Optimal BS Deployment and User Association for 5G Millimeter Wave Communication Networks

Yue Zhang, *Member, IEEE*, Lin Dai, *Senior Member, IEEE*,
and Eric W. M. Wong, *Senior Member, IEEE*

***Abstract*—Although millimeter wave (mmWave) communications can well support high-data-rate transmissions, the inherent shortcomings, e.g., high path loss and sensitivity to blockage, may cause severe outage problems if the network is not configured properly. This paper aims to minimize the long-term outage probability of an mmWave communication network by optimizing the base station (BS) deployment and user association. For the BS deployment problem, existing works usually assumed that the positions of users are fixed and formulated it as a deterministic optimization problem. With the time-varying nature of positions of user equipments (UEs) taken into account, we establish a stochastic optimization framework for BS deployment optimization. The objective is to maximize the average number of physically accessible BSs of each UE under an inaccessible probability constraint, and a cooperative stochastic approximation (CSA)-based algorithm is developed to effectively search the optimal positions of BSs. For user association, our focus is to properly associate UEs with BSs to minimize the outage probability with balanced workloads among BSs. Combined with the proposed user association scheme, the proposed BS deployment scheme can significantly improve the network outage probability in the long term, especially when the aggregation degree of UEs is large.**

***Index Terms*—Millimeter wave (mmWave) networks, base-station (BS) deployment, user association, outage probability, stochastic optimization.**

I. Introduction

To meet the booming data traffic demand, the millimeter wave (mmWave) communication networks, which adopt the mmWave bands from 30 to 300 GHz, have become an indispensable part of the Fifth Generation (5G) communication systems [1]–[3]. Despite abundant spectrum resources, however, mmWave transmissions suffer from severe signal attenuation and are susceptible to blockage. As a result, the coverage areas of mmWave base stations (BSs) usually have irregular shapes and are closely determined by the layout of obstacles, which necessitates delicate BS deployment schemes for mmWave networks [4], [5]. Moreover, to combat the high path loss, mmWave BSs are usually densely deployed, with which the connectivity between users and mmWave BSs would change quickly as users move around. How to properly associate users with BSs is, therefore, another critical problem for mmWave networks [6], [7].

BS deployment and user association are two important issues in wireless communication systems and have been extensively studied in the past decades [8]–[15]. In traditional cellular systems, 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 [8]–[11]. However, such distance-based BS deployment 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 deployment 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 results in significant reduction of co-channel interference. Consequently, user association metrics for interference-limited wireless networks [12]–[15] are not well suited to noise-limited mmWave networks. New user association approaches should be customized for mmWave networks without considering interference coordination.

A. Literature Review

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

1) BS Deployment: The main challenge of mmWave BS deployment originates from the severe blockage effect, which requires the consideration of the layout of obstacles when deploying 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 [16] or connectivity probability [17] of a typical user. Based on the system model in [17], the optimal density of BSs for minimizing BS deployment cost under a certain connectivity probability constraint was derived in [18].

To further determine the optimal positions of BSs, a static blockage model was adopted in [19]–[22], where the positions, shapes and orientations of obstacles are fixed. To maximize the line-of-sight (LoS) coverage of the mmWave network, various assumptions on the feasible positions of BSs were made in [19], [20]. In [19], the region under consideration was divided into quadrilateral or triangle areas, and the mmWave BSs were

Manuscript received March 13, 2020; revised September 30, 2020; accepted November 29, 2020. This work was supported in part by the Research Grants Council (RGC) of Hong Kong under GRF Grant CityU 11210219, and in part by the Innovation and Technology Fund of the Hong Kong Special Administrative Region, China under Grant ITS/359/17. The associate editor coordinating the review of this paper and approving it for publication was N. Prasad. (*Corresponding author: Yue Zhang.*)

The authors are with Department of Electrical Engineering, City University of Hong Kong, Hong Kong (E-mail: yzhang742-c@my.cityu.edu.hk; lindai@cityu.edu.hk; eeewong@cityu.edu.hk).

deployed at the intersection points of areas. Different from [19] where mmWave BSs were placed in the open space, in [20], 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 [19], [20] only focused on maximizing the coverage area, the proposed symmetric BS deployments may lead to unbalanced workloads among BSs if users' positions are asymmetric. The user-position-dependent BS deployment schemes were further studied in [21] and [22], where users are represented by reference points (RPs) with given positions. By generating the candidate sites (CSs) of BSs based on the BS deployment scheme proposed in [20], the optimal subsets of CSs were searched for covering the maximum number of RPs [21], or satisfying the received signal strength requirement at each RP with the minimum number of BSs [22].

Note that in [21] and [22], the BS deployment 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 deployment 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 our recent work [23], with the time-varying nature of users' positions taken into account, a stochastic optimization algorithm was proposed for Wi-Fi networks to optimize the positions and coverage radius of access points, and shown to significantly improve the average network throughput. In this paper, we will further establish a stochastic optimization framework to optimize the BS deployment for mmWave networks. Different from [23], where the objective is to optimize the throughput performance of Wi-Fi networks, here we focus on optimizing the physical accessibility between user equipments (UEs) and BSs, which is important for mmWave networks where the blockage effect is significant.

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 [24], that is, each user associates with the access point 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 [25]–[28], where the workload of each BS, e.g., the number of users that can be served by one BS, is strictly bounded. In [25], for instance, a distributed algorithm was proposed to minimize the maximum workload among the access points. In [26], user association and spectrum resource allocation are jointly optimized for maximizing the downlink sum-rate and minimizing the re-allocation cost of handovers simultaneously. In [27], 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 [28], 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 [16]–[18], which nevertheless, has seldom been considered when optimizing the user association scheme. In this paper, by considering the maximum workload constraint, we propose a low-complexity user association scheme to minimize the outage probability of an mmWave communication network.

B. Our Contributions

In this paper, we focus on the optimization of outage performance of an mmWave communication network with the Manhattan-type geometry by properly designing the BS deployment scheme and user association scheme. For the BS deployment problem, we aim at maximizing the long-term average number of physically accessible BSs of each UE under a constraint of inaccessible probability, i.e., the long-term proportion of UEs which have no physically accessible BS. The problem is formulated as a stochastic optimization problem by taking into account the time-varying nature of UEs' positions. As both the objective and constraint functions are in the form of expectation, which can not be handled by traditional stochastic approximation (SA) algorithms [29]–[31], we propose a novel algorithm based on the recently proposed cooperative SA (CSA) [32] to effectively search the optimal deployment of BSs.

With the optimized positions of BSs, we then 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. Based on the virtual BS splitting technique proposed in [15], the user association problem can be reformulated as a bipartite maximum weight matching problem, which can be optimally solved using the Hungarian algorithm [33]. As the complexity of the Hungarian algorithm is high when the number of BSs or number of UEs is large, we further propose a low-complexity outage mitigation user association scheme. Simulation results corroborate that the proposed scheme can achieve similar outage performance to the Hungarian-algorithm-based optimal user association scheme, but with much less running time. The outage performance of the proposed user association scheme combined with the proposed BS deployment scheme is also examined under various aggregation degrees of UEs, and significant gains are demonstrated over the previous representative BS deployment scheme in [21].

The rest of this paper is organized as below. Section II presents the system model and formulates the BS deployment problem and user association problem. In Section III, a CSA-based optimal BS deployment scheme is proposed to solve the BS deployment problem. The user association issue is addressed in Section IV, where a low-complexity outage mitigation user association scheme is proposed. Simulation results are presented in Section V to illustrate the performance

$$\begin{aligned}\mathcal{I}_{m,n}^{\text{block}} =& I\Big(\sum_{b=1}^{B} I\big((d_{m,n}\sin\theta_{m,n}+E_{\text{bui}})(x_{m,n,b}\sin\theta_{m,n}-y_{m,n,b}\cos\theta_{m,n}+\tfrac{d_{m,n}\sin\theta_{m,n}}{2}+\tfrac{E_{\text{bui}}}{2})\big)\\ &\cdot I\big((d_{m,n}\cos\theta_{m,n}+E_{\text{bui}})(x_{m,n,b}\cos\theta_{m,n}+y_{m,n,b}\sin\theta_{m,n}+\tfrac{d_{m,n}\cos\theta_{m,n}}{2}+\tfrac{E_{\text{bui}}}{2})\big)\\ &\cdot I\big((\sin\theta_{m,n}+\cos\theta_{m,n})(y_{m,n,b}+\tfrac{E_{\text{bui}}}{2}\sin\theta_{m,n}+\tfrac{E_{\text{bui}}}{2}\cos\theta_{m,n})\big)\\ &\cdot I\big((d_{m,n}\sin\theta_{m,n}+E_{\text{bui}})(-x_{m,n,b}\sin\theta_{m,n}+y_{m,n,b}\cos\theta_{m,n}+\tfrac{d_{m,n}\sin\theta_{m,n}}{2}+\tfrac{E_{\text{bui}}}{2})\big)\\ &\cdot I\big((d_{m,n}\cos\theta_{m,n}+E_{\text{bui}})(-x_{m,n,b}\cos\theta_{m,n}-y_{m,n,b}\sin\theta_{m,n}+\tfrac{d_{m,n}\cos\theta_{m,n}}{2}+\tfrac{E_{\text{bui}}}{2})\big)\\ &\cdot I\big((\sin\theta_{m,n}+\cos\theta_{m,n})(-y_{m,n,b}+\tfrac{E_{\text{bui}}}{2}\sin\theta_{m,n}+\tfrac{E_{\text{bui}}}{2}\cos\theta_{m,n})\big)\Big).\end{aligned} \tag{5}$$

of the proposed BS deployment scheme and user association scheme. Finally, Section VI concludes this paper.

Notation: Throughout this paper, $\mathbb{E}[\cdot]$ denotes the expectation operator. $\|\mathbf{x}\|$ denotes the Euclidean norm of vector $\mathbf{x}$. $|\mathcal{K}|$ denotes the cardinality of set $\mathcal{K}$.

II. System Model and Problem Formulation

We consider an mmWave network with multiple BSs serving for a group of UEs, denoted by $\mathcal{M}$ and $\mathcal{N}$, respectively, in a 2-dimensional Manhattan-type geometry [21]. One example is shown in Fig. 1, where multiple square blocks (i.e., buildings) with edge length E_{bui} are separated by streets with width W_{str}.[1] The numbers of BSs, UEs and blocks are given by M, N and B, respectively. We consider the Cartesian coordinate and denote the positions of UE n, BS m and the center of block b as $\mathbf{r}_n^{\text{UE}} = [x_n^{\text{UE}}, y_n^{\text{UE}}], n = 1, \cdots, N$, $\mathbf{r}_m^{\text{BS}} = [x_m^{\text{BS}}, y_m^{\text{BS}}], m = 1, \cdots, M$, 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 C UEs, which is referred to as the *maximum workload*. And each UE can only be served by one BS.

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

A. BS Deployment Problem

For the BS deployment problem, we are concerned with how to properly place the BSs to optimize the physical accessibility. The physical accessibility between a BS and a UE is determined by two factors: coverage and blockage. Let us define two Boolean indicators $\mathcal{I}_{m,n}^{\text{out}}$ and $\mathcal{I}_{m,n}^{\text{block}}$ as

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

[1]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.

