone network for inter-AP transmissions. Such a network structure is widely known as the cellular structure, and has been adopted by many prevalent infrastructure-based wireless networks such as cellular mobile networks and Wi-Fi networks.

In sharp contrast to wired systems which use point-to-point links for data transmissions, in a wireless network, one AP may serve multiple users and communicate with them through a single shared wireless channel. To avoid severe intra-cell interference, proper radio resource control and transmission coordination among users should be adopted in each cell. On the other hand, as the transmissions in different cells are uncoordinated, in order to suppress inter-cell interference, neighboring cells in traditional cellular mobile networks and Wi-Fi networks could operate in different frequency bands under various frequency reuse schemes. The drawback of conventional frequency reuse schemes is the low bandwidth utilization efficiency as spectrum recourses are split among cells. To meet

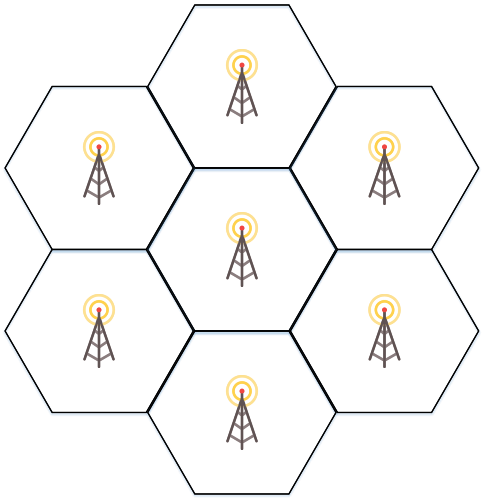

FIGURE 1.2: Graphic illustration of an infrastructure-based wireless network with regular AP placement scheme.

the ever-increasing data traffic demand, the universal frequency reuse scheme, i.e., all APs operate in the same frequency band, has become more popular nowadays and has been adopted in the latest cellular systems [3]. In that case, inter-cell interference could be significant and should be mitigated by other techniques such as appropriate AP placement design.

1.1.3 AP Placement: A Key Issue of Infrastructure-based Wireless Networks

In view of the limited coverage range of each AP, a fundamental conflict in infrastructure-based wireless networks arises between the limited budget for AP deployment and the pursuit of ubiquitous network access. In early cellular systems and Wi-Fi networks, the main concern was the network coverage performance. Therefore, network designers usually place APs regularly on a hexagonal grid to extend the network coverage, as shown in Fig. 1.2. However, in practice, APs may not be placed in a regular manner due to complicated geographic conditions. Instead, random AP layouts, where APs are uniformly placed, were considered in [4–10]. In these works, the focus was shifted to studying the impact of AP density on network performance such as energy efficiency [4], average user throughput [5], received signal quality [6] and connectivity/coverage probability [7–10]. In [5–8], it was shown that the increase of AP density is always beneficial to the networks, while in [4, 9, 10], the authors showed that when the AP density is beyond a certain threshold, the network performance becomes even worse.

It should be noted that the results in [4–10] were all based on the assumption that users are uniformly distributed. If users are non-uniformly distributed, e.g., gathering around several hotspots, placing APs regularly or uniformly in the region is not economical, as some APs which are far away from user-intensive areas could remain idle most of the time, leading to a waste of resources. Even with APs placed in the user-intensive areas, the network performance may not always be improved if their sites are selected improperly. For instance, it has been shown in [11] that for Wi-Fi networks which adopt the carrier sense multiple access (CSMA) protocol for user access control, if all the APs share the same frequency band and have largely overlapping coverage areas, increasing the number of APs cannot benefit the network throughput at all. Furthermore, the results in [9, 10] illustrate that when the AP density exceeds a certain value, the probability that a typical user's signal-to-interference-plus-noise ratio (SINR) is above a pre-designated threshold even decreases with the AP density. The intuition behind this finding is that network densification not only reduces the average access distance between one user and its associated AP, but also decreases the average distance between the user and the interfering APs. As APs are assumed to be randomly deployed without optimization in [9, 10], the inter-cell interference can be further reduced by optimizing the AP placement, e.g., minimizing the size of the overlapping area of different APs. In fact, it has been shown in [12] that optimized AP placement can achieve identical downlink coverage performance to that of random AP placement, but with much fewer APs required.

The above studies have shown that AP placement has great impacts on the performance of infrastructure-based wireless networks. In addition to [12], various performance gains brought about by AP placement optimization have also been demonstrated in the literature [13–40]. In the next section, we present an overview of the existing works on AP placement optimization for various infrastructure-based wireless networks.

1.2 AP Placement Optimization

1.2.1 Previous Work

The study of the optimal AP placement of infrastructure-based wireless networks dates back to the 1970s, when the first cellular system was launched. In

that system, the sites of APs were designed to form a hexagonal pattern as shown in Fig. 1.2. It was shown in [41] that this AP placement is the optimal one to maximize the network coverage under the assumption that the coverage range of each AP is an equal-sized disc.

In spite of its simplicity, the regular AP placement scheme may not be cost-effective when users are non-uniformly distributed, as some of the APs could rarely be active. A more economical placement scheme should take into account the distribution of users, which has been studied in the literature [12–36]. In these works, the distribution of users is represented by a collection of reference points (RPs) with fixed locations, with some aiming at minimizing the implementation cost for the given requirements regarding the traffic demand or signal quality at each RP [12–22], and others focusing on the placement of a given number of APs for optimizing the network performance, including received signal quality [23–30], positioning accuracy [31, 32], total network throughput [33, 34] and energy efficiency [35, 36]. For simplicity, most of these works discretized the radio coverage region into a finite number of candidate sites (CSs), and the AP placement problems become finding the optimal subset of CSs. Due to the NP-hardness of the optimization problems, various heuristic algorithms were proposed to find sub-optimal solutions.

In the aforementioned literature, the focus is usually placed on the downlink coverage performance, in which case it is reasonable to use a group of RPs to represent the users' positions, as a small deviation of users' positions has negligible influence on the network coverage performance. However, for the uplink, access performance of users is crucially determined by their positions. Different from the downlink case where users can receive signals from an AP as long as they are within its coverage range, in the uplink, users have to contend for the limited radio resources, and may not be able to access the AP even within its coverage when the contention level is high. A slight change of users' positions may also lead to different association results of users and APs, which greatly affect the uplink access performance. Therefore, AP placement optimized based on a group of fixed-location RPs would quickly become obsolete when the users' positions change.

To further take into account the time-varying nature of users' positions, performance metrics that are averaged over the spatial distribution of users should be adopted. This kind of AP placement problems was considered in [37–40], where

the objectives were to maximize the uplink sum rate/capacity averaged over users' spatial distribution. Due to the lack of closed-form expressions of the objective functions, various approximations have been proposed. Specifically, the objectives were replaced by minimizing the average access distance between a user and its closest AP in [37] or maximizing bounds of uplink average capacity/rate in [38–40].

In all the aforementioned existing works [12–40], optimal AP placement is usually formulated as a deterministic optimization problem, where users are either represented by a group of fixed-location RPs [12–36], or assumed to follow a certain spatial distribution [37–40]. For AP placement optimization with user-distribution-sensitive objectives such as optimizing the uplink access efficiency or maximizing the uplink capacity/rate, the derivation of average performance of users is usually found to be extremely challenging. In fact, even if a closed-form expression of the objective function could be developed and optimized, it has to be based on accurate characterization of users' theoretical spatial distribution, which is usually unavailable in practice. Instead, sample positions of mobile users that are constantly measured by APs and stored in the central server are abundant [42, 43], and could be well exploited for optimizing the placement of APs. That motivated us to resort to stochastic optimization approaches. As we demonstrate in the next section, instead of requiring complete knowledge of the distribution of the random variables involved in the problems, stochastic optimization methods only require samples of the random variables to solve the problems. This property makes stochastic optimization methods highly appealing for solving AP placement problems, as samples of users' positions can be easily obtained in practice. In the next section, we briefly introduce stochastic optimization and demonstrate how it can be applied for solving AP placement problems.

1.2.2 Stochastic Optimization Methods

Stochastic optimization is an umbrella term that covers a great number of approaches. Generally, there is no single stochastic optimization method that works well for all stochastic optimization problems. The most important factor that should be considered when selecting the optimization method is the solution type of problems. Based on different types of solutions, stochastic optimization problems can be roughly divided into two classes: single-stage problem and multi-stage problem [44, 45]. A single-stage stochastic optimization problem tries to find

a single, optimal solution, which can be obtained without subsequent recourse. One example is finding the optimal parameters of an artificial neural network to minimize the loss function. In contrast, a multi-stage stochastic optimization problem tries to find a sequence of solutions that maximizes a cumulative reward. Usually the solutions in the future will be affected by the past ones. One instance is investing in stocks over a certain period with a given amount of assets. In this thesis, we focus on the deployment of APs, which is a one-time setup. Therefore, we pay more attention to the methods for solving single-stage problems in the remaining part of this section.

For illustration, let us present a canonical single-stage stochastic optimization problem as follows:

$$\max_{\mathbf{x}\in\mathcal{X}} \mathbb{E}_{\mathbf{y}}[F(\mathbf{x},\mathbf{y})], \tag{1.1}$$

where the expectation objective cannot be directly calculated because either the distribution of random variable $\mathbf{y}$ is unknown or the expression of the function $F(\mathbf{x},\mathbf{y})$ is overwhelmingly complicated. Only a collection of samples $\{\mathbf{y}(t) : t = 1,\cdots,T\}$ is available, which can be used to calculate $F(\mathbf{x},\mathbf{y}(t)), t = 1,\cdots,T$ for any $\mathbf{x} \in \mathcal{X}$. Stochastic optimization methods then exploit this collection of samples to produce statistical estimates of the random variables and solve the problem.

Among the various stochastic optimization methods, the most straightforward way is using sample-average $\frac{1}{T}\sum_{t=1}^{T} F(\mathbf{x},\mathbf{y}(t))$ to replace the expectation objective in (1.1). Such an approach is known as the sample-average approximation (SAA) technique [46] or batch method [47]. However, in order to accurately approximate the expectation, the number of samples should be large, which would result in high computational complexity. A category of methods possessing lower complexity is called stochastic search methods, which solve Problem (1.1) by iteratively updating the decision variable $\mathbf{x}$. In each iteration, the updating direction is calculated with one sample of $\mathbf{y}$.

Typical stochastic search methods include metaheuristic algorithms [48], stochastic approximation (SA) methods [49, 50], the stochastic quasi-Newton method [51], and stochastic successive convex approximation [52, 53]. Metaheuristic algorithms are a series of model-free algorithms such as ant colony algorithm and simulated annealing algorithm. The advantage of these algorithms is that they

can solve Problem (1.1) even if the expression of $F(\mathbf{x}, \mathbf{y})$ is unknown and its value is generated by a "black box". However, they usually require a large number of iterations for convergence, as they cannot exploit the structural information of $F(\mathbf{x}, \mathbf{y})$. SA methods, also known as stochastic gradient methods in stochastic optimization, solve the problem by iteratively updating $\mathbf{x}$ along a stochastic gradient of function $F(\mathbf{x}, \mathbf{y}(t))$, calculated with one sample $\mathbf{y}(t)$ of $\mathbf{y}$. The problem with SA methods is that they may be trapped into a sub-optimal stationary point if the objective function is non-convex. Compared to SA methods, the stochastic quasi-Newton method and stochastic successive convex approximation can more easily escape from sub-optimal stationary points and can converge with a smaller number of iterations, as they exploit higher order information of the function $F(\mathbf{x}, \mathbf{y})$. Yet these two methods entail higher per-iteration computational cost than SA methods.

Based on the above elaboration, we can solve the user-distribution-sensitive AP placement problems with stochastic optimization methods. In particular, we can take $\mathbf{x}$, $\mathbf{y}$ and $F(\mathbf{x}, \mathbf{y})$ in Problem (1.1) as the positions of APs, the positions of users and the network utility function such as network sum capacity or average received signal strength, respectively. Then Problem (1.1) becomes a typical AP placement problem. Compared to deterministic optimization whose performance relies heavily on accurate characterization of the spatial distribution of users, stochastic optimization only requires samples of users' positions that are usually available in practice, making it more appropriate for solving AP placement problems. In this thesis, we adopt SA as the main tool for developing various AP placement algorithms because of its high computational efficiency.

1.3 Thesis Contributions and Outline

This thesis is devoted to designing optimal AP placement schemes for various infrastructure-based wireless networks with different objectives. We formulate the AP placement problems as stochastic optimization problems and develop various SA-based algorithms to solve them. The contributions of this thesis are summarized as below.

We start from establishing a stochastic optimization framework for AP placement optimization. In Chapter 2, a general AP placement problem is first formulated, whose objective and constraint functions are both in the form of expectation. Two SA methods, stochastic gradient descent (SGD) and cooperative SA (CSA), which can be used for solving unconstrained and expectation-constrained stochastic optimization problems, respectively, are then introduced in detail. Several key issues of SA algorithm design for solving various AP placement problems are also discussed.

Based on the proposed stochastic optimization framework, we focus on various AP placement problems in different types of infrastructure-based wireless networks, including distributed antenna systems (DASs), millimeter wave (mmWave) cellular networks, and Wi-Fi networks in Chapters 3–5, respectively. Specifically, in Chapter 3, we focus on optimizing the AP placement for maximizing the average ergodic sum capacity of a multi-user DAS. An SGD-based algorithm is developed to solve the optimization problem. To apply SGD, we need to calculate the gradient of the ergodic sum capacity in terms of the positions of APs, which is, nevertheless, difficult to obtain due to the lack of closed-form expression of ergodic sum capacity. To address this issue, we propose two novel closed-form approximations of the ergodic sum capacity, which are shown to be much more accurate than the previous results. Simulation results corroborate that the proposed SGD-based AP placement algorithm significantly outperforms the existing representative AP placement schemes, especially when the aggregation degree of users or the number of APs is large.

In Chapter 4, we aim to minimize the long-term outage probability of an mmWave communication network by optimizing the AP placement and user association. Note that mmWave signals experience high path loss and are sensitive to blockage. Therefore, our goal is to optimize the physical accessibility between users and APs. With the time-varying nature of users' positions taken into account, the AP placement problem is formulated as maximizing the average number of accessible APs of each user under a constraint of inaccessible probability, i.e., the long-term proportion of users which have no physically accessible AP. A novel CSA-based algorithm is proposed to effectively determine the optimal placement of APs. With the optimized positions of APs, we then focus on the user association problem in a given time slot, with the objective of minimizing the outage

probability under the constraint of the maximum workload for each AP. A low-complexity outage mitigation user association scheme is proposed to solve the problem. Simulation results show that the proposed scheme can achieve similar outage performance to the Hungarian-algorithm-based optimal user association scheme, but is much more time-efficient. The outage performance of the proposed user association scheme combined with the proposed AP placement scheme is also examined under various aggregation degrees of users, and significant gains are demonstrated over the previous representative AP placement scheme in [29].

In Chapter 5, the optimization framework is further extended to jointly optimize the positions and coverage radius of APs for a multi-user IEEE 802.11 network. The aim is to maximize the average network throughput under a long-term outage probability constraint, and a new algorithm based on CSA is proposed to solve the problem. The simulation results illustrate that the proposed algorithm can quickly converge and significantly outperform the representative grid installation AP placement scheme in various scenarios. The performance gain becomes more significant when the outage constraint becomes loose or the aggregation degree of users increases.

The remainder of this thesis is organized as follows. Chapter 2 presents the formulation of a general AP placement problem and introduces SA methods. In Chapter 3, closed-form approximations of the ergodic sum capacity of a multi-user DAS are derived, and an SGD-based AP placement scheme is developed to maximize the average ergodic sum capacity. A CSA-based AP placement scheme and a low-complexity user association scheme are proposed to minimize the long-term outage probability of an mmWave communication network in Chapter 4. Chapter 5 shows how to jointly optimize the coverage and placement of APs in order to maximize the average network throughput of a multi-user IEEE 802.11 network under a long-term outage constraint based on CSA. Finally, conclusions are drawn and suggestions for future work are provided in Chapter 6.

Chapter 2

Problem Formulation and Stochastic Approximation-based Optimal AP Placement

---

In this chapter, the AP placement problem in wireless communication networks is formulated. The main mathematical tool, SA, for solving the problem is also introduced in detail. This chapter is organized as follows. Section 2.1 formulates a general AP placement problem. Section 2.2 introduces SA methods and demonstrates how the methods can be used for optimizing the AP placement.

2.1 Problem Formulation

Consider an infrastructure-based wireless communication network with L APs serving K users. All the APs and users are located in a 2-dimensional plane. Let $\mathbf{r}_l^{\text{AP}}, l = 1, \cdots, L$ and $\mathbf{r}_k^{\text{user}}, k = 1, \cdots, K$ denote the positions of AP l and user k, respectively. In this thesis, with the time-varying nature of users' positions taken into account, we aim at optimizing the AP placement of a wireless network for improving its long-term performance. A general AP placement problem can be

formulated as

$$(\text{P1}): \max_{\mathbf{r}^{\text{AP}}} \ f(\mathbf{r}^{\text{AP}}) = \mathbb{E}_{\mathbf{r}^{\text{user}}}[F(\mathbf{r}^{\text{AP}}, \mathbf{r}^{\text{user}})] \tag{2.1}$$

$$\text{s.t. } g(\mathbf{r}^{\text{AP}}) = \mathbb{E}_{\mathbf{r}^{\text{user}}}[G(\mathbf{r}^{\text{AP}}, \mathbf{r}^{\text{user}})] \leq \bar{G}^*, \tag{2.2}$$

$$\mathbf{r}_l^{\text{AP}} \in \mathcal{A}, l = 1, \cdots, L, \tag{2.3}$$

where $\mathbf{r}^{\text{AP}} = [\mathbf{r}_1^{\text{AP}}, \cdots, \mathbf{r}_L^{\text{AP}}]$ denotes the positions of all APs and $\mathbf{r}^{\text{user}} = [\mathbf{r}_1^{\text{user}}, \cdots, \mathbf{r}_K^{\text{user}}]$ denotes the positions of all users. $F(\mathbf{r}^{\text{AP}}, \mathbf{r}^{\text{user}})$ is the network utility function which is desired to be maximized. In Chapters 3–5, it is defined as the ergodic sum capacity of a DAS, the average number of accessible APs of each user in an mmWave network, and the network throughput of a Wi-Fi network, respectively. $G(\mathbf{r}^{\text{AP}}, \mathbf{r}^{\text{user}})$ represents the network constraint function whose expectation should be kept below a certain value $\bar{G}^*$. In Chapters 4 and 5, it is defined as the proportion of users which have no physically accessible AP, and the percentage of users falling outside the coverage areas of all APs, respectively. $\mathcal{A}$ denotes the area within which L APs are deployed. Such area can be a cell of regular shape, e.g., circular cell or square cell, as shown in Chapters 3 and 5. With specific practical geographic constraints taken into account, $\mathcal{A}$ could also have irregular shapes. One example is presented in Chapter 4, where Manhattan-type geometry is considered and APs are confined to being placed outside the buildings.

In Problem P1, both $F(\mathbf{r}^{\text{AP}}, \mathbf{r}^{\text{user}})$ and $G(\mathbf{r}^{\text{AP}}, \mathbf{r}^{\text{user}})$ are functions of the positions of APs $\mathbf{r}^{\text{AP}}$ and positions of users $\mathbf{r}^{\text{user}}$. Due to the attenuation nature of wireless signals, the received signal strength at one user is determined by the distance between it and the transmitting APs. In order to improve the network performance, we should hence optimize $\mathbf{r}^{\text{AP}}$ based on $\mathbf{r}^{\text{user}}$. In practice, however, the positions of users are time-varying due to the mobility, while the positions of APs will not change for a long period once they are deployed. Therefore, it would be more reasonable to optimize $\mathbf{r}^{\text{AP}}$ for improving the long-term performance of the network averaged over the positions of users. That explains why we formulate Problem P1 as maximizing the long-term utility function $\mathbb{E}_{\mathbf{r}^{\text{user}}}[F(\mathbf{r}^{\text{AP}}, \mathbf{r}^{\text{user}})]$ with the long-term constraint function $\mathbb{E}_{\mathbf{r}^{\text{user}}}[G(\mathbf{r}^{\text{AP}}, \mathbf{r}^{\text{user}})]$ bounded by a certain value $\bar{G}^*$.

One way to solve Problem P1 is to calculate $\mathbb{E}_{\mathbf{r}^{\text{user}}}[F(\mathbf{r}^{\text{AP}}, \mathbf{r}^{\text{user}})]$ and $\mathbb{E}_{\mathbf{r}^{\text{user}}}[G(\mathbf{r}^{\text{AP}}, \mathbf{r}^{\text{user}})]$ and turn the problem into a deterministic optimization problem, which can then be solved by many well-developed optimization techniques, e.g., interior-point

methods or quasi-Newton approaches [54]. However, it is extremely challenging to derive $\mathbb{E}_{\mathbf{r}^{\text{user}}}[F(\mathbf{r}^{\text{AP}}, \mathbf{r}^{\text{user}})]$ and $\mathbb{E}_{\mathbf{r}^{\text{user}}}[G(\mathbf{r}^{\text{AP}}, \mathbf{r}^{\text{user}})]$ in practice because of 1) the difficulty of characterizing the theoretical spatial distribution of users and 2) the complicated expressions of $F(\mathbf{r}^{\text{AP}}, \mathbf{r}^{\text{user}})$ and $G(\mathbf{r}^{\text{AP}}, \mathbf{r}^{\text{user}})$.

As demonstrated in Section 1.2.2, although the distribution of users is unknown, samples of users' positions are usually available and can be used for optimizing the placement of APs with stochastic optimization approaches, among which we adopt SA methods to solve Problem P1. In the next section, we introduce two SA methods, SGD and CSA, in detail and show how they can be used to solve Problem P1.

2.2 SA Methods

SA methods are a family of approaches for solving root-finding problems or optimization problems which involve uncertainty [49, 50, 55–59]. The original SA method was proposed by Robbins and Monro in the 1950s for finding the root of a function which cannot be calculated directly, while only noise-corrupted observations whose expectation is equal to the function are available [55]. The SA method in [55] solves the problem by using a stochastic difference equation, which iteratively updates a candidate solution with the noisy observations. SA was later applied for searching for the maximizer of a regression function by Kiefer and Wolfowitz in [56], where solving the optimization problem is equivalent to finding the root of the derivative of the objective function. Furthermore, SA methods were extended to solve multivariate root-finding problems and optimization problems in [57].

After the pioneer works [55–57] which built the theoretical foundation of SA methods, the methods have found many successful applications in various disciplines, including biology [60], economics [61], cybernetics [62], and communication [63]. Particularly, with the proliferation of machine learning in recent years, SA methods have been greatly developed and widely applied to solve large-scale learning problems thanks to its high computational efficiency [64–66]. The substantial development of SA has made it one of the most important mathematical tools for solving stochastic optimization problems [49, 50, 58]. Among the SA methods for solving optimization problems, SGD is one of the most important. Similar to the

original SA method, SGD solves the optimization problem by iteratively updating the decision variable along a stochastic gradient, which is an unbiased estimation of the gradient of the objective function and is calculated with one sample of the random variables involved. However, traditional SGD cannot handle Problem P1 which contains an expectation constraint. Recently, a new SA algorithm called CSA was proposed in [59] to solve expectation-constrained stochastic optimization problems. In the following, let us show how to solve Problem P1, with or without the expectation constraint, based on SGD and CSA.

2.2.1 SGD

Specifically, without the expectation constraint (2.2), Problem P1 can be solved by SGD with the following update [49, 50, 58]:

$$\mathbf{r}^{\mathrm{AP}}(t+1) = \Pi_{\mathcal{A}^L}\big(\mathbf{r}^{\mathrm{AP}}(t) + \tilde{\boldsymbol{\omega}}_F(t)\eta(t)\big), \tag{2.4}$$

where $\tilde{\boldsymbol{\omega}}_F(t)$ denotes the gradient of F in terms of $\mathbf{r}^{\mathrm{AP}}(t)$, which is calculated with one sample of users' positions $\mathbf{r}^{\mathrm{user}}(t)$. $\Pi_{\mathcal{X}}(\mathbf{x})$ is a Euclidean projection of a vector $\mathbf{x}$ on a set $\mathcal{X}$, i.e.,

$$\Pi_{\mathcal{X}}(\mathbf{x}) = \arg\min_{\tilde{\mathbf{x}}}\{\|\mathbf{x}-\tilde{\mathbf{x}}\| \big| \tilde{\mathbf{x}} \in \mathcal{X}\}. \tag{2.5}$$

$\eta(t) > 0$ is the step size.

It has been shown in [58] that if 1) the objective function is a convex function, 2) the term $\|\tilde{\boldsymbol{\omega}}_F(t)\|^2$ is bounded by a finite number for all t, and 3) the step size $\eta(t)$ is square-summable but not summable, i.e.,

$$\sum_{t=0}^{\infty}\eta(t)^2 < \infty, \ \sum_{t=0}^{\infty}\eta(t) = \infty, \tag{2.6}$$

then the SGD algorithm can converge to the global optimum with probability 1.

In general, SGD cannot handle problems with expectation constraints, because in each iteration, SGD projects the decision variables over the feasible set, which is intractable if the constraints are in the form of expectation without closed-form expressions. One possible way to make SGD applicable for expectation-constrained stochastic optimization problems is the penalty-based or primal-dual

approach, which is commonly used to solve deterministic constrained optimization problems [67]. For instance, Problem P1 can be reformulated as $\max_{\mathbf{r}^{\mathrm{AP}} \in \mathcal{A}^L} f(\mathbf{r}^{\mathrm{AP}}) + \mho \min\{0, \bar{G}^* - g(\mathbf{r}^{\mathrm{AP}})\}$ where $\mho > 0$ is a penalty parameter, and can then be handled by SGD. However, the performance of this method relies heavily on the value of the penalty parameter or dual multiplier, which is difficult to be determined in a stochastic optimization problem. In the following, we introduce the CSA algorithm in [59], which can solve problem P1 without introducing an unknown penalty parameter or dual multiplier.

2.2.2 CSA

Similar to SGD, the CSA algorithm can iteratively solve Problem P1 by using the following update [59]:

$$\mathbf{r}^{\mathrm{AP}}(t+1) = \Pi_{\mathcal{A}^L}\big(\mathbf{r}^{\mathrm{AP}}(t) + \tilde{\boldsymbol{\omega}}(t)\eta(t)\big). \tag{2.7}$$

The main difference is the stochastic gradient $\tilde{\boldsymbol{\omega}}(t)$, which is given by

$$\tilde{\boldsymbol{\omega}}(t) = \begin{cases} \tilde{\boldsymbol{\omega}}_F(t) & \text{if } \hat{G}(t) \leq \bar{G}^* \\ -\tilde{\boldsymbol{\omega}}_G(t) & \text{otherwise,} \end{cases} \tag{2.8}$$

where $\tilde{\boldsymbol{\omega}}_G(t)$ denotes the gradient of $G(\mathbf{r}^{\mathrm{AP}}(t), \mathbf{r}^{\mathrm{user}}(t))$ calculated with one sample of users' positions $\mathbf{r}^{\mathrm{user}}(t)$. $\hat{G}(t)$ is an unbiased estimation of $\mathbb{E}_{\mathbf{r}^{\mathrm{user}}}[G(\mathbf{r}^{\mathrm{AP}}, \mathbf{r}^{\mathrm{user}})]$. One way to obtain $\hat{G}(t)$ is to generate J independent and identically distributed (i.i.d.) samples of $\mathbf{r}^{\mathrm{user}}$, $\{\mathbf{r}^{\mathrm{user}}_{(1)}, \cdots, \mathbf{r}^{\mathrm{user}}_{(J)}\}$, and evaluate $\hat{G}(t)$ by $\hat{G}(t) = \frac{1}{J}\sum_{j=1}^{J} G(\mathbf{r}^{\mathrm{AP}}(t), \mathbf{r}^{\mathrm{user}}_{(j)})$. After T iterations, the output of CSA is

$$\bar{\mathbf{r}}^{\mathrm{AP}}_{T,s} = \left(\sum_{t\in\mathcal{D}} \eta(t)\right)^{-1} \left(\sum_{t\in\mathcal{D}} \eta(t)\mathbf{r}^{\mathrm{AP}}(t)\right),$$

where $\mathcal{D} = \{s \leq t \leq T | \hat{G}(t) \leq \bar{G}^*\}$ for some $1 \leq s \leq T$.

We can see from (2.7) and (2.8) that the decision variable $\mathbf{r}^{\mathrm{AP}}$ is updated along either the gradient of the objective function or the opposite of the gradient of the constraint function, depending on whether the estimated constraint is satisfied. By doing so, CSA can maximize the objective function, as well as control the violation of the expectation constraint.

It was shown in [59] that for convex objective function $f(\mathbf{r}^{\text{AP}})$ and convex constraint function $g(\mathbf{r}^{\text{AP}})$ with bounded expectations $\mathbb{E}_{\mathbf{r}^{\text{user}}}[\exp(\|\tilde{\boldsymbol{\omega}}_F(\mathbf{r}^{\text{AP}}, \mathbf{r}^{\text{user}})\|^2)]$, $\mathbb{E}_{\mathbf{r}^{\text{user}}}[\exp(\|\tilde{\boldsymbol{\omega}}_G(\mathbf{r}^{\text{AP}}, \mathbf{r}^{\text{user}})\|^2)]$, $\mathbb{E}_{\mathbf{r}^{\text{user}}}[\exp(\|G(\mathbf{r}^{\text{AP}}, \mathbf{r}^{\text{user}}) - g(\mathbf{r}^{\text{AP}})\|^2)]$, and a constant step size $\eta(t)$, the CSA algorithm exhibits the optimal $\mathcal{O}(1/\Delta^2)$ rate of convergence in terms of both optimality gap and constraint violation Δ. Specifically, if the optimal solution $\mathbf{r}^{\text{AP}*}$ of P1 exists, then for any $\Lambda \in (0, 1)$, we have $\text{Prob}\{f(\bar{\mathbf{r}}^{\text{AP}}_{T,s}) - f(\mathbf{r}^{\text{AP}*}) \leq \iota\} \geq (1 - \Lambda)^2$ and $\text{Prob}\{g(\bar{\mathbf{r}}^{\text{AP}}_{T,s}) \leq \sigma\} \geq 1 - \Lambda$ with the total number of iterations T and the number of samples J bounded by $\mathcal{O}(\max\{\frac{1}{\iota^2}(\log\frac{1}{\Lambda})^2, \frac{1}{\sigma^2}\})$ and $\mathcal{O}(\max\{\frac{1}{\iota^2}(\log\frac{1}{\Lambda})^2, \frac{1}{\sigma^2}\log\frac{1}{\Lambda^3}\})$, respectively.

2.3 SA-based Optimal AP Placement

In Sections 2.2.1 and 2.2.2, we showed that AP placement problem P1, with or without the expectation constraint (2.2), can be solved by SGD and CSA, respectively. However, it should be noted that the performance of SA algorithms is closely related to the parameter setting, e.g., the initial AP placement $\mathbf{r}^{\text{AP}}(0)$ and the step size $\eta(t)$ [66]. Furthermore, in order to derive the stochastic gradient $\tilde{\boldsymbol{\omega}}(t)$, $F(\mathbf{r}^{\text{AP}}, \mathbf{r}^{\text{user}})$ and $G(\mathbf{r}^{\text{AP}}, \mathbf{r}^{\text{user}})$ are required to be differentiable with respect to $\mathbf{r}^{\text{AP}}$, which may not always be satisfied in practice. In the following, we discuss the key issues of SA algorithm design for solving various AP placement problems in Chapters 3–5.

- *Step Size*: The step size (also referred to as learning rate in the context of machine learning [66]) $\eta(t)$ has a great impact on the convergence performance of SA algorithms. In traditional SA, $\eta(t)$ is usually a scalar and a diminishing function of the iteration index t. The convergence of SA is conditioned on certain diminishing rates of $\eta(t)$, as shown in (2.6). Furthermore, $\eta(t)$ can also influence the convergence speed of SA algorithms. Generally, an undersized $\eta(t)$ will lead to slow convergence, while an oversized $\eta(t)$ can cause instability of the algorithm.

  Despite its simplicity, the scalar step size models may result in poor convergence performance if the objective function or constraint function is sensitive to some directions in the variable space [66], where elements of the stochastic gradient would have widely varying magnitudes, rendering it difficult to select a proper scalar step size. In our AP placement problems, a similar

situation occurs if users are non-uniformly distributed and the initial positions of APs $\mathbf{r}^{\mathrm{AP}}(0)$ are randomly generated. In that case, a vectorized step size model would be more appropriate, which adapts its value to the stochastic gradient elementwisely during iterations. In Chapter 3, we adopt the vectorized step size model based on the RMSProp algorithm in [66] for improving the convergence performance of the proposed base station (BS) antenna placement algorithm.

However, the vectorized step size model is not applicable to the gradient with zero elements, as in this model each element of the step size should be finite. Therefore, the vectorized step size cannot help the algorithm escape from the sub-optimal stationary points where some elements of the gradient are zero. In Chapters 4 and 5, the gradients of the objective and constraint functions could have zero elements as both of them involve the sigmoid function, which has a large saturated region where the first derivative is negligibly small. Therefore, we adopt the scalar step size model in these two chapters and use other techniques to improve the convergence performance of the proposed algorithms.

- *Initial Point*: One necessary condition for CSA/SGD to converge to the global optimum is the convexity of the objective function $F(\mathbf{r}^{\mathrm{AP}}, \mathbf{r}^{\mathrm{user}})$ and constraint function $G(\mathbf{r}^{\mathrm{AP}}, \mathbf{r}^{\mathrm{user}})$. Otherwise, the algorithm may be trapped into a sub-optimal stationary point. Whether this undesirable situation occurs or not is crucially determined by the selection of initial point. As we show in Chapter 3, the sensitivity to the initial positions of APs can be mitigated by using the vectorized step size. In Chapters 4 and 5, since the scalar step size model is used, we consider various initial placement schemes of APs and explore their effects on the convergence performance of the proposed AP placement algorithms.

- *Output Solution*: For the output, an ergodic mean of $\mathbf{x}(t)$ over the set $\mathcal{D}$ was adopted in CSA, and the convergence results were established based on a constant step size and the maximum size of set $\mathcal{D}$, i.e., $\mathcal{D} = \{1 \leq t \leq T | \hat{G}(t) \leq \bar{G}^*\}$ [59]. In our cases, by carefully choosing the step size, the proposed algorithms can quickly converge even with $|\mathcal{D}| = 1$, as we show in Chapters 4 and 5. Therefore, we only output the final solution that satisfies the constraint.

- *Approximate Function*: Based on CSA/SGD, one key step to solve Problem P1 is the calculation of the stochastic gradients $\tilde{\boldsymbol{\omega}}_F(t)$ and $\tilde{\boldsymbol{\omega}}_G(t)$. However, in some cases, $F(\mathbf{r}^{\mathrm{AP}}, \mathbf{r}^{\mathrm{user}})$ and/or $G(\mathbf{r}^{\mathrm{AP}}, \mathbf{r}^{\mathrm{user}})$ may be non-differentiable in terms of $\mathbf{r}^{\mathrm{AP}}$, and the elements of $\tilde{\boldsymbol{\omega}}_F(t)$ and $\tilde{\boldsymbol{\omega}}_G(t)$ may have zero magnitude for some inputs which could lead to slow convergence of the algorithm. Similar problems have arisen in neural network training problems when the activation function of each neuron is non-differentiable and has large saturation regions where partial derivatives become zero. A common solution to this problem is replacing the activation functions by their "soft" versions, e.g., replacing indicator or rectifier by sigmoid or softplus, respectively, which are differentiable and have smaller saturation regions [66]. Enlightened by this approach, in Chapters 4 and 5, where the original objective and constraint functions are non-differentiable, we approximate them by differentiable functions in our algorithms.

Under the stochastic optimization framework established in this chapter, in the following chapters, various SA-based algorithms are developed to solve the optimal AP placement problems in different scenarios.

Chapter 3

Optimal BS Antenna Placement for Multi-user DASs

In this chapter, closed-form approximations of the ergodic sum capacity of a multi-user DAS are first derived. By exploiting the zero-order approximation of the ergodic sum capacity, an SGD-based algorithm is then developed to optimize the BS antenna placement for maximizing the average ergodic sum capacity.[1] This chapter is organized as follows. Section 3.1 reviews some previous related studies. Section 3.2 presents the system model and problem formulation. Closed-form approximations of the asymptotic ergodic sum capacity are proposed in Section 3.3. An SGD-based algorithm is proposed in Section 3.4, and its performance is illustrated in Section 3.5 via simulations. Finally, a summary is given in Section 3.6.

3.1 Previous Work

In contrast to the traditional cellular architecture with co-located BS antennas in each cell, in a DAS, many low-power remote antenna ports are geographically distributed over a large area and connected to a central processor by fiber. For the next-generation mobile communication systems with a massive amount of BS antennas to support the growing demand of high data rate, the system performance is crucially determined by how the remote antenna ports are deployed. In order to optimize the BS antenna placement, the effect of BS antenna layout on the

[1] In this chapter, AP is referred to as BS antenna cluster.

system capacity performance should be first characterized. In the following, extensive literature review on the ergodic capacity characterization and BS antenna placement optimization for multi-user DASs is provided.

3.1.1 Asymptotic Ergodic Capacity of DAS

To derive the ergodic capacity of a large-scale DAS, the asymptotic analysis based on random matrix theory has been widely adopted [68, 69]. Different from a point-to-point multiple-input-multiple-output (MIMO) system where entries of the channel matrix can be regarded as i.i.d. random variables, for a multi-user DAS, the large-scale fading coefficient between each BS antenna and each user varies with their locations, leading to non-identically distributed entries of the channel matrix. By assuming that the numbers of user antennas and BS antennas go to infinity but with a constant ratio, the asymptotic ergodic capacity of a multi-user DAS has been characterized in various scenarios [70–75]. For instance, for the simplest uncorrelated Rayleigh fading case, the limiting singular value distribution of the channel matrix has been given by Girko's theorem [76], based on which the asymptotic ergodic capacity can be obtained [70, 71]. The analysis was further extended to more complicated channel types, including correlated Rayleigh fading [72], uncorrelated Rician fading [73], Kronecker correlated Rician fading [74], and Weichselberger correlated Rician fading [75].