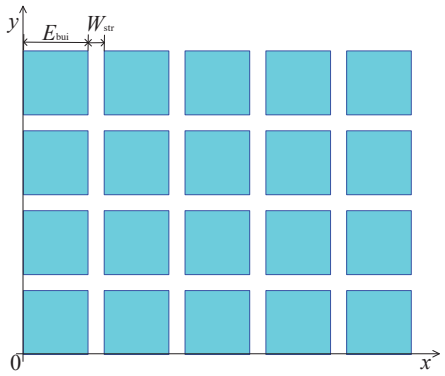


Fig. 1: 2-dimensional Manhattan-type geometry.

$$\mathcal{I}_{m,n}^{\text{block}} = \begin{cases} 1 & \text{if there is no LoS path between} \\ & \text{BS } m \text{ and UE } n \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

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

$$I(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

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

$$\mathcal{I}_{m,n}^{\text{out}} = I(d_{m,n} - R), \tag{4}$$

where R is the coverage radius of each BS and $d_{m,n}$ denotes the distance between BS m and UE n. For $\mathcal{I}_{m,n}^{\text{block}}$, it is shown in Appendix A that it can be derived as (5) at the top of this page, where $x_{m,n,b}$, $y_{m,n,b}$ and $\theta_{m,n}$ are given by

$$\begin{aligned} x_{m,n,b} =& \left(x_b^{\text{block}} - \tfrac{x_n^{\text{UE}}+x_m^{\text{BS}}}{2}\right)\cos\arctan\tfrac{y_n^{\text{UE}}-y_m^{\text{BS}}}{x_n^{\text{UE}}-x_m^{\text{BS}}} \\ &+ \left(y_b^{\text{block}} - \tfrac{y_n^{\text{UE}}+y_m^{\text{BS}}}{2}\right)\sin\arctan\tfrac{y_n^{\text{UE}}-y_m^{\text{BS}}}{x_n^{\text{UE}}-x_m^{\text{BS}}}, \end{aligned} \tag{6}$$

$$\begin{aligned} y_{m,n,b} =& \left(y_b^{\text{block}} - \tfrac{y_n^{\text{UE}}+y_m^{\text{BS}}}{2}\right)\cos\arctan\tfrac{y_n^{\text{UE}}-y_m^{\text{BS}}}{x_n^{\text{UE}}-x_m^{\text{BS}}} \\ &- \left(x_b^{\text{block}} - \tfrac{x_n^{\text{UE}}+x_m^{\text{BS}}}{2}\right)\sin\arctan\tfrac{y_n^{\text{UE}}-y_m^{\text{BS}}}{x_n^{\text{UE}}-x_m^{\text{BS}}}, \end{aligned} \tag{7}$$

and

$$\theta_{m,n} = -\arctan\tfrac{y_n^{\text{UE}}-y_m^{\text{BS}}}{x_n^{\text{UE}}-x_m^{\text{BS}}} + \tfrac{\pi}{2} I\left(\arctan\tfrac{y_n^{\text{UE}}-y_m^{\text{BS}}}{x_n^{\text{UE}}-x_m^{\text{BS}}}\right), \tag{8}$$

respectively.

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

$$\bar{M}_{\text{ac}} = \frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{M}\left(1 - I\left(\mathcal{I}_{m,n}^{\text{out}} + \mathcal{I}_{m,n}^{\text{block}}\right)\right) \tag{9}$$

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

$$P_{\rm iac} = \frac{1}{N}\sum_{n=1}^{N}\prod_{m=1}^{M} I\left(\mathcal{I}_{m,n}^{\rm out} + \mathcal{I}_{m,n}^{\rm block}\right) \tag{10}$$

as the *inaccessible probability*. As both $\mathcal{I}_{m,n}^{\rm out}$ and $\mathcal{I}_{m,n}^{\rm block}$ are functions of the positions of BSs $\{\mathbf{r}_m^{\rm BS}\}$ and positions of UEs $\{\mathbf{r}_n^{\rm UE}\}$ according to (4)–(5), the average number of accessible BSs $\bar{M}_{\rm ac}$ and inaccessible probability $P_{\rm iac}$ are also crucially dependent on the positions of UEs and BSs. In practice, the positions of UEs are time-varying due to the mobility. We are therefore interested in the long-term average performance, i.e., $\mathbb{E}_{\{\mathbf{r}_n^{\rm UE}\}}[\bar{M}_{\rm ac}]$ and $\mathbb{E}_{\{\mathbf{r}_n^{\rm UE}\}}[P_{\rm iac}]$. Our objective is to optimize the positions of BSs for maximizing the long-term average number of accessible BSs $\mathbb{E}_{\{\mathbf{r}_n^{\rm UE}\}}[\bar{M}_{\rm ac}]$, under the constraint that the long-term average inaccessible probability $\mathbb{E}_{\{\mathbf{r}_n^{\rm UE}\}}[P_{\rm iac}]$ is bounded by $P_{\rm iac}^*$.

Finally, we can formulate the BS deployment problem as

$$\begin{aligned}
({\rm P1}):\ \max_{\{\mathbf{r}_m^{\rm BS}\}}\ & \mathbb{E}_{\{\mathbf{r}_n^{\rm UE}\}}[\bar{M}_{\rm ac}] && (11)\\
{\rm s.t.}\ & \mathbb{E}_{\{\mathbf{r}_n^{\rm UE}\}}[P_{\rm iac}] \le P_{\rm iac}^*, && (12)\\
& \mathbf{r}_m^{\rm BS} \in \mathcal{A}, m = 1, \cdots, M, && (13)
\end{aligned}$$

where $\mathcal{A}$ denotes the area in which M BSs and N UEs are located.

Problem P1 is a stochastic optimization problem with both objective function and constraint function in the form of expectation. One possible method to solve it is the penalty-based or primal-dual approach [34], which can turn Problem P1 into a stochastic optimization problem without expectation constraint. However, the performance of this method relies heavily on the value of penalty parameter or dual multiplier, which is difficult to derive in stochastic optimization problems. Another alternative method is the sample-average approximation (SAA) technique [35], which uses sample-averages to replace the expectations and turns the problem into a deterministic one. However, the computational complexity of the SAA method is high when the number of samples is large. Recently, a low-complexity algorithm called CSA was proposed in [32] to solve expectation-constrained stochastic optimization problems. In Section III, a CSA-based algorithm will be proposed to solve Problem P1.

B. User Association Problem

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

$$\mathcal{I}_{m,n}^{\rm associate} = \begin{cases} 1 & \text{if UE } n \text{ associates with BS } m \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

Note that UE n can associate with BS m only if BS m is physically accessible for UE n. We then have $\mathcal{I}_{m,n}^{\rm associate} \le 1 - I\left(\mathcal{I}_{m,n}^{\rm out} + \mathcal{I}_{m,n}^{\rm block}\right)$. Moreover, each BS can serve at most C UEs, which can be written as $\sum_{n=1}^{N}\mathcal{I}_{m,n}^{\rm associate} \le C, m = 1, \cdots, M$. For each UE, it can only associate with one BS, which indicates $\sum_{m=1}^{M}\mathcal{I}_{m,n}^{\rm associate} \le 1, n = 1, \cdots, N$. We aim at minimizing the outage probability, i.e., the proportion of UEs that cannot associate with any BSs, which can be written as

$$P_o = 1 - \frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{M}\mathcal{I}_{m,n}^{\rm associate}. \tag{15}$$

Finally, we can formulate the user association problem as

$$\begin{aligned}
({\rm P2}):\ \min_{\{\mathcal{I}_{m,n}^{\rm associate}\}}\ & P_o && (16)\\
{\rm s.t.}\ & \mathcal{I}_{m,n}^{\rm associate} \le 1 - I\left(\mathcal{I}_{m,n}^{\rm out} + \mathcal{I}_{m,n}^{\rm block}\right), && (17)\\
& \sum_{n=1}^{N}\mathcal{I}_{m,n}^{\rm associate} \le C, m = 1, \cdots, M, && (18)\\
& \sum_{m=1}^{M}\mathcal{I}_{m,n}^{\rm associate} \le 1, n = 1, \cdots, N, && (19)\\
& \mathcal{I}_{m,n}^{\rm associate} \in \{0, 1\}. && (20)
\end{aligned}$$

Problem P2 is a deterministic integer linear programming problem, which can be optimally solved by the Branch-and-Bound (B&B) method [36]. However, the B&B method has been shown with exponential time complexity in the worst case [36]. Based on the virtual BS splitting technique recently proposed in [15], it will be demonstrated in Section IV that P2 can be reformulated as a bipartite maximum weight matching problem, which can be optimally solved using the Hungarian algorithm in polynomial time [33]. In view of the high complexity of Hungarian algorithm when the number of BSs or UEs is large, a low-complexity user association algorithm will further be proposed to solve Problem P2.

III. BS Deployment Optimization

In this section, we will propose a CSA-based algorithm to solve Problem P1. Let us start from introducing the main steps of the CSA algorithm.

A. CSA Algorithm

Consider a stochastic optimization problem in the following form:

$$\begin{aligned}
({\rm P3}):\ \min_{\mathbf{x}}\ & f(\mathbf{x}) = \mathbb{E}_{\mathbf{y}}[F(\mathbf{x}, \mathbf{y})]\\
{\rm s.t.}\ & g(\mathbf{x}) = \mathbb{E}_{\mathbf{y}}[G(\mathbf{x}, \mathbf{y})] \le 0,\\
& \mathbf{x} \in \mathcal{X},
\end{aligned}$$

where $F, G : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ are continuous and differentiable functions of $\mathbf{x}$, and $\mathbf{y}$ is a random variable supported on a set $\mathcal{Y}$. The CSA algorithm solves Problem P3 by applying the following update [32]:

$$\mathbf{x}(t+1) = \Pi_{\mathcal{X}}\big(\mathbf{x}(t) - \tilde{\boldsymbol{\omega}}(t)\eta(t)\big), \tag{21}$$

where

$$\tilde{\boldsymbol{\omega}}(t) = \begin{cases} \tilde{\boldsymbol{\omega}}_F(t) & \text{if } \hat{G}(t) \le 0 \\ \tilde{\boldsymbol{\omega}}_G(t) & \text{otherwise} \end{cases} \tag{22}$$

with $\tilde{\boldsymbol{\omega}}_F(t)$ and $\tilde{\boldsymbol{\omega}}_G(t)$ denoting the gradients of $F(\mathbf{x}(t), \mathbf{y}(t))$ and $G(\mathbf{x}(t), \mathbf{y}(t))$, respectively, which are calculated based on a sample $\mathbf{y}(t)$ of the random variable $\mathbf{y}$. $\Pi_{\mathcal{X}}(\mathbf{x})$ denotes a Euclidean projection of a vector $\mathbf{x}$ on a set $\mathcal{X}$, i.e.,

$$\Pi_{\mathcal{X}}(\mathbf{x}) = \arg\min_{\mathbf{x}'}\{\|\mathbf{x} - \mathbf{x}'\| \big| \mathbf{x}' \in \mathcal{X}\}. \tag{23}$$

$\eta(t) > 0$ denotes the step size. $\hat{G}(t)$ is an unbiased estimation of $\mathbb{E}_{\mathbf{y}}[G(\mathbf{x}, \mathbf{y})]$. One way to obtain $\hat{G}(t)$ is

to generate J independent and identically distributed samples of $\mathbf{y}$, $\{\mathbf{y}_1, \cdots, \mathbf{y}_J\}$, and evaluate $\hat{G}(t)$ by $\hat{G}(t) = \frac{1}{J}\sum_{j=1}^{J} G(\mathbf{x}(t), \mathbf{y}_j)$. After T iterations, the output of CSA is

$$\bar{\mathbf{x}}_{T,s} = \left(\sum_{t\in\mathcal{C}} \eta(t)\right)^{-1} \left(\sum_{t\in\mathcal{C}} \eta(t)\mathbf{x}(t)\right), \tag{24}$$

where $\mathcal{C} = \{s \le t \le T | \hat{G}(t) \le 0\}$ for some $1 \le s \le T$.