In all the above cases, to calculate the asymptotic ergodic capacity, a system of fixed-point equations need to be jointly solved. Take the simplest uncorrelated Rayleigh fading channel as an example. For a K-user DAS with L geographically distributed antenna clusters, to obtain the asymptotic ergodic capacity, $L + K$ fixed-point equations need to be jointly solved, which results in high computational complexity if the number of users K or the number of BS antenna clusters L is large. For a large-scale DAS with high densities of users and BS antennas, a closed-form expression of the asymptotic ergodic capacity would be highly desirable, which is, nevertheless, shown to be available only when the variance profile matrix is mean doubly-regular [70, 71], or the total number of BS antennas grows unboundedly while the amount of user antennas keeps finite [39].

In Section 3.3, we will propose closed-form approximations of the ergodic capacity for a multi-user DAS. By showing that the fixed-point equations can be solved using an iterative method with guaranteed convergence, two closed-form

approximate expressions, i.e., zero-order and first-order, are derived, where the first-order approximation provides better accuracy than the zero-order one by including one more iteration. Simulation results demonstrate that both proposed approximations are more accurate than the existing ones [39, 70, 71] under various system topologies.

3.1.2 BS Antenna Placement Optimization

As the ergodic capacity is a function of the large-scale fading coefficients that vary with the locations of users and distributed BS antennas, the capacity performance of a DAS can be significantly improved by optimizing the BS antenna layout. There have been extensive studies on the optimal BS antenna placement for DASs [37–40, 77, 78]. Most of them focus on the capacity/rate performance of DASs, which is averaged over the spatial distribution of users as users' positions vary with time. Due to the lack of closed-form expressions of the objective functions, various approximations have been proposed. In [37, 77], for instance, the objective functions were replaced by minimizing the average access distance between user and its closest BS antenna and maximizing the downlink expected signal-to-noise-ratio (SNR) at each user, respectively. In [38–40, 78], bounds of uplink average ergodic capacity [39] and uplink/downlink average ergodic rate with zero-forcing detector [38, 40] or zero-forcing beamforming [78] were developed under the assumption that the total number of BS antennas is much larger than the total number of user antennas. Constraints on BS topology were also introduced to reduce the dimensionality of variable space. In [38, 39, 77], for instance, BS antennas were restricted to be evenly placed on the perimeter of a circle, where the optimization variable becomes the radius of the circle.

In these studies, by formulating the optimal BS antenna placement as a deterministic optimization problem, the key lies in the derivation of average capacity/rate for given spatial distribution of users, which nevertheless has been found to be extremely challenging. Even if a closed-form expression of the average capacity/rate could be developed and optimized, it has to be based on an accurate characterization of users' spatial distribution, which is usually unavailable in practice. Instead, sample positions of mobile users are usually available and could be well exploited for optimizing the placement of BS antennas. That motivates us to resort to stochastic optimization approaches.

In this chapter, we focus on optimizing the BS antenna layout of a DAS for maximizing the ergodic sum capacity averaged over the positions of users. We formulate it as a stochastic optimization problem and propose a computationally efficient algorithm to solve it based on SGD. In the proposed algorithm, the gradient of the ergodic sum capacity is derived based on the proposed zero-order approximation of ergodic sum capacity. Simulation results corroborate that the proposed algorithm significantly outperforms the existing representative BS antenna placement schemes, especially when the aggregation degree of users or the number of BS antenna clusters is large.

3.2 System Model and Problem Formulation

Consider a multi-user DAS with L BS antenna clusters, serving for K users. Each user has N_u antennas, and each BS antenna cluster is equipped with N_c antennas. Let $\mathbf{G} \in \mathbb{C}^{M\times N}$ denote the channel gain matrix between $N = KN_u$ user antennas and $M = LN_c$ BS antennas, which can be written as

$$\mathbf{G} = \mathbf{\Gamma} \circ \mathbf{H}, \tag{3.1}$$

where $\circ$ denotes the Hadamard product. $\mathbf{H} \in \mathbb{C}^{M\times N}$ denotes the small-scale fading matrix, whose entries are modelled as i.i.d. circularly symmetric complex Gaussian random variables with zero mean and unit variance, and $\mathbf{\Gamma} \in \mathbb{C}^{M\times N}$ is the large-scale fading matrix. The large-scale fading coefficient $\gamma_{l,k}$ between the lth BS antenna cluster and the kth user can be modelled as

$$\gamma_{l,k} = \|\mathbf{r}_l^{\mathrm{AP}} - \mathbf{r}_k^{\mathrm{user}}\|^{-\alpha/2}, \tag{3.2}$$

where $\mathbf{r}_l^{\mathrm{AP}}$ and $\mathbf{r}_k^{\mathrm{user}}$ denote the positions of BS antenna cluster l and user k, respectively. $\alpha > 2$ denotes the path-loss factor. In this chapter, we set $\alpha = 4$, which corresponds to the urban scenario with rich-scattering environment [79].

When users are unaware of the instantaneous channel state information and transmit at equal power, the uplink normalized ergodic sum capacity can be written as [80]

$$C = \frac{1}{M}\mathbb{E}_{\mathbf{H}}[\log_2 \det(\mathbf{I}_M + \frac{\rho_0}{N_u}\mathbf{G}\mathbf{G}^\dagger)], \tag{3.3}$$

where ρ_0 is the transmit SNR at each user.

It has been shown in [70, 71] that as $N_u \to \infty$, $N_c \to \infty$ and $\frac{N_u}{N_c} \to \beta$, the asymptotic normalized ergodic sum capacity $\mathcal{C}=\lim\limits_{N_u\to\infty,\ N_c\to\infty,\ \frac{N_u}{N_c}\to\beta} C$ can be obtained as

$$\mathcal{C} = \frac{1}{L}\sum_{l=1}^{L}\log_2 U_l^* + \frac{\beta}{L}\sum_{k=1}^{K}\log_2 W_k^* - \frac{\rho_0 \log_2 e}{L}\sum_{l=1}^{L}\sum_{k=1}^{K}\gamma_{l,k}^2 U_l^{*-1} W_k^{*-1}, \tag{3.4}$$

where $\{U_1^*,\cdots,U_L^*\}$ and $\{W_1^*,\cdots,W_K^*\}$ are the solutions of the following fixed-point equations:

$$U_l = 1 + \rho_0 \sum_{k=1}^{K}\gamma_{l,k}^2 W_k^{-1},\ l=1,\cdots,L, \tag{3.5}$$

$$W_k = 1 + \frac{\rho_0}{\beta}\sum_{l=1}^{L}\gamma_{l,k}^2 U_l^{-1},\ k=1,\cdots,K. \tag{3.6}$$

Note that if the variance profile matrix $\mathbf{V}=\mathbf{\Gamma}\circ\mathbf{\Gamma}$ is mean doubly-regular,[2] then explicit expressions of $\{U_1^*,\cdots,U_L^*\}$, $\{W_1^*,\cdots,W_K^*\}$ and $\mathcal{C}$ can be obtained as [70, 71]

$$U_{l\_DR}^* = \Big(1 - \frac{\beta}{\rho_0 L\Phi}\mathcal{F}\big(\frac{\rho_0}{\beta}L\Phi, \frac{K\beta}{L}\big)\Big)^{-1},\ l=1,\cdots,L, \tag{3.7}$$

$$W_{k\_DR}^* = \Big(1 - \frac{1}{\rho_0 K\Phi}\mathcal{F}\big(\frac{\rho_0}{\beta}L\Phi, \frac{K\beta}{L}\big)\Big)^{-1},\ k=1,\cdots,K, \tag{3.8}$$

and

$$\begin{aligned}\mathcal{C}_{DR} = {} & \frac{K\beta}{L}\log_2\Big(1+\frac{\rho_0}{\beta}L\Phi - \mathcal{F}\big(\frac{\rho_0}{\beta}L\Phi, \frac{K\beta}{L}\big)\Big) + \log_2\Big(1+\rho_0 K\Phi - \mathcal{F}\big(\frac{\rho_0}{\beta}L\Phi, \frac{K\beta}{L}\big)\Big) \\ & - \frac{\beta\log_2 e}{L\rho_0\Phi}\mathcal{F}\big(\frac{\rho_0}{\beta}L\Phi, \frac{K\beta}{L}\big),\end{aligned} \tag{3.9}$$

where

$$\mathcal{F}(x,y) = \frac{1}{4}\left(\sqrt{x(1+\sqrt{y})^2+1} - \sqrt{x(1-\sqrt{y})^2+1}\right)^2, \tag{3.10}$$

and

$$\Phi = \frac{1}{LK}\sum_{l=1}^{L}\sum_{k=1}^{K}\gamma_{l,k}^2. \tag{3.11}$$

In general, $\{U_1^*,\cdots,U_L^*\}$ and $\{W_1^*,\cdots,W_K^*\}$ can only be numerically obtained by jointly solving $L+K$ fixed-point equations given in (3.5)–(3.6), and the

[2] An $N\times M$ matrix $\mathbf{P}$ is mean column-regular if $\frac{1}{N}\sum_{n=1}^{N}p_{n,m}$ is independent of m. It is mean doubly-regular if both $\mathbf{P}$ and $\mathbf{P}'$ are mean column-regular [70].

computational complexity increases as the number of users K or the number of BS antenna clusters L grows.

From (3.2) and (3.4), we can see that the sum capacity is closely determined by the positions of users and BS antenna clusters. Due to mobility, the positions of users are time-varying. Therefore, the normalized ergodic sum capacity averaged over users' positions, i.e., $\bar{C} = \mathbb{E}_{\{\mathbf{r}_k^{\text{user}}\}}[C]$, is of more practical interest, which is critically dependent on the placement of BS antennas.

In this chapter, we are interested at maximizing the average normalized ergodic sum capacity $\bar{C}$ by optimizing the positions of BS antenna clusters $\{\mathbf{r}_l^{\text{AP}}\}$. The optimization problem can be formulated as

$$(\text{P2}): \max_{\{\mathbf{r}_l^{\text{AP}}\}} \bar{C} = \mathbb{E}_{\{\mathbf{r}_k^{\text{user}}\}}[C] \tag{3.12}$$

$$\text{s.t. } \mathbf{r}_l^{\text{AP}} \in \mathcal{A}, l = 1, \cdots, L, \tag{3.13}$$

where $\mathcal{A}$ denotes the area in which the BS antenna clusters are deployed. In this chapter, we assume that $\mathcal{A}$ is a circular cell with radius $R_{\text{area}} = 1$.

Problem P2 is a stochastic optimization problem without expectation constraint. Therefore, based on the SA methods introduced in Section 2.2.1, we can solve it based on SGD. A key step is to calculate the gradient of C in terms of $[\mathbf{r}_1^{\text{AP}}, \cdots, \mathbf{r}_L^{\text{AP}}]$, which is, nevertheless, difficult to obtain due to the lack of closed-form expression of ergodic sum capacity C. As shown in (3.4)–(3.6), even in the asymptotic conditions, the sum capacity can only be numerically obtained by jointly solving $L + K$ fixed-point equations. Although in the special case that $\mathbf{V} = \mathbf{\Gamma} \circ \mathbf{\Gamma}$ is mean doubly-regular, a closed-form expression of the asymptotic normalized ergodic sum capacity can be obtained, such a condition on $\mathbf{\Gamma}$ does not hold in general. Therefore, (3.9) only serves as an approximation, which could be inaccurate in various scenarios. In the following section, an iterative method to solve (3.5)–(3.6) is first introduced, and then effective approximations of the asymptotic normalized ergodic sum capacity $\mathcal{C}$ are proposed.

3.3 Approximate Asymptotic Normalized Ergodic Sum Capacity

Based on (3.5)–(3.6), we can rewrite (3.4) as

$$\begin{aligned}\mathcal{C} =& f_{CU}(U_1^*,\cdots,U_L^*) = \frac{1}{L}\sum_{l=1}^{L}\log_2 U_l^* + \frac{\beta}{L}\sum_{k=1}^{K}\log_2(1+\frac{\rho_0}{\beta}\sum_{l=1}^{L}\gamma_{l,k}^2 U_l^{*-1}) \\ &- \frac{\rho_0\log_2 e}{L}\sum_{l=1}^{L}\sum_{k=1}^{K}\frac{\gamma_{l,k}^2}{U_l^*(1+\frac{\rho_0}{\beta}\sum_{i=1}^{L}\gamma_{i,k}^2 U_i^{*-1})},\end{aligned} \tag{3.14}$$

where $\{U_1^*,\cdots,U_L^*\}$ is the solution of the following fixed-point equations:

$$U_l = 1+\rho_0\sum_{k=1}^{K}\frac{\gamma_{l,k}^2}{1+\frac{\rho_0}{\beta}\sum_{i=1}^{L}\gamma_{i,k}^2 U_i^{-1}},\ l=1,\cdots,L. \tag{3.15}$$

One way to solve (3.15) is using the fixed-point iteration algorithm [81]. We can start from an initial point $\{U_1^{(0)},\cdots,U_L^{(0)}\}$ and substitute it into (3.15) to iteratively solve it. The iteration process can be described by

$$U_l^{(\tau+1)} = 1+\rho_0\sum_{k=1}^{K}\frac{\gamma_{l,k}^2}{1+\frac{\rho_0}{\beta}\sum_{i=1}^{L}\gamma_{i,k}^2 U_i^{(\tau)^{-1}}},\ l=1,\cdots,L. \tag{3.16}$$

Alternatively, $\mathcal{C}$ can be also written as

$$\begin{aligned}\mathcal{C} =& f_{CW}(W_1^*,\cdots,W_K^*) = \frac{\beta}{L}\sum_{k=1}^{K}\log_2 W_k^* + \frac{1}{L}\sum_{l=1}^{L}\log_2(1+\rho_0\sum_{k=1}^{K}\gamma_{l,k}^2 W_k^{*-1}) \\ &- \frac{\rho_0\log_2 e}{L}\sum_{l=1}^{L}\sum_{k=1}^{K}\frac{\gamma_{l,k}^2}{W_k^*(1+\rho_0\sum_{j=1}^{K}\gamma_{l,j}^2 W_j^{*-1})},\end{aligned} \tag{3.17}$$

where $\{W_1^*,\cdots,W_K^*\}$ is the solution of the following fixed-point equations:

$$W_k = 1+\frac{\rho_0}{\beta}\sum_{l=1}^{L}\frac{\gamma_{l,k}^2}{1+\rho_0\sum_{j=1}^{K}\gamma_{l,j}^2 W_j^{-1}},\ k=1,\cdots,K. \tag{3.18}$$

We can start from an initial point $\{W_1^{(0)}, \cdots, W_K^{(0)}\}$ and use the following fixed-point iteration process to solve (3.18):

$$W_k^{(\tau+1)} = 1 + \frac{\rho_0}{\beta} \sum_{l=1}^{L} \frac{\gamma_{l,k}^2}{1 + \rho_0 \sum_{j=1}^{K} \gamma_{l,j}^2 W_j^{(\tau)^{-1}}}, \ k = 1, \cdots, K. \tag{3.19}$$

3.3.1 Convergence

Lemma 3.1 shows that starting from any positive initial point, (3.16) and (3.19) will always converge to the unique solutions of (3.15) and (3.18), respectively.

Lemma 3.1. $\lim_{\tau\to\infty} \{U_1^{(\tau)}, \cdots, U_L^{(\tau)}\} \to \{U_1^*, \cdots, U_L^*\}$ *and* $\lim_{\tau\to\infty} \{W_1^{(\tau)}, \cdots, W_K^{(\tau)}\} \to \{W_1^*, \cdots, W_K^*\}$ *for any positive initial values of* $\{U_1^{(0)}, \cdots, U_L^{(0)}\}$ *and* $\{W_1^{(0)}, \cdots, W_K^{(0)}\}$.

Proof. See Appendix A. □

For brevity, we denote $f_{CU}(U_1^{(\tau)}, \cdots, U_L^{(\tau)})$ and $f_{CW}(W_1^{(\tau)}, \cdots, W_K^{(\tau)})$ as $f_{CU}^{(\tau)}$ and $f_{CW}^{(\tau)}$, respectively. Since both $f_{CU}^{(\tau)}$ and $f_{CW}^{(\tau)}$ are continuous functions, we have

$$\lim_{\tau\to\infty} f_{CU}^{(\tau)} \to \mathcal{C}, \text{ and } \lim_{\tau\to\infty} f_{CW}^{(\tau)} \to \mathcal{C}, \tag{3.20}$$

according to Lemma 3.1.

Note that although both $f_{CU}^{(\tau)}$ and $f_{CW}^{(\tau)}$ converge to the same limit, their convergence speeds are different. For illustration, let us set the initial point for $f_{CU}^{(\tau)}$ according to (3.7) as

$$U_l^{(0)} = \Big(1 - \frac{\beta}{\rho_0 L\Phi} \mathcal{F}\big(\frac{\rho_0}{\beta} L\Phi, \frac{K\beta}{L}\big)\Big)^{-1}, \ l = 1, \cdots, L. \tag{3.21}$$

And the initial point for the iteration process $f_{CW}^{(\tau)}$ is set according to (3.8) as

$$W_k^{(0)} = \Big(1 - \frac{1}{\rho_0 K\Phi} \mathcal{F}\big(\frac{\rho_0}{\beta} L\Phi, \frac{K\beta}{L}\big)\Big)^{-1}, \ k = 1, \cdots, K. \tag{3.22}$$

Fig. 3.1 shows how $f_{CU}^{(\tau)}$ and $f_{CW}^{(\tau)}$ vary with the number of iterations τ under different parameter settings. Specifically, we consider two cases: $L > K\beta$ and $L < K\beta$, in Fig. 3.1a and Fig. 3.1b, respectively. Fig. 3.1a illustrates that when

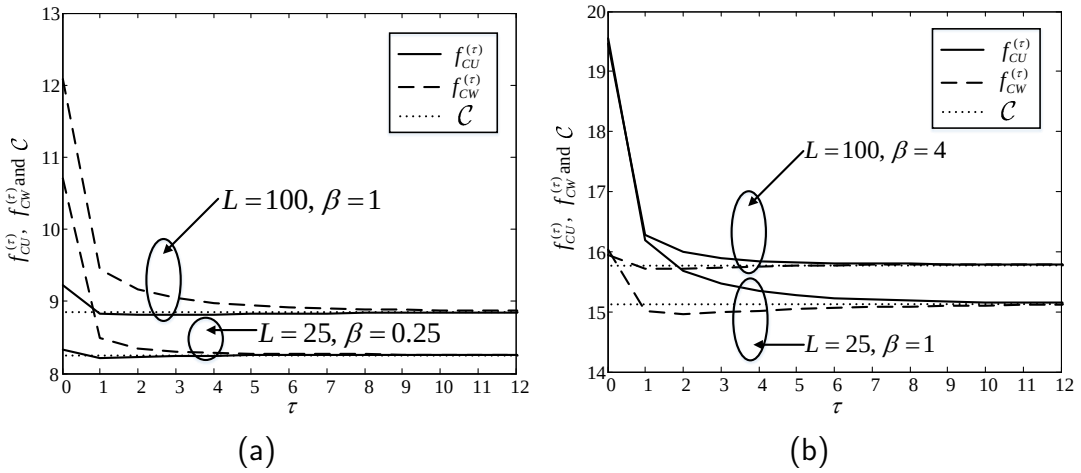

FIGURE 3.1: $f_{CU}^{(\tau)}$ and $f_{CW}^{(\tau)}$ versus number of iterations τ with the initial points given in (3.21) and (3.22), respectively. $K = 50$, $\alpha = 4$, $\rho_0 = 10$ dB. (a) $L > K\beta$. (b) $L < K\beta$.

$L > K\beta$, $f_{CU}^{(\tau)}$ converges faster than $f_{CW}^{(\tau)}$ and approaches the limit more rapidly. On the other hand, $f_{CW}^{(\tau)}$ has a higher convergence speed than $f_{CU}^{(\tau)}$ when $L < K\beta$, as Fig. 3.1b shows.

To understand why $f_{CU}^{(\tau)}$ and $f_{CW}^{(\tau)}$ have different convergence speeds for $L > K\beta$ and $L < K\beta$, recall that $\mathcal{F}(x, y)$ given in (3.10) can be approximated by

$$\mathcal{F}(x,y) \overset{x \gg 1}{\approx} \begin{cases} xy & \text{if } y < 1 \\ x & \text{otherwise.} \end{cases} \tag{3.23}$$

By combining (3.23) with (3.21) and (3.22), we can see that for large $\rho_0 \gg 1$, when $L > K\beta$, the initial point $U_l^{(0)}$ can be approximated as $U_l^{(0)} \approx \frac{L}{L-K\beta}, l = 1, \cdots, L$, while $W_k^{(0)}$ will diverge, $k = 1, \cdots, K$. In contrast, when $L < K\beta$, we have $W_k^{(0)} \approx \frac{K\beta}{K\beta - L}, k = 1, \cdots, K$, yet $U_l^{(0)}$ becomes infinite, $l = 1, \cdots, L$. Therefore, we can expect that $\{U_1^{(\tau)}, \cdots, U_L^{(\tau)}\}$ will converge to the optimal point faster than $\{W_1^{(\tau)}, \cdots, W_K^{(\tau)}\}$ when $L > K\beta$. Otherwise, $\{W_1^{(\tau)}, \cdots, W_K^{(\tau)}\}$ will have a higher convergence speed.

3.3.2 Proposed Approximations

Eq. (3.20) suggests that to obtain the asymptotic normalized ergodic sum capacity $\mathcal{C}$, there exists a tradeoff between accuracy and complexity: with more iterations, $f_{CU}^{(\tau)}$ and $f_{CW}^{(\tau)}$ are closer to the limit $\mathcal{C}$, but with higher computational complexity. We can observe from Fig. 3.1 that with the initial points given in

(3.21) and (3.22), $f_{CU}^{(\tau)}$ and $f_{CW}^{(\tau)}$ can quickly converge. When $L > K\beta$, for instance, $f_{CU}^{(0)}$ is quite close to the limit $\mathcal{C}$, indicating that $f_{CU}^{(0)}$ can well approximate $\mathcal{C}$. Otherwise, $f_{CW}^{(0)}$ serves as a better approximation of $\mathcal{C}$. Accordingly, we propose the following zero-order approximation of the asymptotic normalized ergodic sum capacity $\mathcal{C}$:

$$\tilde{\mathcal{C}}^{(0)} = \begin{cases} f_{CU}^{(0)} & \text{if } L > K\beta \\ f_{CW}^{(0)} & \text{otherwise,} \end{cases} \tag{3.24}$$

where $f_{CU}^{(0)}$ can be obtained by setting $U_l^* = U_l^{(0)}, l = 1, \cdots, L$, and substituting (3.21) into (3.14), and $f_{CW}^{(0)}$ can be obtained by setting $W_k^* = W_k^{(0)}, k = 1, \cdots, K$, and substituting (3.22) into (3.17). The expressions of $f_{CU}^{(0)}$ and $f_{CW}^{(0)}$ are given by

$$\begin{aligned} f_{CU}^{(0)} = & \log_2\big(1-\frac{\beta}{\rho_0 L\Phi}\mathcal{F}(\frac{\rho_0}{\beta}L\Phi, \frac{K\beta}{L})\big)^{-1} + \frac{\beta}{L}\sum_{k=1}^{K}\log_2\Big(1+\big(1-\frac{\beta}{\rho_0 L\Phi} \\ & \cdot \mathcal{F}(\frac{\rho_0}{\beta}L\Phi, \frac{K\beta}{L})\big)\frac{\rho_0}{\beta}\sum_{l=1}^{L}\gamma_{l,k}^2\Big) - \frac{\rho_0 \log_2 e}{L}\big(1-\frac{\beta}{\rho_0 L\Phi}\mathcal{F}(\frac{\rho_0}{\beta}L\Phi, \frac{K\beta}{L})\big) \\ & \cdot \sum_{l=1}^{L}\sum_{k=1}^{K}\gamma_{l,k}^2\Big(1+\big(1-\frac{\beta}{\rho_0 L\Phi}\mathcal{F}(\frac{\rho_0}{\beta}L\Phi, \frac{K\beta}{L})\big)\frac{\rho_0}{\beta}\sum_{i=1}^{L}\gamma_{i,k}^2\Big)^{-1}, \end{aligned} \tag{3.25}$$

and

$$\begin{aligned} f_{CW}^{(0)} = & \frac{K\beta}{L}\log_2\big(1-\frac{1}{\rho_0 K\Phi}\mathcal{F}(\frac{\rho_0}{\beta}L\Phi, \frac{K\beta}{L})\big)^{-1} + \frac{1}{L}\sum_{l=1}^{L}\log_2\Big(1+\big(1-\frac{1}{\rho_0 K\Phi} \\ & \cdot \mathcal{F}(\frac{\rho_0}{\beta}L\Phi, \frac{K\beta}{L})\big)\rho_0\sum_{k=1}^{K}\gamma_{l,k}^2\Big) - \frac{\rho_0 \log_2 e}{L}\big(1-\frac{1}{\rho_0 K\Phi}\mathcal{F}(\frac{\rho_0}{\beta}L\Phi, \frac{K\beta}{L})\big) \\ & \cdot \sum_{l=1}^{L}\sum_{k=1}^{K}\gamma_{l,k}^2\Big(1+\big(1-\frac{1}{\rho_0 K\Phi}\mathcal{F}(\frac{\rho_0}{\beta}L\Phi, \frac{K\beta}{L})\big)\rho_0\sum_{j=1}^{K}\gamma_{l,j}^2\Big)^{-1}, \end{aligned} \tag{3.26}$$

respectively.

The accuracy can be further improved if we perform one more iteration. Specifically, by substituting (3.21) and (3.22) into (3.16) and (3.19), respectively, we have

$$U_l^{(1)} = 1+\rho_0\sum_{k=1}^{K}\gamma_{l,k}^2\Big(1+\frac{\rho_0}{\beta}\sum_{i=1}^{L}\gamma_{i,k}^2\big(1-\frac{\beta}{\rho_0 L\Phi}\mathcal{F}(\frac{\rho_0}{\beta}L\Phi, \frac{K\beta}{L})\big)\Big)^{-1},\ l = 1, \cdots, L, \tag{3.27}$$

$$W_k^{(1)}=1+\frac{\rho_0}{\beta}\sum_{l=1}^{L}\gamma_{l,k}^2\Big(1+\rho_0\sum_{j=1}^{K}\gamma_{l,j}^2\big(1-\frac{1}{\rho_0 K\Phi}\mathcal{F}(\frac{\rho_0}{\beta}L\Phi,\frac{K\beta}{L})\big)\Big)^{-1},\ k=1,\cdots,K. \tag{3.28}$$

Similarly, we can propose the following first-order approximation of $\mathcal{C}$ as

$$\tilde{\mathcal{C}}^{(1)}=\begin{cases} f_{CU}^{(1)}, & \text{if } L>K\beta \\ f_{CW}^{(1)}, & \text{otherwise,} \end{cases} \tag{3.29}$$

where $f_{CU}^{(1)}$ can be obtained by setting $U_l^*=U_l^{(1)}, l=1,\cdots,L$, and substituting (3.27) into (3.14), and $f_{CW}^{(1)}$ can be obtained by setting $W_k^*=W_k^{(1)}, k=1,\cdots,K$, and substituting (3.28) into (3.17).

Now we conduct some simulations to examine the performance of the proposed approximations $\tilde{\mathcal{C}}^{(0)}$ and $\tilde{\mathcal{C}}^{(1)}$. We consider an uplink DAS with K users and L BS antenna clusters uniformly distributed inside $\mathcal{A}$, i.e., a circular cell with unit radius, where each user is equipped with N_u co-located antennas, and each cluster has N_c co-located BS antennas. The system topology, i.e., the locations of BS antenna clusters and users, is generated at random, and for a given system topology, the normalized ergodic sum capacity C is obtained according to (3.3) by averaging over 1000 samples of small-scale fading coefficients.

We also consider two existing approximations as comparison benchmarks for our proposed approximations:

1) Doubly-regular (DR) Approximation: It has been shown in Section 3.2 that when the large-scale fading matrix $\mathbf{\Gamma}$ satisfies a certain condition, i.e., $\mathbf{V}=\mathbf{\Gamma}\circ\mathbf{\Gamma}$ is mean DR, then a closed-form expression of the asymptotic normalized ergodic sum capacity can be obtained [70, 71], which is given in (3.9). In practice, however, such a condition on $\mathbf{\Gamma}$ does not hold in general. Therefore, (3.9) only serves as an approximation, which is referred to as the DR approximation.

2) Asymptotically-orthogonal (AO) Approximation: For massive MIMO systems, a widely-adopted assumption is that the number of BS antennas M tends to infinity while the number of antennas at the user side N keeps finite, with which the users' channels can be regarded as orthogonal to each other [39, 82, 83]. The corresponding asymptotic normalized ergodic sum capacity can be then obtained

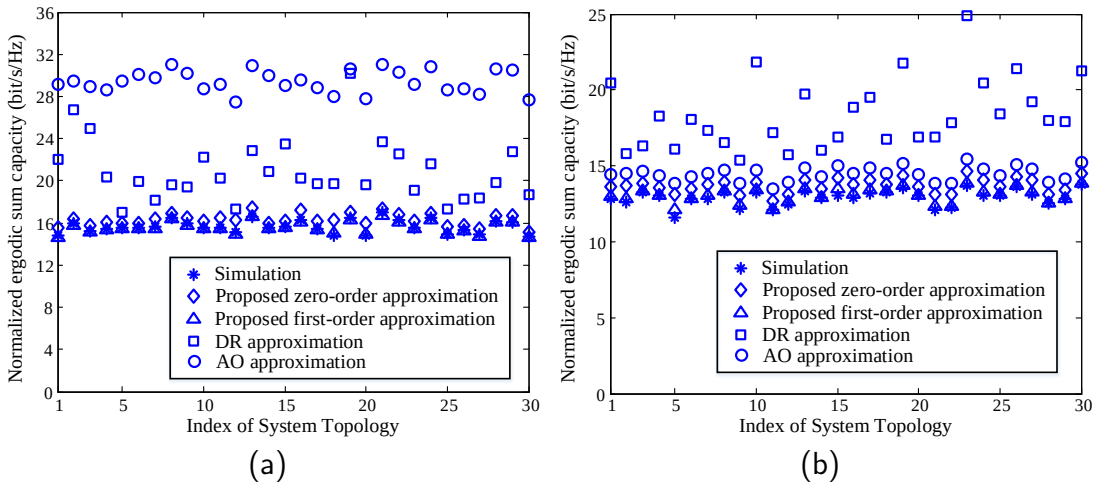

FIGURE 3.2: Normalized ergodic sum capacity under 30 randomly generated system topologies, where both users and BS antenna clusters are uniformly distributed inside $\mathcal{A}$. K=50, N_u=2, N_c=2, ρ_0=10 dB. (a) L=25. (b) L=60.

as [39]

$$\mathcal{C}_{AO} = \frac{\beta}{L}\sum_{k=1}^{K}\log_2\left(1+\frac{\rho_0}{\beta}\sum_{l=1}^{L}\gamma_{l,k}^2\right), \tag{3.30}$$

which is referred to as the AO approximation.

Fig. 3.2 presents the normalized ergodic sum capacity under 30 randomly generated system topologies. In Fig. 3.2, the proposed zero-order and first-order approximations of the asymptotic normalized ergodic sum capacity given in (3.24) and (3.29) are plotted in comparison with the DR approximation given in (3.9) and the AO approximation given in (3.30). It can be seen from Fig. 3.2 that both proposed approximations are close to the normalized ergodic sum capacity, and more accurate than the DR approximation and the AO approximation.

A closer look at Fig. 3.2a shows that when the total amount of BS antennas M=LN_c=50 is smaller than the total amount of user antennas N=KN_u=100, the AO approximation does not work as it is derived under the assumption of $\frac{M}{N} \to \infty$. The DR approximation is also inaccurate as the required condition on the large-scale fading matrix $\mathbf{\Gamma}$ does not hold in general with geographically distributed users and BS antenna clusters. On the other hand, in Fig. 3.2b, with the number of BS antenna clusters L increasing from 25 to 60, the total amount of BS antennas M becomes larger than the total amount of user antennas N, and thus the accuracy of the AO approximation is significantly improved. In contrast, the DR approximation remains inaccurate with discrepancy further enlarged.

We can also see from Fig. 3.2 that the proposed first-order approximation is quite accurate under various parameter settings. By including one more iteration, the computational complexity of the proposed first-order approximation is higher than that of the zero-order one, yet the accuracy is further improved. Our proposed approximations provide a flexible tradeoff between complexity and accuracy.

3.3.3 Approximate Average Normalized Ergodic Sum Capacity

In Section 3.3.2, we have proposed closed-form approximations for the normalized ergodic sum capacity that is closely determined by the large-scale fading coefficients. In practice, the normalized ergodic sum capacity may vary with the system topology, as Fig. 3.2 illustrates. In this subsection, let us assume that both users and BS antenna clusters are uniformly distributed inside $\mathcal{A}$, and further derive a closed-form approximation for the average normalized ergodic sum capacity, which is defined as

$$\bar{\bar{C}} \triangleq \mathbb{E}_{\{\mathbf{r}_l^{\mathrm{AP}}\},\{\mathbf{r}_k^{\mathrm{user}}\}}[C]. \tag{3.31}$$

The approximation not only reveals the effects of key system parameters such as number of BS antenna clusters L and number of users K on the sum capacity, but also serves as a benchmark for comparison to the optimal BS antenna placement.

Specifically, it has been shown in Section 3.3.2 that the asymptotic normalized ergodic sum capacity $\mathcal{C}$ can be well approximated by the proposed zero-order approximation $\tilde{\mathcal{C}}^{(0)}$, which has a closed-form expression. Appendix B further shows that when K users and L BS antenna clusters are uniformly distributed within a circular cell, we have

$$\mathbb{E}_{\{\mathbf{r}_l^{\mathrm{AP}}\},\{\mathbf{r}_k^{\mathrm{user}}\}}[\tilde{\mathcal{C}}^{(0)}] \overset{\rho_0,L,K\gg 1}{\approx} \begin{cases} \frac{K\beta}{L}\log_2\frac{\pi e^{\epsilon-1}\rho_0 L^2}{\beta}+(\frac{K\beta}{L}-1)\log_2(1-\frac{K\beta}{L}) & \text{if } L > K\beta \\ \log_2(\pi e^{\epsilon-1}\rho_0 K^2)+(1-\frac{K\beta}{L})\log_2(1-\frac{L}{K\beta}) & \text{otherwise,} \end{cases} \tag{3.32}$$

where $\epsilon \approx 0.5772$ is the Euler-Mascheroni constant. For finite N_u and N_c, a non-asymptotic approximation of the average normalized ergodic sum capacity $\bar{\bar{C}}$ can

be derived by substituting $\beta = \frac{N_u}{N_c}$ into (3.32) as

$$\bar{\bar{C}}^{(0)} = \begin{cases} \frac{1}{\Omega}\log_2(\pi e^{\epsilon-1}\rho_0 LK\Omega) + (\frac{1}{\Omega}-1)\log_2(1-\frac{1}{\Omega}) & \text{if } \Omega > 1 \\ \log_2(\pi e^{\epsilon-1}\rho_0 K^2) + (1-\frac{1}{\Omega})\log_2(1-\Omega) & \text{otherwise,} \end{cases} \tag{3.33}$$

where $\Omega = \frac{LN_c}{KN_u}$. We can see from (3.33) that with a constant $\Omega > 1$, $\bar{\bar{C}}^{(0)}$ increases with the number of BS antenna clusters L, implying that for a given total number of BS antennas $M = LN_c > KN_u$, the optimal number of BS antenna clusters L to maximize the average per-BS-antenna capacity is equal to M, or equivalently, the optimal BS antenna cluster size N_c is one. In other words, for given amount of BS antennas, a fully distributed BS antenna layout can achieve the highest average per-BS-antenna capacity, which agrees with the observation for the single-user case in [84].

Based on (3.33), we can further obtain a closed-form approximation of the average ergodic *per-user* capacity $\bar{\bar{C}}_u$ as

$$\bar{\bar{C}}_u \approx \frac{M}{K}\bar{\bar{C}}^{(0)} = \begin{cases} N_u\big(\log_2(\pi e^{\epsilon-1}\rho_0 LK\Omega) + (1-\Omega)\log_2(1-\frac{1}{\Omega})\big) & \text{if } \Omega > 1 \\ N_u\big(\Omega\log_2(\pi e^{\epsilon-1}\rho_0 K^2) + (\Omega-1)\log_2(1-\Omega)\big) & \text{otherwise.} \end{cases} \tag{3.34}$$

It can be easily shown that $\bar{\bar{C}}_u$ is a monotone decreasing function of the number of users K. In practice, it is desirable to maintain a constant per-user rate irrespective of the number of users. Appendix C shows that to ensure that the average per-user capacity $\bar{\bar{C}}_u$ is no smaller than a given per-user rate C^*, the total number of BS antennas should be no smaller than

$$M^* = KN_u\left(1 - \frac{\ln(\pi e^{\epsilon-1}\rho_0 K^2) - \frac{C^*\ln 2}{N_u}}{\mathbb{W}_0\Big(\pi e^{\epsilon-1}\rho_0 K^2\big(\ln(\pi e^{\epsilon-1}\rho_0 K^2) - \frac{C^*\ln 2}{N_u}\big)\Big)}\right), \tag{3.35}$$

where $\mathbb{W}_0(\cdot)$ denotes the principal branch of Lambert W function, when the number of users $K > K^* = \sqrt{\frac{2^{C^*/N_u}}{\pi e^{\epsilon-1}\rho_0}}$.[3] As Fig. 3.3 illustrates, a linear increase of the minimum required number of BS antennas M^* with respect to the number of users K can be observed only when K is sufficiently large. For small K, the increasing rate of M^* decreases as K grows. In practice, (3.35) is useful for determining how many BS antennas should be adaptively switched on or off according to the amount of users to fulfill a certain per-user rate requirement.