We can see from (22) and (23) that the decision variable $\mathbf{x}$ is updated along either the gradient of the objective function or the gradient of the constraint function, depending on whether the estimated constraint is satisfied. By doing so, CSA can minimize the objective function, as well as control the violation of the expectation constraint.

It was shown in [32] that for convex objective function $f(\mathbf{x})$ and convex constraint function $g(\mathbf{x})$ with bounded expectations $\mathbb{E}_{\mathbf{y}}[\exp(\|\tilde{\boldsymbol{\omega}}_F(\mathbf{x},\mathbf{y})\|^2)]$, $\mathbb{E}_{\mathbf{y}}[\exp(\|\tilde{\boldsymbol{\omega}}_G(\mathbf{x},\mathbf{y})\|^2)]$ and $\mathbb{E}_{\mathbf{y}}[\exp(\|G(\mathbf{x},\mathbf{y}) - g(\mathbf{x})\|^2)]$ and a constant step size $\eta(t)$, the CSA algorithm exhibits the optimal $\mathcal{O}(1/\xi^2)$ rate of convergence in terms of both optimality gap and constraint violation ξ. Specifically, if the optimal solution $\mathbf{x}^*$ of P3 exists, then for any $\Lambda \in (0,1)$, we have $\text{Prob}\{f(\bar{\mathbf{x}}_{T,s}) - f(\mathbf{x}^*) \le \epsilon\} \ge (1-\Lambda)^2$ and $\text{Prob}\{g(\bar{\mathbf{x}}_{T,s}) \le \vartheta\} \ge 1-\Lambda$ with the total number of iterations T and the number of samples J bounded by $\mathcal{O}(\max\{\frac{1}{\epsilon^2}(\log\frac{1}{\Lambda})^2, \frac{1}{\vartheta^2}\})$ and $\mathcal{O}(\max\{\frac{1}{\epsilon^2}(\log\frac{1}{\Lambda})^2, \frac{1}{\vartheta^2}\log\frac{1}{\Lambda^3}\})$, respectively.

B. CSA-based BS Deployment Algorithm

Based on CSA, we can solve Problem P1 by letting $\mathbf{x} = \mathbf{r}^{\text{BS}}$, $\mathbf{y} = \mathbf{r}^{\text{UE}}$ and $\mathcal{X} = \mathcal{A}^M$, where $\mathbf{r}^{\text{BS}} = [\mathbf{r}_1^{\text{BS}}, \cdots, \mathbf{r}_M^{\text{BS}}]$ and $\mathbf{r}^{\text{UE}} = [\mathbf{r}_1^{\text{UE}}, \cdots, \mathbf{r}_N^{\text{UE}}]$. Then (21) becomes

$$\mathbf{r}^{\text{BS}}(t+1) = \Pi_{\mathcal{A}^M}\left(\mathbf{r}^{\text{BS}}(t) - \tilde{\boldsymbol{\omega}}_{\text{BS}}(t)\eta(t)\right), \tag{25}$$

where

$$\tilde{\boldsymbol{\omega}}_{\text{BS}}(t) = \begin{cases} -\tilde{\boldsymbol{\omega}}_{\bar{M}_{\text{ac}}}(t) & \text{if } \hat{P}_{\text{iac}}(t) - P^*_{\text{iac}} \le 0 \\ \tilde{\boldsymbol{\omega}}_{P_{\text{iac}}}(t) & \text{otherwise,} \end{cases} \tag{26}$$

with $\tilde{\boldsymbol{\omega}}_{\bar{M}_{\text{ac}}}(t)$ and $\tilde{\boldsymbol{\omega}}_{P_{\text{iac}}}(t)$ denoting the gradients of $\bar{M}_{\text{ac}}(t)$ and $P_{\text{iac}}(t)$ in terms of $\mathbf{r}^{\text{BS}}(t)$, respectively. $\hat{P}_{\text{iac}}(t)$ is an unbiased estimation of $\mathbb{E}_{\mathbf{r}^{\text{UE}}}[P_{\text{iac}}(t)]$ generated with 100 realizations of UEs' positions. Note that there is a negative sign before $\tilde{\boldsymbol{\omega}}_{\bar{M}_{\text{ac}}}(t)$ because Problem P1 is a maximization problem.

From (9) and (10), we can see that the expressions of $\bar{M}_{\text{ac}}(t)$ and $P_{\text{iac}}(t)$ involve the indicator function $I(x)$ given in (3), whose derivative is undefined at $x = 0$. To make $\bar{M}_{\text{ac}}(t)$ and $P_{\text{iac}}(t)$ differentiable, we approximate $I(x)$ by a sigmoid function [37]:

$$I(x) \approx S(x) = \frac{1}{1+\exp(-\beta x)}. \tag{27}$$

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

From (4) and (5), we can see that both the expressions of $\mathcal{I}^{\text{out}}_{m,n}$ and $\mathcal{I}^{\text{block}}_{m,n}$ include the indicator function. Therefore, the term $I\left(\mathcal{I}^{\text{out}}_{m,n} + \mathcal{I}^{\text{block}}_{m,n}\right)$ in (9) and (10) has a multi-layer structure of indicator functions. When calculating the gradient $\tilde{\boldsymbol{\omega}}_{\text{BS}}(t)$ by approximating the indicator function by the sigmoid function, such a multi-layer structure will cause the vanishing gradient problem.[2] 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 $I\left(\mathcal{I}^{\text{out}}_{m,n} + \mathcal{I}^{\text{block}}_{m,n}\right)$ and rewrite $\bar{M}_{\text{ac}}$ and P_{iac} as

$$\bar{M}_{\text{ac}} = M - \frac{1}{N}\sum_{n=1}^{N}\left(\sum_{m=1}^{M}\mathcal{I}^{\text{out}}_{m,n} + \sum_{m\in\mathcal{M}_n}\mathcal{I}^{\text{block}}_{m,n}\right) \tag{28}$$

and

$$P_{\text{iac}} = \frac{1}{N}\sum_{n=1}^{N}\left(\prod_{m=1}^{M}\mathcal{I}^{\text{out}}_{m,n} + \prod_{m\in\mathcal{M}_n}\mathcal{I}^{\text{block}}_{m,n}\right), \tag{29}$$

respectively, where $\mathcal{M}_n = \{m \in \mathcal{M} : d_{m,n} \le R\}$ denotes the set of BSs that include UE n in their coverage areas. The detailed derivations of (28) and (29) are provided in Appendix B.

With (28) and (29), we can obtain the gradient $\tilde{\boldsymbol{\omega}}_{\text{BS}}(t)$ as

$$\tilde{\boldsymbol{\omega}}_{\text{BS}}(t) = \tilde{\boldsymbol{\omega}}^{\text{out}}_{\text{BS}}(t) + \tilde{\boldsymbol{\omega}}^{\text{block}}_{\text{BS}}(t), \tag{30}$$

where

$$\tilde{\boldsymbol{\omega}}^{\text{out}}_{\text{BS}}(t) = \begin{cases} \frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{M}\frac{\partial\mathcal{I}^{\text{out}}_{m,n}(t)}{\partial\mathbf{r}^{\text{BS}}(t)} & \text{if } \hat{P}_{\text{iac}}(t) - P^*_{\text{iac}} \le 0 \\ \frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{M} & \\ \prod_{i=1,i\neq m}^{M}\mathcal{I}^{\text{out}}_{i,n}(t)\frac{\partial\mathcal{I}^{\text{out}}_{m,n}(t)}{\partial\mathbf{r}^{\text{BS}}(t)} & \text{otherwise} \end{cases} \tag{31}$$

and

$$\tilde{\boldsymbol{\omega}}^{\text{block}}_{\text{BS}}(t) = \begin{cases} \frac{1}{N}\sum_{n=1}^{N}\sum_{m\in\mathcal{M}_n}\frac{\partial\mathcal{I}^{\text{block}}_{m,n}(t)}{\partial\mathbf{r}^{\text{BS}}(t)} & \text{if } \hat{P}_{\text{iac}}(t) - P^*_{\text{iac}} \le 0 \\ \frac{1}{N}\sum_{n=1}^{N}\sum_{m\in\mathcal{M}_n} & \\ \prod_{i\in\mathcal{M}_n,i\neq m}\mathcal{I}^{\text{block}}_{i,n}(t)\frac{\partial\mathcal{I}^{\text{block}}_{m,n}(t)}{\partial\mathbf{r}^{\text{BS}}(t)} & \text{otherwise.} \end{cases} \tag{32}$$

$\frac{\partial\mathcal{I}^{\text{out}}_{m,n}(t)}{\partial\mathbf{r}^{\text{BS}}(t)}$ and $\frac{\partial\mathcal{I}^{\text{block}}_{m,n}(t)}{\partial\mathbf{r}^{\text{BS}}(t)}$ can be calculated by combining (54)–(57) and (58)–(65), respectively, with detailed derivations given in Appendix C.

From (4) and (5), we can see that the inputs of the indicator function in $\mathcal{I}^{\text{out}}_{m,n}$ and $\mathcal{I}^{\text{block}}_{m,n}$ have different orders of magnitude. In $\mathcal{I}^{\text{out}}_{m,n}$, the input of the indicator function $I(x)$ is $d_{m,n} - R$. While in $\mathcal{I}^{\text{block}}_{m,n}$, some of the inputs of $I(x)$ include the term $d^2_{m,n}$, which is of a higher order of magnitude. Therefore, when replacing the indicator function $I(x)$ with the sigmoid function $S(x)$ in $\frac{\partial\mathcal{I}^{\text{out}}_{m,n}(t)}{\partial\mathbf{r}^{\text{BS}}(t)}$ and $\frac{\partial\mathcal{I}^{\text{block}}_{m,n}(t)}{\partial\mathbf{r}^{\text{BS}}(t)}$, we should use a smaller value of parameter β for $\frac{\partial\mathcal{I}^{\text{block}}_{m,n}(t)}{\partial\mathbf{r}^{\text{BS}}(t)}$ 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 $\tilde{\boldsymbol{\omega}}_{\text{BS}}(t)$, the parameter β of the sigmoid function (27) is set as 1 for $\tilde{\boldsymbol{\omega}}^{\text{out}}_{\text{BS}}(t)$ and 0.001 for $\tilde{\boldsymbol{\omega}}^{\text{block}}_{\text{BS}}(t)$.

[2]More specifically, as the sigmoid function has a large saturated region where the first derivative is negligibly small, if the approximations of $\bar{M}_{\text{ac}}(t)$ 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 $\tilde{\boldsymbol{\omega}}_{\text{BS}}(t)$, 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 [38], [39].