[3]Note that $K^* \ll 1$ for high SNR $\rho_0 \gg 1$.

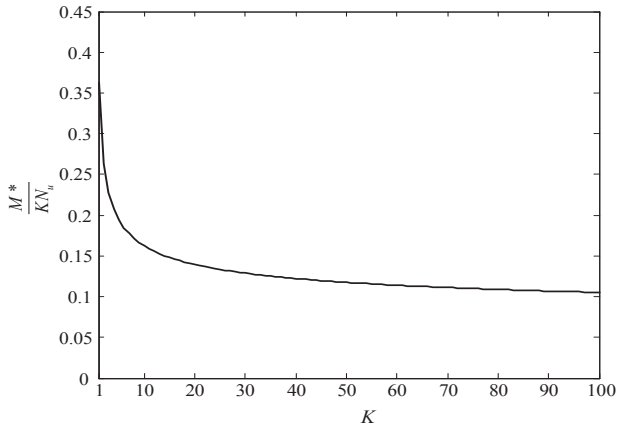

FIGURE 3.3: Ratio of the minimum required number of BS antennas M^* for maintaining a constant per-user rate C^* to the total number of user antennas KN_u versus the number of users K. N_u=5, N_c=2, C^*=10 bit/s/Hz, ρ_0=10 dB.

3.4 SGD-based BS Antenna Placement

In Section 3.3, we have proposed closed-form approximations, i.e., zero-order and first-order ones, of the asymptotic ergodic sum capacity $\mathcal{C}$. In this section, we will propose an SGD-based algorithm to solve the BS antenna placement problem P2 by exploiting the proposed zero-order approximation (3.24).

For Problem P2, the objective function can be written as $\max_{\mathbf{r}^{\mathrm{AP}}} \bar{C} = \mathbb{E}_{\mathbf{r}^{\mathrm{user}}}[C(\mathbf{r}^{\mathrm{AP}}, \mathbf{r}^{\mathrm{user}})]$, where $\mathbf{r}^{\mathrm{AP}} = [\mathbf{r}_1^{\mathrm{AP}}, \cdots, \mathbf{r}_L^{\mathrm{AP}}]$ denotes the positions of all the BS antenna clusters and $\mathbf{r}^{\mathrm{user}} = [\mathbf{r}_1^{\mathrm{user}}, \cdots, \mathbf{r}_K^{\mathrm{user}}]$ denotes the positions of all the users. To apply SGD, we need to calculate the gradient of C in terms of $\mathbf{r}^{\mathrm{AP}}$, which is, nevertheless, difficult to obtain due to the lack of closed-form expression of ergodic sum capacity C. In the last section, closed-form approximate expressions of the asymptotic ergodic sum capacity $\mathcal{C}$ were derived, and shown to be accurate in various scenarios. Based on the proposed zero-order approximation of $\mathcal{C}$ given in (3.24), we derive the gradient $\tilde{\boldsymbol{\omega}}_{\mathcal{C}}$ as

$$\tilde{\boldsymbol{\omega}}_{\mathcal{C}} = \begin{cases} \frac{\partial f_C(\{U_l\},\{W_k\})}{\partial \mathbf{r}^{\mathrm{AP}}}\Big|_{U_l=U_l^{(0)},W_k=\tilde{W}_k^{(0)}} & \text{if } L > K\beta \\ \frac{\partial f_C(\{U_l\},\{W_k\})}{\partial \mathbf{r}^{\mathrm{AP}}}\Big|_{U_l=\tilde{U}_l^{(0)},W_k=W_k^{(0)}} & \text{otherwise,} \end{cases} \tag{3.36}$$

where

$$\frac{\partial f_C(\{U_l\},\{W_k\})}{\partial \mathbf{r}^{\mathrm{AP}}}=\frac{1}{L}\sum_{l=1}^{L}\frac{\log_2 e}{U_l}\cdot\frac{\partial U_l}{\partial \mathbf{r}^{\mathrm{AP}}}+\frac{\beta}{L}\sum_{k=1}^{K}\frac{\log_2 e}{W_k}\cdot\frac{\partial W_k}{\partial \mathbf{r}^{\mathrm{AP}}}$$
$$-\frac{\rho_0\log_2 e}{L}\sum_{l=1}^{L}\sum_{k=1}^{K}\frac{\gamma_{l,k}^2}{U_l W_k}\Big(\frac{1}{\gamma_{l,k}^2}\cdot\frac{\gamma_{l,k}^2}{\partial \mathbf{r}^{\mathrm{AP}}}-\frac{1}{U_l}\cdot\frac{\partial U_l}{\partial \mathbf{r}^{\mathrm{AP}}}-\frac{1}{W_k}\cdot\frac{\partial W_k}{\partial \mathbf{r}^{\mathrm{AP}}}\Big) \tag{3.37}$$

with $\frac{\partial U_l^{(0)}}{\partial \mathbf{r}^{\mathrm{AP}}}$, $\frac{\partial W_k^{(0)}}{\partial \mathbf{r}^{\mathrm{AP}}}$, $\frac{\partial \tilde{U}_l^{(0)}}{\partial \mathbf{r}^{\mathrm{AP}}}$ and $\frac{\partial \tilde{W}_k^{(0)}}{\partial \mathbf{r}^{\mathrm{AP}}}$ given in (D.2), (D.3), (D.6) and (D.7), respectively. The detailed derivation of $\tilde{\boldsymbol{\omega}}_{\mathcal{C}}$ is presented in Appendix D.

The convergence performance of SGD is crucially determined by the step size $\eta(t)$. As illustrated in Section 2.2.2, $\eta(t)$ is usually a scalar, which may result in poor convergence performance if the objective function is sensitive to some directions in variable space [66]. In that case, it would be more appropriate to adopt a vectorized step size, which adapts its value to the stochastic gradient elementwisely during iterations.

In our problem, if users are non-uniformly distributed, then the movements of the BS antenna clusters located in the user-intensive areas would bring more significant changes to the ergodic sum capacity than those far away from users, leading to a large difference in magnitude of elements of the stochastic gradient $\tilde{\boldsymbol{\omega}}_{\mathcal{C}}$. That motivates us to adopt the vectorized step size model based on the RMSProp algorithm [66]. Specifically, we set the step size of our algorithm as

$$\boldsymbol{\eta}(t)=\Big[\frac{\varsigma(t)}{\sqrt{\zeta+\psi_1(t)}},\cdots,\frac{\varsigma(t)}{\sqrt{\zeta+\psi_L(t)}}\Big], \tag{3.38}$$

where $\varsigma(t)=\frac{a}{t^b}$ is the global step size with $a>0$ and $b>0.5$ [85], and $\zeta=10^{-12}$ is a small positive constant used to stabilize division by small numbers. $\psi_l(t)$ is the l-th element of the accumulate squared gradient $\boldsymbol{\psi}(t)$, which is given by

$$\boldsymbol{\psi}(t)=\kappa\boldsymbol{\psi}(t-1)+(1-\kappa)\tilde{\boldsymbol{\omega}}_{\mathcal{C}}(t)\circ\tilde{\boldsymbol{\omega}}_{\mathcal{C}}(t), \tag{3.39}$$

where $\kappa\in(0,1)$ is the decay factor, and $\psi_l(-1)=0, l=1,\cdots,L$.[4] We can

[4]Note that the parameters a and b of the global step size $\varsigma(t)$ and the decay factor κ of the accumulate squared gradient $\boldsymbol{\psi}(t)$ are critical to the convergence performance and should be carefully chosen. Parameter a determines the global step size in the first few iterations, and parameters b and κ control the decay rates of $\varsigma(t)$ and $\boldsymbol{\psi}(t)$, respectively. An excessively small or large a or an oversized κ would lead to slow convergence, and an oversized b would cause premature. In the simulations, the values of a, b, and κ are fine-tuned to be 0.1, 0.8 and 0.5, respectively.

Algorithm 1 SGD-based BS Antenna Placement Scheme

Require: Initial positions of BS antenna clusters $\mathbf{r}^{\mathrm{AP}}(0)$.

1: **for** $t = 0, \cdots, T-1$ **do**
2: Compute the stochastic gradient $\tilde{\omega}_{\mathcal{C}}(t)$ based on (3.36) and the step size $\boldsymbol{\eta}(t)$ based on (3.38) with a sample of users' positions $\mathbf{r}^{\mathrm{user}}(t)$.
3: Update the positions of BS antenna clusters $\mathbf{r}^{\mathrm{AP}}(t+1)$ based on (3.40).
4: **end for**
5: $\mathbf{r}^{\mathrm{AP}*} \leftarrow \mathbf{r}^{\mathrm{AP}}(T)$.

Ensure: The optimized positions of BS antenna clusters $\mathbf{r}^{\mathrm{AP}*}$.

see from (3.38)–(3.39) that with this vectorized step size model, the BS antenna clusters far away from users with smaller magnitudes of partial derivative would have larger step sizes, and thus can more easily converge to the user-intensive areas. While the BS antenna clusters in user-intensive areas would make more conservative movements due to smaller step sizes.

By letting $\tilde{\boldsymbol{\omega}}_F = \tilde{\boldsymbol{\omega}}_{\mathcal{C}}$ and setting the step size based on (3.38), (2.4) becomes

$$\mathbf{r}^{\mathrm{AP}}(t+1) = \Pi_{\mathcal{A}^L}(\mathbf{r}^{\mathrm{AP}}(t) + \tilde{\boldsymbol{\omega}}_{\mathcal{C}}(t) \circ \boldsymbol{\eta}(t)). \tag{3.40}$$

Note that with the projection operator $\Pi_{\mathcal{A}^L}(\cdot)$, the proposed algorithm can be applied to different shapes of area $\mathcal{A}$ with practical geographic constraints. Based on the above illustration, we summarize our SGD-based BS antenna placement scheme in Algorithm 1.

3.5 Simulation Results

In this section, simulations are conducted to demonstrate the performance of the proposed BS antenna placement scheme. In particular, we consider an uplink DAS with K users and L BS antenna clusters distributed inside $\mathcal{A}$, a circular cell with radius $R_{\mathrm{area}} = 1$, where each user is equipped with N_u antennas, and each BS antenna cluster has N_c co-located antennas. We generate the users' positions in a similar way to the traffic generation method in [86], where users gather around some hotspots with a certain degree of aggregation. We start from randomly generated positions of users and hotspots, then update the position of each user by moving towards its closest hotspot with a Gaussian distributed distance with mean $\mu_\delta d_0$ and variance $(\frac{0.5-|\mu_\delta-0.5|}{3} d_0)^2$, where d_0 is the distance between each user and its closest hotspot. When $\mu_\delta = 0$, all the users are uniformly distributed

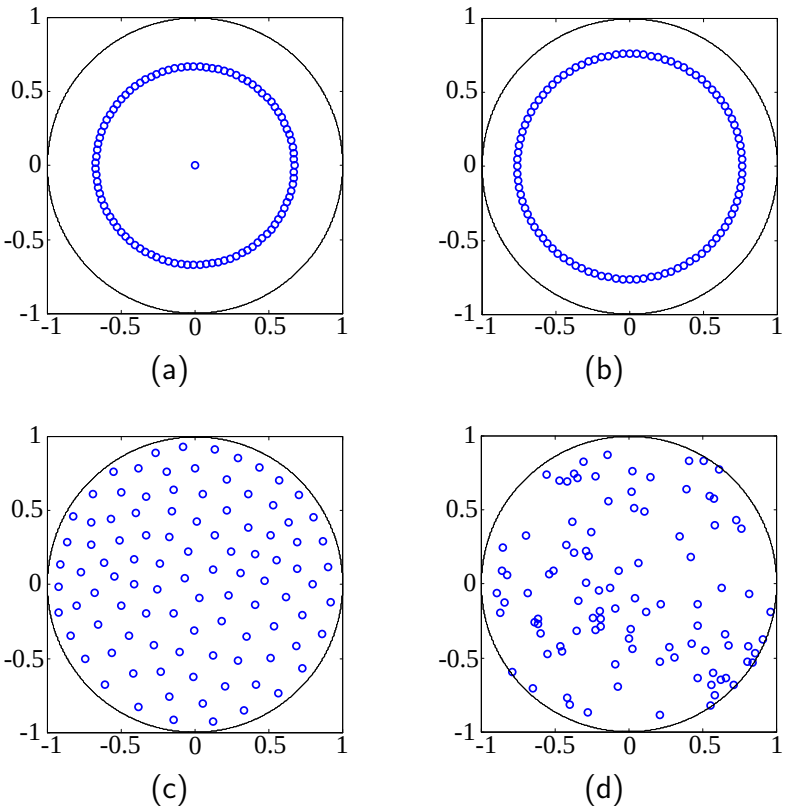

FIGURE 3.4: Positions of BS antenna clusters with (a) the circular plus centre placement in [77], (b) the circular placement in [38, 39], (c) the even placement in [37, 87], and (d) the uniform placement in [88, 89]. $L = 100$ BS antenna clusters are represented by circles.

within the area. The aggregation level increases with μ_δ, and users are located exactly at the hotspots when $\mu_\delta = 1$. The number of hotspots is set as 5.

In simulations, the normalized ergodic sum capacity C for a given system topology is evaluated according to (3.3) by averaging over 100 samples of small-scale fading coefficients. The average normalized ergodic sum capacity $\bar{C}$ in Figs. 3.5 and 3.6 is obtained by further averaging C over 100 samples of users' positions generated with a realization of hotspots. In Figs. 3.7 and 3.9, $\bar{C}$ is obtained by averaging C over 10000 samples of users' positions, where every 100 samples are generated with a common realization of hotspots.

We consider the following representative BS antenna placement schemes as the performance benchmarks:

- **Circular plus centre placement**: In [77], one BS antenna cluster is placed at the cell centre, while the others are evenly placed on a concentric circle

of the cell. The radius of the circular layout is optimized to maximize the downlink expected SNR as $R_{\text{opt1}} = \frac{2R_{\text{area}}(1-L^{-3/2})(\sin(\pi/(L-1)))}{3\pi(1-1/L)/(L-1)}$. A graphic illustration of this scheme is given in Fig. 3.4a.

- **Circular placement**: In [38] and [39], all the BS antenna clusters are evenly placed on a concentric circle of the cell, as shown in Fig. 3.4b. The radius of the circular antenna layout is optimized to maximize the asymptotic uplink average ergodic sum capacity under the assumption that $M \gg N$, and is derived as $R_{\text{opt2}} = \frac{R_{\text{area}}}{\sqrt{1+v}}$, where v is the solution of the equation $x^{3+\frac{2}{\alpha-2}} + 2x^{2+\frac{2}{\alpha-2}} - 1 = 0$. With the path-loss factor $\alpha = 4$, $v = 0.717$.

- **Even placement**: In [87], BS antennas are evenly placed by following the Quadrature Amplitude Modulation (QAM) constellation. For an arbitrary number of BS antenna clusters, the Lloyd algorithm can be applied to generate such an antenna placement [37], as shown in Fig. 3.4c.

- **Uniform placement**: In practice, it may not be feasible to place BS antennas in the above regular manners due to complicated geographic conditions. In [88] and [89], BS antenna clusters are assumed to be uniformly distributed. One sample is shown in Fig. 3.4d.

3.5.1 Convergence Performance

As demonstrated in Section 2.3, one advantage of the vectorized step size model is to make SA-based algorithms insensitive to the initial positions of BS antenna clusters. To show this advantage, we consider the following scalar step size model as a comparison benchmark:

$$\eta(t) = \frac{0.5}{t^{0.8}\|\tilde{\boldsymbol{\omega}}_{\mathcal{C}}(t)\|}. \tag{3.41}$$

We also consider four different initial placements of BS antennas, which are the four benchmark schemes shown in Fig. 3.4.

Fig. 3.5 demonstrates the convergence performance of the proposed algorithm with various initial placement schemes under the vectorized and scalar step size models, where users are non-uniformly distributed with aggregation factor $\mu_\delta = 0.7$. We can see that the convergence of the algorithm is sensitive to the initial placement schemes when the scalar step size model given in (3.41) is adopted.

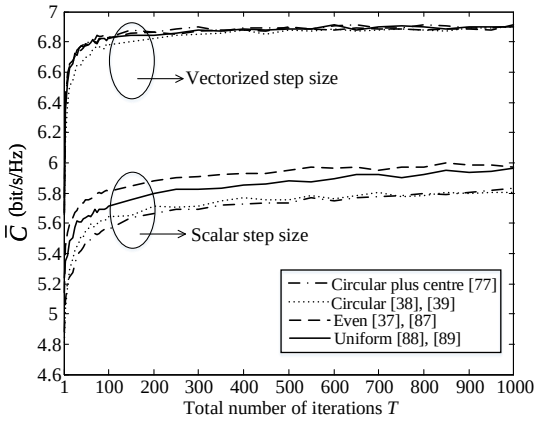

FIGURE 3.5: Average normalized ergodic sum capacity $\bar{C}$ (bit/s/Hz) versus the total number of iterations T with different initial placements of BS antenna clusters and step size models (3.38) and (3.41). $L = 100$, $K = 100$, $N_c = 4$, $N_u = 1$, $\mu_\delta = 0.7$, $\rho_0 = 10$ dB.

While when the proposed vectorized step size model given in (3.38) is adopted, the algorithm converges to the same value of average ergodic sum capacity with different initial placements of BS antenna clusters.

As demonstrated in Section 3.4, with non-uniformly distributed users, the average ergodic sum capacity is much more sensitive to the movements of the BS antenna clusters in the user-intensive areas than those far away from users. Correspondingly, the values of the partial derivatives of the ergodic sum capacity with respect to different BS antenna clusters' positions would vary widely, making it difficult to adopt a proper scalar step size for updating the positions of all the BS antenna clusters. For instance, for the BS antenna clusters located in the user-intensive areas, a small step size is preferable in order to keep them in these areas. However, with a small step size, the BS antenna clusters far away from users would make overly conservative movement, and cannot move to the user-intensive areas. In that case, the performance of the algorithm would be crucially dependent on the initial placement of BS antenna clusters, i.e., how many of them are initially far away from the user-intensive areas.

The drawback of scalar step size model can be overcome by instead adopting the vectorized step size model (3.38), which adapts its value to the stochastic gradient elementwisely. As a result, the algorithm becomes insensitive to the initial placement of BS antennas. In the following simulations, we will generate the initial positions of BS antenna clusters $\mathbf{r}^{\mathrm{AP}}(0)$ in Algorithm 1 following the uniform distribution.

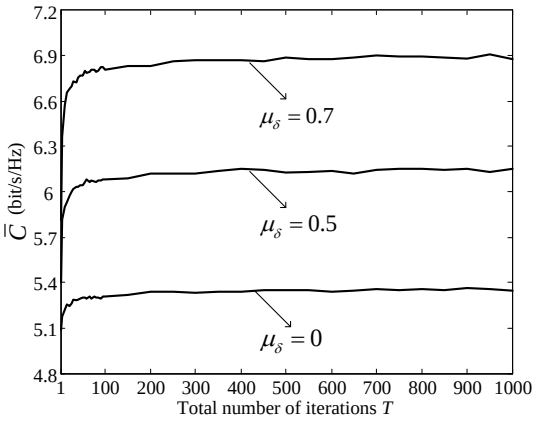

FIGURE 3.6: Average normalized ergodic sum capacity $\bar{C}$ (bit/s/Hz) versus the total number of iterations T. $L = 100$, $K = 100$, $N_c = 4$, $N_u = 1$, $\rho_0 = 10$ dB.

Fig. 3.6 demonstrates the convergence performance of the proposed algorithm with various values of aggregation factor μ_δ. We can see that the proposed algorithm converges fast regardless of the aggregation degree of users. Based on the convergence results, in Figs. 3.7–3.9, we set the total number of iterations T as 500.

3.5.2 Comparison with the Benchmarks

Fig. 3.7 shows how the average normalized ergodic sum capacity $\bar{C}$ varies with the aggregation factor μ_δ under different BS antenna placement schemes. We can see that the proposed scheme significantly outperforms the benchmarks especially when μ_δ is large. The reason is that an underlying assumption for the benchmarks is that users are uniformly distributed. As a result, when the aggregation factor μ_δ increases, i.e., users become more clustered, the average access distance between each user and the BS antennas increases. In contrast, with the proposed scheme, the placement of BS antennas is optimized based on samples of users' locations. As Fig. 3.8 illustrates, when $\mu_\delta = 0$, i.e., users are uniformly distributed, the BS antenna clusters are evenly scattered over the cell, which forms a similar pattern to the even scheme. While when users are clustered with $\mu_\delta = 0.5$, BS antenna clusters are placed in the user-intensive areas, with which the average access distance between users and BS antennas is reduced. That is why the average normalized ergodic sum capacity with the proposed placement scheme increases with μ_δ, leading to more substantial gains over the benchmarks.

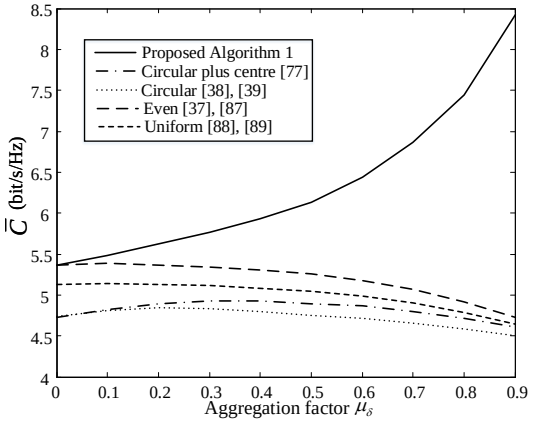

FIGURE 3.7: Average normalized ergodic sum capacity $\bar{C}$ (bit/s/Hz) versus the aggregation factor μ_δ. $L = 100$, $K = 100$, $N_c = 4$, $N_u = 1$, $\rho_0 = 10$ dB.

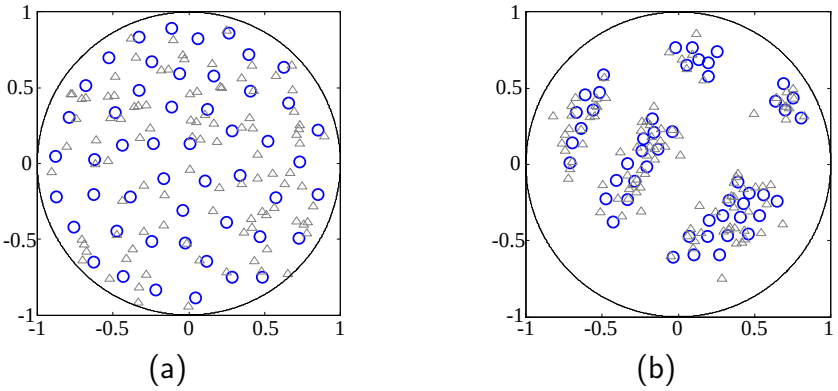

FIGURE 3.8: Positions of BS antenna clusters with the proposed Algorithm 1. Snapshots of users' positions are also presented. $L = 50$ BS antenna clusters and $K = 100$ users are represented by circles and triangles, respectively. (a) $\mu_\delta = 0$. (b) $\mu_\delta = 0.5$.

Note that in Fig. 3.7, we fix the aggregation factor μ_δ when calculating $\bar{C}$ to explore the effect of μ_δ on the capacity performance of different BS antenna placement schemes. In practice, μ_δ may be time-varying, as users' aggregation level may change with time. In Fig. 3.9, we set μ_δ as a uniform random variable between 0.3 and 0.7 and demonstrate that the capacity gain brought by the proposed scheme also increases as the number of BS antenna clusters L grows. Specifically, with L increasing, the sum capacity with the circular placement schemes quickly becomes saturated, and may even be worse than the case with uniform distribution. By restricting the BS antennas to a one-dimensional circle, the increase of L can hardly reduce the average access distance between users and BS antennas when L is large. Improvements can be observed only when the BS antenna clusters are scattered over the cell, e.g., with the even or uniformly distributed placement.

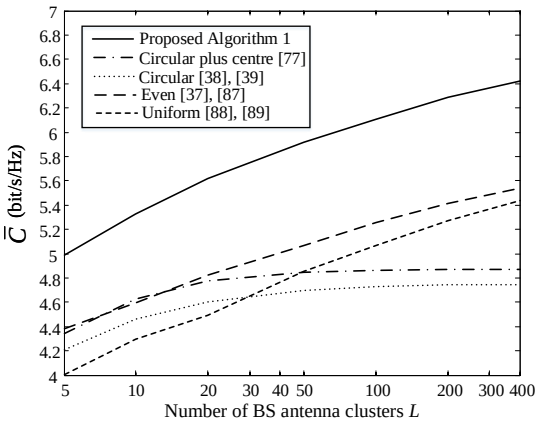

FIGURE 3.9: Average normalized ergodic sum capacity $\bar{C}$ (bit/s/Hz) versus the number of BS antenna clusters L. $K = 100$, $N_u = 1$, $M = 400$, $\rho_0 = 10$ dB, μ_δ is uniformly distributed between 0.3 and 0.7.

Nevertheless, both of them are significantly inferior to the proposed placement due to the lack of consideration of users' spatial distribution.

3.6 Summary

In this chapter, we propose a BS antenna placement scheme to maximize the uplink average ergodic sum capacity of a multi-user DAS. Starting from the derivation of the asymptotic ergodic sum capacity, we demonstrate that by solving the fixed-point equations using an iterative method, closed-form approximate expressions of the asymptotic ergodic sum capacity can be obtained, where the tradeoff between accuracy and computational complexity can be flexibly controlled by adjusting the number of iterations. For illustration, zero-order and first-order approximations are proposed, and compared against two representative ones, i.e., DR approximation and AO approximation. Simulation results show that both proposed approximations provide much better accuracy than the DR approximation and AO approximation under various system settings.

As it is difficult to solve the BS antenna placement problem with a deterministic approach, we formulate it as a stochastic optimization problem and propose an SGD-based algorithm to effectively solve it. In the proposed algorithm, the stochastic gradient is calculated based on the zero-order approximation of the ergodic sum capacity. Simulation results show that compared with existing representative BS antenna placement schemes, the proposed algorithm can achieve a

much higher average ergodic sum capacity, and the improvement increases with the aggregation degree of users and the number of BS antenna clusters.

In the next chapter, the stochastic optimization framework is further extended to optimize the BS placement of an mmWave cellular network. As the BS placement problem involves an expectation constraint, we will develop a novel algorithm to solve it based on CSA.

Chapter 4

Optimal BS Placement and User Association for Millimeter Wave Communication Networks

In this chapter, the stochastic optimization framework is further extended to optimize the BS placement scheme of an mmWave cellular network for improving the long-term physical accessibility between users and BSs.[1] A low-complexity user association scheme is also proposed to minimize the instantaneous outage probability. Combined with the proposed user association scheme, the proposed BS placement scheme outperforms the previous representative BS placement scheme in terms of long-term outage probability. This chapter is organized as follows. Section 4.1 presents a literature review on previous related work. Section 4.2 presents the system model and formulates the BS placement problem and user association problem. In Section 4.3, a CSA-based optimal BS placement scheme is proposed to solve the BS placement problem. The user association issue is addressed in Section 4.4, where a low-complexity outage mitigation user association scheme is proposed. Simulation results are presented in Section 4.5 to illustrate the performance of the proposed BS placement scheme and user association scheme. Finally, Section 4.6 summarizes this chapter.

[1] In this chapter, AP is referred to as BS.

4.1 Previous Work

BS placement and user association are two important issues in wireless communication systems and have been extensively studied in the past decades [14, 16, 21, 25, 30, 90–93]. In traditional cellular system, BSs are usually deployed to achieve seamless coverage of the network. Whether a user can be covered by one BS is determined by the distance between them [14, 16, 21, 25, 30]. However, such distance-based BS placement schemes are not suitable for mmWave networks, where the blockage effect becomes more pronounced. If mmWave BSs are deployed by only considering the distances to users, it can be expected that the transmissions between some BSs and users could be easily blocked by the surrounding obstacles. Therefore, there is an essential need to re-design BS placement schemes for mmWave networks. Furthermore, in order to compensate for the high path loss, mmWave BSs usually employ a massive number of antennas to form narrow beams, which result in significant reduction of co-channel interference. Consequently, user association metrics for interference-limited wireless networks [90–93] are not well suited to noise-limited mmWave networks. New user association approaches should be customized for mmWave networks without considering interference coordination.

There have been many studies on the design of BS placement and user association for mmWave networks. In the following, let us review each issue separately.

4.1.1 BS Placement

The main challenge of mmWave BS placement originates from the severe blockage effect, which requires the consideration of the layout of obstacles when placing BSs. Some of the existing works assumed that both obstacles and mmWave BSs are randomly distributed, and characterized the impact of BS density on the received signal quality [6] or connectivity probability [7] of a typical user. Based on the system model in [7], the optimal density of BSs for minimizing BS placement cost under a certain connectivity probability constraint was derived in [8].

To further determine the optimal positions of BSs, a static blockage model was adopted in [22, 29, 94, 95], where the positions, shapes and orientations of obstacles are fixed. To maximize the line-of-sight (LoS) coverage of the mmWave networks,

various assumptions on the feasible positions of BSs were made in [94, 95]. In [94], the region under consideration was divided into quadrilateral or triangle areas, and the mmWave BSs were placed at the intersection points of areas. Different from [94] where mmWave BSs were placed in the open space, in [95], the BSs were assumed to be deployed on the facade of buildings. The shapes of buildings were approximated by simple polygons, and the optimal positions of BSs were determined by applying the computational geometry theory.

As [94, 95] only focused on maximizing the coverage area, the proposed symmetric BS placements may lead to unbalanced workloads among BSs if users' positions are asymmetric. The user-position-dependent BS placement schemes were further studied in [22] and [29], where users are represented by RPs with given positions. By generating the CSs of BSs based on the BS placement scheme proposed in [95], the optimal subsets of CSs were searched for satisfying the received signal strength requirement at each RP with the minimum number of BSs [22], or covering the maximum number of RPs [29].

Note that in [22] and [29], the BS placement problem is formulated as a deterministic optimization problem, where the objective function is calculated based on a given set of fixed users' positions. In practice, however, as the positions of users vary with time, the optimal BS placement based on such a deterministic optimization framework would quickly become obsolete when the users' positions change. A stochastic optimization approach could be more appropriate. In this chapter, the stochastic optimization framework established in Chapter 2 will be further extended to optimize the BS placement for mmWave networks. Specifically, we aim at maximizing the long-term average number of physically accessible BSs of each user under a constraint of inaccessible probability, i.e., the long-term proportion of users who has no physically accessible BS. As both the objective and constraint functions are in the form of expectation, we propose a novel CSA-based algorithm to effectively search the optimal placement of BSs.

4.1.2 User Association

In current standards for mmWave Wi-Fi networks, such as IEEE 802.11ad and IEEE 802.15.3c, the received-signal-strength-indicator (RSSI) criterion is adopted for user association [96], that is, each user associates with the AP that provides the strongest signal strength. Despite the simplicity, the RSSI-based user association

scheme may lead to unbalanced workloads of BSs when users are not uniformly distributed. In recent years, the load-balancing user association schemes for mmWave networks have gained much attention and have been studied in [97–100], where the workload of each BS, e.g., the number of users that can be served by one BS, is strictly bounded. In [97], for instance, a distributed algorithm was proposed to minimize the maximum workload among the APs. In [98], user association and spectrum resource allocation are jointly optimized for maximizing the downlink sum-rate and minimizing the reallocation cost of handovers simultaneously. In [99], a load balancing user association scheme was proposed to maximize the network utility function, which can be the downlink sum-rate or the minimum downlink rate among the users. In [100], a heuristic algorithm was proposed to jointly maximize the number of users that can associate with one BS and minimize the total number of time-frequency resource blocks consumed by the BSs.

In the above studies, the focus is usually placed on balancing the workloads of BSs. The outage probability, i.e., the proportion of the users that cannot be served by any BSs, is an important performance metric for mmWave networks [6–8], which nevertheless has seldom been considered when optimizing the user association scheme. In this chapter, with the optimized positions of BSs, we focus on the user association problem in a given time slot, with the objective of minimizing the outage probability under a constraint of the maximum workload for each BS. We propose a low-complexity outage mitigation user association scheme, which can achieve similar performance to the Hungarian-algorithm-based optimal user association scheme, but with much less running time.

4.2 System Model and Problem Formulation

We consider an mmWave network with multiple BSs serving for a group of users, denoted by $\mathcal{L}$ and $\mathcal{K}$, respectively, in a 2-dimensional Manhattan-type geometry [29]. One example is shown in Fig. 4.1, where multiple square blocks (i.e., buildings) with edge length E_{bui} are separated by streets with width W_{str}.[2] The

[2]It should be noted that this geometry can be easily extended to the scenario with irregular-shape buildings by approximating their shapes with a series of squares of different sizes. A tradeoff between computational complexity and approximate accuracy can be obtained by adjusting the number of blocks.

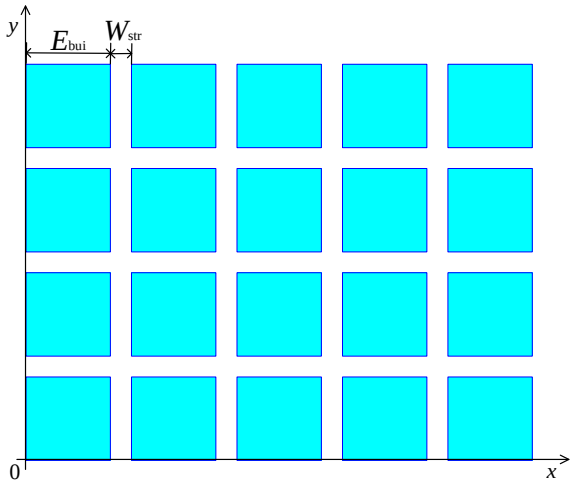

FIGURE 4.1: 2-dimensional Manhattan-type geometry.

numbers of BSs, users and blocks are given by L, K and B, respectively. We consider the Cartesian coordinate and denote the positions of user k, BS l and the center of block b as $\mathbf{r}_k^{\text{user}} = [x_k^{\text{user}}, y_k^{\text{user}}], k = 1, \cdots, K$, $\mathbf{r}_l^{\text{AP}} = [x_l^{\text{AP}}, y_l^{\text{AP}}], l = 1, \cdots, L$, and $\mathbf{r}_b^{\text{block}} = [x_b^{\text{block}}, y_b^{\text{block}}], b = 1, \cdots, B$, respectively. We also assume that each BS can only serve at most $K_{\max}$ users, which is referred to as the *maximum workload*. And each user can only be served by one BS.

In this chapter, we are interested at the outage probability of the mmWave network, which is defined as the proportion of users that cannot associate with any BSs. Generally, whether a user can associate with one BS is determined by both the user association scheme and the physical accessibility between the user and the BS, i.e., the user is within the coverage area of the BS and there exists an LoS path between them. The physical accessibility between a BS and a user is closely related to their positions. Therefore, in order to improve the network outage performance, both the BS placement and user association should be optimized. In the following, we will formulate the BS placement problem and user association problem.

4.2.1 BS Placement Problem

For the BS placement problem, we are concerned with how to properly place the BSs to optimize the physical accessibility. The physical accessibility between a BS and a user is determined by two factors: coverage and blockage. Let us define

two Boolean indicators $\mathcal{I}_{l,k}^{\text{out}}$ and $\mathcal{I}_{l,k}^{\text{block}}$ as

$$\mathcal{I}_{l,k}^{\text{out}} = \begin{cases} 1 & \text{if user } k \text{ is outside the coverage range of BS } l \\ 0 & \text{otherwise,} \end{cases} \tag{4.1}$$

$$\mathcal{I}_{l,k}^{\text{block}} = \begin{cases} 1 & \text{if there is no LoS path between BS } l \text{ and user } k \\ 0 & \text{otherwise.} \end{cases} \tag{4.2}$$

We can see that user k can physically access BS l if and only if $I\left(\mathcal{I}_{l,k}^{\text{out}} + \mathcal{I}_{l,k}^{\text{block}}\right) = 0$, where

$$I(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.3}$$

denotes the indicator function. With $I(x)$, the expression of $\mathcal{I}_{l,k}^{\text{out}}$ can be written as

$$\mathcal{I}_{l,k}^{\text{out}} = I(d_{l,k} - R_{\text{B}}), \tag{4.4}$$

where R_{B} is the coverage radius of each BS and $d_{l,k}$ denotes the distance between BS l and user k. For $\mathcal{I}_{l,k}^{\text{block}}$, it is shown in Appendix E that

$$\begin{aligned} \mathcal{I}_{l,k}^{\text{block}} = & I\left(\sum_{b=1}^{B} I\big((d_{l,k}\sin\theta_{l,k}+E_{\text{bui}})(x_{l,k,b}\sin\theta_{l,k}-y_{l,k,b}\cos\theta_{l,k}+\frac{d_{l,k}\sin\theta_{l,k}}{2}+\frac{E_{\text{bui}}}{2})\big)\right. \\ & \cdot I\big((d_{l,k}\cos\theta_{l,k}+E_{\text{bui}})(x_{l,k,b}\cos\theta_{l,k}+y_{l,k,b}\sin\theta_{l,k}+\frac{d_{l,k}\cos\theta_{l,k}}{2}+\frac{E_{\text{bui}}}{2})\big) \\ & \cdot I\big((\sin\theta_{l,k}+\cos\theta_{l,k})(y_{l,k,b}+\frac{E_{\text{bui}}}{2}\sin\theta_{l,k}+\frac{E_{\text{bui}}}{2}\cos\theta_{l,k})\big) \\ & \cdot I\big((d_{l,k}\sin\theta_{l,k}+E_{\text{bui}})(-x_{l,k,b}\sin\theta_{l,k}+y_{l,k,b}\cos\theta_{l,k}+\frac{d_{l,k}\sin\theta_{l,k}}{2}+\frac{E_{\text{bui}}}{2})\big) \\ & \cdot I\big((d_{l,k}\cos\theta_{l,k}+E_{\text{bui}})(-x_{l,k,b}\cos\theta_{l,k}-y_{l,k,b}\sin\theta_{l,k}+\frac{d_{l,k}\cos\theta_{l,k}}{2}+\frac{E_{\text{bui}}}{2})\big) \\ & \left. \cdot I\big((\sin\theta_{l,k}+\cos\theta_{l,k})(-y_{l,k,b}+\frac{E_{\text{bui}}}{2}\sin\theta_{l,k}+\frac{E_{\text{bui}}}{2}\cos\theta_{l,k})\big)\right), \end{aligned} \tag{4.5}$$

where $x_{l,k,b}$, $y_{l,k,b}$ and $\theta_{l,k}$ are given by

$$x_{l,k,b}=\left(x_b^{\text{block}}-\frac{x_k^{\text{user}}+x_l^{\text{AP}}}{2}\right)\cos\arctan\frac{y_k^{\text{user}}-y_l^{\text{AP}}}{x_k^{\text{user}}-x_l^{\text{AP}}}+\left(y_b^{\text{block}}-\frac{y_k^{\text{user}}+y_l^{\text{AP}}}{2}\right)\sin\arctan\frac{y_k^{\text{user}}-y_l^{\text{AP}}}{x_k^{\text{user}}-x_l^{\text{AP}}}, \tag{4.6}$$

$$y_{l,k,b}=\left(y_b^{\text{block}}-\frac{y_k^{\text{user}}+y_l^{\text{AP}}}{2}\right)\cos\arctan\frac{y_k^{\text{user}}-y_l^{\text{AP}}}{x_k^{\text{user}}-x_l^{\text{AP}}}-\left(x_b^{\text{block}}-\frac{x_k^{\text{user}}+x_l^{\text{AP}}}{2}\right)\sin\arctan\frac{y_k^{\text{user}}-y_l^{\text{AP}}}{x_k^{\text{user}}-x_l^{\text{AP}}}, \tag{4.7}$$

and

$$\theta_{l,k}=-\arctan\frac{y_k^{\text{user}}-y_l^{\text{AP}}}{x_k^{\text{user}}-x_l^{\text{AP}}}+\frac{\pi}{2}I\left(\arctan\frac{y_k^{\text{user}}-y_l^{\text{AP}}}{x_k^{\text{user}}-x_l^{\text{AP}}}\right), \tag{4.8}$$

respectively.