Algorithm 1 CSA-based Optimal BS Deployment Scheme

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

for $t = 0, \cdots, T-1$ **do**
  Compute the stochastic gradient $\tilde{\boldsymbol{\omega}}_{\mathrm{BS}}(t)$ based on (30) and the step sizes $\eta^{\mathrm{out}}(t)$ and $\eta^{\mathrm{block}}(t)$ based on (34) and (35) with a sample of UEs' positions $\mathbf{r}^{\mathrm{UE}}(t)$.
  Update the positions of BSs $\mathbf{r}^{\mathrm{BS}}(t+1)$ based on (33).
  if $\hat{P}_{\mathrm{iac}}(t+1) \leq P^*_{\mathrm{iac}}$ **then**
    $\mathbf{r}^{\mathrm{BS}*} \leftarrow \mathbf{r}^{\mathrm{BS}}(t+1)$.
  end if
end for

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

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

$$\mathbf{r}^{\mathrm{BS}}(t+1) = \Pi_{\mathcal{A}^M}\Big(\mathbf{r}^{\mathrm{BS}}(t) - \big(\tilde{\boldsymbol{\omega}}^{\mathrm{out}}_{\mathrm{BS}}(t)\eta^{\mathrm{out}}(t) + \tilde{\boldsymbol{\omega}}^{\mathrm{block}}_{\mathrm{BS}}(t)\eta^{\mathrm{block}}(t)\big)\Big). \tag{33}$$

In particular, we adopt the step size model proposed in [31] and further normalize it by the norm of the gradient as

$$\eta^{\mathrm{out}}(t) = \frac{\kappa}{t^b \|\tilde{\boldsymbol{\omega}}^{\mathrm{out}}_{\mathrm{BS}}(t)\|}, \tag{34}$$

and

$$\eta^{\mathrm{block}}(t) = \frac{\kappa}{t^b \|\tilde{\boldsymbol{\omega}}^{\mathrm{block}}_{\mathrm{BS}}(t)\|}, \tag{35}$$

where $\kappa \in (0, 100]$ 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 [30]. 2) We normalize $\tilde{\boldsymbol{\omega}}^{\mathrm{out}}_{\mathrm{BS}}(t)$ and $\tilde{\boldsymbol{\omega}}^{\mathrm{block}}_{\mathrm{BS}}(t)$ separately to ensure that $\tilde{\boldsymbol{\omega}}^{\mathrm{out}}_{\mathrm{BS}}(t)\eta^{\mathrm{out}}(t)$ and $\tilde{\boldsymbol{\omega}}^{\mathrm{block}}_{\mathrm{BS}}(t)\eta^{\mathrm{block}}(t)$ have identical scale, and hence balance the effects of the movements of BSs on the coverage and blockage performance of the network. κ mainly determines the step size in the first few iterations while b determines the asymptotic diminishing rate. The algorithm may converge quickly to a sub-optimal solution with a small κ or an over-sized b. While large κ or small b may cause instability of the algorithm. In the simulation parts, the values of κ 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 $\mathcal{C}$, and the convergence results were established based on a constant step size and $\mathcal{C}$ with the maximum size, i.e., $\mathcal{C} = \{1 \leq t \leq T | \hat{G}(t) \leq 0\}$ [32]. In our case, we select the diminishing step sizes (34) and (35) to improve the convergence performance of the algorithm. As we will show in Section V, with these step size models, the proposed algorithm can quickly converge even with $|\mathcal{C}| = 1$. Therefore, we only output the final solution that satisfies the constraint.

Now we can summarize our CSA-based BS deployment algorithm for solving Problem P1 in Algorithm 1. It should be noted that as $\bar{M}_{\mathrm{ac}}$ and P_{iac} are not convex functions of $\mathbf{r}^{\mathrm{BS}}$, Algorithm 1 may be trapped into a sub-optimal stationary point. Whether this undesirable situation occurs or not is crucially determined by the initial positions of BSs $\mathbf{r}^{\mathrm{BS}}(0)$. In Section V, we will explore the effect of $\mathbf{r}^{\mathrm{BS}}(0)$ on the convergence performance of Algorithm 1 via simulations.

IV. User Association Optimization

In this section, we first show that by using the virtual BS splitting technique in [15], Problem P2 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 P2.

A. Optimal User Association Scheme

Let us split each BS into C virtual BSs and denote the set of virtual BSs as $\mathcal{M}_{\mathrm{V}} = \{\mathsf{m}_1, \cdots, \mathsf{m}_{MC}\}$, where m_q is the q-th virtual BS which is split from BS m if $q \in [C(m-1)+1, Cm]$, $q = 1, \cdots, MC$. Let $\mathcal{N} = \{\mathsf{n}_1, \cdots, \mathsf{n}_N\}$ denote the set of UEs. The mmWave network can be then modelled as a bipartite graph $\mathcal{G}(\mathcal{M}_{\mathrm{V}} \cup \mathcal{N}, \underline{\mathcal{E}})$, where $\underline{\mathcal{E}} = \{(\mathsf{m}_q, \mathsf{n}_n) : \mathsf{m}_q \in \mathcal{M}_{\mathrm{V}}, \mathsf{n}_n \in \mathcal{N}\}$ is the edge set. Denote the weight matrix of the edge set $\underline{\mathcal{E}}$ as $\mathbf{W} \in \mathbb{R}^{MC \times N}$, whose (q,n)-th element is given by

$$W_{q,n} = 1 - I\left(\mathcal{I}^{\mathrm{out}}_{m,n} + \mathcal{I}^{\mathrm{block}}_{m,n}\right). \tag{36}$$

(36) indicates that an edge $(\mathsf{m}_q, \mathsf{n}_n)$ has weight one if and only if UE n can physically access BS m. Otherwise, the edge weight is zero. Define $\underline{\mathcal{E}}^{(q)} = \{(\mathsf{m}_q, \mathsf{n}_n) : \mathsf{n}_n \in \mathcal{N}\}$ as the set of edges between virtual BS q and all UEs, and $\underline{\mathcal{E}}_{(n)} = \{(\mathsf{m}_q, \mathsf{n}_n) : \mathsf{m}_q \in \mathcal{M}_{\mathrm{V}}\}$ as the set of edges between UE n and all virtual BSs. We can then reformulate Problem P2 as the following maximum weight matching problem:

$$(\mathrm{P4}): \max_{\mathcal{E} \subset \underline{\mathcal{E}}} \sum_{(\mathsf{m}_q, \mathsf{n}_n) \in \mathcal{E}} W_{q,n}, \tag{37}$$

$$\text{s.t. } |\mathcal{E} \cap \underline{\mathcal{E}}^{(q)}| \leq 1, \forall q = 1, \cdots, MC, \tag{38}$$

$$|\mathcal{E} \cap \underline{\mathcal{E}}_{(n)}| \leq 1, \forall n = 1, \cdots, N. \tag{39}$$

Problem P4 can be optimally solved by using the Hungarian algorithm in polynomial time [33]. With the optimal weight matching $\mathcal{E}^*$, the optimal user association scheme $\{\mathcal{I}^{\mathrm{associate}*}_{m,n}\}$ can be obtained as

$$\mathcal{I}^{\mathrm{associate}*}_{m,n} = \begin{cases} 1 & \text{if } (\mathsf{m}_q, \mathsf{n}_n) \in \mathcal{E}^* \\ 0 & \text{otherwise.} \end{cases} \tag{40}$$

B. Low-Complexity User Association Scheme

Note that the complexity of the Hungarian algorithm can be as high as $\mathcal{O}(\max\{(MC)^3, N^3\})$ [33], which quickly grows with the number of BSs M and the number of UEs N. 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 P2 iteratively. In each iteration, one UE is selected to associate with one BS. The algorithm is based on the rational that a UE 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 UEs, we choose the UE, which 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.

Algorithm 2 Low-Complexity Outage Mitigation User Association Scheme

Input: Blockage indicator $\{\mathcal{I}_{m,n}^{\text{block}}\}$, coverage indicator $\{\mathcal{I}_{m,n}^{\text{out}}\}$, maximum workload of each BS C.

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

Output: The user association indicators $\{\mathcal{I}_{m,n}^{\text{associate}}\}$.

Furthermore, we should balance the workload of each BS, i.e., the number of UEs that served by one BS, to reduce the chance that a UE cannot associate with one BS due to the maximum workload constraint. As a result, for the selected UE, among its physically accessible BSs, we choose the BS that has the minimum number of associated UEs to be associated with.

The detailed steps of the proposed algorithm are presented in Algorithm 2. It starts from initiating a candidate BS set $\mathcal{M}_0$ and a remaining UE set $\mathcal{N}_0$ by letting $\mathcal{M}_0 = \mathcal{M}$ and $\mathcal{N}_0$ be the set of UEs which have at least one physically accessible BS, as shown in Steps 2 and 3 of Algorithm 2. Then each UE in $\mathcal{N}_0$ is selected to associate with one BS iteratively. The UE with the smallest number of physically accessible BSs is selected in each iteration, as indicated in Step 5 of Algorithm 2. It is associated to one of its physically accessible BSs, which has the smallest number of associated UEs, in the candidate set $\mathcal{M}_0$, as shown in Step 7. Step 12 implies that if a BS has been associated with C UEs, it will be removed from the candidate set $\mathcal{M}_0$. The algorithm terminates when $\mathcal{N}_0$ or $\mathcal{M}_0$ becomes empty, in which case all the UEs have been considered or there are no available BSs for UEs to associate with.

It can be clearly seen from Algorithm 2 that there is at most $\min\{N, MC\}$ iterations. Furthermore, as finding the minimum variable over a set with cardinality n, also known as sorting algorithms, requires $n \log n$ comparisons [40], the complexity of Step 5 and Step 7 are $\mathcal{O}(N \log N)$ and $\mathcal{O}(M \log M)$, respectively. Therefore, the computational complexity of the proposed user association algorithm is $\mathcal{O}\big(\min\{N, MC\}(N \log N + M \log M)\big)$, which is much lower than the complexity of the Hungarian algorithm. Furthermore, based on the submodular optimization theory [41], [42], Problem P2 can be solved by Algorithm 2 with a constant-factor $\frac{1}{3}$ approximation guarantee. The detailed derivation of this approximation guarantee is presented in Appendix D. In Section V, we will illustrate the performance of Algorithm 2 by comparing it with the optimal user association scheme in terms of outage performance and time efficiency via simulations.

V. Simulation Results

In this section, simulations are conducted to verify the performance of the proposed CSA-based optimal BS deployment scheme, i.e., Algorithm 1, and low-complexity outage mitigation user association scheme, i.e., Algorithm 2, under different distributions of UEs' positions and various parameter settings.

A. Simulation Setting

In particular, we consider a 2-dimensional Manhattan-type geometry with $B = 20$ square blocks as shown in Fig. 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 as 100 meters. Unless otherwise stated, the number of BSs M, the number of UEs N and the maximum workload C are set as 15, 450 and 30, respectively. We set the maximum inaccessible probability P_{iac}^* of Problem P1 as 0.05, i.e., the percentage of the UEs which do not have any accessible BSs should not exceed 5%.

We generate the UEs' positions in a similar way to the traffic generation method in [43], where UEs gather around some hotspots with a certain degree of aggregation. We start from randomly generated positions of UEs and hotspots, then update the position of each UE 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 UE and its closest hotspot. The UEs which fall inside the blocks are projected to the street area by using the function $\Pi_{\mathcal{A}^N}(\cdot)$ given in (23). The aggregation factor $\mu_\delta \in [0, 1]$ determines the aggregation degree of UEs, that is, each UE has a higher chance of approaching its closest hotspot with a larger μ_δ. When $\mu_\delta = 0$, UEs are uniformly distributed in the area. With $\mu_\delta = 1$, UEs are located exactly at the hotspots. Fig. 2 shows two samples of UEs' positions with $\mu_\delta = 0$ and $\mu_\delta = 0.5$ with 5 hotspots.