To provide more freedom for user association, it is desirable to maximize the number of accessible BSs for each user. Meanwhile, to ensure fairness among users, the proportion of users that cannot associate with any BSs should be bounded. Define

$$\bar{L}_{\text{ac}}=\frac{1}{K}\sum_{k=1}^{K}\sum_{l=1}^{L}\left(1-I\left(\mathcal{I}_{l,k}^{\text{out}}+\mathcal{I}_{l,k}^{\text{block}}\right)\right) \tag{4.9}$$

as the *average number of accessible BSs* of each user, and

$$P_{\text{iac}}=\frac{1}{K}\sum_{k=1}^{K}\prod_{l=1}^{L}I\left(\mathcal{I}_{l,k}^{\text{out}}+\mathcal{I}_{l,k}^{\text{block}}\right) \tag{4.10}$$

as the *inaccessible probability*. As both $\mathcal{I}_{l,k}^{\text{out}}$ and $\mathcal{I}_{l,k}^{\text{block}}$ are functions of the positions of BSs $\{\mathbf{r}_l^{\text{AP}}\}$ and positions of users $\{\mathbf{r}_k^{\text{user}}\}$ according to (4.4)–(4.5), the average number of accessible BSs $\bar{L}_{\text{ac}}$ and inaccessible probability P_{iac} are also crucially dependent on the positions of users and BSs. In practice, the positions of users are time-varying due to the mobility. We are therefore interested in the long-term average performance, i.e., $\mathbb{E}_{\{\mathbf{r}_k^{\text{user}}\}}[\bar{L}_{\text{ac}}]$ and $\mathbb{E}_{\{\mathbf{r}_k^{\text{user}}\}}[P_{\text{iac}}]$. Our objective is to optimize the positions of BSs for maximizing the long-term average number of accessible BSs $\mathbb{E}_{\{\mathbf{r}_k^{\text{user}}\}}[\bar{L}_{\text{ac}}]$, under the constraint that the long-term average inaccessible probability $\mathbb{E}_{\{\mathbf{r}_k^{\text{user}}\}}[P_{\text{iac}}]$ is bounded by P_{iac}^*.

Finally, we can formulate the BS placement problem as

$$(\text{P3}): \max_{\{\mathbf{r}_l^{\text{AP}}\}} \mathbb{E}_{\{\mathbf{r}_k^{\text{user}}\}}[\bar{L}_{\text{ac}}] \tag{4.11}$$

$$\text{s.t. } \mathbb{E}_{\{\mathbf{r}_k^{\text{user}}\}}[P_{\text{iac}}] \leq P_{\text{iac}}^*, \tag{4.12}$$

$$\mathbf{r}_l^{\text{AP}} \in \mathcal{A}, l = 1, \cdots, L, \tag{4.13}$$

where $\mathcal{A}$ denotes the area in which L BSs and K users are located.

Problem P3 is a stochastic optimization problem with both objective function and constraint function in the form of expectation. As demonstrated in Chapter 2, such type of problems can be solved by CSA. In Section 4.3, a CSA-based algorithm will be proposed to solve Problem P3.

4.2.2 User Association Problem

Given the optimized positions of BSs, here we are concerned with how to properly associate users with BSs to optimize the outage performance. Define a user association indicator as

$$\mathcal{I}_{l,k}^{\text{associate}} = \begin{cases} 1 & \text{if user } k \text{ associates with BS } l \\ 0 & \text{otherwise.} \end{cases} \tag{4.14}$$

Note that user k can associate with BS l only if BS l is physically accessible for user k. We then have $\mathcal{I}_{l,k}^{\text{associate}} \leq 1 - I\left(\mathcal{I}_{l,k}^{\text{out}} + \mathcal{I}_{l,k}^{\text{block}}\right)$. Moreover, each BS can serve at most $K_{\max}$ users, which can be written as $\sum_{k=1}^{K} \mathcal{I}_{l,k}^{\text{associate}} \leq K_{\max}, l = 1, \cdots, L$. For each user, it can only associate with one BS, which indicates $\sum_{l=1}^{L} \mathcal{I}_{l,k}^{\text{associate}} \leq 1, k = 1, \cdots, K$. We aim at minimizing the outage probability, i.e., the proportion of users that cannot associate with any BSs, which can be written as

$$P_{\text{outB}} = 1 - \frac{1}{K} \sum_{k=1}^{K} \sum_{l=1}^{L} \mathcal{I}_{l,k}^{\text{associate}}. \tag{4.15}$$

Finally, we can formulate the user association problem as

$$(\text{P4}): \min_{\{\mathcal{I}_{l,k}^{\text{associate}}\}} \quad P_{\text{outB}} \tag{4.16}$$

$$\text{s.t. } \mathcal{I}_{l,k}^{\text{associate}} \leq 1 - I\left(\mathcal{I}_{l,k}^{\text{out}} + \mathcal{I}_{l,k}^{\text{block}}\right), \tag{4.17}$$

$$\sum_{k=1}^{K} \mathcal{I}_{l,k}^{\text{associate}} \leq K_{\max}, l = 1, \cdots, L, \tag{4.18}$$

$$\sum_{l=1}^{L} \mathcal{I}_{l,k}^{\text{associate}} \leq 1, k = 1, \cdots, K, \tag{4.19}$$

$$\mathcal{I}_{l,k}^{\text{associate}} \in \{0, 1\}. \tag{4.20}$$

Problem P4 is a deterministic integer linear programming problem, which can be optimally solved by the Branch-and-Bound (B&B) method [102]. However, the B&B method has been shown with exponential time complexity in the worst case [102]. Based on the virtual BS splitting technique recently proposed in [93], it will be demonstrated in Section 4.4 that P4 can be reformulated as a bipartite maximum weight matching problem, which can be optimally solved using the Hungarian algorithm in polynomial time [103]. In view of the high complexity of Hungarian algorithm when the number of BSs or users is large, a low-complexity user association algorithm will further be proposed to solve Problem P4.

4.3 CSA-based BS Placement Optimization

Based on the CSA introduced in Section 2.2.2 and by letting $\mathbf{r}^{\text{AP}} = [\mathbf{r}_1^{\text{AP}}, \cdots, \mathbf{r}_L^{\text{AP}}]$, we can solve Problem P3 with the following update

$$\mathbf{r}^{\text{AP}}(t+1) = \Pi_{\mathcal{A}^L}\left(\mathbf{r}^{\text{AP}}(t) + \tilde{\boldsymbol{\omega}}_{\text{B}}(t)\eta(t)\right), \tag{4.21}$$

where

$$\tilde{\boldsymbol{\omega}}_{\text{B}}(t) = \begin{cases} \tilde{\boldsymbol{\omega}}_{\bar{L}_{\text{ac}}}(t) & \text{if } \hat{P}_{\text{iac}}(t) - P_{\text{iac}}^{*} \leq 0 \\ -\tilde{\boldsymbol{\omega}}_{P_{\text{iac}}}(t) & \text{otherwise,} \end{cases} \tag{4.22}$$

with $\tilde{\boldsymbol{\omega}}_{\bar{L}_{\text{ac}}}(t)$ and $\tilde{\boldsymbol{\omega}}_{P_{\text{iac}}}(t)$ denoting the gradients of $\bar{L}_{\text{ac}}(t)$ and $P_{\text{iac}}(t)$ in terms of $\mathbf{r}^{\text{AP}}(t)$, respectively. $\hat{P}_{\text{iac}}(t)$ is an unbiased estimation of $\mathbb{E}_{\{\mathbf{r}_k^{\text{user}}\}}[P_{\text{iac}}(t)]$ generated with 100 realizations of users' positions.

From (4.9) and (4.10), we can see that the expressions of $\bar{L}_{\text{ac}}(t)$ and $P_{\text{iac}}(t)$ involve the indicator function $I(x)$ given in (4.3), whose derivative is undefined at $x = 0$. To make $\bar{L}_{\text{ac}}(t)$ and $P_{\text{iac}}(t)$ differentiable, we approximate $I(x)$ by a sigmoid function [104]:

$$I(x) \approx S(x) = \frac{1}{1 + \exp(-\xi x)}. \tag{4.23}$$

The parameter ξ controls the accuracy and steepness of $S(x)$. $S(x)$ is more accurate in approximating $I(x)$ with a larger ξ and $\lim_{\xi \to \infty} S(x) = I(x)$. However, a large ξ also leads to flat gradient and early stop of the algorithm on a sub-optimal point.

From (4.4) and (4.5), we can see that both the expressions of $\mathcal{I}_{l,k}^{\text{out}}$ and $\mathcal{I}_{l,k}^{\text{block}}$ include the indicator function. Therefore, the term $I\left(\mathcal{I}_{l,k}^{\text{out}} + \mathcal{I}_{l,k}^{\text{block}}\right)$ in (4.9) and (4.10) has a multi-layer structure of indicator functions. When calculating the gradient $\tilde{\boldsymbol{\omega}}_{\text{B}}(t)$ by approximating the indicator function by the sigmoid function, such a multi-layer structure will cause the vanishing gradient problem.[3] Moreover, the approximation error also increases with the number of indicator functions involved. In order to reduce approximation error and avoid vanishing gradient, we remove the outer-layer indicator function of $I\left(\mathcal{I}_{l,k}^{\text{out}} + \mathcal{I}_{l,k}^{\text{block}}\right)$ and rewrite $\bar{L}_{\text{ac}}$ and P_{iac} as

$$\bar{L}_{\text{ac}} = L - \frac{1}{K} \sum_{k=1}^{K} \left(\sum_{l=1}^{L} \mathcal{I}_{l,k}^{\text{out}} + \sum_{l \in \mathcal{L}_k} \mathcal{I}_{l,k}^{\text{block}} \right) \tag{4.24}$$

and

$$P_{\text{iac}} = \frac{1}{K} \sum_{k=1}^{K} \left(\prod_{l=1}^{L} \mathcal{I}_{l,k}^{\text{out}} + \prod_{l \in \mathcal{L}_k} \mathcal{I}_{l,k}^{\text{block}} \right), \tag{4.25}$$

respectively, where $\mathcal{L}_k = \{l \in \mathcal{L} : d_{l,k} \leq R_{\text{B}}\}$ denotes the set of BSs that include user k in their coverage areas. The detailed derivations of (4.24) and (4.25) are provided in Appendix F.

[3] More specifically, as the sigmoid function has a large saturated region where the first derivative is negligibly small, if the approximations of $\bar{L}_{\text{ac}}(t)$ and $P_{\text{iac}}(t)$ include multi-layer sigmoid functions, the multiplicative effect of the chain rule will lead to a vanishingly small magnitude of the gradient $\tilde{\boldsymbol{\omega}}_{\text{B}}(t)$, preventing the positions of BSs from changing their values. Such a problem commonly arises in the context of machine learning when training a deep neural network with sigmoid activation function, and has been referred to as vanishing gradient problem in [105, 106].

With (4.24) and (4.25), we can obtain the gradient $\tilde{\boldsymbol{\omega}}_{\mathrm{B}}(t)$ as

$$\tilde{\boldsymbol{\omega}}_{\mathrm{B}}(t) = \tilde{\boldsymbol{\omega}}_{\mathrm{B}}^{\mathrm{out}}(t) + \tilde{\boldsymbol{\omega}}_{\mathrm{B}}^{\mathrm{block}}(t), \tag{4.26}$$

where

$$\tilde{\boldsymbol{\omega}}_{\mathrm{B}}^{\mathrm{out}}(t) = \begin{cases} -\frac{1}{K}\sum_{k=1}^{K}\sum_{l=1}^{L}\frac{\partial \mathcal{I}_{l,k}^{\mathrm{out}}(t)}{\partial \mathbf{r}^{\mathrm{AP}}(t)} & \text{if } \hat{P}_{\mathrm{iac}}(t) - P_{\mathrm{iac}}^{*} \leq 0 \\ -\frac{1}{K}\sum_{k=1}^{K}\sum_{l=1}^{L}\prod_{i=1,i\neq l}^{L}\mathcal{I}_{i,k}^{\mathrm{out}}(t)\frac{\partial \mathcal{I}_{l,k}^{\mathrm{out}}(t)}{\partial \mathbf{r}^{\mathrm{AP}}(t)} & \text{otherwise} \end{cases} \tag{4.27}$$

and

$$\tilde{\boldsymbol{\omega}}_{\mathrm{B}}^{\mathrm{block}}(t) = \begin{cases} -\frac{1}{K}\sum_{k=1}^{K}\sum_{l\in\mathcal{L}_k}\frac{\partial \mathcal{I}_{l,k}^{\mathrm{block}}(t)}{\partial \mathbf{r}^{\mathrm{AP}}(t)} & \text{if } \hat{P}_{\mathrm{iac}}(t) - P_{\mathrm{iac}}^{*} \leq 0 \\ -\frac{1}{K}\sum_{k=1}^{K}\sum_{l\in\mathcal{L}_k}\prod_{i\in\mathcal{L}_k,i\neq l}\mathcal{I}_{i,k}^{\mathrm{block}}(t)\frac{\partial \mathcal{I}_{l,k}^{\mathrm{block}}(t)}{\partial \mathbf{r}^{\mathrm{AP}}(t)} & \text{otherwise.} \end{cases} \tag{4.28}$$

$\frac{\partial \mathcal{I}_{l,k}^{\mathrm{out}}(t)}{\partial \mathbf{r}^{\mathrm{AP}}(t)}$ and $\frac{\partial \mathcal{I}_{l,k}^{\mathrm{block}}(t)}{\partial \mathbf{r}^{\mathrm{AP}}(t)}$ can be calculated by combining (G.1)–(G.4) and (G.5)–(G.12), respectively, with detailed derivations given in Appendix G.

From (4.4) and (4.5), we can see that the inputs of the indicator function in $\mathcal{I}_{l,k}^{\mathrm{out}}$ and $\mathcal{I}_{l,k}^{\mathrm{block}}$ have different orders of magnitude. In $\mathcal{I}_{l,k}^{\mathrm{out}}$, the input of the indicator function $I(x)$ is $d_{l,k} - R_{\mathrm{B}}$. While in $\mathcal{I}_{l,k}^{\mathrm{block}}$, some of the inputs of $I(x)$ include the term $d_{l,k}^2$, which is of a higher order of magnitude. Therefore, when replacing the indicator function $I(x)$ with the sigmoid function $S(x)$ in $\frac{\partial \mathcal{I}_{l,k}^{\mathrm{out}}(t)}{\partial \mathbf{r}^{\mathrm{AP}}(t)}$ and $\frac{\partial \mathcal{I}_{l,k}^{\mathrm{block}}(t)}{\partial \mathbf{r}^{\mathrm{AP}}(t)}$, we should use a smaller value of parameter ξ for $\frac{\partial \mathcal{I}_{l,k}^{\mathrm{block}}(t)}{\partial \mathbf{r}^{\mathrm{AP}}(t)}$ to reduce the steepness of the sigmoid function and avoid the input of the sigmoid function falling in the saturated region, which leads to zero gradient and early stop of the algorithm. Specifically, in the simulation part, when calculating $\tilde{\boldsymbol{\omega}}_{\mathrm{B}}(t)$, the parameter ξ of the sigmoid function (4.23) is set as 1 for $\tilde{\boldsymbol{\omega}}_{\mathrm{B}}^{\mathrm{out}}(t)$ and 0.001 for $\tilde{\boldsymbol{\omega}}_{\mathrm{B}}^{\mathrm{block}}(t)$.

Furthermore, we update $\tilde{\boldsymbol{\omega}}_{\mathrm{B}}^{\mathrm{out}}(t)$ and $\tilde{\boldsymbol{\omega}}_{\mathrm{B}}^{\mathrm{block}}(t)$ with different step sizes $\eta^{\mathrm{out}}(t)$ and $\eta^{\mathrm{block}}(t)$ in each iteration:

$$\mathbf{r}^{\mathrm{AP}}(t+1) = \Pi_{\mathcal{A}^L}\Big(\mathbf{r}^{\mathrm{AP}}(t) + \big(\tilde{\boldsymbol{\omega}}_{\mathrm{B}}^{\mathrm{out}}(t)\eta^{\mathrm{out}}(t) + \tilde{\boldsymbol{\omega}}_{\mathrm{B}}^{\mathrm{block}}(t)\eta^{\mathrm{block}}(t)\big)\Big). \tag{4.29}$$

In particular, we adopt the step size model proposed in [85] and further normalize it by the norm of the gradient as

$$\eta^{\text{out}}(t) = \frac{a}{t^b \|\tilde{\boldsymbol{\omega}}_{\text{B}}^{\text{out}}(t)\|}, \tag{4.30}$$

and

$$\eta^{\text{block}}(t) = \frac{a}{t^b \|\tilde{\boldsymbol{\omega}}_{\text{B}}^{\text{block}}(t)\|}, \tag{4.31}$$

where $a \in (0, 100]$ and $b \in (0.5, 1]$ are two positive constants. The reasons for using these normalized step size models are two folds: 1) The normalization can stabilize the algorithm and is sometimes referred to as "gradient clipping" in the context of machine learning [66]. 2) We normalize $\tilde{\boldsymbol{\omega}}_{\text{B}}^{\text{out}}(t)$ and $\tilde{\boldsymbol{\omega}}_{\text{B}}^{\text{block}}(t)$ separately to ensure that $\tilde{\boldsymbol{\omega}}_{\text{B}}^{\text{out}}(t)\eta^{\text{out}}(t)$ and $\tilde{\boldsymbol{\omega}}_{\text{B}}^{\text{block}}(t)\eta^{\text{block}}(t)$ have identical scale, and hence balance the effects of the movements of BSs on the coverage and blockage performance of the network. As demonstrated in Section 3.4, the step size model $\frac{a}{t^b}$ is adopted according to [85]. a mainly determines the step size in the first few iterations while b determines the asymptotic diminishing rate. The algorithm may converge slowly or even become premature with a small a or an over-sized b. While large a or small b may cause instability of the algorithm. In the simulation parts, the values of a and b are fine-tuned to achieve good convergence performance of the proposed algorithm in different scenarios.

The output of CSA is an ergodic mean of $\mathbf{x}(t)$ over the set $\mathcal{D}$, and the convergence results were established based on a constant step size and $\mathcal{D}$ with the maximum size, i.e., $\mathcal{D} = \{1 \leq t \leq T | \hat{G}(t) \leq \bar{G}^*\}$ [59]. In our case, we select the diminishing step sizes (4.30) and (4.31) to improve the convergence performance of the algorithm. As we will show in Section 4.5, with these step size models, the proposed algorithm can quickly converge even with $|\mathcal{D}| = 1$. Therefore, we only output the final solution that satisfies the constraint.

Now we can summarize our CSA-based BS placement algorithm for solving Problem P3 in Algorithm 2. It should be noted that as $\bar{L}_{\text{ac}}(t)$ and $P_{\text{iac}}(t)$ are not convex functions of $\mathbf{r}^{\text{AP}}$, Algorithm 2 may be trapped into a sub-optimal stationary point. Whether this undesirable situation occurs or not could be crucially determined by the initial positions of BSs $\mathbf{r}^{\text{AP}}(0)$. In Section 4.5, we will explore the effect of $\mathbf{r}^{\text{AP}}(0)$ on the convergence performance of Algorithm 2 via simulations.

Algorithm 2 CSA-based Optimal BS Placement Scheme

Require: Initial positions of BSs $\mathbf{r}^{\mathrm{AP}}(0)$.

1: **for** $t = 0, \cdots, T-1$ **do**
2: Compute the stochastic gradient $\tilde{\boldsymbol{\omega}}_{\mathrm{B}}(t)$ based on (4.26) and the step sizes $\eta^{\mathrm{out}}(t)$ and $\eta^{\mathrm{block}}(t)$ based on (4.30) and (4.31) with a sample of users' positions $\mathbf{r}^{\mathrm{user}}(t)$.
3: Update the positions of BSs $\mathbf{r}^{\mathrm{AP}}(t+1)$ based on (4.29).
4: **if** $\hat{P}_{\mathrm{iac}}(t+1) \leq P^{*}_{\mathrm{iac}}$ **then**
5: $\mathbf{r}^{\mathrm{AP}*} \leftarrow \mathbf{r}^{\mathrm{AP}}(t+1)$.
6: **end if**
7: **end for**

Ensure: The optimized positions of BSs $\mathbf{r}^{\mathrm{AP}*}$.

4.4 User Association Optimization

In this section, we first show that by using the virtual BS splitting technique in [93], Problem P4 can be reformulated as a maximum weight matching problem, which can be optimally solved by the Hungarian algorithm, but at the price of high computational complexity. To further reduce the computational complexity, we propose a sub-optimal low-complexity user association algorithm to solve Problem P4.

4.4.1 Optimal User Association Scheme

Let us split each BS into $K_{\max}$ virtual BSs and denote the set of virtual BSs as $\mathcal{L}_{\mathrm{V}} = \{\mathsf{l}_1, \cdots, \mathsf{l}_{LK_{\max}}\}$, where l_q is the q-th virtual BS which is split from BS l if $q \in [K_{\max}(l-1)+1, K_{\max}l]$, $q = 1, \cdots, LK_{\max}$. Let $\mathcal{K} = \{\mathsf{k}_1, \cdots, \mathsf{k}_K\}$ denote the set of users. The mmWave network can be then modelled as a bipartite graph $G(\mathcal{L}_{\mathrm{V}} \cup \mathcal{K}, \underline{\mathcal{E}})$, where $\underline{\mathcal{E}} = \{(\mathsf{l}_q, \mathsf{k}_k) : \mathsf{l}_q \in \mathcal{L}_{\mathrm{V}}, \mathsf{k}_k \in \mathcal{K}\}$ is the edge set. Denote the weight matrix of the edge set $\underline{\mathcal{E}}$ as $\mathbf{W} \in \mathbb{R}^{LK_{\max} \times K}$, whose (q,k)-th element is given by

$$W_{q,k} = 1 - I\left(\mathcal{I}^{\mathrm{out}}_{l,k} + \mathcal{I}^{\mathrm{block}}_{l,k}\right). \tag{4.32}$$

(4.32) indicates that an edge $(\mathsf{l}_q, \mathsf{k}_k)$ has weight one if and only if user k can physically access BS l. Otherwise, the edge weight is zero. Define $\underline{\mathcal{E}}^{(q)} = \{(\mathsf{l}_q, \mathsf{k}_k) : \mathsf{k}_k \in \mathcal{K}\}$ as the set of edges between virtual BS q and all users, and $\underline{\mathcal{E}}_{(k)} = \{(\mathsf{l}_q, \mathsf{k}_k) : \mathsf{l}_q \in \mathcal{L}_{\mathrm{V}}\}$ as the set of edges between user k and all virtual BSs. We can then

reformulate Problem P4 as the following maximum weight matching problem:

$$(\text{P5}): \max_{\mathcal{E}\subset\underline{\mathcal{E}}} \sum_{(\mathsf{l}_q,\mathsf{k}_k)\in\mathcal{E}} W_{q,k}, \tag{4.33}$$

$$\text{s.t. } |\mathcal{E}\cap\underline{\mathcal{E}}^{(q)}| \leq 1, \forall q = 1,\cdots,LK_{\max}, \tag{4.34}$$

$$|\mathcal{E}\cap\underline{\mathcal{E}}_{(k)}| \leq 1, \forall k = 1,\cdots,K. \tag{4.35}$$

Problem P5 can be optimally solved by using the Hungarian algorithm in polynomial time [103]. With the optimal weight matching $\mathcal{E}^*$, the optimal user association scheme $\{\mathcal{I}_{l,k}^{\text{associate}*}\}$ can be obtained as

$$\mathcal{I}_{l,k}^{\text{associate}*} = \begin{cases} 1 & \text{if } (\mathsf{l}_q,\mathsf{k}_k)\in\mathcal{E}^* \\ 0 & \text{otherwise.} \end{cases} \tag{4.36}$$

4.4.2 Low-Complexity User Association Scheme

Note that the complexity of the Hungarian algorithm can be as high as $\mathcal{O}(\max\{(LK_{\max})^3, K^3\})$ [103], which quickly grows with the number of BSs L and the number of users K. To further reduce the complexity of the Hungarian-algorithm-based optimal user association scheme, we now propose a sub-optimal low-complexity user association algorithm to solve Problem P4 iteratively. In each iteration, one user is selected to associate with one BS. The algorithm is based on the rational that a user with more physically accessible BSs has a higher chance to be able to associate with one BS. Therefore, in order to maximize the number of associated users, we choose the user, who has the minimum number of physically accessible BSs, to associate with one BS in each iteration, as it is less likely to be able to associate with one BS in the subsequent iterations. Furthermore, we should balance the workload of each BS, i.e., the number of users that served by one BS, to reduce the chance that a user cannot associate with one BS due to the maximum workload constraint. As a result, for the selected user, among its physically accessible BSs, we choose the BS that has the minimum number of associated users to be associated with.

The detailed steps of the proposed algorithm are presented in Algorithm 3. It starts from initiating a candidate BS set $\mathcal{L}_0$ and a remaining user set $\mathcal{K}_0$ by letting $\mathcal{L}_0 = \mathcal{L}$ and $\mathcal{K}_0$ be the set of users who have at least one physically accessible BS, as shown in Steps 2 and 3 of Algorithm 3. Then each user in $\mathcal{K}_0$ is selected to

Algorithm 3 Low-Complexity Outage Mitigation User Association Scheme

Require: Blockage indicator $\{\mathcal{I}_{l,k}^{\text{block}}\}$, coverage indicator $\{\mathcal{I}_{l,k}^{\text{out}}\}$, maximum workload of each BS $K_{\max}$.

1: Initialize the user association indicators $\mathcal{I}_{l,k}^{\text{associate}} \leftarrow 0, \forall l, k$.
2: $\mathcal{L}_0 \leftarrow \mathcal{L}$.
3: $\mathcal{K}_0 \leftarrow \{k \in \mathcal{K} : \prod_{l=1}^{L} I(\mathcal{I}_{l,k}^{\text{out}} + \mathcal{I}_{l,k}^{\text{block}}) = 0\}$.
4: **while** $\mathcal{K}_0 \neq \emptyset$ and $\mathcal{L}_0 \neq \emptyset$ **do**
5: $\quad k^* = \arg\min_{k \in \mathcal{K}_0} \sum_{l=1}^{L} \big(1 - I(\mathcal{I}_{l,k}^{\text{out}} + \mathcal{I}_{l,k}^{\text{block}})\big)$.
6: $\quad$ **if** $\mathcal{L}_0 \cap \{i : I(\mathcal{I}_{i,k^*}^{\text{out}} + \mathcal{I}_{i,k^*}^{\text{block}}) = 0\} \neq \emptyset$ **then**
7: $\quad\quad l^* = \arg\min_{l \in \mathcal{L}_0 \cap \{i : I(\mathcal{I}_{i,k^*}^{\text{out}} + \mathcal{I}_{i,k^*}^{\text{block}}) = 0\}} \sum_{k=1}^{K} \mathcal{I}_{l,k}^{\text{associate}}$.
8: $\quad\quad \mathcal{I}_{l^*,k^*}^{\text{associate}} \leftarrow 1$.
9: $\quad$ **end if**
10: $\quad \mathcal{K}_0 \leftarrow \mathcal{K}_0 \backslash \{k^*\}$.
11: $\quad$ **if** $\sum_{k=1}^{K} \mathcal{I}_{l^*,k}^{\text{associate}} = K_{\max}$ **then**
12: $\quad\quad \mathcal{L}_0 \leftarrow \mathcal{L}_0 \backslash \{l^*\}$.
13: $\quad$ **end if**
14: **end while**

Ensure: The user association indicators $\{\mathcal{I}_{l,k}^{\text{associate}}\}$.

associate with one BS iteratively. The user with the smallest number of physically accessible BSs is selected in each iteration, as indicated in Step 5 of Algorithm 3. It is associated to one of its physically accessible BSs, which has the smallest number of associated users, in the candidate set $\mathcal{L}_0$, as shown in Step 7. Step 12 implies that if a BS has been associated with $K_{\max}$ users, it will be removed from the candidate set $\mathcal{L}_0$. The algorithm terminates when $\mathcal{K}_0$ or $\mathcal{L}_0$ becomes empty, in which case all the users have been considered or there are no available BSs for users to associate with.

It can be clearly seen from Algorithm 3 that there is at most $\min\{K, LK_{\max}\}$ iterations. Furthermore, as finding the minimum variable over a set with cardinality l, also known as sorting algorithms, requires $l \log l$ comparisons [107], the complexity of Step 5 and Step 7 are $\mathcal{O}(K \log K)$ and $\mathcal{O}(L \log L)$, respectively. Therefore, the computational complexity of the proposed user association algorithm is $\mathcal{O}\big(\min\{K, LK_{\max}\}(K \log K + L \log L)\big)$, which is much lower than the complexity of the Hungarian algorithm. Furthermore, based on the submodular optimization theory [108, 109], Problem P2 can be solved by Algorithm 3 with a constant-factor $\frac{1}{3}$ approximation guarantee. The detailed derivation of this approximation guarantee is presented in Appendix H. In the next section, we will illustrate the performance of Algorithm 3 by comparing it with the optimal user

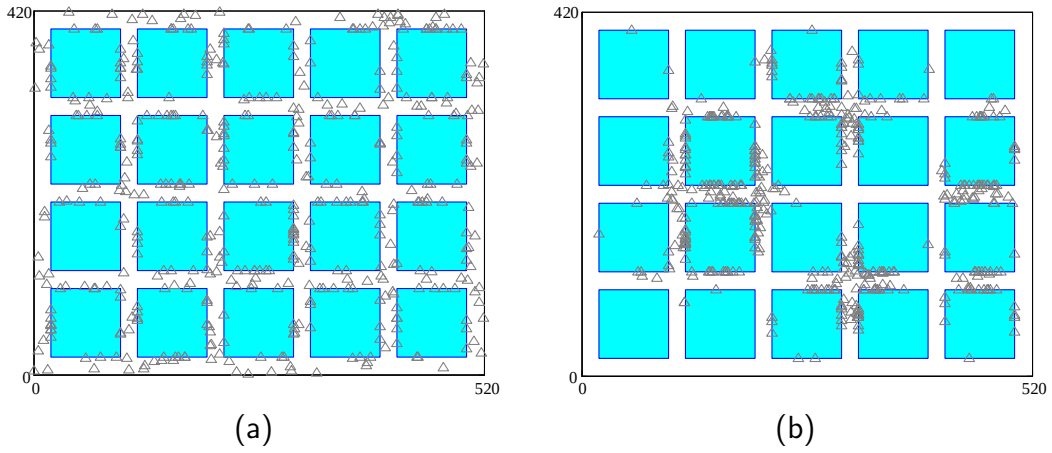

FIGURE 4.2: Snapshots of users' positions with different aggregation factors μ_δ. (a) $\mu_\delta = 0$. (b) $\mu_\delta = 0.5$.

association scheme in terms of outage performance and time efficiency via simulations.

4.5 Simulation Results and Discussions

In this section, simulations are conducted to verify the performance of the proposed CSA-based optimal BS placement scheme, i.e., Algorithm 2, and low-complexity outage mitigation user association scheme, i.e., Algorithm 3, under different distributions of users' positions and various parameter settings.

4.5.1 Simulation Setting

In particular, we consider a 2-dimensional Manhattan-type geometry with $B = 20$ square blocks as shown in Fig. 4.2, where the side length of each block E_{bui} and the street width W_{str} are set as 80 meters and 20 meters, respectively. We also fix the coverage radius of each BS R_{B} as 100 meters. Unless otherwise stated, the number of BSs L, the number of users K and the maximum workload $K_{\max}$ are set as 15, 450 and 30, respectively. We set the maximum inaccessible probability P^*_{iac} of Problem P3 as 0.05, i.e., the percentage of the users who do not have any accessible BSs should not exceed 5%.

We generate the users' positions in a similar way to the user position generation method in Section 3.5, i.e., each user is attracted by its closest hotspot.

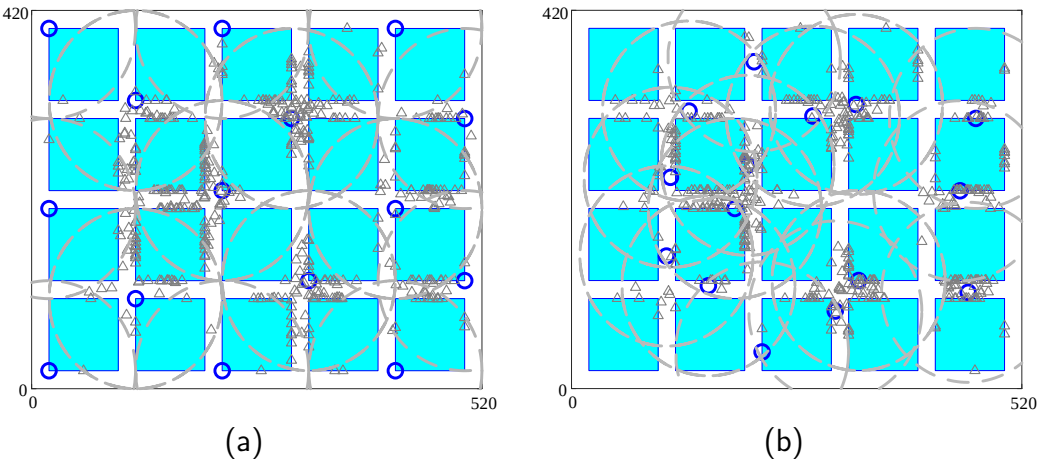

FIGURE 4.3: Positions and coverage of BSs with (a) the MC initial BS placement scheme and (b) the KM initial BS placement scheme. A snapshot of users' positions with $\mu_\delta = 0.5$ and 5 hotspots is also presented. $L = 15$ BSs and $K = 450$ users are represented by circles and triangles, respectively. The dashed line represents the coverage of each BS.

The aggregation degree of users is controlled by an aggregation factor $\mu_\delta \in [0, 1]$. The aggregation level increases with μ_δ, and users are uniformly distributed in the area when $\mu_\delta = 0$. The number of hotspots is set as 5. Furthermore, the users who fall inside the blocks are projected to the street area by using the function $\Pi_{\mathcal{A}^K}(\cdot)$ given in (2.5). Fig. 4.2 shows two samples of users' positions with $\mu_\delta = 0$ and $\mu_\delta = 0.5$.

4.5.2 Performance of the Proposed CSA-based Optimal BS Placement Algorithm

As pointed out in Section 4.3, the convergence of the proposed CSA-based optimal BS placement scheme, i.e., Algorithm 2, depends on the initial placement scheme due to the non-convexity of the objective function and constraint function. For illustration, we consider two types of initial placement schemes of BSs, which are listed as follows:

- **Maximum Coverage (MC) Scheme**: The MC scheme is a BS placement scheme proposed in [29] to maximize the LoS coverage of the mmWave network. The positions of BSs are selected from a set of CSs, which are placed on corners and middle of block sides for a Manhattan-type geometry [29]. A set of RPs are also evenly placed in the street area. The optimal BS positions

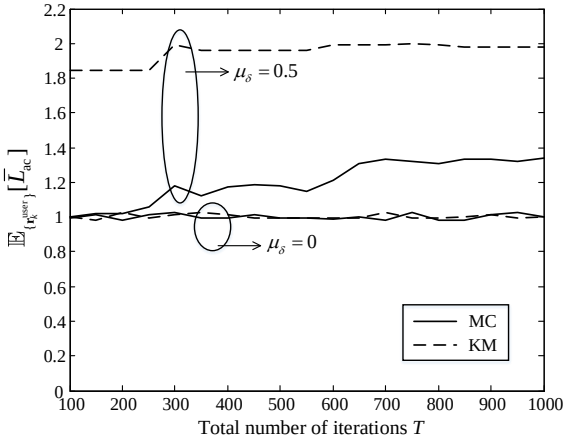

FIGURE 4.4: Long-term average number of accessible BSs $\mathbb{E}_{\{\mathbf{r}_k^{\text{user}}\}}[\bar{L}_{\text{ac}}]$ versus the total number of iterations T of Algorithm 2, the proposed CSA-based optimal BS placement scheme, with two initial placement schemes of BSs, MC and KM.

are then obtained by solving the problem of maximizing the number of RPs that can be covered by at least one BS. A graphic illustration of this scheme is shown in Fig. 4.3a.