B. Performance of the Proposed CSA-based Optimal BS Deployment Algorithm

As pointed out in Section III, the convergence of the proposed CSA-based optimal BS deployment scheme, i.e., Algorithm 1, depends on the initial deployment scheme due to the non-convexity of the objective function and constraint function. For illustration, we consider two types of initial deployment schemes of BSs, which are listed as follows:

- **Maximum Coverage (MC) Scheme**: The MC scheme is a BS deployment scheme proposed in [21] 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 [21]. A set of RPs are also evenly placed in the street area. The optimal BS positions are then obtained by solving the problem of maximizing the number of

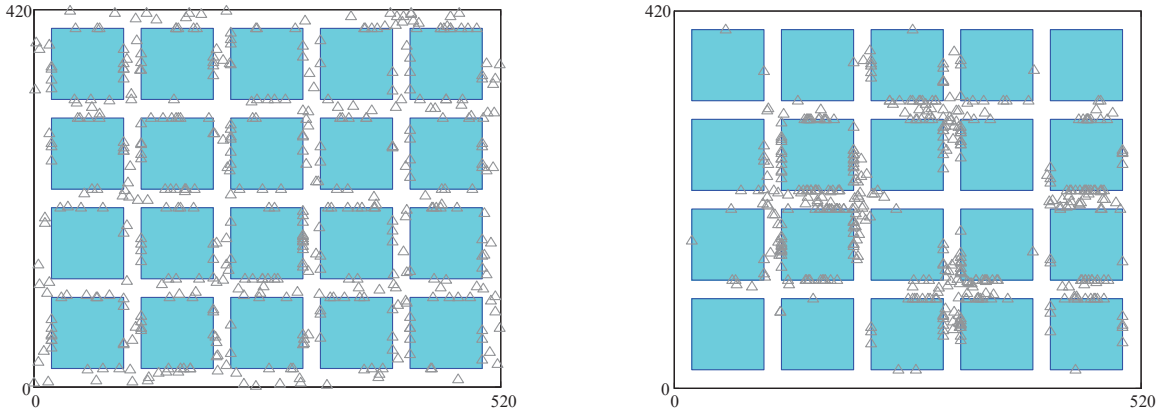


Fig. 2: Snapshots of UEs' positions with different aggregation factors μ_δ. (a) $\mu_\delta = 0$. (b) $\mu_\delta = 0.5$.

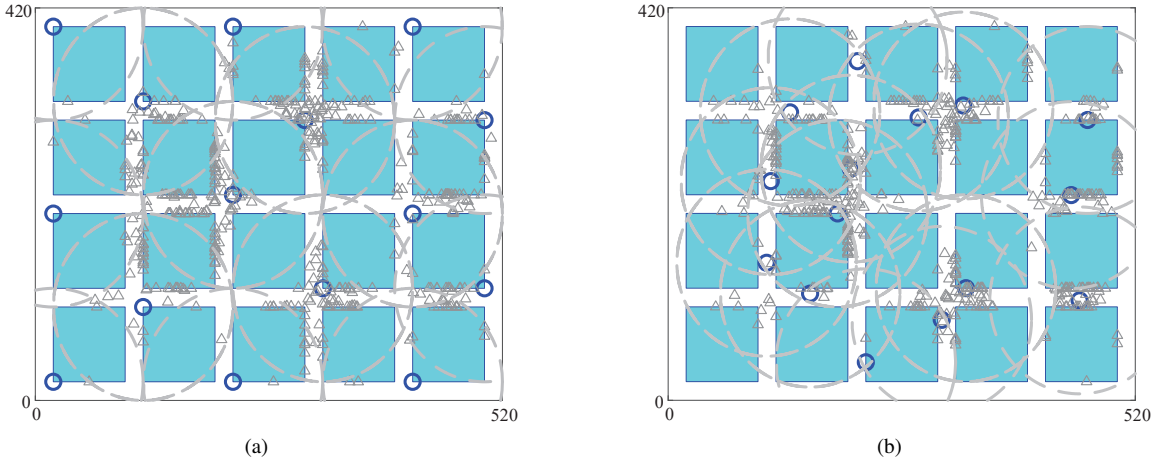


Fig. 3: Positions and coverage of BSs with (a) the MC initial BS deployment scheme and (b) the KM initial BS deployment scheme. A snapshot of UEs' positions with $\mu_\delta = 0.5$ and 5 hotspots is also presented. $M = 15$ BSs and $N = 450$ UEs are represented by circles and triangles, respectively. The dashed line represents the coverage of each BS.

RPs that can be covered by at least one BS. A graphic illustration of this scheme is shown in Fig. 3a.

- **K-means (KM) Scheme**: Randomly generate one sample of UEs' positions. Apply the K-means method [44] to group UEs into $M = 15$ clusters. Set the initial positions of BSs as the centroids of clusters. A graphic illustration of this scheme is shown in Fig. 3b.

Fig. 4 shows how the average number of accessible BSs with the optimized positions of BSs $\mathbf{r}^{\mathrm{BS}*}$ varies with the total number of iterations T under the above initial deployment schemes. We can see that in both cases, the proposed algorithm converges after $T \geq 700$. With $\mu_\delta = 0$, i.e., UEs are uniformly distributed in the area, the algorithm is insensitive to the initial deployment of BSs. In contrast, in the clustered-UE case with $\mu_\delta = 0.5$, the algorithm converges to a larger value with the KM initial deployment scheme than that with the MC initial deployment 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 $\mathcal{I}^{\mathrm{out}}_{m,n}$ with respect to each BS's position, $\frac{\partial \mathcal{I}^{\mathrm{out}}_{m,n}}{\partial \mathbf{r}^{\mathrm{BS}}_m}$, is crucially determined by the number of UEs close to the edge of the

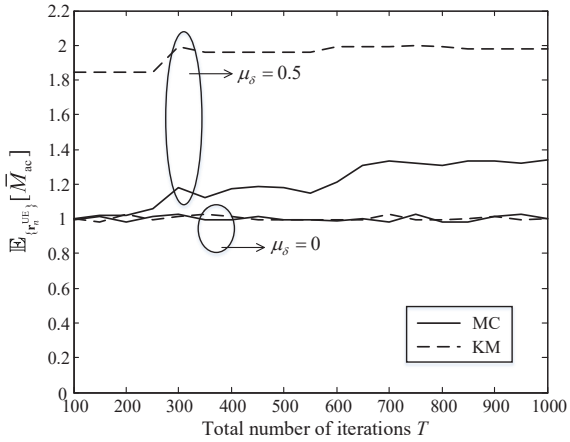


Fig. 4: Long-term average number of accessible BSs $\mathbb{E}_{\{\mathbf{r}_n^{\mathrm{UE}}\}}[\bar{M}_{\mathrm{ac}}]$ versus the total number of iterations T of Algorithm 1, the proposed CSA-based optimal BS deployment scheme, with two initial deployment schemes of BSs, MC and KM.

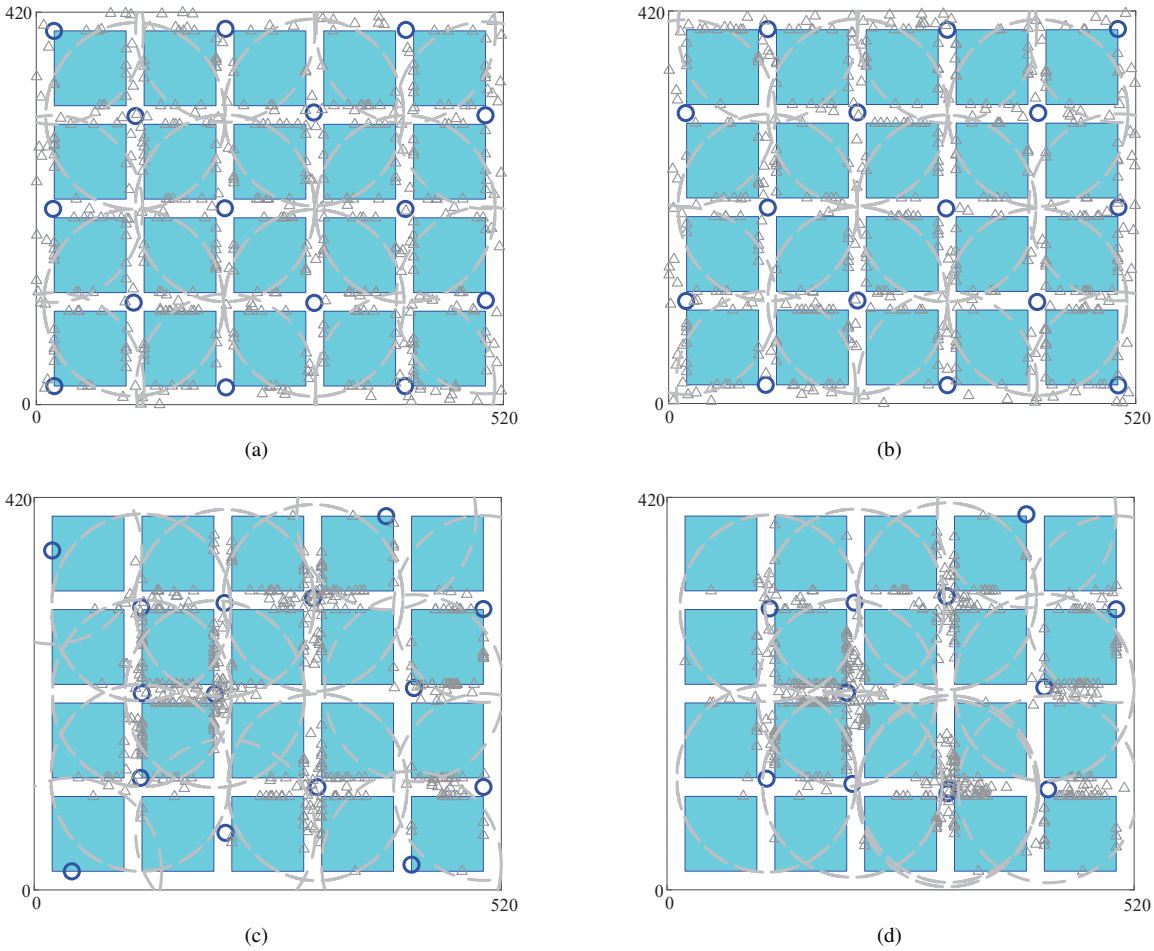


Fig. 5: Positions and coverage of BSs with Algorithm 1, the proposed CSA-based optimal BS deployment scheme. Snapshots of UEs' positions with aggregation factors $\mu_\delta \in \{0, 0.5\}$ and 5 hotspots are also presented. BSs and UEs are represented by circles and triangles, respectively. The dashed line represents the coverage of each BS. (a) MC initial deployment, $\mu_\delta = 0$. (b) KM initial deployment, $\mu_\delta = 0$. (c) MC initial deployment, $\mu_\delta = 0.5$. (d) KM initial deployment, $\mu_\delta = 0.5$.

BS's coverage area. As all the BSs' positions are updated with the same step size, those with few edge-UEs would make overly conservative movements. With the MC initial deployment scheme, the number of BSs with no UEs in the coverage area could be much higher than that with the KM initial deployment scheme when UEs are clustered. Therefore, it is much more likely to be trapped into sub-optimal positions. As Fig. 5c shows, with the MC initial deployment, some of the BSs cannot move to the UE-intensive area, leading to a significantly lower number of accessible BSs compared to that with the KM initial deployment.