- **K-means (KM) Scheme**: Randomly generate one sample of users' positions. Apply the K-means method [110] to group users into $L = 15$ clusters. Set the initial positions of BSs as the centroids of clusters. A graphic illustration of this scheme is shown in Fig. 4.3b.

Fig. 4.4 shows how the average number of accessible BSs with the optimized positions of BSs $\mathbf{r}^{\text{AP}*}$ varies with the total number of iterations T under the above initial placement schemes. We can see that in both cases, the proposed algorithm converges after $T \geq 700$. With $\mu_\delta = 0$, i.e., users are uniformly distributed in the area, the algorithm is insensitive to the initial placement of BSs. In contrast, in the clustered-user case with $\mu_\delta = 0.5$, the algorithm converges to a larger value with the KM initial placement scheme than that with the MC initial placement scheme, indicating that it is trapped into a sub-optimal stationary point in the latter case. The reason is that the partial derivative of the coverage indicator $\mathcal{I}_{l,k}^{\text{out}}$ with respect to each BS's position, $\frac{\partial \mathcal{I}_{l,k}^{\text{out}}}{\partial \mathbf{r}_l^{\text{AP}}}$, is crucially determined by the number of users close to the edge of the BS's coverage area. As all the BSs' positions are updated with the same step size, those with few edge-users would make overly conservative movements. With the MC initial placement scheme, the number of BSs with no users in the coverage area could be much higher than that with the KM initial placement scheme when users are clustered. Therefore, it is much more

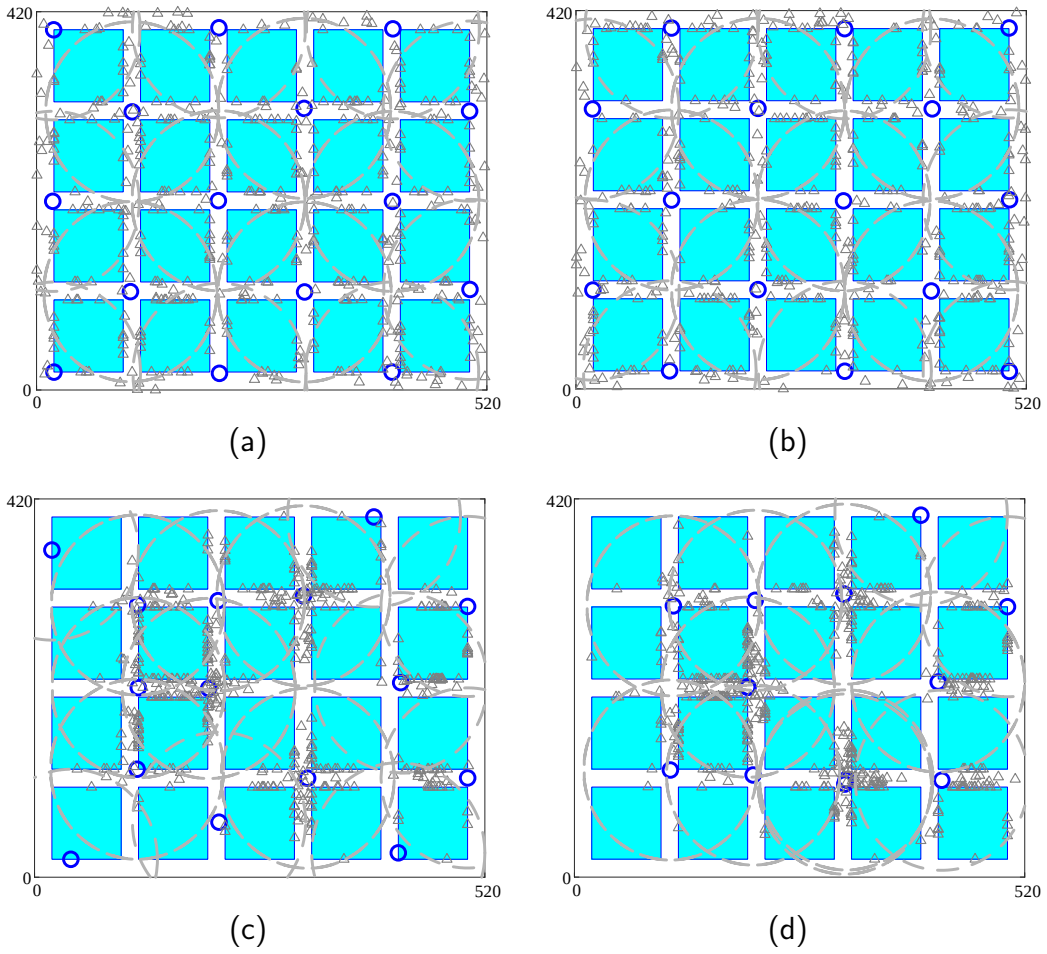

FIGURE 4.5: Positions and coverage of BSs with Algorithm 2, the proposed CSA-based optimal BS placement scheme. Snapshots of users' positions with aggregation factors $\mu_\delta \in \{0, 0.5\}$ and 5 hotspots are also presented. BSs and users are represented by circles and triangles, respectively. The dashed line represents the coverage of each BS. (a) MC initial placement, $\mu_\delta = 0$. (b) KM initial placement, $\mu_\delta = 0$. (c) MC initial placement, $\mu_\delta = 0.5$. (d) KM initial placement, $\mu_\delta = 0.5$.

likely to be trapped into sub-optimal positions. As Fig. 4.5c shows, with the MC initial placement, some of the BSs cannot move to the user-intensive area, leading to a significantly lower number of accessible BSs compared to that with the KM initial placement.

We can also see from Fig. 4.5 that after optimization, most of the BSs will move to the intersection of two streets. Intuitively, placing a BS at the intersection can maximize the number of LoS paths between the BS and the users in the two streets, and thus increase the average number of accessible BSs. Fig. 4.5 also shows that some of the BSs are located in the middle of a street, which may not

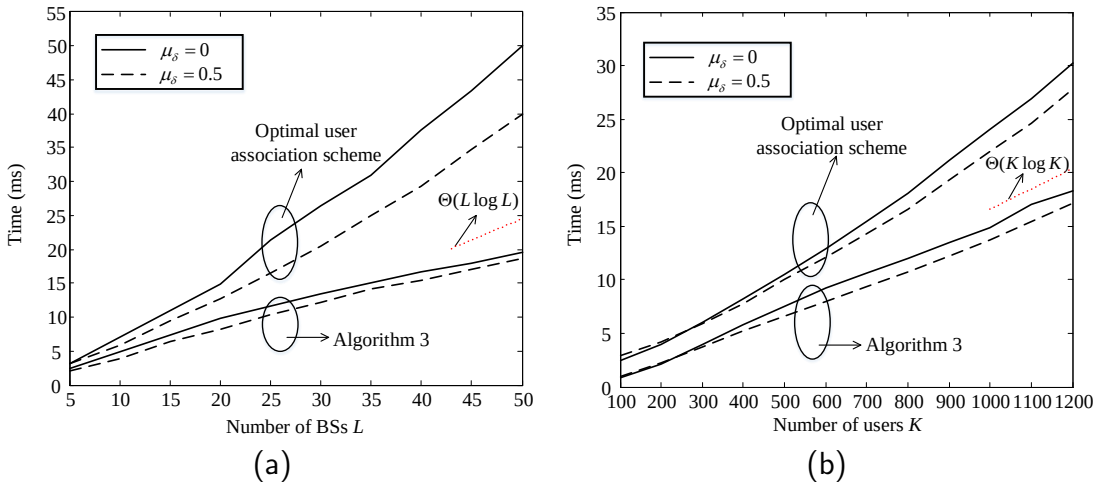

FIGURE 4.6: Average running time (in unit of milliseconds) of Algorithm 3, the proposed low-complexity outage mitigation user association scheme, and the optimal user association algorithm versus (a) the number of BSs L, and (b) the number of users K.

be feasible in practice. In Section 4.5.4, we will further project the optimized positions of BSs to the closest corners of buildings.

4.5.3 Performance of the Proposed Low-Complexity Outage Mitigation User Association Algorithm

In this subsection, we illustrate the performance of the proposed low-complexity outage mitigation user association scheme, i.e., Algorithm 3, by comparing it with the Hungarian-algorithm-based optimal user association scheme.

Fig. 4.6 shows the running time of Algorithm 3 and the optimal user association scheme, which is averaged over 1000 realizations of users' and BSs' positions. The simulation is performed on the platform of MATLAB R2020a using a 3.4 GHz Intel Core i7. The Hungarian algorithm is implemented by using the '`matchpairs`' function of the MATLAB platform.

The complexity analysis in Section 4.4 has shown that the complexity of Algorithm 3 scales with the number of BSs L and the number of users K as $\mathcal{O}(L \log L)$ and $\mathcal{O}(K \log K)$, respectively, which is verified by the simulation results presented in Fig. 4.6. Fig. 4.6 also shows that compared with the optimal user association scheme, the running time of Algorithm 3 is significantly reduced, and the gain increases as the number of BSs L or the number of users K becomes

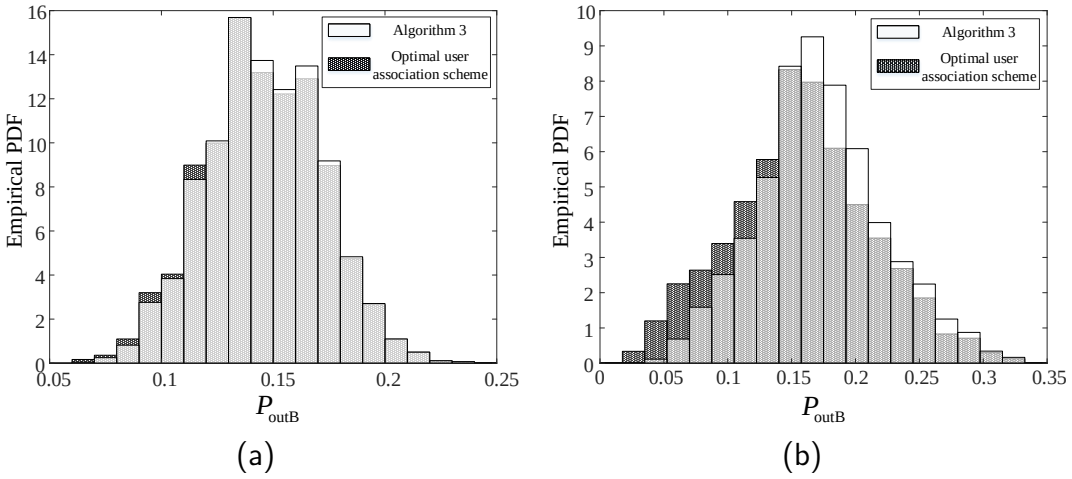

FIGURE 4.7: Empirical PDF of outage probability P_{outB} with Algorithm 3, the proposed low-complexity outage mitigation user association scheme, and the optimal user association scheme. (a) $\mu_\delta = 0$. (b) $\mu_\delta = 0.5$.

larger. For instance, in Fig. 4.6a, when the aggregation factor $\mu_\delta = 0$, the running time of the proposed algorithm is 65.59% of that of the optimal user association scheme with $L = 20$. It is further reduced to 39.10% with $L = 50$.

Fig. 4.7 presents empirical probability density function (PDF) of the outage probability P_{outB} with both the proposed user association scheme and the optimal user association scheme. The empirical PDF is generated with 10000 realizations of users' and BSs' positions. We can see from Fig. 4.7a that when $\mu_\delta = 0$, i.e., users are uniformly distributed, the proposed scheme can achieve almost identical outage performance with the optimal user association scheme. While in Fig. 4.7b where $\mu_\delta = 0.5$, the percentage for small value of P_{outB} with the proposed user association scheme is considerably lower than that with the optimal user association scheme, indicating that the network with the proposed user association scheme is more likely to have a larger outage probability.

For the uniformly-distributed-user case, i.e., $\mu_\delta = 0$, we can see from Figs. 4.5a and 4.5b that for most users, each of them has only one physically accessible BS. This can be also seen from Fig. 4.4, where $\mathbb{E}_{\{\mathbf{r}_k^{\text{user}}\}}[\bar{L}_{\text{ac}}]$ is close to 1 when $\mu_\delta = 0$. In that case, with both the optimal user association scheme and the proposed Algorithm 3, most users are associated with their single physically accessible BS, and thus both algorithms achieve similar outage performance. In contrast, when $\mu_\delta = 0.5$, as users are clustered, Fig. 4.5d shows that many users have more than one physically accessible BSs. For Algorithm 3 which associates one user with one

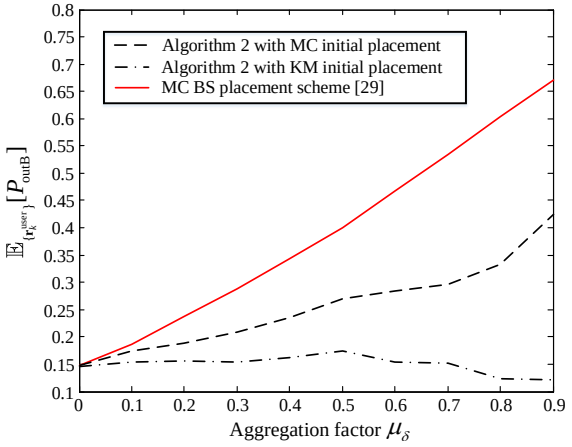

FIGURE 4.8: Average outage probability $\mathbb{E}_{\{\mathbf{r}_k^{\text{user}}\}}[P_{\text{outB}}]$ versus the aggregation factor μ_δ.

BS in each iteration in a greedy manner, it is not guaranteed that each user can always make a globally optimal choice. Therefore, some users may not be able to associate with their physically accessible BSs that have become saturated, leading to a higher outage probability than the optimal user association scheme.

4.5.4 Outage Performance

In this subsection, we evaluate the outage performance of the proposed CSA-based optimal BS placement combined with the proposed low-complexity outage mitigation user association scheme. For the proposed CSA-based optimal BS placement, both MC and KM initial placement schemes are considered. The total number of iterations is set as $T = 700$, and the optimized positions of BSs are further projected to the nearest corners of buildings. For comparison, we use the MC BS placement scheme [29] combined with the proposed low-complexity outage mitigation user association scheme as the performance benchmark.

Fig. 4.8 presents the long-term average outage probability $\mathbb{E}_{\{\mathbf{r}_k^{\text{user}}\}}[P_{\text{outB}}]$ with different BS placement schemes versus the aggregation factor of users' distribution μ_δ. We can see that the proposed CSA-based optimal BS placement scheme significantly outperforms the benchmark MC BS placement scheme when μ_δ>0, and the performance gap is enlarged with the increase of μ_δ. Intuitively, with symmetrically placed BSs, the number of users that fall into the coverage of each BS becomes more unbalanced as μ_δ increases, i.e., users become more clustered. Consequently, there may not be sufficient BSs near the hotspots to support the

excessive number of users, leading to high outage probability in the long term. In contrast, the proposed CSA-based optimal BS placement scheme optimizes the BSs' positions based on users' spatial distribution, and therefore can better balance the workloads of the BSs. It can also be seen from Fig. 4.8 that the KM initial placement scheme leads to a smaller average outage probability than the MC initial placement scheme, especially when μ_δ is large. This is consistent with the convergence results in Fig. 4.4, where Algorithm 2 with the KM initial placement converges to a larger long-term average number of accessible BSs $\mathbb{E}_{\{\mathbf{r}_k^{\text{user}}\}}[\bar{L}_{\text{ac}}]$ than that with the MC initial placement. With more accessible BSs, users have more freedom to select a BS to associate with, thus leading to a lower average outage probability $\mathbb{E}_{\{\mathbf{r}_k^{\text{user}}\}}[P_{\text{outB}}]$.

4.6 Summary

In this chapter, we propose a BS placement scheme and a user association scheme to minimize the long-term average outage probability of an mmWave communication network in a Manhattan-type geometry. For the BS placement problem, as the physical accessibility between users and BSs is closely related to the positions of users, which may vary with time, the problem is formulated as a stochastic optimization problem aiming at maximizing the long-term average number of physically accessible BSs of each user under an inaccessible probability constraint, and a CSA-based algorithm is developed to effectively search the optimal positions of BSs. For the user association problem, with the maximum workload constraint on each BS taken into account, a low-complexity outage mitigation user association scheme was proposed, which achieves similar outage performance to the optimal user association scheme, but with much less running time. Gains over the representative MC BS placement scheme are also demonstrated, which increase as the users' spatial distribution becomes more clustered.

Note that in this chapter, the coverage radius of each BS is fixed. In the next chapter, we will focus on the AP placement in IEEE 802.11 networks, where the positions and coverage radius of APs are jointly optimized. The main concern is to maximize the network throughput under a long-term outage probability constraint. A novel CSA-based algorithm is also developed to solve the problem.

Chapter 5

Optimal Placement and Coverage of APs for IEEE 802.11 Networks

In this chapter, the stochastic optimization method is further applied to jointly optimize the positions and coverage radius of APs in IEEE 802.11 networks (also referred to as Wi-Fi networks). This chapter is organized as follows. Section 5.1 reviews the previous related work. Section 5.2 presents the system model and problem formulation. A CSA-based algorithm is proposed in Section 5.3, and the performance is demonstrated by simulation results provided in Section 5.4. Finally, concluding remarks are summarized in Section 5.5.

5.1 Previous Work

The AP placement problem in Wi-Fi networks shares similarity with the BS placement problem in cellular networks, which has been extensively studied. As mentioned in Section 4.1, the main focus of BS placement is usually on the downlink coverage performance, where the radio coverage region is discretized into a finite number of CSs for BSs and RPs for receivers, and the optimal subset of CSs is selected based on the signal quality at RPs. Due to the NP-hardness of the optimization problem, heuristic algorithms have been proposed based on tabu search [14], randomized greedy search [16] or simulated annealing [25]. Similarly, for Wi-Fi networks, various greedy algorithms have been developed to minimize the AP installation cost [17, 19], improve the network performance [33, 34] or positioning accuracy [31, 32].

Note that in the above studies, a deterministic optimization problem is usually formulated, where performance metrics are calculated based on a given set of fixed users' positions, i.e., RPs [14, 16, 17, 19, 25, 31–34]. As pointed out in Section 1.2.1, the optimal AP placement based on such a deterministic optimization framework would quickly become obsolete when the users' positions change. A stochastic optimization approach could be more appropriate. Moreover, for the next-generation Wi-Fi networks with a massive number of users, the uplink access performance would become crucial as the deluge of access requests may easily paralyze the network. Apparently, the access performance is crucially determined by the positions and coverage radius of APs. Yet it has been little understood how to properly place and configure the APs to optimize the access performance of Wi-Fi networks.

The challenge originates from the lack of characterization of access performance. For a multi-AP Wi-Fi network, most studies focused on estimation of throughput performance [33, 34, 111, 112]. An accurate closed-form throughput expression was only recently derived in [11]. It shows that the throughput performance of a multi-AP Wi-Fi network is closely dependent on the number of users that can be heard by multiple APs, which is determined by the users' positions, APs' positions and their coverage radius. With the time-varying positions of users, the long-term average throughput performance, i.e., averaged over users' positions, would be a more pertinent objective when optimizing the placement and configuration of APs.

In this chapter, we aim at optimizing the positions and coverage radius of APs for maximizing the average network throughput. We also consider a constraint of service outage probability, i.e., the percentage of users falling outside the coverage areas of all APs, which should be kept within a certain level in practice. Specifically, we formulate the problem as a stochastic optimization problem by taking into account the time-varying nature of users' positions. As both the objective and constraint functions are in the expectation form, we propose a new algorithm based on the CSA introduced in Section 2.2.2 to effectively search the optimal placement and coverage radius of APs. In the next section, the system model will be first introduced and the optimization problem will be formulated.

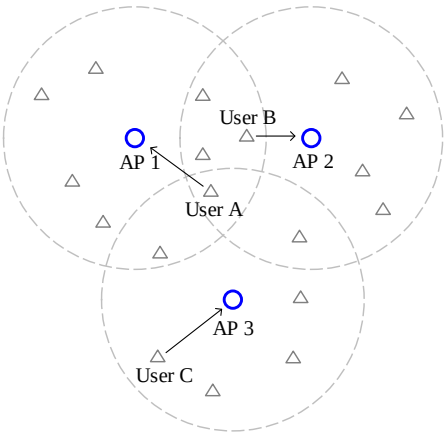

FIGURE 5.1: Graphic illustration of the users in different groups in a three-BSS network. APs and users are represented by circles and triangles, respectively.

5.2 System Model and Problem Formulation

Consider an IEEE 802.11 distributed coordination function (DCF) network with L basic service sets (BSSs), where in each BSS, multiple users transmit to a single AP. Assume that universal frequency reuse is adopted among BSSs, that is, all the BSSs share the same spectrum. Let $\mathcal{L}$ and $\mathcal{K}$ denote the sets of APs and users, respectively, with $|\mathcal{L}|=L$ and $|\mathcal{K}|=K$. Suppose that the coverage area of each AP is a disc with radius R_{A}. For each user, it can be heard[1] by an AP if and only if it is within the coverage area of one AP. With multiple APs, their coverage areas may overlap. Users who fall into the overlapping areas can be heard by multiple APs, among which they choose one to associate with. Denote $\mathcal{G}_{i,\mathcal{S}}$ as the group of users, which associate with AP i and can be heard by the APs in the set $\mathcal{S}$. As Fig. 5.1 illustrates, User A associates with AP 1, but can be heard by APs 1, 2 and 3. So it belongs to Group $\mathcal{G}_{1,\{1,2,3\}}$. Similarly, User B is in Group $\mathcal{G}_{2,\{1,2\}}$, and User C is in Group $\mathcal{G}_{3,\{3\}}$.

Let $\mathbf{r}_l^{\mathrm{AP}}, l = 1, \cdots, L$ and $\mathbf{r}_k^{\mathrm{user}}, k = 1, \cdots, K$ denote the positions of AP l and user k, respectively. Then the number of users in Group $\mathcal{G}_{i,\mathcal{S}}$ can be written as

$$k^{(i,\mathcal{S})} = \sum_{k \in \mathcal{K}_i^*} \prod_{l \in \mathcal{S}} I(R_{\mathrm{A}} - d_{l,k}) \prod_{\tilde{l} \in \mathcal{L} \setminus \mathcal{S}} I(d_{\tilde{l},k} - R_{\mathrm{A}}), \tag{5.1}$$

[1]Specifically, "hear" means that the average received power of the received packets is higher than the threshold for successful decoding. In the IEEE 802.11n standard, for instance, the minimum requirement of received power for an AP to successfully decode a packet is to detect the start of a valid transmission signal at a power level of at least -82/-79 dBm for a 20/40 MHz channel [113].

where

$$d_{l,k} = \|\mathbf{r}_l^{\mathrm{AP}} - \mathbf{r}_k^{\mathrm{user}}\| \tag{5.2}$$

is the distance between AP l and user k. $\mathcal{K}_i^*$ represents the set of users who associate with AP i. $I(x)$ is the indicator function given in (4.3).

In this chapter, we focus on the network throughput performance, which is defined as the average number of successfully transmitted packets of the network in each time slot. Denote $\lambda^{(i,\mathcal{S})}$ as the throughput of a user from Group $\mathcal{G}_{i,\mathcal{S}}$. Then the network throughput can be written as

$$\hat{\lambda} = \sum_{i\in\mathcal{L}} \sum_{\substack{\mathcal{S}=\tilde{\mathcal{S}}\cup\{i\} \\ \tilde{\mathcal{S}}\subseteq\mathcal{L}\setminus\{i\}}} k^{(i,\mathcal{S})}\lambda^{(i,\mathcal{S})}. \tag{5.3}$$

It has been shown in [11] that for a saturated IEEE 802.11 DCF network, with universal frequency reuse, the user throughput $\lambda^{(i,\mathcal{S})}$ can be obtained as

$$\begin{aligned}\lambda^{(i,\mathcal{S})} =& \varepsilon^{(i,\mathcal{S})}\tau_T \Bigg/ \Bigg(\varepsilon^{(i,\mathcal{S})} \Big(\tau_T + \tau_F \cdot \frac{1-p^{(i)}}{p^{(i)}}\Big) + \frac{1}{2p^{(i)}} \Bigg(1+\vartheta \\ & \cdot \Big(\frac{p^{(i)}}{2p^{(i)}-1} + \big(1 - \frac{p^{(i)}}{2p^{(i)}-1}\big)(2-2p^{(i)})^{\varrho} \Big) \Bigg) \Bigg),\end{aligned} \tag{5.4}$$

where ϑ and ϱ are the initial backoff window size and cutoff phase of each user, respectively. Typical values of ϑ and ϱ for IEEE 802.11n are $\vartheta = 16$ and $\varrho = 6$ [113]. τ_T and τ_F are the successful transmission time and collision time, respectively, which can be obtained based on the system parameter setting according to Fig. 5 in [11]. $\varepsilon^{(i,\mathcal{S})}$ is the probability of sensing the channels of BSS set $\mathcal{S}$ idle by users in Group $\mathcal{G}_{i,\mathcal{S}}$, which is given by

$$\varepsilon^{(i,\mathcal{S})} = \prod_{j\in\mathcal{S}} \frac{1}{1+\tau_F - \tau_F p^{(j)} - (\tau_T - \tau_F)p^{(j)} \ln p^{(j)}}, \tag{5.5}$$

where $p^{(j)}$ is the limiting probability of successful transmission of head-of-line packets of BSS j given that the channel of BSS j is idle, which can be obtained

by jointly solving the following L fixed-point equations

$$p^{(j)}=\exp\left\{\sum_{i\in\mathcal{L}}\sum_{\substack{\mathcal{S}=\tilde{\mathcal{S}}\cup\{i,j\}\\ \tilde{\mathcal{S}}\subseteq\mathcal{L}\setminus\{i,j\}}}-\frac{k^{(i,\mathcal{S})}\varepsilon^{(i,\mathcal{S})}}{\varepsilon^{(j,\{j\})}p^{(j)}}\Big/\left(\varepsilon^{(i,\mathcal{S})}\left(\tau_T+\tau_F\cdot\frac{1-p^{(i)}}{p^{(i)}}\right)\right.\right.$$
$$\left.\left.+\frac{1}{2p^{(i)}}\left(1+\vartheta\left(\frac{p^{(i)}}{2p^{(i)}-1}+\left(1-\frac{p^{(i)}}{2p^{(i)}-1}\right)(2-2p^{(i)})^{\varrho}\right)\right)\right)\right\}, \tag{5.6}$$

$j\in\mathcal{L}$.

From (5.1) and (5.3), we can see that the network throughput $\hat{\lambda}$ is closely determined by the group size $k^{(i,\mathcal{S})}$, which is a function of the positions of APs, $\{\mathbf{r}_l^{\mathrm{AP}}\}$, the coverage radius of each AP, R_{A}, and the positions of users, $\{\mathbf{r}_k^{\mathrm{user}}\}$. In practice, due to the mobility of users, $\{\mathbf{r}_k^{\mathrm{user}}\}$ vary from time to time. Therefore, we are interested at the long-term average network throughput $\mathbb{E}_{\{\mathbf{r}_k^{\mathrm{user}}\}}[\hat{\lambda}]$, which depends on the positions and coverage radius of APs, $\{\mathbf{r}_l^{\mathrm{AP}}\}$ and R_{A}.

In this chapter, we aim at maximizing the average network throughput by optimally choosing the coverage area and positions of APs. Note that in addition to boosting the throughput performance, it is also important to ensure that most users are covered by the APs. Define K_{out} as the number of users outside the coverage areas of all the APs, i.e.,

$$K_{\mathrm{out}}=\sum_{k=1}^{K}\prod_{l=1}^{L}I(d_{l,k}-R_{\mathrm{A}}). \tag{5.7}$$

The percentage of users falling outside the coverage areas indicates the service outage probability, which should be kept at a low level. Denote $\bar{P}_{\mathrm{outA}}$ as the maximum allowable outage probability. Then we have $\frac{1}{K}\mathbb{E}_{\{\mathbf{r}_k^{\mathrm{user}}\}}[K_{\mathrm{out}}]\leq\bar{P}_{\mathrm{outA}}$.

Finally, we can formulate the problem of outage probability constrained average network throughput maximization as

$$(\mathrm{P6}):\max_{R_{\mathrm{A}},\{\mathbf{r}_l^{\mathrm{AP}}\}}\ \mathbb{E}_{\{\mathbf{r}_k^{\mathrm{user}}\}}[\hat{\lambda}] \tag{5.8}$$
$$\text{s.t. } R_{\mathrm{A}}\in\mathbb{R}_+, \tag{5.9}$$
$$\mathbf{r}_l^{\mathrm{AP}}\in\mathcal{A}, l=1,\cdots,L, \tag{5.10}$$
$$\frac{1}{K}\mathbb{E}_{\{\mathbf{r}_k^{\mathrm{user}}\}}[K_{\mathrm{out}}]\leq\bar{P}_{\mathrm{outA}}, \tag{5.11}$$

where $\mathcal{A}$ is the area in which the APs are deployed.

Problem P6 is a stochastic optimization problem with both objective function and constraint function in the form of expectation. As mentioned in Section 2.2.2, this kind of problems can be solved by CSA. In the following section, we will propose a CSA-based algorithm to solve Problem P6.

5.3 Joint Coverage and Placement Optimization of APs

Recall the CSA algorithm introduced in Section 2.2.2 for solving Problem P1, we can solve Problem P6 by replacing $\mathbf{r}^{\mathrm{AP}}$ and $\mathcal{A}^L$ in (2.7) by $[\mathbf{r}^{\mathrm{AP}}, R_{\mathrm{A}}]$ and $\mathcal{A}^L \times \mathbb{R}_+$, respectively, where $\mathbf{r}^{\mathrm{AP}} = [\mathbf{r}_1^{\mathrm{AP}}, \cdots, \mathbf{r}_L^{\mathrm{AP}}]$. Then (2.7) becomes

$$[\mathbf{r}^{\mathrm{AP}}(t+1), R_{\mathrm{A}}(t+1)] = \Pi_{\mathcal{A}^L \times \mathbb{R}_+}([\mathbf{r}^{\mathrm{AP}}(t), R_{\mathrm{A}}(t)] + \tilde{\boldsymbol{\omega}}_{\mathrm{A}}(t)\eta(t)), \tag{5.12}$$

where

$$\tilde{\boldsymbol{\omega}}_{\mathrm{A}}(t) = \begin{cases} \tilde{\boldsymbol{\omega}}_{\hat{\lambda}}(t) & \text{if } \frac{\tilde{K}_{\mathrm{out}}(t)}{K} \leq \bar{P}_{\mathrm{outA}} \\ -\tilde{\boldsymbol{\omega}}_{K_{\mathrm{out}}}(t) & \text{otherwise.} \end{cases} \tag{5.13}$$

$\tilde{\boldsymbol{\omega}}_{\hat{\lambda}}(t)$ and $\tilde{\boldsymbol{\omega}}_{K_{\mathrm{out}}}(t)$ denote the gradients of $\hat{\lambda}(t)$ and $K_{\mathrm{out}}(t)$ in terms of $[\mathbf{r}^{\mathrm{AP}}(t), R_{\mathrm{A}}(t)]$, respectively. $\tilde{K}_{\mathrm{out}}(t)$ is an unbiased estimation of $\mathbb{E}_{\{\mathbf{r}_k^{\mathrm{user}}\}}[K_{\mathrm{out}}]$ generated at time t based on 500 samples of users' positions.

From (5.1)–(5.3) and (5.7), we can see that the expressions of both the network throughput, $\hat{\lambda}$, and the number of users outside the coverage areas of all the APs, K_{out}, involve the indicator function $I(x)$ given in (4.3), whose derivative is undefined at $x = 0$. To make $\hat{\lambda}$ and K_{out} differentiable, we approximate the indicator function (4.3) by the sigmoid function given in (4.23), i.e., $I(x) \approx S(x) = \frac{1}{1+\exp(-\xi x)}$, as what we have done for deriving the gradients of $\bar{L}_{\mathrm{ac}}(t)$ and $P_{\mathrm{iac}}(t)$ in Section 4.3. Based on the approximation (4.23), the gradients $\tilde{\boldsymbol{\omega}}_{K_{\mathrm{out}}}(t)$ and $\tilde{\boldsymbol{\omega}}_{\hat{\lambda}}(t)$ can be calculated by combining (I.1)–(I.2), (G.2)–(G.4) and (I.3)–(I.14), respectively, with detailed derivations given in Appendix I.

Algorithm 4 Joint Optimization of AP Placement and Coverage Radius

Require: Initial positions of APs $\mathbf{r}^{\mathrm{AP}}(0)$ and coverage radius $R_{\mathrm{A}}(0)$.

1: **for** $t = 0, \cdots, T-1$ **do**
2: Compute the stochastic gradient $\tilde{\boldsymbol{\omega}}_{\mathrm{A}}(t)$ based on (5.13) and the step size $\eta(t)$ based on (5.14) with a sample of users' positions $\mathbf{r}^{\mathrm{user}}(t)$.
3: Update the positions of APs $\mathbf{r}^{\mathrm{AP}}(t+1)$ and radius of coverage $R_{\mathrm{A}}(t+1)$ based on (5.12).
4: **if** $\frac{\tilde{K}_{\mathrm{out}}(t+1)}{K} \leq \bar{P}_{\mathrm{outA}}$ **then**
5: $[\mathbf{r}^{\mathrm{AP}*}, R_{\mathrm{A}}^*] \leftarrow [\mathbf{r}^{\mathrm{AP}}(t+1), R_{\mathrm{A}}(t+1)]$.
6: **end if**
7: **end for**

Ensure: The optimized positions of APs and coverage radius $[\mathbf{r}^{\mathrm{AP}*}, R_{\mathrm{A}}^*]$.

For the step size $\eta(t)$, we set it following the model adopted in (4.30) and (4.31) with $a = 0.5$ and $b = 0.6$, i.e.,

$$\eta(t) = \frac{0.5}{t^{0.6}\|\tilde{\boldsymbol{\omega}}_{\mathrm{A}}(t)\|}. \tag{5.14}$$

For the output, similar to the setting in Section 4.3, we only output the final solution that satisfies the constraint.

We summarize the detailed steps of our algorithm for solving Problem P6 in Algorithm 4. Similar to Algorithm 2 in Section 4.3, the convergence performance of Algorithm 4 could be crucially determined by the selection of the initial positions of APs, $\mathbf{r}^{\mathrm{AP}}(0)$, as the objective function of Problem P6 is not a convex function of $\mathbf{r}^{\mathrm{AP}}$ and R_{A}. In the next section, we will consider various initial placement schemes of APs and explore their effects on the convergence performance of Algorithm 4.

5.4 Simulation Results and Discussions

In this section, simulations are conducted to illustrate the performance of the proposed Algorithm 4.

5.4.1 Simulation Setting

In particular, we consider a multi-BSS network with $L = 16$ APs and $K = 100$ users inside a square area with unit side length. Each user is associated with

TABLE 5.1: System Parameter Setting [113]

Packet Payload	1024*8 bits
Channel Bit Rate	57.8 Mbps
MAC header	288 bits
Basic Rate	1 Mbps
PHY header	20 μs
ACK	112 bits + PHY header
RTS	160 bits + PHY header
CTS	112 bits + PHY header
DIFS	34 μs
SIFS	16 μs
Slot Time	9 μs
Initial backoff window size ϑ	16
Cutoff phase ϱ	6

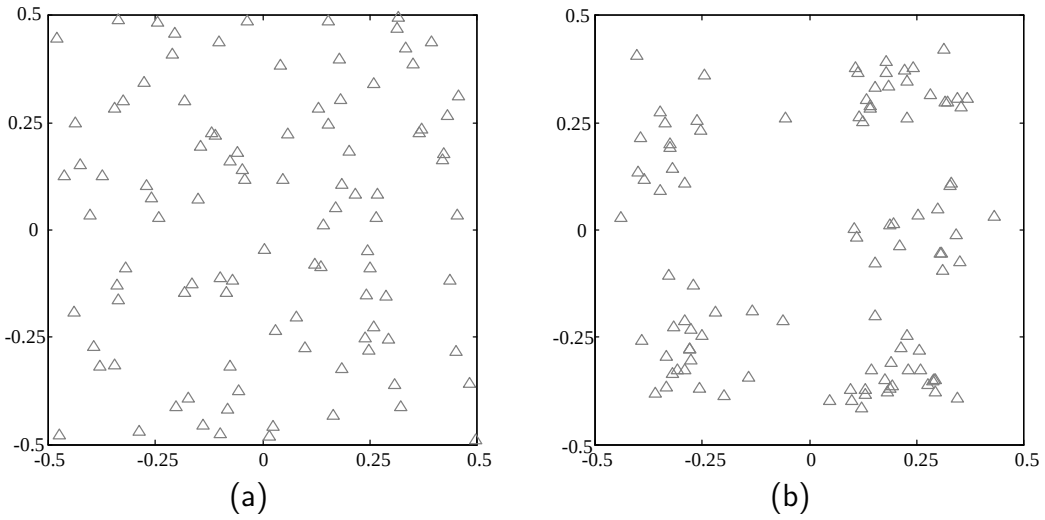

FIGURE 5.2: Snapshots of users' positions with different aggregation factors μ_δ. (a) $\mu_\delta = 0$. (b) $\mu_\delta = 0.5$.

the closest AP, and the Request-to-Send/Clear-to-Send (RTS/CTS) mechanism is adopted. The system parameters are summarized in Table 5.1 according to the IEEE 802.11n standard [113], based on which the successful transmission time τ_T and collision time τ_F can be calculated as 77 time slots and 23.8 time slots, respectively.

Samples of the users' positions are generated in a similar way to the user position generation method in Section 3.5 and Section 4.5.1, i.e., each user is attracted by its closest hotspot. The aggregation degree of users is controlled by an aggregation factor $\mu_\delta \in [0, 1]$. With $\mu_\delta = 0$, all the users are uniformly distributed within the area. The aggregation level increases with μ_δ. The number of hotspots is set as 5. Fig. 5.2 illustrates two snapshots of users' positions with $\mu_\delta = 0$ and $\mu_\delta = 0.5$.

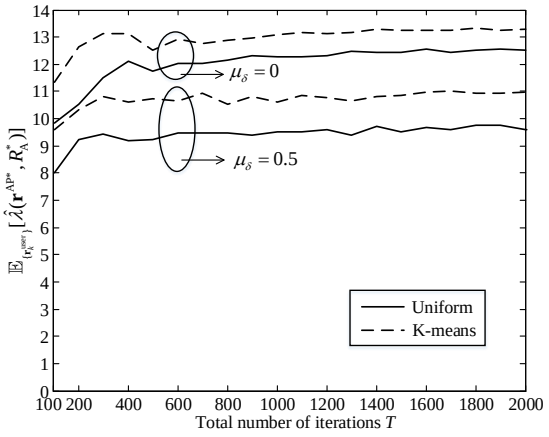

FIGURE 5.3: Average network throughput with the optimized positions of APs and coverage radius $[\mathbf{r}^{\mathrm{AP}*}, R_{\mathrm{A}}^*]$ versus the total number of iterations T under two initial placement schemes of APs, uniform and K-means. $\bar{P}_{\mathrm{outA}} = 10^{-3}$.

5.4.2 Effect of Initial Positions of APs

As we mentioned in Section 5.3, due to the non-convex nature of the objective function, the performance of the proposed algorithm could be crucially dependent on the initial positions of APs. Here we consider two types of initial placement schemes of APs, which are listed as follows:

- **Uniform**: Randomly generate the initial positions of APs following the uniform distribution within the area.
- **K-means**: Randomly generate one sample of users' positions. Apply the K-means method [110] to group users into $L = 16$ clusters. Set the initial positions of APs as the centroids of clusters.

The initial coverage radius $R_{\mathrm{A}}(0)$ is set as 1, and the maximum allowable outage probability $\bar{P}_{\mathrm{outA}}$ is set as 10^{-3}.

Fig. 5.3 illustrates how the average network throughput with the optimized positions of APs and coverage radius $[\mathbf{r}^{\mathrm{AP}*}, R_{\mathrm{A}}^*]$ varies with the total number of iterations T under various values of μ_δ and initial placement schemes. It can be seen from Fig. 5.3 that the proposed algorithm converges after $T \geq 1000$. It converges to a larger value with the K-means initial placement scheme than that with the uniform initial placement scheme, indicating that the algorithm is trapped into a sub-optimal stationary point in the latter case. The reason is that the partial derivative of the network throughput with respect to each AP's position, $\frac{\partial \hat{\lambda}}{\partial \mathbf{r}_l^{\mathrm{AP}}}$, is

crucially determined by the number of users close to the edge of the AP's coverage area. As all the APs' positions are updated with the same step size, those with few edge-users would make overly conservative movements. With the uniform initial placement scheme, the number of APs with no users in the coverage area could be much higher than that with the K-means initial placement scheme. Therefore, it is much more likely to be trapped into sub-optimal stationary positions, especially when users are clustered. It can be clearly seen from Fig. 5.3 that the throughput gap between the two initial placement schemes is enlarged as the aggregation factor μ_δ increases from 0 to 0.5.

5.4.3 Comparison with the Benchmark

In practice, the grid installation AP placement scheme, illustrated in Fig. 5.4a, has been widely adopted in Wi-Fi networks to achieve full coverage of the area [114, 115]. For comparison, Fig. 5.4b presents the optimized positions and coverage of APs by applying the proposed Algorithm 4 with the K-means initial placement scheme and the total number of iterations $T = 1000$. It can be clearly seen from Fig. 5.4 that with the optimized AP placement, the number of users associated with each AP is more balanced, and the overlapping areas of different APs' coverage are significantly reduced. It has been pointed out in [11] that the network throughput would degrade with more users falling into the overlapping areas. It can thus be expected that substantial throughput gains can be achieved by the proposed algorithm.

As Fig. 5.5 shows, the proposed algorithm significantly outperforms the grid installation scheme, especially when the maximum allowable outage probability $\bar{P}_{\text{outA}}$ or the aggregation factor μ_δ is large. Intuitively, by relaxing the outage constraint or with a more clustered users' distribution, the coverage radius of each AP can be greatly reduced with the APs' positions optimally placed, leading to much smaller overlapping coverage areas of different APs. As we can see from Fig. 5.5, the throughput gain over the benchmark grid installation scheme is considerably enlarged as the maximum allowable outage probability $\bar{P}_{\text{outA}}$ or the aggregation factor μ_δ increases. For instance, with $\mu_\delta = 0.5$, the network throughput is improved by 23.2% when $\bar{P}_{\text{outA}} = 0.005$, and the gain is nearly doubled when $\bar{P}_{\text{outA}}$ is further relaxed to 0.05.

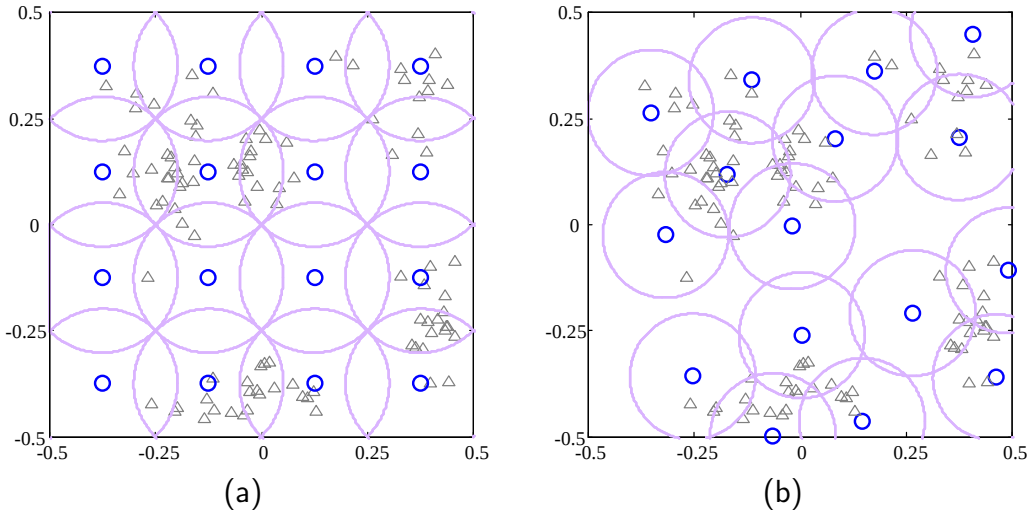

FIGURE 5.4: Positions and coverage of APs with (a) the grid installation scheme and (b) the proposed Algorithm 4 with the K-means initial placement scheme. $\bar{P}_{\text{outA}} = 10^{-2}$. A snapshot of users' positions with $\mu_\delta = 0.5$ and 5 hotspots is also presented. $L = 16$ APs and $K = 100$ users are represented by circles and triangles, respectively.

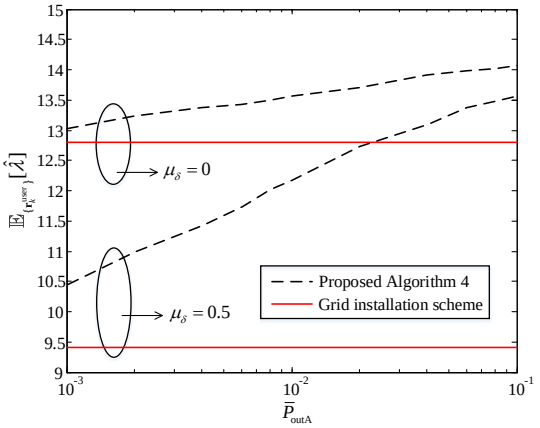

FIGURE 5.5: Average network throughput $\mathbb{E}_{\{\mathbf{r}_k^{\text{user}}\}}[\hat{\lambda}]$ versus the maximum allowable outage probability $\bar{P}_{\text{outA}}$.

5.5 Summary

In this chapter, we study how to properly place and configure the APs to optimize the throughput performance of IEEE 802.11 networks. As the network throughput is crucially dependent on users' positions, which may vary with time, we adopt a stochastic optimization approach to maximize the average network throughput under an outage probability constraint. A CSA-based algorithm is developed to effectively search the optimal positions and coverage radius of APs. Simulation results demonstrate that the proposed algorithm can quickly converge,

and gains over the representative grid installation AP placement scheme increase as the users' spatial distribution becomes more clustered.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

In response to the skyrocketing data traffic demand, network operators are aggressively densifying network infrastructure to extract the spatial reuse gains. However, simply increasing the number of APs could be less beneficial or even detrimental to network performance, if the APs are placed improperly. Although many efforts have been devoted to AP placement optimization, they either represent users by a collection of fixed-location RPs, or require accurate characterization of users' spatial distribution, which is usually unavailable in practice. To optimize the placement of APs with the time-varying nature of users' positions taken into account, a stochastic optimization framework for AP placement has been established in this thesis.

Instead of full knowledge of users' spatial distribution, the proposed stochastic optimization methods only require samples of users' positions that can be easily obtained in practice. The effectiveness and efficiency of the proposed optimal AP placement methods are then illustrated by the successful applications in three representative infrastructure-based wireless networks, which are DASs, mmWave cellular networks and Wi-Fi networks, for improving their capacity, outage, and access performances, respectively.

Specifically, the optimization framework is first applied to optimize the BS antenna placement of a multi-user DAS. With the aim of maximizing the uplink average ergodic sum capacity, an SGD-based algorithm is proposed to effectively search for the optimal solution. To enhance the computational efficiency of the

proposed SGD-based algorithm, two closed-form approximations of the ergodic sum capacity are also derived, which are shown to be much more accurate than the previous results. Simulation results corroborate that the proposed BS antenna placement scheme can achieve a much higher uplink average ergodic sum capacity than several existing representative BS antenna placement schemes. The capacity gains become more prominent when the number of BS antenna clusters or the aggregation degree of users is large.

The proposed optimization framework is further extended to optimize the BS placement of an mmWave communication network for maximizing the long-term average number of physically accessible BSs of each user under an inaccessible probability constraint. As both the objective and constraint functions are in the form of expectation, a novel algorithm is developed based on the recently proposed CSA. With the optimized positions of BSs, the user association problem in a given time slot is also investigated. With the maximum workload constraint on each BS considered, a low-complexity outage mitigation user association scheme is further proposed, which is shown to achieve similar outage performance to that of the optimal user association scheme, but is much more time-efficient. Simulation results demonstrate that combined with the proposed user association scheme, the proposed CSA-based BS placement scheme can greatly improve the long-term outage performance of the network, especially when users' spatial distribution becomes more clustered.

Finally, the optimization framework is applied to a Wi-Fi network to jointly optimize the positions and coverage radius of APs for maximizing the average network throughput under an outage probability constraint. A CSA-based algorithm is designed to solve the AP placement and configuration problem. Simulation results show that the proposed algorithm can quickly converge, and gains over the representative grid installation AP placement scheme increase with the aggregation degree of users and the relaxation level of the outage probability constraint.

In summary, the proposed SA-based stochastic optimization methods can be applied to solve AP placement problems in various wireless networks. Compared to the existing AP placement schemes, significant performance gains have been demonstrated, especially in asymmetric user distribution cases. The proposed methods only require samples of users' positions for AP placement optimization, rather than the theoretical spatial distribution of users. Such a feature makes

the proposed optimal AP placement methods very suitable for practical network topology design.

6.2 Future Work

The work in this thesis only provides a starting point. In the future, more realistic assumptions should be considered. For instance, in Chapter 4, we only consider the blockage effect caused by buildings. In practice, mmWave signals may also be blocked by moving obstacles such as vehicles and pedestrians [116]. Therefore, when formulating the mmWave BS deployment problem, the randomness of the surrounding obstacles should also be taken into account. Furthermore, in some literature [117, 118], it was assumed that even if the LoS path between one user and an mmWave BS is blocked, the user can still access the BS via a first-order reflection path. It would be more practical to consider the reflection paths when optimizing the placement of mmWave BSs.

Note that in Chapter 3, all the distributed BS antennas jointly decode signals from all the users, which could lead to high computational cost when the number of BS antennas or the number of users is large. Recently, a more computationally efficient network architecture for DAS called virtual cell was studied in [119, 120], where each user is only served by several closest distributed BS antennas. With such a network architecture, the proposed BS antenna placement scheme in Chapter 3 may no longer be optimal for average capacity/rate maximization. It would be interesting to explore how the precoder/detector design or virtual cell size, i.e., the number of BS antennas serving one user, affects the optimal placement of BS antennas.

Moreover, in Chapter 5, each AP is assumed to have an identical coverage radius, indicating that each user has identical transmit power and each AP has an identical carrier sense threshold (CST), which are adopted in traditional IEEE 802.11 networks. However, in the latest IEEE 802.11ax protocol, both of these two parameters are allowed to be distinct for different users and APs. In fact, in order to cope with the severe interference problem in overlapping BSSs, IEEE 802.11ax employs three mechanisms, which are 1) dynamic CST, 2) dynamic transmit power control, and 3) BSS coloring [121]. In the BSS coloring scheme, an AP can distinguish inter- and intra-BSS frames via the information contained in the PHY

header. For all inter-BSS frames, the CST will be set higher. By doing so, for one BSS, users belonging to other BSSs are less likely to cause collision in it and more parallel transmissions could hence be created. In the IEEE 802.11ax protocol, only the range of CST and transmit power are discussed. A consensus has not yet been reached on how to optimally tune the parameters of the above three mechanisms based on the contention level or spatial distribution of users. It would be of great practical significance to jointly optimize the placement of APs and the parameter setting of the above three mechanisms for improving the long-term performance of a Wi-Fi network.

Last but not least, in this thesis, we focus on the scenarios where multiple APs are deployed to serve the users in a given area without considering the existing infrastructure. In practice, however, an area may already be covered by a group of APs, which cannot meet the rapidly increasing data traffic demand, and more APs are hence required to be deployed. How to optimize the placement of the newly added APs with the existing ones taken into account is an interesting yet much more complicated problem, as the new APs may also cause interference to the existing ones. Furthermore, radio resources may need to be re-allocated among all the APs, and the optimal resource allocation scheme could be closely determined by the placement of new APs. How to address this new challenge by extending the methodology in this thesis is an interesting topic that deserves much attention in the future study.

Appendix A

Proof of Lemma 3.1

Proof. For (3.15), define a function $\mathbf{f}_U(U_1, \cdots, U_L)$ with $\mathbf{f}_U = (f_U^1, \cdots, f_U^L)$, and

$$f_U^l(U_1, \cdots, U_L) = 1 + \rho_0 \sum_{k=1}^{K} \frac{\gamma_{l,k}^2}{1 + \frac{\rho_0}{\beta} \sum_{i=1}^{L} \gamma_{i,k}^2 U_i^{-1}}, \tag{A.1}$$

$l = 1, \cdots, L$. It can be easily shown that $\mathbf{f}_U(U_1, \cdots, U_L)$ is a standard function according to Definition 6.2 in [69]. Further note that if $U_l > 0$ for all l, then $f_U^l(U_1, \cdots, U_L)$ is upper-bounded by

$$f_U^l(U_1, \cdots, U_L) < 1 + \rho_0 \sum_{k=1}^{K} \gamma_{l,k}^2. \tag{A.2}$$

Therefore, if we set $U_l \geq 1 + \rho_0 \sum_{k=1}^{K} \gamma_{l,k}^2, l = 1, \cdots, L$, then we have $U_l \geq f_U^l(U_1, \cdots, U_L)$ for all l. According to Theorem 6.18 in [69], we can conclude that $\lim_{\tau \to \infty} \{U_1^{(\tau)}, \cdots, U_L^{(\tau)}\} \to \{U_1^*, \cdots, U_L^*\}$ for any positive initial values of $\{U_1, \cdots, U_L\}$.

Similarly, we can prove $\lim_{\tau \to \infty} \{W_1^{(\tau)}, \cdots, W_K^{(\tau)}\} \to \{W_1^*, \cdots, W_K^*\}$ for any positive initial values of $\{W_1, \cdots, W_K\}$ by 1) defining a function $\mathbf{f}_W(W_1, \cdots, W_K)$ with $\mathbf{f}_W = (f_W^1, \cdots, f_W^K)$ and

$$f_W^k(W_1, \cdots, W_K) = 1 + \frac{\rho_0}{\beta} \sum_{l=1}^{L} \frac{\gamma_{l,k}^2}{1 + \rho_0 \sum_{j=1}^{K} \gamma_{l,j}^2 W_j^{-1}}, \tag{A.3}$$

$k = 1, \cdots, K$, according to (3.18), 2) showing that $\mathbf{f}_W(W_1, \cdots, W_K)$ is a standard function, and 3) noting that $W_k \geq f_W^k(W_1, \cdots, W_K)$ for all k if W_k is set as $W_k \geq 1 + \frac{\rho_0}{\beta} \sum_{l=1}^{L} \gamma_{l,k}^2, k = 1, \cdots, K$. □

Appendix B

Derivation of (3.32)

Let us start from the asymptotic result of $\lim_{\rho_0\to\infty} \tilde{\mathcal{C}}^{(0)}$. By combining

$$\lim_{\rho_0\to\infty} \frac{\beta}{\rho_0 L\Phi}\mathcal{F}\left(\frac{\rho_0}{\beta}L\Phi, \frac{K\beta}{L}\right) = \frac{K\beta}{L}, \tag{B.1}$$

when $L > K\beta$, with

$$\begin{aligned}&\lim_{\rho_0\to\infty} \frac{\beta}{L}\sum_{k=1}^{K}\log_2\left(1+\left(1-\frac{\beta}{\rho_0 L\Phi}\mathcal{F}(\frac{\rho_0}{\beta}L\Phi, \frac{K\beta}{L})\right)\frac{\rho_0}{\beta}\sum_{l=1}^{L}\gamma_{l,k}^2\right)\\ =&\frac{K\beta}{L}\log_2\left((1-\frac{K\beta}{L})\frac{1}{\beta}\right)+\frac{\beta}{L}\sum_{k=1}^{K}\log_2\sum_{l=1}^{L}\gamma_{l,k}^2+\lim_{\rho_0\to\infty}\frac{K\beta}{L}\log_2\rho_0\end{aligned} \tag{B.2}$$

and

$$\begin{aligned}&\lim_{\rho_0\to\infty}\frac{\rho_0}{L}\sum_{l=1}^{L}\sum_{k=1}^{K}\gamma_{l,k}^2\left(1+\left(1-\frac{\beta}{\rho_0 L\Phi}\mathcal{F}(\frac{\rho_0}{\beta}L\Phi, \frac{K\beta}{L})\right)\frac{\rho_0}{\beta}\sum_{i=1}^{L}\gamma_{i,k}^2\right)^{-1}\\ =&\lim_{\rho_0\to\infty}\frac{\beta}{L}\left(1-\frac{\beta}{\rho_0 L\Phi}\mathcal{F}(\frac{\rho_0}{\beta}L\Phi, \frac{K\beta}{L})\right)^{-1}\sum_{k=1}^{K}\sum_{l=1}^{L}\frac{\gamma_{l,k}^2}{\sum_{i=1}^{L}\gamma_{i,k}^2} = \frac{K\beta}{L}\left(1-\frac{K\beta}{L}\right)^{-1},\end{aligned} \tag{B.3}$$

we can obtain from (3.25) that

$$\lim_{\rho_0\to\infty} f_{CU}^{(0)} = \left(\frac{K\beta}{L}-1\right)\log_2\left(1-\frac{K\beta}{L}\right)-\frac{K\beta}{L}\log_2(\beta e)+\frac{\beta}{L}\sum_{k=1}^{K}\log_2\sum_{l=1}^{L}\gamma_{l,k}^2+\lim_{\rho_0\to\infty}\frac{K\beta}{L}\log_2\rho_0. \tag{B.4}$$

Similarly, we have

$$\lim_{\rho_0\to\infty} f_{CW}^{(0)} = \left(1-\frac{K\beta}{L}\right)\log_2\left(1-\frac{L}{K\beta}\right) - \log_2 e + \frac{1}{L}\sum_{l=1}^{L}\log_2\sum_{k=1}^{K}\gamma_{l,k}^2 + \lim_{\rho_0\to\infty}\log_2\rho_0, \tag{B.5}$$

by noting that

$$\lim_{\rho_0\to\infty}\frac{1}{\rho_0 K\Phi}\mathcal{F}\left(\frac{\rho_0}{\beta}L\Phi, \frac{K\beta}{L}\right) = \frac{L}{K\beta}, \tag{B.6}$$

when $L \leq K\beta$. For large $\rho_0 \gg 1$, a high-SNR approximation of $\tilde{\mathcal{C}}^{(0)}$ can be obtained from (B.4)–(B.5) as

$$\tilde{\mathcal{C}}^{(0)} \overset{\rho_0\gg 1}{\approx} \begin{cases} \frac{\beta}{L}\sum_{k=1}^{K}\log_2(\sum_{l=1}^{L}\gamma_{l,k}^2) - (1-\frac{K\beta}{L})\log_2(1-\frac{K\beta}{L}) + \frac{K\beta}{L}\log_2\frac{\rho_0}{\beta e} & \text{if } L > K\beta \\ \frac{1}{L}\sum_{l=1}^{L}\log_2(\sum_{k=1}^{K}\gamma_{l,k}^2) + (1-\frac{K\beta}{L})\log_2(1-\frac{L}{K\beta}) + \log_2\frac{\rho_0}{e} & \text{otherwise.} \end{cases} \tag{B.7}$$

To further obtain $\mathbb{E}_{\{\mathbf{r}_l^{\mathrm{AP}}\},\{\mathbf{r}_k^{\mathrm{user}}\}}[\tilde{\mathcal{C}}^{(0)}]$, let us ignore the edge effect. The large-scale fading coefficient $\gamma_{l,k}^2$ between BS antenna cluster l and user k then follows the Pareto distribution with the PDF of [122]

$$f_{\gamma_{l,k}^2}(x) = \frac{1}{2}x^{-\frac{3}{2}}, \tag{B.8}$$

when the path-loss factor $\alpha = 4$. Let $S_L = \frac{1}{L^2}\sum_{l=1}^{L}\gamma_{l,k}^2$ and $S_K = \frac{1}{K^2}\sum_{k=1}^{K}\gamma_{l,k}^2$. Based on the generalized central limit theorem, the limiting distributions of S_L and S_K when the number of users K and the number of BS antenna clusters L go to infinity weakly converge to a stable distribution, which has the PDF of [122]

$$f_{\mathrm{st}}(x) = \frac{1}{2}x^{-3/2}\exp\left(-\frac{\pi}{4x}\right). \tag{B.9}$$

We then have

$$\begin{aligned} &\lim_{L\to\infty}\mathbb{E}_{\{\gamma_{l,k}:l=1,\cdots,L\}}\left[\log_2\left(\frac{1}{L^2}\sum_{l=1}^{L}\gamma_{l,k}^2\right)\right] \\ &= \lim_{K\to\infty}\mathbb{E}_{\{\gamma_{l,k}:k=1,\cdots,K\}}\left[\log_2\left(\frac{1}{K^2}\sum_{k=1}^{K}\gamma_{l,k}^2\right)\right] \\ &= \int_0^\infty \log_2 x f_{\mathrm{st}}(x)dx = \log_2(\pi e^{\epsilon}), \end{aligned} \tag{B.10}$$

where ϵ is the Euler-Mascheroni constant. Therefore, when L and K are large, we have

$$\sum_{k=1}^{K} \mathbb{E}_{\{\gamma_{l,k}:l=1,\cdots,L\}} \left[\log_2 \sum_{l=1}^{L} \gamma_{l,k}^2 \right] \overset{L\gg 1}{\approx} K \log_2(\pi e^{\epsilon} L^2), \tag{B.11}$$

and

$$\sum_{l=1}^{L} \mathbb{E}_{\{\gamma_{l,k}:k=1,\cdots,K\}} \left[\log_2 \sum_{k=1}^{K} \gamma_{l,k}^2 \right] \overset{K\gg 1}{\approx} L \log_2(\pi e^{\epsilon} K^2). \tag{B.12}$$

Finally, by combining (B.7), (B.11) and (B.12), (3.32) can be derived.

Appendix C

Derivation of (3.35)

Let us first denote Ω^* as the root of $\bar{\bar{C}}_u = C^*$. When $\Omega \leq 1$, $\bar{\bar{C}}_u = C^*$ can be written as

$$N_u\big(\Omega \log_2(\pi e^{\epsilon-1}\rho_0 K^2) + (\Omega - 1)\log_2(1-\Omega)\big) = C^*, \tag{C.1}$$

according to (3.34), with the non-zero root of

$$\Omega^* = 1 - \frac{\ln(\pi e^{\epsilon-1}\rho_0 K^2) - \frac{C^* \ln 2}{N_u}}{\mathbb{W}_0\Big(\pi e^{\epsilon-1}\rho_0 K^2\big(\ln(\pi e^{\epsilon-1}\rho_0 K^2) - \frac{C^* \ln 2}{N_u}\big)\Big)}. \tag{C.2}$$

To ensure that $\Omega^* \leq 1$, it can be obtained from (C.2) that $K \geq K^* = \sqrt{\frac{2^{C^*/Nu}}{\pi e^{\epsilon-1}\rho_0}}$. Finally, (3.35) can be obtained by noting that $M^* = \Omega^* \cdot KN_u$.

Appendix D

Derivation of $\tilde{\boldsymbol{\omega}}_{\mathcal{C}}$

Let us denote $\mathcal{C}$ in (3.4) as $\mathcal{C} = f_C(\{U_l^*\}, \{W_k^*\})$. The zero-order approximation of $\mathcal{C}$ is given by

$$\mathcal{C} \approx \begin{cases} f_C(\{U_l^{(0)}\}, \{\tilde{W}_k^{(0)}\}) & \text{if } L > K\beta \\ f_C(\{\tilde{U}_l^{(0)}\}, \{W_k^{(0)}\}) & \text{otherwise,} \end{cases} \tag{D.1}$$

where $U_l^{(0)}$ and $W_k^{(0)}$ are given in (3.21) and (3.22), respectively. $\tilde{U}_l^{(0)}$ and $\tilde{W}_k^{(0)}$ are obtained by substituting $W_k = W_k^{(0)}$ into (3.5) and $U_l = U_l^{(0)}$ into (3.6), respectively.

Based on the approximation in (D.1), the gradient $\tilde{\boldsymbol{\omega}}_{\mathcal{C}}$ can be written as (3.36). $\frac{\partial U_l^{(0)}}{\partial \mathbf{r}^{\mathrm{AP}}}$ and $\frac{\partial W_k^{(0)}}{\partial \mathbf{r}^{\mathrm{AP}}}$ in (3.36) can be obtained from (3.21) and (3.22) as

$$\frac{\partial U_l^{(0)}}{\partial \mathbf{r}^{\mathrm{AP}}} = -\frac{\beta {U_l^{(0)}}^2}{\rho_0 L \Phi^2} \left(\mathcal{F}\left(\frac{\rho_0}{\beta} L\Phi, \frac{K\beta}{L}\right) - \Phi \cdot \frac{\partial \mathcal{F}\left(\frac{\rho_0}{\beta} L\Phi, \frac{K\beta}{L}\right)}{\partial \Phi} \right) \frac{\partial \Phi}{\partial \mathbf{r}^{\mathrm{AP}}}, \tag{D.2}$$

and

$$\frac{\partial W_k^{(0)}}{\partial \mathbf{r}^{\mathrm{AP}}} = -\frac{{W_k^{(0)}}^2}{\rho_0 K \Phi^2} \left(\mathcal{F}\left(\frac{\rho_0}{\beta} L\Phi, \frac{K\beta}{L}\right) - \Phi \cdot \frac{\partial \mathcal{F}\left(\frac{\rho_0}{\beta} L\Phi, \frac{K\beta}{L}\right)}{\partial \Phi} \right) \frac{\partial \Phi}{\partial \mathbf{r}^{\mathrm{AP}}}, \tag{D.3}$$

where

$$\frac{\partial \mathcal{F}(x,y)}{\partial x} = \frac{1+y}{2} - \frac{x(y-1)^2+1+y}{2\sqrt{x(1+\sqrt{y})^2+1}\sqrt{x(1-\sqrt{y})^2+1}}, \tag{D.4}$$

and

$$\frac{\partial \Phi}{\partial \mathbf{r}^{\mathrm{AP}}} = \frac{1}{LK} \sum_{k=1}^{K} \sum_{l=1}^{L} \frac{\partial \gamma_{l,k}^2}{\partial \mathbf{r}^{\mathrm{AP}}}. \tag{D.5}$$

$\frac{\partial \tilde{U}_l^{(0)}}{\partial \mathbf{r}^{\mathrm{AP}}}$ and $\frac{\partial \tilde{W}_k^{(0)}}{\partial \mathbf{r}^{\mathrm{AP}}}$ in (3.36) can be obtained from (3.5)–(3.6) as

$$\frac{\partial \tilde{U}_l^{(0)}}{\partial \mathbf{r}^{\mathrm{AP}}} = \rho_0 \sum_{k=1}^{K} \frac{1}{W_k^{(0)}} \cdot \frac{\partial \gamma_{l,k}^2}{\partial \mathbf{r}^{\mathrm{AP}}} - \frac{\gamma_{l,k}^2}{W_k^{(0)\,2}} \cdot \frac{\partial W_k^{(0)}}{\partial \mathbf{r}^{\mathrm{AP}}}, \tag{D.6}$$

and

$$\frac{\partial \tilde{W}_k^{(0)}}{\partial \mathbf{r}^{\mathrm{AP}}} = \frac{\rho_0}{\beta} \sum_{l=1}^{L} \frac{1}{U_l^{(0)}} \cdot \frac{\partial \gamma_{l,k}^2}{\partial \mathbf{r}^{\mathrm{AP}}} - \frac{\gamma_{l,k}^2}{U_l^{(0)\,2}} \cdot \frac{\partial U_l^{(0)}}{\partial \mathbf{r}^{\mathrm{AP}}}, \tag{D.7}$$

where $\frac{\partial U_l^{(0)}}{\partial \mathbf{r}^{\mathrm{AP}}}$ and $\frac{\partial W_k^{(0)}}{\partial \mathbf{r}^{\mathrm{AP}}}$ are given in (D.2) and (D.3), respectively.

Now let us focus on $\frac{\partial \gamma_{l,k}^2}{\partial \mathbf{r}^{\mathrm{AP}}}$. If we consider the Cartesian coordinate with $\mathbf{r}_l^{\mathrm{AP}} = [x_l^{\mathrm{AP}}, y_l^{\mathrm{AP}}]$ and $\mathbf{r}_k^{\mathrm{user}} = [x_k^{\mathrm{user}}, y_k^{\mathrm{user}}]$, then we have

$$\frac{\partial \gamma_{l,k}^2}{\partial \mathbf{r}^{\mathrm{AP}}} = \Big[\frac{\partial \gamma_{l,k}^2}{\partial \mathbf{r}_1^{\mathrm{AP}}}, \cdots, \frac{\partial \gamma_{l,k}^2}{\partial \mathbf{r}_L^{\mathrm{AP}}}\Big], \tag{D.8}$$

where

$$\frac{\partial \gamma_{l,k}^2}{\partial \mathbf{r}_j^{\mathrm{AP}}}\Big|_{j \neq l} = [0, 0], \text{ and } \frac{\partial \gamma_{l,k}^2}{\partial \mathbf{r}_l^{\mathrm{AP}}} = \Big[\frac{\partial \gamma_{l,k}^2}{\partial x_l^{\mathrm{AP}}}, \frac{\partial \gamma_{l,k}^2}{\partial y_l^{\mathrm{AP}}}\Big], \tag{D.9}$$

with

$$\frac{\partial \gamma_{l,k}^2}{\partial x_l^{\mathrm{AP}}} = \alpha(x_k^{\mathrm{user}} - x_l^{\mathrm{AP}})\Big((x_l^{\mathrm{AP}} - x_k^{\mathrm{user}})^2 + (y_l^{\mathrm{AP}} - y_k^{\mathrm{user}})^2\Big)^{-\alpha/2-1}, \tag{D.10}$$

and

$$\frac{\partial \gamma_{l,k}^2}{\partial y_l^{\mathrm{AP}}} = \alpha(y_k^{\mathrm{user}} - y_l^{\mathrm{AP}})\Big((x_l^{\mathrm{AP}} - x_k^{\mathrm{user}})^2 + (y_l^{\mathrm{AP}} - y_k^{\mathrm{user}})^2\Big)^{-\alpha/2-1}. \tag{D.11}$$

The gradient $\tilde{\boldsymbol{\omega}}_{\mathcal{C}}$ can be then obtained by combining (3.36)–(3.37) and (D.2)–(D.11).

Appendix E

Derivation of (4.5)

Let us first rewrite $\mathcal{I}_{l,k}^{\text{block}}$ in (4.2) as

$$\mathcal{I}_{l,k}^{\text{block}} = I\left(\sum_{b=1}^{B} \mathcal{I}_{l,k,b}^{\text{block}}\right), \tag{E.1}$$

where

$$\mathcal{I}_{l,k,b}^{\text{block}} = \begin{cases} 1 & \text{if the LoS path between BS } m \text{ and user } n \text{ is blocked by block } b \\ 0 & \text{otherwise.} \end{cases} \tag{E.2}$$

In the following, we will derive the expression of $\mathcal{I}_{l,k,b}^{\text{block}}$.

Let us map the positions of BS l and user k from the coordinate system in Fig. 4.1 to the one in Fig. E.1, where the origin is the center of the line segment connecting BS l and user k. The coordinates of BS l and user k in

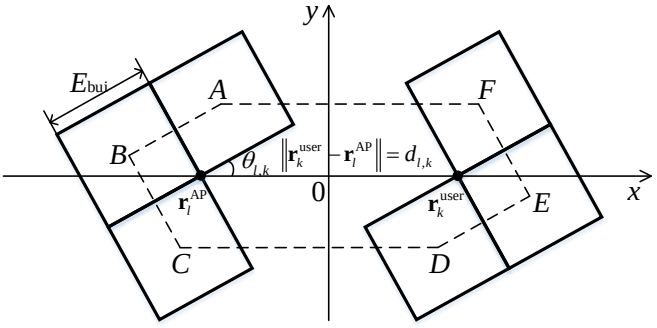

FIGURE E.1: Region $ABCDEF$ that the center of block b falling in will block the LoS path between BS l and user k.

the new coordinate system are $(-\frac{d_{l,k}}{2}, 0)$ and $(\frac{d_{l,k}}{2}, 0)$, respectively. Then the new coordinate of the center of block b are $(x_{l,k,b}, y_{l,k,b})$ with $x_{l,k,b}$ and $y_{l,k,b}$ given in (4.6) and (4.7), respectively. If block b blocks the LoS path between BS l and user k, $(x_{l,k,b}, y_{l,k,b})$ should fall in the region $ABCDEF$ [123]. The coordinates of the points A, B, C, D, E, F in the new coordinate system can be obtained as

- A: $\left(\frac{\sqrt{2}}{2}E_{\text{bui}}\cos(\theta_{l,k}+\frac{\pi}{4})-\frac{d_{l,k}}{2}, \frac{\sqrt{2}}{2}E_{\text{bui}}\sin(\theta_{l,k}+\frac{\pi}{4})\right)$,
- B: $\left(\frac{\sqrt{2}}{2}E_{\text{bui}}\cos(\theta_{l,k}+\frac{3\pi}{4})-\frac{d_{l,k}}{2}, \frac{\sqrt{2}}{2}E_{\text{bui}}\sin(\theta_{l,k}+\frac{3\pi}{4})\right)$,
- C: $\left(\frac{\sqrt{2}}{2}E_{\text{bui}}\cos(\theta_{l,k}+\frac{5\pi}{4})-\frac{d_{l,k}}{2}, \frac{\sqrt{2}}{2}E_{\text{bui}}\sin(\theta_{l,k}+\frac{5\pi}{4})\right)$,
- D: $\left(\frac{\sqrt{2}}{2}E_{\text{bui}}\cos(\theta_{l,k}+\frac{5\pi}{4})+\frac{d_{l,k}}{2}, \frac{\sqrt{2}}{2}E_{\text{bui}}\sin(\theta_{l,k}+\frac{5\pi}{4})\right)$,
- E: $\left(\frac{\sqrt{2}}{2}E_{\text{bui}}\cos(\theta_{l,k}+\frac{7\pi}{4})+\frac{d_{l,k}}{2}, \frac{\sqrt{2}}{2}E_{\text{bui}}\sin(\theta_{l,k}+\frac{7\pi}{4})\right)$,
- F: $\left(\frac{\sqrt{2}}{2}E_{\text{bui}}\cos(\theta_{l,k}+\frac{\pi}{4})+\frac{d_{l,k}}{2}, \frac{\sqrt{2}}{2}E_{\text{bui}}\sin(\theta_{l,k}+\frac{\pi}{4})\right)$,

where $\theta_{l,k}$ is given in (4.8). Let us define the functions of the six edges of the polygon $ABCDEF$ as $f_{\overline{AB}}(x,y)=0$, $f_{\overline{BC}}(x,y)=0$, $f_{\overline{CD}}(x,y)=0$, $f_{\overline{DE}}(x,y)=0$, $f_{\overline{EF}}(x,y)=0$ and $f_{\overline{FA}}(x,y)=0$. Then based on the coordinates of A, B, C, D, E, F, we can obtain

$$\begin{cases} f_{\overline{AB}}(x,y) = x\sin\theta_{l,k} - y\cos\theta_{l,k} + \frac{1}{2}d_{l,k}\sin\theta_{l,k} + \frac{1}{2}E_{\text{bui}} \\ f_{\overline{BC}}(x,y) = x\cos\theta_{l,k} + y\sin\theta_{l,k} + \frac{1}{2}d_{l,k}\cos\theta_{l,k} + \frac{1}{2}E_{\text{bui}} \\ f_{\overline{CD}}(x,y) = y + \frac{1}{2}E_{\text{bui}}\sin\theta_{l,k} + \frac{1}{2}E_{\text{bui}}\cos\theta_{l,k} \\ f_{\overline{DE}}(x,y) = -x\sin\theta_{l,k} + y\cos\theta_{l,k} + \frac{1}{2}d_{l,k}\sin\theta_{l,k} + \frac{1}{2}E_{\text{bui}} \\ f_{\overline{EF}}(x,y) = -x\cos\theta_{l,k} - y\sin\theta_{l,k} + \frac{1}{2}d_{l,k}\cos\theta_{l,k} + \frac{1}{2}E_{\text{bui}} \\ f_{\overline{FA}}(x,y) = -y + \frac{1}{2}E_{\text{bui}}\sin\theta_{l,k} + \frac{1}{2}E_{\text{bui}}\cos\theta_{l,k}. \end{cases} \tag{E.3}$$

If the point $(x_{l,k,b}, y_{l,k,b})$ is within the area $ABCDEF$, it and the origin must be in the same half planes divided by the six edge functions given above. We then have

$$\begin{aligned} \mathcal{I}_{l,k,b}^{\text{block}} =& I\left(f_{\overline{AB}}(0,0)\cdot f_{\overline{AB}}(x_{l,k,b},y_{l,k,b})\right)\cdot I\left(f_{\overline{BC}}(0,0)\cdot f_{\overline{BC}}(x_{l,k,b},y_{l,k,b})\right) \\ &\cdot I\left(f_{\overline{CD}}(0,0)\cdot f_{\overline{CD}}(x_{l,k,b},y_{l,k,b})\right)\cdot I\left(f_{\overline{DE}}(0,0)\cdot f_{\overline{DE}}(x_{l,k,b},y_{l,k,b})\right) \\ &\cdot I\left(f_{\overline{EF}}(0,0)\cdot f_{\overline{EF}}(x_{l,k,b},y_{l,k,b})\right)\cdot I\left(f_{\overline{FA}}(0,0)\cdot f_{\overline{FA}}(x_{l,k,b},y_{l,k,b})\right). \end{aligned} \tag{E.4}$$

(4.5) can then be obtained by combining (E.1), (E.3) and (E.4).