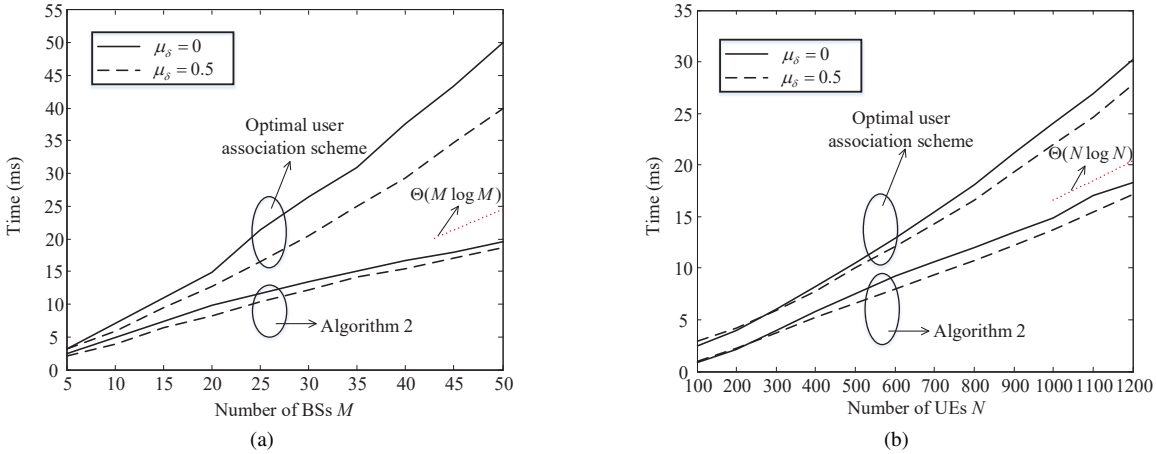
We can also see from Fig. 5 that after optimization, most of the BSs will move to the intersection of two streets. Intuitively, placing a BS at the intersection can maximize the number of LoS paths between the BS and the UEs in the two streets, and thus increase the average number of accessible BSs. Fig. 5 also shows that some of the BSs are located in the middle of a street, which may not be feasible in practice. In Section V-D, we will further project the optimized positions of BSs to the closest corners of buildings.

C. 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 2, by comparing it with the Hungarian-algorithm-based optimal user association scheme.

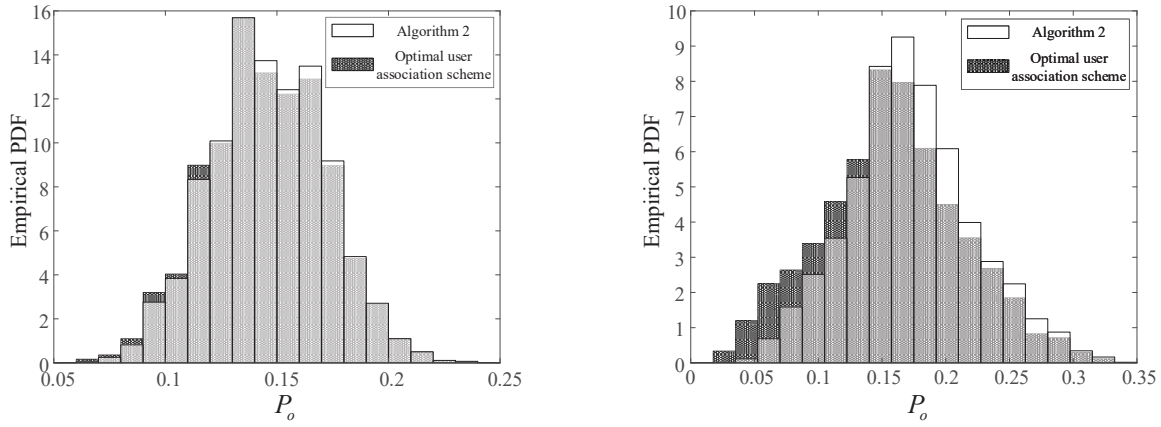
Fig. 6 shows the running time of Algorithm 2 and the optimal user association scheme, which is averaged over 1000 realizations of UEs' and BSs' positions. The simulation is performed on the platform of MATLAB R2020a using a 3.4 GHz Intel Core i7. The Hungarian algorithm is implemented by using the '`matchpairs`' function of the MATLAB platform.

The complexity analysis in Section IV has shown that the complexity of Algorithm 2 scales with the number of BSs M and the number of UEs N as $\mathcal{O}(M \log M)$ and $\mathcal{O}(N \log N)$, respectively, which is verified by the simulation results presented in Fig. 6. Fig. 6 also shows that compared with the optimal user association scheme, the running time of Algorithm 2 is significantly reduced, and the gain increases as the number of BSs M or the number of UEs N becomes larger. For instance, in Fig. 6a, when the aggregation factor



(a) (b)

Fig. 6: Average running time (in unit of milliseconds) of Algorithm 2, the proposed low-complexity outage mitigation user association scheme, and the optimal user association scheme versus (a) the number of BSs M, and (b) the number of UEs N.



(a) (b)

Fig. 7: Empirical PDF of outage probability P_o with Algorithm 2, the proposed low-complexity outage mitigation user association scheme, and the optimal user association scheme. (a) $\mu_\delta = 0$. (b) $\mu_\delta = 0.5$.

$\mu_\delta = 0$, the running time of the proposed algorithm is 65.59% of that of the optimal user association scheme with $M = 20$. It is further reduced to 39.10% with $M = 50$.

Fig. 7 presents empirical probability density function (PDF) of the outage probability P_o with both the proposed user association scheme and the optimal user association scheme. The empirical PDF is generated with 10000 realizations of UEs' and BSs' positions. We can see from Fig. 7a that when $\mu_\delta = 0$, i.e., UEs are uniformly distributed, the proposed scheme can achieve almost identical outage performance with the optimal user association scheme. While in Fig. 7b where $\mu_\delta = 0.5$, the percentage for small value of P_o 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-UE case, i.e., $\mu_\delta = 0$, we can see from Figs. 5a and 5b that for most UEs, each of them has only one physically accessible BS. This can be also seen from Fig. 4, where $\mathbb{E}_{\{\mathbf{r}_n^{\mathrm{UE}}\}}[\bar{M}_{\mathrm{ac}}]$ is close to 1 when $\mu_\delta = 0$. In that case, with both the optimal user association scheme and the proposed Algorithm 2, most UEs are associated with their single physically accessible BS, and thus both algorithms achieve similar outage performance. In contrast, when $\mu_\delta = 0.5$, as UEs are clustered, Fig. 5d shows that many UEs have more than one physically accessible BSs. For Algorithm 2 which associates one UE with one BS in each iteration in a greedy manner, it is not guaranteed that each UE can always make a globally optimal choice. Therefore, some UEs 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.

D. Outage Performance

In this subsection, we evaluate the outage performance of the proposed CSA-based optimal BS deployment combined with the proposed low-complexity outage mitigation user association scheme. For the proposed CSA-based optimal BS deployment, both MC and KM initial deployment 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

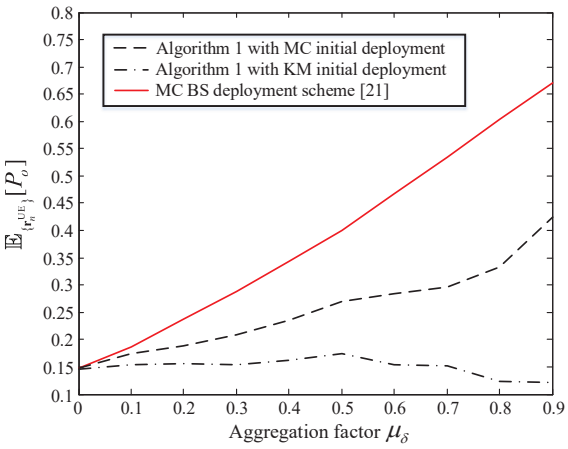


Fig. 8: Average outage probability $\mathbb{E}_{\{\mathbf{r}_n^{\mathrm{UE}}\}}[P_o]$ versus the aggregation factor μ_δ.

MC BS deployment scheme [21] combined with the proposed low-complexity outage mitigation user association scheme as the performance benchmark.

Fig. 8 presents the long-term average outage probability $\mathbb{E}_{\{\mathbf{r}_n^{\mathrm{UE}}\}}[P_o]$ with different BS deployment schemes versus the aggregation factor of UEs' distribution μ_δ. We can see that the proposed CSA-based optimal BS deployment scheme significantly outperforms the benchmark MC BS deployment scheme when $\mu_\delta>0$, and the performance gap is enlarged with the increase of μ_δ. Intuitively, with symmetrically placed BSs, the number of UEs that fall into the coverage of each BS becomes more unbalanced as μ_δ increases, i.e., UEs become more clustered. Consequently, there may not be sufficient BSs near the hotspots to support the excessive number of UEs, leading to high outage probability in the long term. In contrast, the proposed CSA-based optimal BS deployment scheme optimizes the BSs' positions based on UEs' spatial distribution, and therefore can better balance the workloads of the BSs. It can also be seen from Fig. 8 that the KM initial deployment scheme leads to a smaller average outage probability than the MC initial deployment scheme, especially when μ_δ is large. This is consistent with the convergence results in Fig. 4, where Algorithm 1 with the KM initial deployment converges to a larger long-term average number of accessible BSs $\mathbb{E}_{\{\mathbf{r}_n^{\mathrm{UE}}\}}[\bar{M}_{\mathrm{ac}}]$ than that with the MC initial deployment. With more accessible BSs, UEs have more freedom to select a BS to associate with, thus leading to a lower average outage probability $\mathbb{E}_{\{\mathbf{r}_n^{\mathrm{UE}}\}}[P_o]$.

VI. Conclusion

In this paper, we proposed a BS deployment 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 deployment problem, as the physical accessibility between UEs and BSs is closely related to the positions of UEs, which may vary with time, a stochastic optimization framework was established for maximizing the long-term average number of physically accessible BSs of each UE under an inaccessible probability constraint, based on which a CSA-based algorithm was 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 deployment scheme are also demonstrated, which increase as the UEs' spatial distribution becomes more clustered.

Note that in this paper, the BS deployment problem is formulated as a stochastic optimization problem with one expectation constraint and multiple deterministic constraints, which can be well handled by the proposed CSA-based algorithm. However, for more complicated BS deployment problems which involve multiple expectation constraints, CSA may not be applicable as it can only control the violation of one expectation constraint. Moreover, in an mmWave communication network, the system topology affects not only the connectivity between UEs and BSs, but also the network capacity. It is of great practical significance to further study how to optimize the positions of BSs for improving the mmWave network capacity under a certain outage probability constraint. Last but not least, in practice, mmWave networks alone can hardly provide ubiquitous network coverage due to the inherent limitations of mmWave propagation. One promising solution to enhance the robustness of mmWave networks is combining them with the conventional cellular systems [4]. How to jointly optimize the deployment of mmWave BSs and conventional BSs is an interesting topic that deserves much attention in the future study.

Appendix A
Derivation of (5)

Let us first rewrite $\mathcal{I}_{m,n}^{\mathrm{block}}$ in (2) as

$$\mathcal{I}_{m,n}^{\mathrm{block}} = I\left(\sum_{b=1}^{B}\mathcal{I}_{m,n,b}^{\mathrm{block}}\right), \tag{41}$$

where

$$\mathcal{I}_{m,n,b}^{\mathrm{block}} = \begin{cases} 1 & \text{if the LoS path between BS } m \\ & \text{and UE } n \text{ is blocked by block } b \\ 0 & \text{otherwise.} \end{cases} \tag{42}$$

In the following, we will derive the expression of $\mathcal{I}_{m,n,b}^{\mathrm{block}}$.