Appendix F

Derivations of (4.24) and (4.25)

Let us rewrite the term $\sum_{l=1}^{L} I(\mathcal{I}_{l,k}^{\text{cover}} + \mathcal{I}_{l,k}^{\text{block}})$ in (4.9) as

$$\sum_{l=1}^{L} I(\mathcal{I}_{l,k}^{\text{cover}} + \mathcal{I}_{l,k}^{\text{block}}) = \sum_{l \in \mathcal{L}_k} I(\mathcal{I}_{l,k}^{\text{cover}} + \mathcal{I}_{l,k}^{\text{block}}) + \sum_{l \in \mathcal{L} \backslash \mathcal{L}_k} I(\mathcal{I}_{l,k}^{\text{cover}} + \mathcal{I}_{l,k}^{\text{block}}). \tag{F.1}$$

Based on (4.1) and the definition of $\mathcal{L}_k$, i.e., the set of BSs which have user k in each of their coverage areas, we have

$$\mathcal{I}_{l,k}^{\text{cover}} = \begin{cases} 1 & \text{if } l \in \mathcal{L} \backslash \mathcal{L}_k \\ 0 & \text{otherwise.} \end{cases} \tag{F.2}$$

Therefore, we can obtain

$$\sum_{l \in \mathcal{L}_k} I(\mathcal{I}_{l,k}^{\text{cover}} + \mathcal{I}_{l,k}^{\text{block}}) = \sum_{l \in \mathcal{L}_k} I(\mathcal{I}_{l,k}^{\text{block}}) = \sum_{l \in \mathcal{L}_k} \mathcal{I}_{l,k}^{\text{block}} \tag{F.3}$$

and

$$\sum_{l \in \mathcal{L} \backslash \mathcal{L}_k} I(\mathcal{I}_{l,k}^{\text{cover}} + \mathcal{I}_{l,k}^{\text{block}}) = \sum_{l \in \mathcal{L} \backslash \mathcal{L}_k} I(\mathcal{I}_{l,k}^{\text{cover}}) = \sum_{l \in \mathcal{L} \backslash \mathcal{L}_k} \mathcal{I}_{l,k}^{\text{cover}} = \sum_{l=1}^{L} \mathcal{I}_{l,k}^{\text{cover}}. \tag{F.4}$$

By substituting (F.3) and (F.4) into (F.1), we have

$$\sum_{l=1}^{L} I(\mathcal{I}_{l,k}^{\text{cover}} + \mathcal{I}_{l,k}^{\text{block}}) = \sum_{l=1}^{L} \mathcal{I}_{l,k}^{\text{cover}} + \sum_{l \in \mathcal{L}_k} \mathcal{I}_{l,k}^{\text{block}}. \tag{F.5}$$

Similarly, we can express the term $\prod_{l=1}^{L} I\left(\mathcal{I}_{l,k}^{\text{cover}} + \mathcal{I}_{l,k}^{\text{block}}\right)$ in (4.10) as

$$\prod_{l=1}^{L} I\left(\mathcal{I}_{l,k}^{\text{cover}} + \mathcal{I}_{l,k}^{\text{block}}\right) = \prod_{l\in\mathcal{L}_k} I\left(\mathcal{I}_{l,k}^{\text{cover}} + \mathcal{I}_{l,k}^{\text{block}}\right) \cdot \prod_{l\in\mathcal{L}\setminus\mathcal{L}_k} I\left(\mathcal{I}_{l,k}^{\text{cover}} + \mathcal{I}_{l,k}^{\text{block}}\right). \quad \text{(F.6)}$$

By combining (F.2) and (F.6), we have

$$\prod_{l=1}^{L} I\left(\mathcal{I}_{l,k}^{\text{cover}} + \mathcal{I}_{l,k}^{\text{block}}\right) = \prod_{l\in\mathcal{L}_k} \mathcal{I}_{l,k}^{\text{block}}, \text{ if } \mathcal{L}_k \neq \emptyset. \quad \text{(F.7)}$$

If $\mathcal{L}_k$ is empty, which means user k is not in the coverage area of any BSs, i.e., $\mathcal{I}_{l,k}^{\text{cover}} = 1, \forall l$ then we have

$$\prod_{l=1}^{L} I\left(\mathcal{I}_{l,k}^{\text{cover}} + \mathcal{I}_{l,k}^{\text{block}}\right) = \prod_{l=1}^{L} \mathcal{I}_{l,k}^{\text{cover}}. \quad \text{(F.8)}$$

Based on (F.7), (F.8) and $\prod_{l=1}^{L} \mathcal{I}_{l,k}^{\text{cover}} = 0$ if $\mathcal{L}_k \neq \emptyset$, we have

$$\prod_{l=1}^{L} I\left(\mathcal{I}_{l,k}^{\text{cover}} + \mathcal{I}_{l,k}^{\text{block}}\right) = I\left(\prod_{l\in\mathcal{L}_k} \mathcal{I}_{l,k}^{\text{block}} + \prod_{l=1}^{L} \mathcal{I}_{l,k}^{\text{cover}}\right) \overset{(a)}{=} \prod_{l\in\mathcal{L}_k} \mathcal{I}_{l,k}^{\text{block}} + \prod_{l=1}^{L} \mathcal{I}_{l,k}^{\text{cover}}, \quad \text{(F.9)}$$

where (a) follows from the fact that $\left(\prod_{l\in\mathcal{L}_k} \mathcal{I}_{l,k}^{\text{block}} + \prod_{l=1}^{L} \mathcal{I}_{l,k}^{\text{cover}}\right) \in \{0, 1\}$. (4.24) and (4.25) can be then obtained by inserting (F.5) and (F.9) into (4.9) and (4.10), respectively.

Appendix G

Derivations of $\frac{\partial \mathcal{I}_{l,k}^{\text{cover}}(t)}{\partial \mathbf{r}^{\text{AP}}(t)}$ and $\frac{\partial \mathcal{I}_{l,k}^{\text{block}}(t)}{\partial \mathbf{r}^{\text{AP}}(t)}$

In the following, we drop the iteration index t for brevity.

G.1 Derivation of $\frac{\partial \mathcal{I}_{l,k}^{\text{cover}}}{\partial \mathbf{r}^{\text{AP}}}$

By replacing the indicator function in (4.4) with (4.23), $\frac{\partial \mathcal{I}_{l,k}^{\text{cover}}}{\partial \mathbf{r}^{\text{AP}}}$ can be derived as

$$\frac{\partial \mathcal{I}_{l,k}^{\text{cover}}}{\partial \mathbf{r}^{\text{AP}}} \approx S'(d_{l,k} - R_{\text{B}}) \cdot \frac{\partial d_{l,k}}{\partial \mathbf{r}^{\text{AP}}}. \tag{G.1}$$

where $S'(x) = \xi S(x)(1 - S(x))$ is the derivative of the sigmoid function $S(x)$. As $d_{l,k} = \|\mathbf{r}_l^{\text{AP}} - \mathbf{r}_k^{\text{user}}\|$, the partial derivative $\frac{\partial d_{l,k}}{\partial \mathbf{r}_l^{\text{AP}}}, l = 1, \cdots, L$ is given by

$$\frac{\partial d_{l,k}}{\partial \mathbf{r}_j^{\text{AP}}}\Big|_{j \neq l} = [0, 0], \text{ and } \frac{\partial d_{l,k}}{\partial \mathbf{r}_l^{\text{AP}}} = \Big[\frac{\partial d_{l,k}}{\partial x_l^{\text{AP}}}, \frac{\partial d_{l,k}}{\partial y_l^{\text{AP}}}\Big], \tag{G.2}$$

where

$$\frac{\partial d_{l,k}}{\partial x_l^{\text{AP}}} = (x_l^{\text{AP}} - x_k^{\text{user}}) \Big((x_l^{\text{AP}} - x_k^{\text{user}})^2 + (y_l^{\text{AP}} - y_k^{\text{user}})^2 \Big)^{-1/2}, \tag{G.3}$$

and

$$\frac{\partial d_{l,k}}{\partial y_l^{\text{AP}}} = (y_l^{\text{AP}} - y_k^{\text{user}}) \Big((x_l^{\text{AP}} - x_k^{\text{user}})^2 + (y_l^{\text{AP}} - y_k^{\text{user}})^2 \Big)^{-1/2}. \tag{G.4}$$

G.2 Derivation of $\frac{\partial \mathcal{I}_{l,k}^{\text{block}}}{\partial \mathbf{r}^{\text{AP}}}$

By replacing the indicator function in (E.1) with (4.23), $\frac{\partial \mathcal{I}_{l,k}^{\text{block}}}{\partial \mathbf{r}^{\text{AP}}}$ can be approximated by

$$\frac{\partial \mathcal{I}_{l,k}^{\text{block}}}{\partial \mathbf{r}^{\text{AP}}} \approx S'\left(\sum_{b=1}^{B} \mathcal{I}_{l,k,b}^{\text{block}}\right) \cdot \sum_{b=1}^{B} \frac{\partial \mathcal{I}_{l,k,b}^{\text{block}}}{\partial \mathbf{r}^{\text{AP}}}, \tag{G.5}$$

where

$$\frac{\partial \mathcal{I}_{l,k,b}^{\text{block}}}{\partial \mathbf{r}^{\text{AP}}} = \Big[\frac{\partial \mathcal{I}_{l,k,b}^{\text{block}}}{\partial \mathbf{r}_1^{\text{AP}}}, \cdots, \frac{\partial \mathcal{I}_{l,k,b}^{\text{block}}}{\partial \mathbf{r}_L^{\text{AP}}}\Big]. \tag{G.6}$$

$\frac{\partial \mathcal{I}_{l,k,b}^{\text{block}}}{\partial \mathbf{r}_j^{\text{AP}}}$ is given by

$$\frac{\partial \mathcal{I}_{l,k,b}^{\text{block}}}{\partial \mathbf{r}_j^{\text{AP}}}\Big|_{j\neq l} = [0,0], \text{ and } \frac{\partial \mathcal{I}_{l,k,b}^{\text{block}}}{\partial \mathbf{r}_l^{\text{AP}}} = \sum_{i=1}^{6} \prod_{j=1, j\neq i}^{6} \mathcal{I}_{l,k,b}^{(j)} \frac{\partial \mathcal{I}_{l,k,b}^{(i)}}{\partial \mathbf{r}_l^{\text{AP}}}, \tag{G.7}$$

where

$$\begin{cases} \mathcal{I}_{l,k,b}^{(1)} = I\big(f_{\overline{AB}}(0,0) \cdot f_{\overline{AB}}(x_{l,k,b}, y_{l,k,b})\big) \\ \mathcal{I}_{l,k,b}^{(2)} = I\big(f_{\overline{BC}}(0,0) \cdot f_{\overline{BC}}(x_{l,k,b}, y_{l,k,b})\big) \\ \mathcal{I}_{l,k,b}^{(3)} = I\big(f_{\overline{CD}}(0,0) \cdot f_{\overline{CD}}(x_{l,k,b}, y_{l,k,b})\big) \\ \mathcal{I}_{l,k,b}^{(4)} = I\big(f_{\overline{DE}}(0,0) \cdot f_{\overline{DE}}(x_{l,k,b}, y_{l,k,b})\big) \\ \mathcal{I}_{l,k,b}^{(5)} = I\big(f_{\overline{EF}}(0,0) \cdot f_{\overline{EF}}(x_{l,k,b}, y_{l,k,b})\big) \\ \mathcal{I}_{l,k,b}^{(6)} = I\big(f_{\overline{FA}}(0,0) \cdot f_{\overline{FA}}(x_{l,k,b}, y_{l,k,b})\big) \end{cases} \tag{G.8}$$

according to (E.4). By replacing the indicator function in (G.8) with (4.23), $\frac{\partial \mathcal{I}_{l,k,b}^{(i)}}{\partial \mathbf{r}_l^{\mathrm{AP}}}, i = 1, \cdots, 6$ can be approximately obtained as

$$\begin{cases} \frac{\partial \mathcal{I}_{l,k,b}^{(1)}}{\partial \mathbf{r}_l^{\mathrm{AP}}} \approx S'\big(f_{\overline{AB}}(0,0) f_{\overline{AB}}(x_{l,k,b}, y_{l,k,b})\big)\big(\frac{\partial f_{\overline{AB}}(0,0)}{\partial \mathbf{r}_l^{\mathrm{AP}}} f_{\overline{AB}}(x_{l,k,b}, y_{l,k,b}) + \frac{\partial f_{\overline{AB}}(x_{l,k,b}, y_{l,k,b})}{\partial \mathbf{r}_l^{\mathrm{AP}}} f_{\overline{AB}}(0,0)\big) \\ \frac{\partial \mathcal{I}_{l,k,b}^{(2)}}{\partial \mathbf{r}_l^{\mathrm{AP}}} \approx S'\big(f_{\overline{BC}}(0,0) f_{\overline{BC}}(x_{l,k,b}, y_{l,k,b})\big)\big(\frac{\partial f_{\overline{BC}}(0,0)}{\partial \mathbf{r}_l^{\mathrm{AP}}} f_{\overline{BC}}(x_{l,k,b}, y_{l,k,b}) + \frac{\partial f_{\overline{BC}}(x_{l,k,b}, y_{l,k,b})}{\partial \mathbf{r}_l^{\mathrm{AP}}} f_{\overline{BC}}(0,0)\big) \\ \frac{\partial \mathcal{I}_{l,k,b}^{(3)}}{\partial \mathbf{r}_l^{\mathrm{AP}}} \approx S'\big(f_{\overline{CD}}(0,0) f_{\overline{CD}}(x_{l,k,b}, y_{l,k,b})\big)\big(\frac{\partial f_{\overline{CD}}(0,0)}{\partial \mathbf{r}_l^{\mathrm{AP}}} f_{\overline{CD}}(x_{l,k,b}, y_{l,k,b}) + \frac{\partial f_{\overline{CD}}(x_{l,k,b}, y_{l,k,b})}{\partial \mathbf{r}_l^{\mathrm{AP}}} f_{\overline{CD}}(0,0)\big) \\ \frac{\partial \mathcal{I}_{l,k,b}^{(4)}}{\partial \mathbf{r}_l^{\mathrm{AP}}} \approx S'\big(f_{\overline{DE}}(0,0) f_{\overline{DE}}(x_{l,k,b}, y_{l,k,b})\big)\big(\frac{\partial f_{\overline{DE}}(0,0)}{\partial \mathbf{r}_l^{\mathrm{AP}}} f_{\overline{DE}}(x_{l,k,b}, y_{l,k,b}) + \frac{\partial f_{\overline{DE}}(x_{l,k,b}, y_{l,k,b})}{\partial \mathbf{r}_l^{\mathrm{AP}}} f_{\overline{DE}}(0,0)\big) \\ \frac{\partial \mathcal{I}_{l,k,b}^{(5)}}{\partial \mathbf{r}_l^{\mathrm{AP}}} \approx S'\big(f_{\overline{EF}}(0,0) f_{\overline{EF}}(x_{l,k,b}, y_{l,k,b})\big)\big(\frac{\partial f_{\overline{EF}}(0,0)}{\partial \mathbf{r}_l^{\mathrm{AP}}} f_{\overline{EF}}(x_{l,k,b}, y_{l,k,b}) + \frac{\partial f_{\overline{EF}}(x_{l,k,b}, y_{l,k,b})}{\partial \mathbf{r}_l^{\mathrm{AP}}} f_{\overline{EF}}(0,0)\big) \\ \frac{\partial \mathcal{I}_{l,k,b}^{(6)}}{\partial \mathbf{r}_l^{\mathrm{AP}}} \approx S'\big(f_{\overline{FA}}(0,0) f_{\overline{FA}}(x_{l,k,b}, y_{l,k,b})\big)\big(\frac{\partial f_{\overline{FA}}(0,0)}{\partial \mathbf{r}_l^{\mathrm{AP}}} f_{\overline{FA}}(x_{l,k,b}, y_{l,k,b}) + \frac{\partial f_{\overline{FA}}(x_{l,k,b}, y_{l,k,b})}{\partial \mathbf{r}_l^{\mathrm{AP}}} f_{\overline{FA}}(0,0)\big), \end{cases} \tag{G.9}$$

where

$$\begin{cases} \frac{\partial f_{\overline{AB}}(0,0)}{\partial \mathbf{r}_l^{\mathrm{AP}}} = \frac{\partial f_{\overline{DE}}(0,0)}{\partial \mathbf{r}_l^{\mathrm{AP}}} = \frac{1}{2}(\sin\theta_{l,k} \frac{\partial d_{l,k}}{\partial \mathbf{r}_l^{\mathrm{AP}}} + d_{l,k} \cos\theta_{l,k} \frac{\partial \theta_{l,k}}{\partial \mathbf{r}_l^{\mathrm{AP}}}) \\ \frac{\partial f_{\overline{BC}}(0,0)}{\partial \mathbf{r}_l^{\mathrm{AP}}} = \frac{\partial f_{\overline{EF}}(0,0)}{\partial \mathbf{r}_l^{\mathrm{AP}}} = \frac{1}{2}(\cos\theta_{l,k} \frac{\partial d_{l,k}}{\partial \mathbf{r}_l^{\mathrm{AP}}} - d_{l,k} \sin\theta_{l,k} \frac{\partial \theta_{l,k}}{\partial \mathbf{r}_l^{\mathrm{AP}}}) \\ \frac{\partial f_{\overline{CD}}(0,0)}{\partial \mathbf{r}_l^{\mathrm{AP}}} = \frac{\partial f_{\overline{FA}}(0,0)}{\partial \mathbf{r}_l^{\mathrm{AP}}} = \frac{E_{\mathrm{bui}}}{2}(\cos\theta_{l,k} - \sin\theta_{l,k}) \frac{\partial \theta_{l,k}}{\partial \mathbf{r}_l^{\mathrm{AP}}} \\ \frac{\partial f_{\overline{AB}}(x_{l,k,b}, y_{l,k,b})}{\partial \mathbf{r}_l^{\mathrm{AP}}} = (x_{l,k,b} \cos\theta_{l,k} + y_{l,k,b} \sin\theta_{l,k} + \frac{d_{l,k}\cos\theta_{l,k}}{2}) \frac{\partial \theta_{l,k}}{\partial \mathbf{r}_l^{\mathrm{AP}}} \\ \qquad + \sin\theta_{l,k} \frac{\partial x_{l,k,b}}{\partial \mathbf{r}_l^{\mathrm{AP}}} - \cos\theta_{l,k} \frac{\partial y_{l,k,b}}{\partial \mathbf{r}_l^{\mathrm{AP}}} + \frac{\sin\theta_{l,k}}{2} \frac{\partial d_{l,k}}{\partial \mathbf{r}_l^{\mathrm{AP}}} \\ \frac{\partial f_{\overline{BC}}(x_{l,k,b}, y_{l,k,b})}{\partial \mathbf{r}_l^{\mathrm{AP}}} = (y_{l,k,b} \cos\theta_{l,k} - x_{l,k,b} \sin\theta_{l,k} - \frac{d_{l,k}\sin\theta_{l,k}}{2}) \frac{\partial \theta_{l,k}}{\partial \mathbf{r}_l^{\mathrm{AP}}} \\ \qquad + \cos\theta_{l,k} \frac{\partial x_{l,k,b}}{\partial \mathbf{r}_l^{\mathrm{AP}}} + \sin\theta_{l,k} \frac{\partial y_{l,k,b}}{\partial \mathbf{r}_l^{\mathrm{AP}}} + \frac{\cos\theta_{l,k}}{2} \frac{\partial d_{l,k}}{\partial \mathbf{r}_l^{\mathrm{AP}}} \\ \frac{\partial f_{\overline{CD}}(x_{l,k,b}, y_{l,k,b})}{\partial \mathbf{r}_l^{\mathrm{AP}}} = \frac{E_{\mathrm{bui}}}{2}\big(\cos\theta_{l,k} - \sin\theta_{l,k}\big) \frac{\partial \theta_{l,k}}{\partial \mathbf{r}_l^{\mathrm{AP}}} + \frac{\partial y_{l,k,b}}{\partial \mathbf{r}_l^{\mathrm{AP}}} \\ \frac{\partial f_{\overline{DE}}(x_{l,k,b}, y_{l,k,b})}{\partial \mathbf{r}_l^{\mathrm{AP}}} = (-x_{l,k,b} \cos\theta_{l,k} - y_{l,k,b} \sin\theta_{l,k} + \frac{d_{l,k}\cos\theta_{l,k}}{2}) \frac{\partial \theta_{l,k}}{\partial \mathbf{r}_l^{\mathrm{AP}}} \\ \qquad - \sin\theta_{l,k} \frac{\partial x_{l,k,b}}{\partial \mathbf{r}_l^{\mathrm{AP}}} + \cos\theta_{l,k} \frac{\partial y_{l,k,b}}{\partial \mathbf{r}_l^{\mathrm{AP}}} + \frac{\sin\theta_{l,k}}{2} \frac{\partial d_{l,k}}{\partial \mathbf{r}_l^{\mathrm{AP}}} \\ \frac{\partial f_{\overline{EF}}(x_{l,k,b}, y_{l,k,b})}{\partial \mathbf{r}_l^{\mathrm{AP}}} = (-y_{l,k,b} \cos\theta_{l,k} + x_{l,k,b} \sin\theta_{l,k} - \frac{d_{l,k}\sin\theta_{l,k}}{2}) \frac{\partial \theta_{l,k}}{\partial \mathbf{r}_l^{\mathrm{AP}}} \\ \qquad - \cos\theta_{l,k} \frac{\partial x_{l,k,b}}{\partial \mathbf{r}_l^{\mathrm{AP}}} - \sin\theta_{l,k} \frac{\partial y_{l,k,b}}{\partial \mathbf{r}_l^{\mathrm{AP}}} + \frac{\cos\theta_{l,k}}{2} \frac{\partial d_{l,k}}{\partial \mathbf{r}_l^{\mathrm{AP}}} \\ \frac{\partial f_{\overline{FA}}(x_{l,k,b}, y_{l,k,b})}{\partial \mathbf{r}_l^{\mathrm{AP}}} = \frac{E_{\mathrm{bui}}}{2}\big(\cos\theta_{l,k} - \sin\theta_{l,k}\big) \frac{\partial \theta_{l,k}}{\partial \mathbf{r}_l^{\mathrm{AP}}} - \frac{\partial y_{l,k,b}}{\partial \mathbf{r}_l^{\mathrm{AP}}} \end{cases} \tag{G.10}$$

with

$$\begin{aligned} \frac{\partial \theta_{l,k}}{\partial \mathbf{r}_l^{\mathrm{AP}}} = & \left(\frac{\pi}{2} S'\left(\arctan \frac{y_k^{\mathrm{user}} - y_l^{\mathrm{AP}}}{x_k^{\mathrm{user}} - x_l^{\mathrm{AP}}}\right) - 1\right) \\ & \cdot \left[\left(1 + \left(\frac{y_k^{\mathrm{user}} - y_l^{\mathrm{AP}}}{x_k^{\mathrm{user}} - x_l^{\mathrm{AP}}}\right)^2\right)^{-1} \frac{y_k^{\mathrm{user}} - y_l^{\mathrm{AP}}}{(x_k^{\mathrm{user}} - x_l^{\mathrm{AP}})^2}, \left(1 + \left(\frac{y_k^{\mathrm{user}} - y_l^{\mathrm{AP}}}{x_k^{\mathrm{user}} - x_l^{\mathrm{AP}}}\right)^2\right)^{-1} \frac{1}{x_l^{\mathrm{AP}} - x_k^{\mathrm{user}}}\right] \end{aligned} \tag{G.11}$$

based on (4.8) and (E.3). According to (4.6) and (4.7), the partial derivatives $\frac{\partial x_{l,k,b}}{\partial \mathbf{r}_l^{\mathrm{AP}}}$ and $\frac{\partial y_{l,k,b}}{\partial \mathbf{r}_l^{\mathrm{AP}}}$ are given by $\frac{\partial x_{l,k,b}}{\partial \mathbf{r}_l^{\mathrm{AP}}} = [\frac{\partial x_{l,k,b}}{\partial x_l^{\mathrm{AP}}}, \frac{\partial x_{l,k,b}}{\partial y_l^{\mathrm{AP}}}]$ and $\frac{\partial y_{l,k,b}}{\partial \mathbf{r}_l^{\mathrm{AP}}} = [\frac{\partial y_{l,k,b}}{\partial x_l^{\mathrm{AP}}}, \frac{\partial y_{l,k,b}}{\partial y_l^{\mathrm{AP}}}]$, respectively, where

$$\begin{cases} \frac{\partial x_{l,k,b}}{\partial x_l^{\mathrm{AP}}} = -\frac{1}{2}\cos\left(\arctan\frac{y_k^{\mathrm{user}}-y_l^{\mathrm{AP}}}{x_k^{\mathrm{user}}-x_l^{\mathrm{AP}}}\right) + \left(1+\left(\frac{y_k^{\mathrm{user}}-y_l^{\mathrm{AP}}}{x_k^{\mathrm{user}}-x_l^{\mathrm{AP}}}\right)^2\right)^{-1} \frac{y_{l,k,b}(y_k^{\mathrm{user}}-y_l^{\mathrm{AP}})}{(x_k^{\mathrm{user}}-x_l^{\mathrm{AP}})^2} \\ \frac{\partial x_{l,k,b}}{\partial y_l^{\mathrm{AP}}} = -\frac{1}{2}\sin\left(\arctan\frac{y_k^{\mathrm{user}}-y_l^{\mathrm{AP}}}{x_k^{\mathrm{user}}-x_l^{\mathrm{AP}}}\right) + \left(1+\left(\frac{y_k^{\mathrm{user}}-y_l^{\mathrm{AP}}}{x_k^{\mathrm{user}}-x_l^{\mathrm{AP}}}\right)^2\right)^{-1} \frac{y_{l,k,b}}{x_l^{\mathrm{AP}}-x_k^{\mathrm{user}}} \\ \frac{\partial y_{l,k,b}}{\partial x_l^{\mathrm{AP}}} = \frac{1}{2}\sin\left(\arctan\frac{y_k^{\mathrm{user}}-y_l^{\mathrm{AP}}}{x_k^{\mathrm{user}}-x_l^{\mathrm{AP}}}\right) - \left(1+\left(\frac{y_k^{\mathrm{user}}-y_l^{\mathrm{AP}}}{x_k^{\mathrm{user}}-x_l^{\mathrm{AP}}}\right)^2\right)^{-1} \frac{x_{l,k,b}(y_k^{\mathrm{user}}-y_l^{\mathrm{AP}})}{(x_k^{\mathrm{user}}-x_l^{\mathrm{AP}})^2} \\ \frac{\partial y_{l,k,b}}{\partial y_l^{\mathrm{AP}}} = -\frac{1}{2}\cos\left(\arctan\frac{y_k^{\mathrm{user}}-y_l^{\mathrm{AP}}}{x_k^{\mathrm{user}}-x_l^{\mathrm{AP}}}\right) - \left(1+\left(\frac{y_k^{\mathrm{user}}-y_l^{\mathrm{AP}}}{x_k^{\mathrm{user}}-x_l^{\mathrm{AP}}}\right)^2\right)^{-1} \frac{x_{l,k,b}}{x_l^{\mathrm{AP}}-x_k^{\mathrm{user}}}. \end{cases} \tag{G.12}$$

Finally, $\frac{\partial \mathcal{I}_{l,k}^{\mathrm{block}}}{\partial \mathbf{r}^{\mathrm{AP}}}$ can be obtained by combining (G.5)–(G.12).

Appendix H

Derivation of the Approximation Guarantee of Algorithm 3

The derivation of the approximation guarantee of Algorithm 3 is mainly based on the following performance guarantee result in [108]: For an optimization problem with a monotone submodular objective and Y matroid constraints, the ratio of the solution obtained by a greedy algorithm to the optimal solution is lower bounded by $\frac{1}{Y+1}$ [108].

For Algorithm 3, it associates one user with one BS in a greedy manner in each iteration to improve the objective function of Problem P4. In the following, we will further show that Problem P4 can be reformulated as an optimization problem with a monotone submodular objective and 2 matroid constraints.

Let us first present the definitions of submodular function and matroid in [109] as follows.

Definition H.1. Let $\mathcal{V}$ be a finite ground set, and $2^{\mathcal{V}}$ be the power set of $\mathcal{V}$. A set function $f(\mathcal{S})$ with the input $\mathcal{S} \in 2^{\mathcal{V}}$ and a real value output, denoted by $f : 2^{\mathcal{V}} \rightarrow \mathbb{R}$, is submodular if

$$f(\mathcal{S} \cup \{v\}) - f(\mathcal{S}) \geq f(\mathcal{T} \cup \{v\}) - f(\mathcal{T}), \tag{H.1}$$

for any $\mathcal{S} \subseteq \mathcal{T} \subseteq \mathcal{V}$ and $v \in \mathcal{V} \backslash \mathcal{T}$, i.e., the marginal gain of adding an extra element in the set decreases or remains unchanged as the size of the set grows.

Furthermore, a set function $f(\mathcal{S})$ is monotone if

$$f(\mathcal{S}) \leq f(\mathcal{T}), \tag{H.2}$$

for any $\mathcal{S} \subseteq \mathcal{T} \subseteq \mathcal{V}$.

Definition H.2. A matroid $\mathfrak{M}$ is a pair $(\mathcal{V}, Z)$, denoted by $\mathfrak{M} = (\mathcal{V}, Z)$, where $\mathcal{V}$ is a finite ground set and $Z \subseteq 2^{\mathcal{V}}$ is a collection of subsets of $\mathcal{V}$ with the following properties:
(1) Z is nonempty;
(2) Z is downward closed, i.e., for each $\mathcal{X} \subseteq \mathcal{Y} \in Z$, we have $\mathcal{X} \in Z$;
(3) If $\mathcal{X}, \mathcal{Y} \in Z$ and $|\mathcal{X}| > |\mathcal{Y}|$, then there exists an element $v \in \mathcal{X} \backslash \mathcal{Y}$ such that $\mathcal{Y} \cup \{v\} \in Z$.
In particular, a partition matroid is a matroid $(\mathcal{V}, Z)$ where the ground set $\mathcal{V}$ is partitioned into some disjoint sets, $\mathcal{V}_1, \mathcal{V}_2, \cdots, \mathcal{V}_L$, and

$$Z = \{\mathcal{X} \subseteq \mathcal{V} : |\mathcal{X} \cap \mathcal{V}_l| \leq \Xi_l, \forall l = 1, 2, \cdots, L\}, \tag{H.3}$$

for some given parameters $\Xi_1, \Xi_2, \cdots, \Xi_L$.

In the user association problem P4, the constraint (4.17) indicates that user k can associate with BS l only if BS l is physically accessible to user k, i.e., $I\left(\mathcal{I}_{l,k}^{\text{out}} + \mathcal{I}_{l,k}^{\text{block}}\right) = 0$. In order to reformulate the objective function (4.16) as a set function, with constraint (4.17) taken into account, we define the ground set $\mathcal{V}$ as

$$\mathcal{V} = \{v_{l,k} : l = 1, \cdots, L; k = 1, \cdots, K; I\left(\mathcal{I}_{l,k}^{\text{out}} + \mathcal{I}_{l,k}^{\text{block}}\right) = 0\}, \tag{H.4}$$

and the user association set $\mathcal{S}_{\text{a}}$ as a subset of $\mathcal{V}$ such that $v_{l,k} \in \mathcal{S}_{\text{a}}$ if and only if user k associates with BS l, i.e., $\mathcal{I}_{l,k}^{\text{associate}} = 1$. Therefore, minimizing the objective (4.16) of Problem P4 becomes equivalent to maximizing the set function $f(\mathcal{S}_{\text{a}}) = |\mathcal{S}_{\text{a}}|$, which is equal to the number of users that can associate with BSs. As the set function $f(\cdot)$ satisfies

$$f(\mathcal{S} \cup \{v\}) - f(\mathcal{S}) = f(\mathcal{T} \cup \{v\}) - f(\mathcal{T}) \text{ and } f(\mathcal{S}) \leq f(\mathcal{T}), \tag{H.5}$$

for any $\mathcal{S} \subseteq \mathcal{T} \subseteq \mathcal{V}$ and $v \in \mathcal{V} \backslash \mathcal{T}$, it is a monotone submodular function according to Definition H.1.

With the ground set $\mathcal{V}$ given in (H.4), the constraint (4.18) can be written as

$$\mathcal{S} \in Z_B, \tag{H.6}$$

where $Z_B = \{\mathcal{X} \subseteq \mathcal{V} : |\mathcal{X} \cap \mathcal{V}_l^B| \leq K_{\max}, \forall l = 1, 2, \cdots, L\}$. $\mathcal{V}_l^B = \{v_{l,k} : k = 1, \cdots, K; I(\mathcal{I}_{l,k}^{\text{out}} + \mathcal{I}_{l,k}^{\text{block}}) = 0\}$ is the set containing all the possible user associations between BS l and all the users, which satisfies $\bigcup_{l=1,\cdots,L} \mathcal{V}_l^B = \mathcal{V}$ and $\mathcal{V}_i^B \cap \mathcal{V}_j^B = \emptyset, \forall i \neq j$. Based on Definition H.2, we can see that $(\mathcal{V}, Z_B)$ is a partition matroid, and hence (H.6) is a matroid constraint [108].

Similarly, the constraint (4.19) can also be written as

$$\mathcal{S} \in Z_U, \tag{H.7}$$

where $Z_U = \{\mathcal{X} \subseteq \mathcal{V} : |\mathcal{X} \cap \mathcal{V}_k^U| \leq 1, \forall k = 1, 2, \cdots, K\}$. $\mathcal{V}_k^U = \{v_{l,k} : l = 1, \cdots, L; I(\mathcal{I}_{l,k}^{\text{out}} + \mathcal{I}_{l,k}^{\text{block}}) = 0\}$ is the set containing all the possible user associations between user k and all the BSs, which satisfies $\bigcup_{k=1,\cdots,K} \mathcal{V}_k^U = \mathcal{V}$ and $\mathcal{V}_i^U \cap \mathcal{V}_j^U = \emptyset, \forall i \neq j$. It is clear from Definition H.2 that $(\mathcal{V}, Z_U)$ is also a partition matroid. Therefore, (H.7) is also a matroid constraint [108].

So far, we have shown that Problem P4 can be reformulated as a submodular optimization problem with two matroid constraints. Therefore, based on the performance guarantee result in [108], P4 can be solved by a greedy algorithm, e.g., Algorithm 3, with a constant-factor $\frac{1}{3}$ approximation guarantee.

Appendix I

Derivations of $\tilde{\boldsymbol{\omega}}_{K_{\text{out}}}(t)$ and $\tilde{\boldsymbol{\omega}}_{\hat{\lambda}}(t)$

In the following derivation, we drop the iteration index t for the sake of conciseness.

I.1 Derivation of $\tilde{\boldsymbol{\omega}}_{K_{\text{out}}}$

By replacing the indicator function in (5.7) with (4.23), $\tilde{\boldsymbol{\omega}}_{K_{\text{out}}}$ can be approximated by

$$\tilde{\boldsymbol{\omega}}_{K_{\text{out}}} \approx \left[\sum_{k=1}^{K} \sum_{l=1}^{L} S'(d_{l,k} - R_{\text{A}}) \prod_{i \in \mathcal{L} \setminus \{l\}} S(d_{i,k} - R_{\text{A}}) \frac{\partial d_{l,k}}{\partial \mathbf{r}^{\text{AP}}}, \right. \\ \left. - \sum_{k=1}^{K} \sum_{l=1}^{L} S'(d_{l,k} - R_{\text{A}}) \prod_{i \in \mathcal{L} \setminus \{l\}} S(d_{i,k} - R_{\text{A}}) \right], \tag{I.1}$$

where $S'(d_{l,k} - R_{\text{A}}) = \xi S(d_{l,k} - R_{\text{A}})(1 - S(d_{l,k} - R_{\text{A}}))$, and $\frac{\partial d_{l,k}}{\partial \mathbf{r}^{\text{AP}}}$ is given by

$$\frac{\partial d_{l,k}}{\partial \mathbf{r}^{\text{AP}}} = \left[\frac{\partial d_{l,k}}{\partial \mathbf{r}_1^{\text{AP}}}, \cdots, \frac{\partial d_{l,k}}{\partial \mathbf{r}_L^{\text{AP}}} \right]. \tag{I.2}$$

In the Cartesian coordinate with $\mathbf{r}_l^{\text{AP}} = [x_l^{\text{AP}}, y_l^{\text{AP}}]$ and $\mathbf{r}_k^{\text{user}} = [x_k^{\text{user}}, y_k^{\text{user}}]$, $\frac{\partial d_{l,k}}{\partial \mathbf{r}_j^{\text{AP}}}, j = 1, \cdots, L$ can be derived as (G.2).