Let us map the positions of BS m and UE n from the coordinate system in Fig. 1 to the one in Fig. 9, where the origin is the center of the line segment connecting BS m and UE n. The coordinates of BS m and UE n in the new coordinate system are $(-\frac{d_{m,n}}{2},0)$ and $(\frac{d_{m,n}}{2},0)$, respectively. Then the new coordinate of the center of block b are $(x_{m,n,b},y_{m,n,b})$ with $x_{m,n,b}$ and $y_{m,n,b}$ given in (6) and (7), respectively. If block b blocks the LoS path between BS m and UE n, $(x_{m,n,b},y_{m,n,b})$ should fall in the region $ABCDEF$ [45]. The coordinates of the points A,B,C,D,E,F in the new

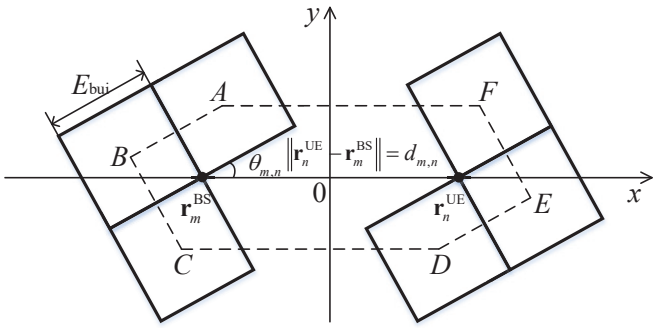


Fig. 9: Region $ABCDEF$ that the center of block b falling in will block the LoS path between BS m and UE n.

coordinate system can be obtained as

$$\begin{aligned}
&A:\big(\tfrac{\sqrt{2}}{2}E_{\text{bui}}\cos(\theta_{m,n}+\tfrac{\pi}{4})-\tfrac{d_{m,n}}{2},\tfrac{\sqrt{2}}{2}E_{\text{bui}}\sin(\theta_{m,n}+\tfrac{\pi}{4})\big),\\
&B:\big(\tfrac{\sqrt{2}}{2}E_{\text{bui}}\cos(\theta_{m,n}+\tfrac{3\pi}{4})-\tfrac{d_{m,n}}{2},\tfrac{\sqrt{2}}{2}E_{\text{bui}}\sin(\theta_{m,n}+\tfrac{3\pi}{4})\big),\\
&C:\big(\tfrac{\sqrt{2}}{2}E_{\text{bui}}\cos(\theta_{m,n}+\tfrac{5\pi}{4})-\tfrac{d_{m,n}}{2},\tfrac{\sqrt{2}}{2}E_{\text{bui}}\sin(\theta_{m,n}+\tfrac{5\pi}{4})\big),\\
&D:\big(\tfrac{\sqrt{2}}{2}E_{\text{bui}}\cos(\theta_{m,n}+\tfrac{5\pi}{4})+\tfrac{d_{m,n}}{2},\tfrac{\sqrt{2}}{2}E_{\text{bui}}\sin(\theta_{m,n}+\tfrac{5\pi}{4})\big),\\
&E:\big(\tfrac{\sqrt{2}}{2}E_{\text{bui}}\cos(\theta_{m,n}+\tfrac{7\pi}{4})+\tfrac{d_{m,n}}{2},\tfrac{\sqrt{2}}{2}E_{\text{bui}}\sin(\theta_{m,n}+\tfrac{7\pi}{4})\big),\\
&F:\big(\tfrac{\sqrt{2}}{2}E_{\text{bui}}\cos(\theta_{m,n}+\tfrac{\pi}{4})+\tfrac{d_{m,n}}{2},\tfrac{\sqrt{2}}{2}E_{\text{bui}}\sin(\theta_{m,n}+\tfrac{\pi}{4})\big),
\end{aligned}$$

where $\theta_{m,n}$ is given in (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

$$\begin{cases}
f_{\overline{AB}}(x,y)=x\sin\theta_{m,n}-y\cos\theta_{m,n}+\frac{1}{2}d_{m,n}\sin\theta_{m,n}+\frac{1}{2}E_{\text{bui}}\\
f_{\overline{BC}}(x,y)=x\cos\theta_{m,n}+y\sin\theta_{m,n}+\frac{1}{2}d_{m,n}\cos\theta_{m,n}+\frac{1}{2}E_{\text{bui}}\\
f_{\overline{CD}}(x,y)=y+\frac{1}{2}E_{\text{bui}}\sin\theta_{m,n}+\frac{1}{2}E_{\text{bui}}\cos\theta_{m,n}\\
f_{\overline{DE}}(x,y)=-x\sin\theta_{m,n}+y\cos\theta_{m,n}+\frac{1}{2}d_{m,n}\sin\theta_{m,n}+\frac{1}{2}E_{\text{bui}}\\
f_{\overline{EF}}(x,y)=-x\cos\theta_{m,n}-y\sin\theta_{m,n}+\frac{1}{2}d_{m,n}\cos\theta_{m,n}+\frac{1}{2}E_{\text{bui}}\\
f_{\overline{FA}}(x,y)=-y+\frac{1}{2}E_{\text{bui}}\sin\theta_{m,n}+\frac{1}{2}E_{\text{bui}}\cos\theta_{m,n}.
\end{cases} \tag{43}$$

If the point $(x_{m,n,b},y_{m,n,b})$ 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

$$\begin{aligned}
\mathcal{I}_{m,n,b}^{\text{block}} &= I\big(f_{\overline{AB}}(0,0)\cdot f_{\overline{AB}}(x_{m,n,b},y_{m,n,b})\big)\cdot I\big(f_{\overline{BC}}(0,0)\\
&\cdot f_{\overline{BC}}(x_{m,n,b},y_{m,n,b})\big)\cdot I\big(f_{\overline{CD}}(0,0)\cdot f_{\overline{CD}}(x_{m,n,b},y_{m,n,b})\big)\cdot\\
&I\big(f_{\overline{DE}}(0,0)\cdot f_{\overline{DE}}(x_{m,n,b},y_{m,n,b})\big)\cdot I\big(f_{\overline{EF}}(0,0)\\
&\cdot f_{\overline{EF}}(x_{m,n,b},y_{m,n,b})\big)\cdot I\big(f_{\overline{FA}}(0,0)\cdot f_{\overline{FA}}(x_{m,n,b},y_{m,n,b})\big).
\end{aligned} \tag{44}$$

(5) can then be obtained by combining (41), (43) and (44).

Appendix B
Derivations of (28) and (29)

Let us rewrite the term $\sum_{m=1}^{M} I(\mathcal{I}_{m,n}^{\text{out}}+\mathcal{I}_{m,n}^{\text{block}})$ in (9) as

$$\begin{aligned}
\sum_{m=1}^{M} I(\mathcal{I}_{m,n}^{\text{out}}+\mathcal{I}_{m,n}^{\text{block}}) &= \sum_{m\in\mathcal{M}_n} I(\mathcal{I}_{m,n}^{\text{out}}+\mathcal{I}_{m,n}^{\text{block}})\\
&+\sum_{m\in\mathcal{M}\backslash\mathcal{M}_n} I(\mathcal{I}_{m,n}^{\text{out}}+\mathcal{I}_{m,n}^{\text{block}}).
\end{aligned} \tag{45}$$

Based on (1) and the definition of $\mathcal{M}_n$, i.e., the set of BSs which have UE n in each of their coverage areas, we have

$$\mathcal{I}_{m,n}^{\text{out}} = \begin{cases} 1 & \text{if } m\in\mathcal{M}\backslash\mathcal{M}_n\\ 0 & \text{otherwise.}\end{cases} \tag{46}$$

Therefore, we can obtain

$$\sum_{m\in\mathcal{M}_n} I(\mathcal{I}_{m,n}^{\text{out}}+\mathcal{I}_{m,n}^{\text{block}}) = \sum_{m\in\mathcal{M}_n} I(\mathcal{I}_{m,n}^{\text{block}}) = \sum_{m\in\mathcal{M}_n}\mathcal{I}_{m,n}^{\text{block}} \tag{47}$$

and

$$\sum_{m\in\mathcal{M}\backslash\mathcal{M}_n} I(\mathcal{I}_{m,n}^{\text{out}}+\mathcal{I}_{m,n}^{\text{block}}) = \sum_{m\in\mathcal{M}\backslash\mathcal{M}_n} I(\mathcal{I}_{m,n}^{\text{out}}) = \sum_{m=1}^{M}\mathcal{I}_{m,n}^{\text{out}}. \tag{48}$$

By substituting (47) and (48) into (45), we have

$$\sum_{m=1}^{M} I(\mathcal{I}_{m,n}^{\text{out}}+\mathcal{I}_{m,n}^{\text{block}}) = \sum_{m=1}^{M}\mathcal{I}_{m,n}^{\text{out}} + \sum_{m\in\mathcal{M}_n}\mathcal{I}_{m,n}^{\text{block}}. \tag{49}$$

Similarly, we can express the term $\prod_{m=1}^{M} I\left(\mathcal{I}_{m,n}^{\text{out}}+\mathcal{I}_{m,n}^{\text{block}}\right)$ in (10) as

$$\begin{aligned}
\prod_{m=1}^{M} I\left(\mathcal{I}_{m,n}^{\text{out}}+\mathcal{I}_{m,n}^{\text{block}}\right) &= \prod_{m\in\mathcal{M}_n} I\left(\mathcal{I}_{m,n}^{\text{out}}+\mathcal{I}_{m,n}^{\text{block}}\right)\\
&\cdot\prod_{m\in\mathcal{M}\backslash\mathcal{M}_n} I\left(\mathcal{I}_{m,n}^{\text{out}}+\mathcal{I}_{m,n}^{\text{block}}\right).
\end{aligned} \tag{50}$$

By combining (46) and (50), we have

$$\prod_{m=1}^{M} I\left(\mathcal{I}_{m,n}^{\text{out}}+\mathcal{I}_{m,n}^{\text{block}}\right) = \prod_{m\in\mathcal{M}_n}\mathcal{I}_{m,n}^{\text{block}},\ \text{if } \mathcal{M}_n\neq\emptyset. \tag{51}$$

If $\mathcal{M}_n$ is empty, which means UE n is not in the coverage area of any BSs, i.e., $\mathcal{I}_{m,n}^{\text{out}}=1,\forall m$ then we have

$$\prod_{m=1}^{M} I\left(\mathcal{I}_{m,n}^{\text{out}}+\mathcal{I}_{m,n}^{\text{block}}\right) = \prod_{m=1}^{M}\mathcal{I}_{m,n}^{\text{out}}. \tag{52}$$

Based on (51), (52) and $\prod_{m=1}^{M}\mathcal{I}_{m,n}^{\text{out}}=0$ if $\mathcal{M}_n\neq\emptyset$, we have

$$\begin{aligned}
\prod_{m=1}^{M} I\left(\mathcal{I}_{m,n}^{\text{out}}+\mathcal{I}_{m,n}^{\text{block}}\right) &= I\left(\prod_{m\in\mathcal{M}_n}\mathcal{I}_{m,n}^{\text{block}} + \prod_{m=1}^{M}\mathcal{I}_{m,n}^{\text{out}}\right)\\
&\overset{(a)}{=} \prod_{m\in\mathcal{M}_n}\mathcal{I}_{m,n}^{\text{block}} + \prod_{m=1}^{M}\mathcal{I}_{m,n}^{\text{out}},
\end{aligned} \tag{53}$$

where (a) follows from the fact that $\left(\prod_{m\in\mathcal{M}_n}\mathcal{I}_{m,n}^{\text{block}}+\prod_{m=1}^{M}\mathcal{I}_{m,n}^{\text{out}}\right)\in\{0,1\}$. (28) and (29) can be then obtained by inserting (49) and (53) into (9) and (10), respectively.