I.2 Derivation of $\tilde{\boldsymbol{\omega}}_{\hat{\lambda}}$

According to (5.3), $\tilde{\boldsymbol{\omega}}_{\hat{\lambda}}$ can be written as

$$\tilde{\boldsymbol{\omega}}_{\hat{\lambda}} = \left[\sum_{i\in\mathcal{L}} \sum_{\substack{\mathcal{S}=\tilde{\mathcal{S}}\cup\{i\} \\ \tilde{\mathcal{S}}\subseteq\mathcal{L}\setminus\{i\}}} \frac{\partial\hat{\lambda}}{\partial k^{(i,\mathcal{S})}} \cdot \frac{\partial k^{(i,\mathcal{S})}}{\partial \mathbf{r}^{\mathrm{AP}}}, \sum_{i\in\mathcal{L}} \sum_{\substack{\mathcal{S}=\tilde{\mathcal{S}}\cup\{i\} \\ \tilde{\mathcal{S}}\subseteq\mathcal{L}\setminus\{i\}}} \frac{\partial\hat{\lambda}}{\partial k^{(i,\mathcal{S})}} \cdot \frac{\partial k^{(i,\mathcal{S})}}{\partial R_{\mathrm{A}}} \right]. \tag{I.3}$$

By replacing the indicator function in (5.1) with (4.23), the partial derivatives $\frac{\partial k^{(i,\mathcal{S})}}{\partial R_{\mathrm{A}}}$ and $\frac{\partial k^{(i,\mathcal{S})}}{\partial \mathbf{r}^{\mathrm{AP}}}$ can be approximated by

$$\begin{aligned} \frac{\partial k^{(i,\mathcal{S})}}{\partial R_{\mathrm{A}}} \approx \sum_{k\in\mathcal{K}_i^*} \Bigg(& \sum_{l\in\mathcal{S}} S'(R_{\mathrm{A}} - d_{l,k}) \prod_{j\in\mathcal{S}\setminus\{l\}} S(C - d_{j,k}) \prod_{q\in\mathcal{L}\setminus\mathcal{S}} S(d_{q,k} - R_{\mathrm{A}}) \\ & - \sum_{\tilde{l}\in\mathcal{L}\setminus\mathcal{S}} S'(d_{\tilde{l},k} - R_{\mathrm{A}}) \prod_{j\in\mathcal{S}} S(R_{\mathrm{A}} - d_{j,k}) \prod_{q\in\mathcal{L}\setminus(\mathcal{S}\cup\{\tilde{l}\})} S(d_{q,k} - R_{\mathrm{A}}) \Bigg), \end{aligned} \tag{I.4}$$

and

$$\begin{aligned} \frac{\partial k^{(i,\mathcal{S})}}{\partial \mathbf{r}^{\mathrm{AP}}} \approx \sum_{k\in\mathcal{K}_i^*} \Bigg(& \sum_{l\in\mathcal{S}} -S'(R_{\mathrm{A}} - d_{l,k}) \prod_{j\in\mathcal{S}\setminus\{l\}} S(R_{\mathrm{A}} - d_{j,k}) \prod_{q\in\mathcal{L}\setminus\mathcal{S}} S(d_{q,k} - R_{\mathrm{A}}) \frac{\partial d_{l,k}}{\partial \mathbf{r}^{\mathrm{AP}}} \\ & + \sum_{\tilde{l}\in\mathcal{L}\setminus\mathcal{S}} S'(d_{\tilde{m},n} - R_{\mathrm{A}}) \prod_{j\in\mathcal{S}} S(R_{\mathrm{A}} - d_{j,k}) \prod_{q\in\mathcal{L}\setminus(\mathcal{S}\cup\{\tilde{l}\})} S(d_{q,k} - R_{\mathrm{A}}) \frac{\partial d_{\tilde{l},k}}{\partial \mathbf{r}^{\mathrm{AP}}} \Bigg), \end{aligned} \tag{I.5}$$

respectively, where $\frac{\partial d_{l,k}}{\partial \mathbf{r}^{\mathrm{AP}}}$ is given in (I.2).

For the partial derivative $\frac{\partial\hat{\lambda}}{\partial k^{(i,\mathcal{S})}}$, according to (5.3), we have

$$\frac{\partial\hat{\lambda}}{\partial k^{(i,\mathcal{S})}} = \lambda^{(i,\mathcal{S})} + \sum_{j\in\mathcal{L}} \sum_{\substack{\mathcal{V}=\tilde{\mathcal{V}}\cup\{j\} \\ \tilde{\mathcal{V}}\subseteq\mathcal{L}\setminus\{j\}}} k^{(j,\mathcal{V})} \frac{\partial\lambda^{(j,\mathcal{V})}}{\partial k^{(i,\mathcal{S})}}, \tag{I.6}$$

where

$$\frac{\partial\lambda^{(j,\mathcal{V})}}{\partial k^{(i,\mathcal{S})}} = \left(\frac{\partial\lambda^{(j,\mathcal{V})}}{\partial \mathbf{p}}\right)' \frac{\partial \mathbf{p}}{\partial k^{(i,\mathcal{S})}} \tag{I.7}$$

with $\mathbf{p}=[p^{(1)},\cdots,p^{(L)}]'$. According to (5.4), the partial derivative $\frac{\partial\lambda^{(j,\mathcal{V})}}{\partial p^{(l)}}$ can be obtained as

$$\frac{\partial\lambda^{(j,\mathcal{V})}}{\partial p^{(l)}}=\begin{cases}\frac{\tau_T\frac{\partial\varepsilon^{(l,\{l\})}}{\partial p^{(l)}}\prod_{q\in\mathcal{V}\setminus\{l\}}\varepsilon^{q,\{q\}}}{\Xi^{(j,\mathcal{V})}}-\frac{\tau_T\frac{\partial\Xi^{(j,\mathcal{V})}}{\partial p^{(l)}}\varepsilon^{(j,\mathcal{V})}}{(\Xi^{(j,\mathcal{V})})^2} & \text{if } l\in\mathcal{V}\\ 0 & \text{otherwise,}\end{cases} \tag{I.8}$$

where

$$\Xi^{(j,\mathcal{V})}=\varepsilon^{(j,\mathcal{V})}\Big(\tau_T+\tau_F\cdot\frac{1-p^{(j)}}{p^{(j)}}\Big)+\frac{1}{2p^{(j)}}\bigg(1+\vartheta\Big(\frac{p^{(j)}}{2p^{(j)}-1}+\big(1-\frac{p^{(j)}}{2p^{(j)}-1}\big)(2-2p^{(j)})^{\varrho}\Big)\bigg), \tag{I.9}$$

and

$$\frac{\partial\Xi^{(j,\mathcal{V})}}{\partial p^{(l)}}=\begin{cases}\frac{\partial\varepsilon^{(l,\{l\})}}{\partial p^{(l)}}\prod_{q\in\mathcal{V}\setminus\{l\}}\varepsilon^{q,\{q\}}\big(\tau_T+\tau_F\cdot\frac{1-p^{(l)}}{p^{(l)}}\big)-\frac{1}{2(p^{(l)})^2}\\ \cdot\Big(2\tau_F\varepsilon^{(l,\mathcal{V})}+1+\vartheta\big(\frac{p^{(l)}}{2p^{(l)}-1}+(1-\frac{p^{(l)}}{2p^{(l)}-1})(2-2p^{(l)})^{\varrho}\big)\Big)\\ +\frac{\vartheta}{2p^{(l)}}\Big(\frac{(2-2p^{(l)})^{\varrho}-1}{(2p^{(l)}-1)^2}-2\varrho(1-\frac{p^{(l)}}{2p^{(l)}-1})(2-2p^{(l)})^{\varrho-1}\Big) & \text{if } l=j\\ \Big(\tau_T+\tau_F\cdot\frac{1-p^{(j)}}{p^{(j)}}\Big)\frac{\partial\varepsilon^{(l,\{l\})}}{\partial p^{(l)}}\prod_{q\in\mathcal{V}\setminus\{l\}}\varepsilon^{q,\{q\}} & \text{otherwise.}\end{cases} \tag{I.10}$$

We further have

$$\frac{\partial\varepsilon^{(l,\{l\})}}{\partial p^{(l)}}=\big(\varepsilon^{(l,\{l\})}\big)^2\big(\tau_F+(\tau_T-\tau_F)(1+\ln p^{(l)})\big) \tag{I.11}$$

according to (5.5). To obtain $\frac{\partial\mathbf{p}}{\partial k^{(i,\mathcal{S})}}$ in (I.7), we rewrite (5.6) as

$$p^{(l)}=\exp\Bigg(-\frac{1}{\tau_T\varepsilon^{(l,\{l\})}p^{(l)}}\sum_{j\in\mathcal{L}}\sum_{\substack{\mathcal{V}=\tilde{\mathcal{V}}\cup\{l,j\}\\ \tilde{\mathcal{V}}\subseteq\mathcal{L}\setminus\{l,j\}}}k^{(j,\mathcal{V})}\lambda^{(j,\mathcal{V})}\Bigg), \tag{I.12}$$

$l\in\mathcal{L}$. By taking derivative with respect to $k^{(i,\mathcal{S})}$ on both sides of (I.12) and after some manipulations, we can obtain $\frac{\partial\mathbf{p}}{\partial k^{(i,\mathcal{S})}}=\mathbf{A}^{-1}\mathbf{b}$, where the (l,k)-th element of matrix $\mathbf{A}\in\mathbb{R}^{L\times L}$ can be derived as:

$$A_{l,k}=\begin{cases}\sum_{j\in\mathcal{L}}\sum_{\substack{\mathcal{V}=\tilde{\mathcal{V}}\cup\{l,j\}\\ \tilde{\mathcal{V}}\subseteq\mathcal{L}\setminus\{l,j\}}}k^{(j,\mathcal{V})}\Big(\frac{\varepsilon^{(l,\{l\})}+p^{(l)}\cdot\partial\varepsilon^{(l,\{l\})}/\partial p^{(l)}}{\tau_T(\varepsilon^{(l,\{l\})})^2p^{(l)}}\lambda^{(j,\mathcal{V})}\\ -\frac{1}{\tau_T\varepsilon^{(l,\{l\})}}\cdot\frac{\partial\lambda^{(j,\mathcal{V})}}{\partial p^{(l)}}\Big)-1 & \text{if } l=k\\ -\frac{1}{\tau_T\varepsilon^{(l,\{l\})}}\sum_{j\in\mathcal{L}}\sum_{\substack{\mathcal{V}=\tilde{\mathcal{V}}\cup\{l,k,j\}\\ \tilde{\mathcal{V}}\subseteq\mathcal{L}\setminus\{l,k,j\}}}k^{(j,\mathcal{V})}\frac{\partial\lambda^{(j,\mathcal{V})}}{\partial p^{(k)}} & \text{otherwise}\end{cases} \tag{I.13}$$

with $\frac{\partial \lambda^{(j,\mathcal{V})}}{\partial p^{(l)}}$ and $\frac{\partial \varepsilon^{(l,\{l\})}}{\partial p^{(l)}}$ given in (I.8) and (I.11), respectively, and the l-th element of vector $\mathbf{b} \in \mathbb{R}^{L\times 1}$ is given by

$$b_l = \begin{cases} \frac{\lambda^{(i,\mathcal{S})}}{\tau_T \varepsilon^{(l,\{l\})}} & \text{if } l \in \mathcal{S} \\ 0 & \text{otherwise.} \end{cases} \tag{I.14}$$

Finally, $\tilde{\boldsymbol{\omega}}_{\hat{\lambda}}$ can be obtained by combining (I.3)–(I.14).

Bibliography

[1] J. Kelly, "Google Employees Will Work From Home Through Next Summer," Forbes, *https://www.forbes.com/sites/jackkelly/2020/07/27/google-employees-will-work-from-home-through-next-summer/#5363a8083268.* Accessed 27 Jul. 2020.

[2] A. Goldsmith, *Wireless Communications.* Cambridge, U.K.: Cambridge Univ. Press, 2005.

[3] A. R. Mishra, *Fundamentals of Network Planning and Optimisation 2G/3G/4G: Evolution to 5G.* Hoboken, NJ, USA: Wiley, 2018.

[4] X. Ge, S. Tu, G. Mao, C.-X. Wang, and T. Han, "5G ultra-dense cellular networks," *IEEE Wireless Commun.,* vol. 23, no. 1, pp. 72–79, Feb. 2016.

[5] D. López-Pérez, M. Ding, H. Claussen, and A. H. Jafari, "Towards 1 Gbps/UE in cellular systems: Understanding ultra-dense small cell deployment," *IEEE Commun. Surveys Tuts.,* vol. 17, no. 4, pp. 2078–2101, 4th Quart. 2015.

[6] S. Singh, M. N. Kulkarni, A. Ghosh, and J. G. Andrews, "Tractable model for rate in self-backhauled millimeter wave cellular networks," *IEEE J. Sel. Areas Commun.,* vol. 33, no. 10, pp. 2196–2211, Oct. 2015.

[7] K. Han, Y. Cui, Y. Wu, and K. Huang, "The connectivity of millimeter wave networks in urban environments modeled using random lattices," *IEEE Trans. Wireless Commun.,* vol. 17, no. 5, pp. 3357–3372, May 2018.

[8] M. Dong, T. Kim, J. Wu, and E. W. M. Wong, "Cost-efficient millimeter wave base station deployment in Manhattan-type geometry," *IEEE Access,* vol. 7, pp. 149959–149970, Oct. 2019.

[9] T. Bai and R. W. Heath Jr., "Coverage and rate analysis for millimeter wave cellular networks," *IEEE Trans. Wireless Commun.,* vol. 14, no. 2, pp. 1100–1114, Oct. 2014.

[10] M. Ding, P. Wang, D. López-Pérez, G. Mao, and Z. Lin, "Performance impact of LoS and NLoS transmissions in dense cellular networks," *IEEE Trans. Wireless Commun.,* vol. 15, no. 3, pp. 2365–2380, Mar. 2016.

[11] Y. Gao, L. Dai, and X. Hei, "Throughput optimization of multi-BSS IEEE 802.11 networks with universal frequency reuse," *IEEE Trans. Commun.,* vol. 65, no. 8, pp. 3399–3414, Aug. 2017.

[12] S. Chatterjee, M. J. Abdel-Rahman , and A. B. MacKenzie, "Optimal base station deployment with downlink rate coverage probability constraint,", *IEEE Wireless Commun. Lett.,* vol. 7, no. 3, pp. 340–343, Jun. 2018.

[13] K. Tutschku, "Demand-based radio network planning of cellular mobile communication systems," in *Proc. IEEE INFOCOM 1998*, pp. 1054–1061.

[14] C. Y. Lee and H. G. Kang, "Cell planning with capacity expansion in mobile communications: a tabu search approach," *IEEE Trans. Veh. Technol.,* vol. 49, no. 5, pp. 1678–1691, Sep. 2000.

[15] S. Hurley, "Planning effective cellular mobile radio networks," *IEEE Trans. Veh. Technol.,* vol. 51, no. 2, Mar. 2002, pp. 243–253.

[16] E. Amaldi, A. Capone, and F. Malucelli, "Planning UMTS base station location: optimization models with power control and algorithms," *IEEE Trans. Wireless Commun.,* vol. 2, no. 5, pp. 939–952, Sep. 2003.

[17] J. Zhang, X. Jia, Z. Zheng, and Y. Zhou, "Minimizing cost of placement of multi-radio and multi-power-level access points with rate adaptation in indoor environment," *IEEE Trans. Wireless Commun.,* vol. 10, no. 7, pp. 2186–2195, Jul. 2011.

[18] Z. Zheng, L. X. Cai, M. Dong, X. Shen, and H. V. Poor, "Constrained energy-aware AP placement with rate adaptation in WLAN mesh networks," in *Proc. IEEE Globecom 2011,* pp. 1–5.

[19] T. Wang, W. Jia, G. Xing, and M. Li, "Exploiting statistical mobility models for efficient Wi-Fi deployment," *IEEE Trans. Veh. Technol.,* vol. 62, no. 1, pp. 360–373, Jan. 2013.

[20] A. Farsi, N. Achir, and K. Boussetta, "WLAN planning: Separate and joint optimization of both access point placement and channel assignment," *Ann. Telecommun.,* vol. 70, nos. 5–6, pp. 263–274, Jun. 2015.

[21] S. Wang and C. Ran, "Rethinking cellular network planning and optimization," *IEEE Wireless Commun.,* vol. 23, no. 2, pp. 118–125, Apr. 2016.

[22] I. Mavromatis, A. Tassi, R. J. Piechocki, and A. Nix, "Efficient millimeter-wave infrastructure placement for city-scale ITS," in *Proc. IEEE VTC 2019*, pp. 1–5.

[23] H. R. Anderson and J. P. McGeehan, "Optimizing microcell base station locations using simulated annealing techniques," in *Proc. IEEE VTC 1994*, pp. 858–862.

[24] H. D. Sherali, C. M. Pendyala, and Theodore S. Rappaport, "Optimal location of transmitters for micro-cellular radio communication system design," *IEEE J. Sel. Areas Commun.,* vol. 14, no. 4, pp. 662–671, May 1996.

[25] M. Unbehaun and M. Kamenetky, "On the deployment of picocellular wireless infrastructure," *IEEE Wireless Commun.,* vol. 10, no. 6, pp. 70–80, Dec. 2003.

[26] K. Jaffres-Runser, J. Gorce, and S. Ubeda, "Mono- and multiobjective formulations for the indoor wireless LAN planning problem," *Computers & Operations Research,* vol. 35, no. 12, pp. 3885–3901, Dec. 2008.

[27] M. Taufiq, M. Abdullah, and D. Choi, "Wireless LAN access point placement based on user mobility," *Wireless Personal Commun.,* vol. 60, no. 3, pp. 431–440, Oct. 2011.

[28] S. Tang, L. Ma, Y. Xu, "A novel AP placement algorithm based on user distribution for indoor WLAN system," *China Commun.,* vol. 13, no. 10, pp. 108–118, Oct. 2016.

[29] N. Palizban, S. Szyszkowicz, and H. Yanikomeroglu, "Automation of millimeter wave network planning for outdoor coverage in dense urban areas using wall-mounted base stations," *IEEE Wireless Commun. Lett.,* vol. 6, no. 2, pp. 206–209, Apr. 2017.

[30] Y. Liu, W. Huangfu, H. Zhang, H. Wang, W. An, and K. Long, "An efficient geometry-induced genetic algorithm for base station placement in cellular networks," *IEEE Access,* vol. 7, pp. 108604–108616, Aug. 2019.

[31] S. Fang and T. Lin, "A novel access point placement approach for WLAN-based location systems," in *Proc. IEEE WCNC 2010,* pp. 1–4.

[32] X. Du and K. Yang, "A map-assisted WiFi AP placement algorithm enabling mobile device's indoor positioning," *IEEE Syst. J.,* vol. 11, no. 3, pp. 1467–1475, Sep. 2017.

[33] X. Ling and K. L. Yeung, "Joint access point placement and channel assignment for 802.11 wireless LANs," *IEEE Trans. Wireless Commun.,* vol. 5, no. 10, pp. 2705–2711, Oct. 2006.

[34] S. Bosio, A. Capone, and M. Cesana, "Radio planning of wireless local area networks," *ACM/IEEE Trans. Netw.,* vol. 15, no. 6, pp. 1414–1427, Dec. 2007.

[35] J. Liu, T. Kou, Q. Chen, and H. D. Sherali, "Femtocell base station deployment in commercial buildings: A global optimization approach," *IEEE J. Sel. Areas Commun.,* vol. 30, no. 3, pp. 652–663, Apr. 2012.

[36] G. Zhao, S. Chen, L. Zhao, and L. Hanzo, "Joint energy-spectral-efficiency optimization of CoMP and BS deployment in dense large-scale cellular networks," *IEEE Trans. Wireless Commun.,* vol. 16, no. 7, pp. 4832–4847, Jul. 2017.

[37] X. Wang, P. Zhu, and M. Chen, "Antenna location design for generalized distributed antenna systems," *IEEE Commun. Lett.,* vol. 13, no. 5, pp. 315–317, May 2009.

[38] A. Yang, Y. Jing, C. Xing, Z. Fei, and J. Kuang, "Performance analysis and location optimization for massive MIMO systems with circularly distributed antennas," *IEEE Trans. Wireless Commun.,* vol. 14, no. 10, pp. 5659–5671, Oct. 2015.

[39] G. N. Kamga, M. Xia, and S. Aïssa, "Spectral-efficiency analysis of massive MIMO systems in centralized and distributed schemes," *IEEE Trans. Commun.,* vol. 64, no. 5, pp. 1930–1941, May 2016.

[40] E. Koyuncu, "Performance gains of optimal antenna deployment in massive MIMO systems", *IEEE Trans. Wireless Commun.,* vol. 17, no. 4, pp. 2633–2644, Apr. 2018.

[41] R. Kershner, "The number of circles covering a set," *Amer. J. Math.,* vol. 61, no. 3, pp. 665–671, Jul. 1939.

[42] Y. T. Chan, W. Y. Tsui, H. C. So, and P. C. Ching, "Time of arrival based localization under NLOS conditions," *IEEE Trans. Veh. Technol.,* vol. 55, no. 1, pp. 17–24, Jan. 2006.

[43] J. He and H. C. So, "A hybrid TDOA-fingerprinting-based localization system for LTE network", *IEEE Sensors J.*, vol. 20, no. 22, pp. 13653–13665, Jun. 2020.

[44] L. A. Hannah, "Stochastic optimization," lecture note, available at *http://www.stat.columbia.edu/ liam/teaching/compstat-spr15/lauren-notes.pdf.*

[45] W. B. Powell, "A unified framework for stochastic optimization," *Eur. J. Oper. Res.,* vol. 275, no. 3, pp. 795–821, Jun. 2019.

[46] J. Linderoth, A. Shapiro, and S. Wright, "The empirical behavior of sampling methods for stochastic programming," *Ann. Oper. Res.,* vol. 142, pp. 215–241, Feb. 2006.

[47] L. Bottou, F. E. Curtis, and J. Nocedal, "Optimization methods for large-scale machine learning," *SIAM Rev.,* vol. 60, no. 2, pp. 223–311, 2018.

[48] L. Bianchi, M. Dorigo, L. M. Gambardella, and W. J. Gutjahr, "A survey on metaheuristics for stochastic combinatorial optimization," *Natural Comput.,* vol. 8, no. 2, pp. 239–287, 2009.

[49] H. J. Kushner and G. G. Yin, *Stochastic Approximation Algorithms and Applications.* 2nd ed. Berlin, Germany: Springer-Verlag, 2003.

[50] J. Spall, *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control.* Hoboken, NJ: Wiley, 2003.

[51] R. H. Byrd, S. L. Hansen, J. Nocedal, and Y. Singer, "A stochastic quasi-Newton method for large-scale optimization," *SIAM J. Optim.,* vol. 26, no. 2, pp. 1008–1031, Jan. 2014.

[52] S. Scardapane and P. Lorenzo, "Stochastic training of neural networks via successive convex approximations," *IEEE Trans. Neural Netw. Learn. Syst.* vol. 29, no. 10, pp. 4947–4956, Oct. 2018.

[53] A. Liu, V. K. N. Lau, and B. Kananian, "Stochastic successive convex approximation for non-convex constrained stochastic optimization," *IEEE Trans. Signal Process.,* vol. 67, no. 16, pp. 4189–4203, Aug. 2019.

[54] J. Nocedal and S. J. Wright. *Numerical Optimization.* Springer New York, Second edition, 2006.

[55] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Statist.,* vol. 22, no. 3, pp. 400–407, 1951.

[56] J. Kiefer and J. Wolfowitz, "Stochastic estimation of a maximum of a regression function," *Ann. Math. Statist.,* vol. 23, no. 3, pp. 462–466, 1952.

[57] J. R. Blum, "Multidimensional stochastic approximation methods," *Ann. Math. Statist.,* vol. 25, no. 4, pp. 737–744, 1954.

[58] S. Boyd, A. Mutapcic and J. Duchi, "Stochastic subgradient methods," lecture note, available at *https://stanford.edu/class/ee364b/lectures/stoch_subgrad_notes.pdf.*

[59] G. Lan and Z. Zhou, "Algorithms for stochastic optimization with function or expectation constraints", *Comput. Optim. Appl.,* vol 76, pp. 461–498, Feb. 2020.

[60] S. Shafir and J. Roughgarden,"The effect of memory length on individual fitness in a lizard," *Adaptive Individuals in Evolving Populations: Models and Algorithms,* pp. 173–186, 1996.

[61] R. Marimon and E. McGrattan, "On adaptive learning in strategic games," *Learning and Rationality in Economics,* pp. 63–101, 1995.

[62] H. J. Kushner and R. Kumar, "Convergence and rate of convergence of a recursive identification and adaptive control method which uses truncated estimators," *IEEE Trans. Automat. Contr.,* vol. 27, no. 4, pp. 146–153 Feb. 1982.

[63] H. J. Kushner and P. A. Whiting, "Convergence of proportional-fair sharing algorithms under general conditions" *IEEE Trans. Wireless Commun.,* vol. 3, no. 4, pp. 1250–1259, Jul. 2004.

[64] L. Bottou, "Online Algorithms and Stochastic Approximations," *Online Learning and Neural Networks,* D. Saad, ed., Cambridge Univ. Press, 1998.

[65] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE,* vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[66] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning.* Cambridge, MA, USA: MIT Press, 2016.

[67] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge, U.K.: Cambridge Univ. Press, 2004.

[68] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications.* Now Publishers Inc., 2004.

[69] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications.* Cambridge, U.K.: Cambridge Univ. Press, 2011.

[70] A. M. Tulino, A. Lozano, and S. Verdú, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Trans. Inf. Theory,* vol. 51, no. 7, pp. 2491–2509, Jul. 2005.

[71] D. Aktas, M. N. Bacha, J. S. Evans, and S. V. Hanly, "Scaling results on the sum capacity of cellular networks with MIMO links," *IEEE Trans. Inf. Theory,* vol. 52, no. 7, pp. 3264–3274 Jul. 2006.

[72] S. Chatzinotas, M. A. Imran, and R. Hoshyar, "On the multicell processing capacity of the cellular MIMO uplink channel in correlated Rayleigh fading environment," *IEEE Trans. Wireless Commun.,* vol. 8, no. 7, pp. 3704–3715, Jul. 2009.

[73] W. Hachem, P. Loubaton, and J. Najim, "Deterministic equivalents for certain functionals of large random matrices," *Annals of Applied Probability,* vol. 17, no. 3, pp. 875–930, Jun. 2007.

[74] J. Zhang, C.-K. Wen, S. Jin, X. Gao, and K.-K. Wong, "On capacity of large-scale MIMO multiple access channels with distributed sets of correlated antennas," *IEEE J. Sel. Areas Commun.,* vol. 31, no. 2, pp. 133–148, Feb. 2013.

[75] A. Lu, X. Gao, and C. Xiao, "Free deterministic equivalents for the analysis of MIMO multiple access channel," *IEEE Trans. Inf. Theory,* vol. 62, no. 8, pp. 4604–4629, Aug. 2016.

[76] V. L. Girko, *Theory of Random Determinants.* Kluwer Academic, 1990.

[77] E. Park, S.-R. Lee, and I. Lee, "Antenna placement optimization for distributed antenna systems," *IEEE Trans. Wireless Commun.,* vol. 11, no. 7, pp. 2468–2477, Jul. 2012.

[78] A. Minasian, R. S. Adve, S. Shahbazpanahi, and G. Boudreau, "On RRH placement for multi-user distributed massive MIMO systems", *IEEE Access,* vol. 6, pp. 70597–70614, Nov. 2018.

[79] M. K. Simon and M. S. Alouini, *Digital Communication over Fading Channels: A Unified Approach to Performance Analysis.* Wiley, 2000.

[80] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication.* Cambridge University Press, May 2005.

[81] L. R. Scott, *Numerical Analysis.* Princeton and Oxford, U.S.: Princeton Univ. Press, 2011.

[82] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

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

[84] Z. Liu and L. Dai, "On the scaling behavior of average ergodic capacity of distributed MIMO systems," in *Proc. IEEE Globecom 2015,* pp. 1–6.

[85] F. Bach and E. Moulines, "Non-asymptotic analysis of stochastic approximation algorithms for machine learning," in *Proc. NeurIPS 2011* pp. 451–459.

[86] M. Mirahsan, R. Schoenen, and H. Yanikomeroglu, "Hethetnets: heterogeneous traffic distribution in heterogeneous wireless cellular networks," *IEEE J. Sel. Areas Commun.,* vol. 33, no. 10, pp. 2252–2265, Oct. 2015.

[87] L. Dai, "Distributed antenna system: Performance analysis in multi-user scenario," in *Proc. IEEE CISS 2008*, pp. 85–89.

[88] L. Dai, "A comparative study on uplink sum capacity with co-located and distributed antennas," *IEEE J. Sel. Areas Commun.,* vol. 29, no. 6, pp 1200–1213, June 2011.

[89] J. Zhang and J. G. Andrews, "Distributed antenna systems with randomness," *IEEE Trans. Wireless Commun.,* vol. 7, no. 9, pp. 3636–3646, Sep. 2008.

[90] K. Son, S. Chong, and G. Veciana, "Dynamic association for load balancing and interference avoidance in multi-cell networks," *IEEE Trans. Wireless Commun.,* vol. 8, no. 7, pp. 3566–3576, Jul. 2009.

[91] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Trans. Wireless Commun.,* vol. 12, no. 6, pp. 2706–2716, Jun. 2013.

[92] K. Shen and W. Yu, "Distributed pricing-based user association for downlink heterogeneous cellular networks," *IEEE J. Sel. Areas Commun.,* vol. 32, no. 6, pp. 1100–1113, Jun. 2014.

[93] N. Prasad, M. Arslan, and S. Rangarajan, "Exploiting cell dormancy and load balancing in LTE HetNets: Optimizing the proportional fairness utility," *IEEE Trans. Commun.,* vol. 62, no. 10, pp. 3706–3722, Oct. 2014.

[94] Y. Lu, H.-W. Hsu, and L.-C. Wang, "Performance model and deployment strategy for mm-wave multi-cellular systems," in *Proc. IEEE WOCC 2016*, pp. 1–4.

[95] S. Szyszkowicz, A. Lou, and H. Yanikomeroglu, "Automated placement of individual millimeter-wave wall-mounted base stations for line-of-sight coverage of outdoor urban areas," *IEEE Wireless Commun. Lett.,* vol. 5, no. 3, pp. 316–319, Jun. 2016.

[96] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications (mmWave) for 5G: opportunities and challenges," *Wireless Netw.,* vol. 21, no. 8, pp. 2657–2676, Nov. 2015.

[97] 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 Trans. Netw.,* vol. 23, no. 3, pp. 836–850, Jun. 2015.

[98] S. Goyal, M. Mezzavilla, S. Rangan, S. Panwar, and M. Zorzi, "User association in 5G mmWave networks," in *Proc. IEEE WCNC 2017*, pp. 1–6.

[99] A. Alizadeh and M. Vu, "Load balancing user association in millimeter wave MIMO networks," *IEEE Trans. Wireless Commun.,* vol. 18, no. 6, pp. 2932–2945, Jun. 2019.

[100] C. Chaieb, Z. Mlika, F. Abdelkefi, and W. Ajib, "On the user association and resource allocation in hetnets with mmWave base stations," in *Proc. IEEE PIMRC 2017*, pp. 1–5.

[101] G. Sierksma and Y. Zwols, *Linear and Integer Programming: Theory and Practice.* Boca Raton: CRC Press, 2015.

[102] S. Boyd, *Branch and Bound Methods.* Standford, CA, USA: Stanford Univ. Press, 2007.

[103] D. Jungnickel, *Graphs, Networks and Algorithms,* 4th ed. Berlin, Germany: Springer-Verlag, 2013.

[104] L. Yan, R. Dodier, M. C. Mozer, and R. Wolniewicz, "Optimizing classifier performance via an approximation to the Wilcoxon-Mann-Whitney statistic," in *Proc. ICML 2003*, pp. 848–855.

[105] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. AISTATS 2011*, pp. 315–323.

[106] E. S. Marquez, J. S. Hare, and M. Niranjan, "Deep cascade learning," *IEEE Trans. Neural Netw. Learn. Syst.,* vol. 29, no. 11, pp. 5475–5485, Nov. 2018.

[107] T. H. Cormen, C E. Leiserson, R. L. Rivest, C. Stein, *Introduction to Algorithms, Third Edition.* Cambridge, MA, USA: MIT Press, 2009.

[108] M. L. Fisher, G. L. Nemhauser, and L. A. Wolsey, "An analysis of approximations for maximizing submodular set functions—II," *Math. Program. Study,* vol. 8, pp. 73–87, 1978.

[109] S. Fujishige, *Submodular Functions and Optimization.* Amsterdam, The Netherlands: Elsevier, 2005.

[110] J. MacQueen, "Some methods for classification and analysis of multivariate observation," in *Proc. Berkeley Symp. Math. Statist. Probability 1967*, pp. 281–297.

[111] Y. Zhu, Q. Zhang, Z. Niu, and J. Zhu, "Leveraging multi-AP diversity for transmission resilience in wireless networks: architecture and performance analysis," *IEEE Trans. Wireless Commun.,* vol. 8, no. 10, pp. 5030–5040, Oct. 2009.

[112] M. Panda and A. Kumar, "Cell-level modeling of IEEE 802.11 WLANs," *Ad Hoc Netw.,* vol. 25, pp. 84–101, Feb. 2015.

[113] *IEEE Standard for Information Technology—Local and Metropolitan Area Networks—Specific Requirements—Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 5: Enhancements for Higher Throughput,* IEEE Standard 802.11n-2009, Oct. 2009.

[114] A. Hills, "Large-scale wireless LAN design," *IEEE Commun. Mag.,* vol. 39, no. 11, pp. 98–104, Nov. 2001.

[115] L. Liao, W. Chen, C. Zhang, L. Zhang, D. Xuan, and W. Jia, "Two birds with one stone: wireless access point deployment for both coverage and localization," *IEEE Trans. Veh. Technol.,* vol. 60, no. 5, pp. 2239–2252, Jun. 2011.

[116] R. J. Weiler, M. Peter, W. Keusgen, K. Sakaguchi, and F. Undi "Environment induced shadowing of urban millimeter-wave access links," *IEEE Wireless Commun. lett.,* vol. 5, no. 4, pp. 440–443, Aug. 2016.

[117] M. Dong and T. Kim, "Interference analysis for millimeter-wave networks with geometry-dependent first-order reflections," *IEEE Trans. Veh. Technol.,* vol. 67, no. 12, pp. 12404–12409, Dec. 2018.

[118] C. E. O'Lone, H. S. Dhillon, and R. M. Buehrer, "Single-anchor localizability in 5G millimeter wave networks," *IEEE Wireless Commun. lett.,* vol. 9, no. 1, pp. 65–69, Jan. 2020.

[119] L. Dai, "An uplink capacity analysis of the distributed antenna system (DAS): From cellular DAS to DAS with virtual cells," *IEEE Trans. Wireless Commun.,* vol. 13, no. 5, pp. 2717–2731, May 2014.

[120] J. Wang and L. Dai, "Downlink rate analysis for virtual-cell based large-scale distributed antenna systems," *IEEE Trans. Wireless Commun.,* vol. 15, no. 3, pp. 1998–2011, Mar. 2016.

[121] A. Valkanis, A. Iossifides, P. Chatzimisios, M. Angelopoulos, and V. Katos, "IEEE 802.11ax spatial reuse improvement: An interference-based channel-access algorithm," *IEEE Veh. Technol. Mag.,* vol. 14, no. 2, pp. 78–84, Jun. 2019.

[122] I. V. Zaliapin, Y. Y. Kagan, and F. P. Schoenberg, "Approximating the distribution of Pareto sums," *Pure Appl. Geophys.,* vol. 162, no. 6, pp. 1187–1228, Jun. 2005.

[123] T. Bai, R. Vaze, and R. W. Heath, Jr., "Analysis of blockage effects on urban cellular networks," *IEEE Trans. Wireless Commun.,* vol. 13, no. 9, pp. 5070–5083, Sep. 2014.

List of Publications

Journal Papers

1. Yue Zhang and Lin Dai, "A Closed-Form Approximation for Uplink Average Ergodic Sum Capacity of Large-Scale Multi-User Distributed Antenna Systems," *IEEE Trans. Veh. Technol.,* vol. 68, no. 2, pp. 1745–1756, Feb. 2019.

2. Yue Zhang and Lin Dai, "On the Optimal Placement of Base Station Antennas for Distributed Antenna Systems," *IEEE Commun. Lett.,* vol. 24, no. 12, pp. 2878–2882, Dec. 2020.

3. Yue Zhang, Lin Dai, and Eric W. M. Wong, "Optimal BS Deployment and User Association for 5G Millimeter Wave Communication Networks," *IEEE Trans. Wireless Commun.*, early access.

Conference Papers

1. Yue Zhang and Lin Dai, "Approximate Ergodic Sum Capacity of Large-Scale Multi-User Distributed Antenna Systems," in *Proc. WCSP 2018*, pp. 1–6.

2. Yue Zhang and Lin Dai, "Joint Optimization of Placement and Coverage of Access Points for IEEE 802.11 Networks," in *Proc. IEEE ICC 2020*, pp. 1–7.