Appendix C
Derivations of $\frac{\partial\mathcal{I}_{m,n}^{\text{out}}(t)}{\partial\mathbf{r}^{\text{BS}}(t)}$ and $\frac{\partial\mathcal{I}_{m,n}^{\text{block}}(t)}{\partial\mathbf{r}^{\text{BS}}(t)}$

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

A. Derivation of $\frac{\partial\mathcal{I}_{m,n}^{\text{out}}}{\partial\mathbf{r}^{\text{BS}}}$

By replacing the indicator function in (4) with (27), $\frac{\partial\mathcal{I}_{m,n}^{\text{out}}}{\partial\mathbf{r}^{\text{BS}}}$ can be derived as

$$\frac{\partial\mathcal{I}_{m,n}^{\text{out}}}{\partial\mathbf{r}^{\text{BS}}} \approx S'(d_{m,n}-R)\cdot\frac{\partial d_{m,n}}{\partial\mathbf{r}^{\text{BS}}}. \tag{54}$$

where $S'(x) = \beta S(x)(1 - S(x))$ is the derivative of the sigmoid function $S(x)$. As $d_{m,n} = \|\mathbf{r}_m^{\mathrm{BS}} - \mathbf{r}_n^{\mathrm{UE}}\|$, the partial derivative $\frac{\partial d_{m,n}}{\partial \mathbf{r}_m^{\mathrm{BS}}}, m = 1, \cdots, M$ is given by

$$\frac{\partial d_{m,n}}{\partial \mathbf{r}_j^{\mathrm{BS}}}\Big|_{j\neq m} = [0,0], \text{ and } \frac{\partial d_{m,n}}{\partial \mathbf{r}_m^{\mathrm{BS}}} = \Big[\frac{\partial d_{m,n}}{\partial x_m^{\mathrm{BS}}}, \frac{\partial d_{m,n}}{\partial y_m^{\mathrm{BS}}}\Big], \tag{55}$$

where

$$\frac{\partial d_{m,n}}{\partial x_m^{\mathrm{BS}}} = (x_m^{\mathrm{BS}} - x_n^{\mathrm{UE}})\Big((x_m^{\mathrm{BS}} - x_n^{\mathrm{UE}})^2 + (y_m^{\mathrm{BS}} - y_n^{\mathrm{UE}})^2\Big)^{-1/2}, \tag{56}$$

and

$$\frac{\partial d_{m,n}}{\partial y_m^{\mathrm{BS}}} = (y_m^{\mathrm{BS}} - y_n^{\mathrm{UE}})\Big((x_m^{\mathrm{BS}} - x_n^{\mathrm{UE}})^2 + (y_m^{\mathrm{BS}} - y_n^{\mathrm{UE}})^2\Big)^{-1/2}. \tag{57}$$

B. Derivation of $\frac{\partial \mathcal{I}_{m,n}^{\mathrm{block}}}{\partial \mathbf{r}^{\mathrm{BS}}}$

By replacing the indicator function in (41) with (27), $\frac{\partial \mathcal{I}_{m,n}^{\mathrm{block}}}{\partial \mathbf{r}^{\mathrm{BS}}}$ can be approximated by

$$\frac{\partial \mathcal{I}_{m,n}^{\mathrm{block}}}{\partial \mathbf{r}^{\mathrm{BS}}} \approx S'\left(\sum_{b=1}^{B} \mathcal{I}_{m,n,b}^{\mathrm{block}}\right) \cdot \sum_{b=1}^{B} \frac{\partial \mathcal{I}_{m,n,b}^{\mathrm{block}}}{\partial \mathbf{r}^{\mathrm{BS}}}, \tag{58}$$

where

$$\frac{\partial \mathcal{I}_{m,n,b}^{\mathrm{block}}}{\partial \mathbf{r}^{\mathrm{BS}}} = \Big[\frac{\partial \mathcal{I}_{m,n,b}^{\mathrm{block}}}{\partial \mathbf{r}_1^{\mathrm{BS}}}, \cdots, \frac{\partial \mathcal{I}_{m,n,b}^{\mathrm{block}}}{\partial \mathbf{r}_M^{\mathrm{BS}}}\Big]. \tag{59}$$

$\frac{\partial \mathcal{I}_{m,n,b}^{\mathrm{block}}}{\partial \mathbf{r}_j^{\mathrm{BS}}}$ is given by

$$\begin{cases} \frac{\partial \mathcal{I}_{m,n,b}^{\mathrm{block}}}{\partial \mathbf{r}_j^{\mathrm{BS}}}\Big|_{j\neq m} = [0,0], \\ \frac{\partial \mathcal{I}_{m,n,b}^{\mathrm{block}}}{\partial \mathbf{r}_m^{\mathrm{BS}}} = \sum_{i=1}^{6} \prod_{j=1, j\neq i}^{6} \mathcal{I}_{m,n,b}^{(j)} \frac{\partial \mathcal{I}_{m,n,b}^{(i)}}{\partial \mathbf{r}_m^{\mathrm{BS}}}, \end{cases} \tag{60}$$

where

$$\begin{cases} \mathcal{I}_{m,n,b}^{(1)} = I\big(f_{\overline{AB}}(0,0) \cdot f_{\overline{AB}}(x_{m,n,b}, y_{m,n,b})\big) \\ \mathcal{I}_{m,n,b}^{(2)} = I\big(f_{\overline{BC}}(0,0) \cdot f_{\overline{BC}}(x_{m,n,b}, y_{m,n,b})\big) \\ \mathcal{I}_{m,n,b}^{(3)} = I\big(f_{\overline{CD}}(0,0) \cdot f_{\overline{CD}}(x_{m,n,b}, y_{m,n,b})\big) \\ \mathcal{I}_{m,n,b}^{(4)} = I\big(f_{\overline{DE}}(0,0) \cdot f_{\overline{DE}}(x_{m,n,b}, y_{m,n,b})\big) \\ \mathcal{I}_{m,n,b}^{(5)} = I\big(f_{\overline{EF}}(0,0) \cdot f_{\overline{EF}}(x_{m,n,b}, y_{m,n,b})\big) \\ \mathcal{I}_{m,n,b}^{(6)} = I\big(f_{\overline{FA}}(0,0) \cdot f_{\overline{FA}}(x_{m,n,b}, y_{m,n,b})\big) \end{cases} \tag{61}$$

according to (44). By replacing the indicator function in (61) with (27), $\frac{\partial \mathcal{I}_{m,n,b}^{(i)}}{\partial \mathbf{r}_m^{\mathrm{BS}}}, i = 1, \cdots, 6$ can be approximately obtained as

$$\begin{cases} \frac{\partial \mathcal{I}_{m,n,b}^{(1)}}{\partial \mathbf{r}_m^{\mathrm{BS}}} \approx S'\big(f_{\overline{AB}}(0,0) f_{\overline{AB}}(x_{m,n,b}, y_{m,n,b})\big)\big(\frac{\partial f_{\overline{AB}}(0,0)}{\partial \mathbf{r}_m^{\mathrm{BS}}} \\ \cdot f_{\overline{AB}}(x_{m,n,b}, y_{m,n,b}) + \frac{\partial f_{\overline{AB}}(x_{m,n,b}, y_{m,n,b})}{\partial \mathbf{r}_m^{\mathrm{BS}}} f_{\overline{AB}}(0,0)\big) \\ \frac{\partial \mathcal{I}_{m,n,b}^{(2)}}{\partial \mathbf{r}_m^{\mathrm{BS}}} \approx S'\big(f_{\overline{BC}}(0,0) f_{\overline{BC}}(x_{m,n,b}, y_{m,n,b})\big)\big(\frac{\partial f_{\overline{BC}}(0,0)}{\partial \mathbf{r}_m^{\mathrm{BS}}} \\ \cdot f_{\overline{BC}}(x_{m,n,b}, y_{m,n,b}) + \frac{\partial f_{\overline{BC}}(x_{m,n,b}, y_{m,n,b})}{\partial \mathbf{r}_m^{\mathrm{BS}}} f_{\overline{BC}}(0,0)\big) \\ \frac{\partial \mathcal{I}_{m,n,b}^{(3)}}{\partial \mathbf{r}_m^{\mathrm{BS}}} \approx S'\big(f_{\overline{CD}}(0,0) f_{\overline{CD}}(x_{m,n,b}, y_{m,n,b})\big)\big(\frac{\partial f_{\overline{CD}}(0,0)}{\partial \mathbf{r}_m^{\mathrm{BS}}} \\ \cdot f_{\overline{CD}}(x_{m,n,b}, y_{m,n,b}) + \frac{\partial f_{\overline{CD}}(x_{m,n,b}, y_{m,n,b})}{\partial \mathbf{r}_m^{\mathrm{BS}}} f_{\overline{CD}}(0,0)\big) \\ \frac{\partial \mathcal{I}_{m,n,b}^{(4)}}{\partial \mathbf{r}_m^{\mathrm{BS}}} \approx S'\big(f_{\overline{DE}}(0,0) f_{\overline{DE}}(x_{m,n,b}, y_{m,n,b})\big)\big(\frac{\partial f_{\overline{DE}}(0,0)}{\partial \mathbf{r}_m^{\mathrm{BS}}} \\ \cdot f_{\overline{DE}}(x_{m,n,b}, y_{m,n,b}) + \frac{\partial f_{\overline{DE}}(x_{m,n,b}, y_{m,n,b})}{\partial \mathbf{r}_m^{\mathrm{BS}}} f_{\overline{DE}}(0,0)\big) \\ \frac{\partial \mathcal{I}_{m,n,b}^{(5)}}{\partial \mathbf{r}_m^{\mathrm{BS}}} \approx S'\big(f_{\overline{EF}}(0,0) f_{\overline{EF}}(x_{m,n,b}, y_{m,n,b})\big)\big(\frac{\partial f_{\overline{EF}}(0,0)}{\partial \mathbf{r}_m^{\mathrm{BS}}} \\ \cdot f_{\overline{EF}}(x_{m,n,b}, y_{m,n,b}) + \frac{\partial f_{\overline{EF}}(x_{m,n,b}, y_{m,n,b})}{\partial \mathbf{r}_m^{\mathrm{BS}}} f_{\overline{EF}}(0,0)\big) \\ \frac{\partial \mathcal{I}_{m,n,b}^{(6)}}{\partial \mathbf{r}_m^{\mathrm{BS}}} \approx S'\big(f_{\overline{FA}}(0,0) f_{\overline{FA}}(x_{m,n,b}, y_{m,n,b})\big)\big(\frac{\partial f_{\overline{FA}}(0,0)}{\partial \mathbf{r}_m^{\mathrm{BS}}} \\ \cdot f_{\overline{FA}}(x_{m,n,b}, y_{m,n,b}) + \frac{\partial f_{\overline{FA}}(x_{m,n,b}, y_{m,n,b})}{\partial \mathbf{r}_m^{\mathrm{BS}}} f_{\overline{FA}}(0,0)\big), \end{cases} \tag{62}$$

where

$$\begin{cases} \frac{\partial f_{\overline{AB}}(0,0)}{\partial \mathbf{r}_m^{\mathrm{BS}}} = \frac{\partial f_{\overline{DE}}(0,0)}{\partial \mathbf{r}_m^{\mathrm{BS}}} = \frac{1}{2}(\sin\theta_{m,n} \frac{\partial d_{m,n}}{\partial \mathbf{r}_m^{\mathrm{BS}}} + d_{m,n} \cos\theta_{m,n} \frac{\partial \theta_{m,n}}{\partial \mathbf{r}_m^{\mathrm{BS}}}) \\ \frac{\partial f_{\overline{BC}}(0,0)}{\partial \mathbf{r}_m^{\mathrm{BS}}} = \frac{\partial f_{\overline{EF}}(0,0)}{\partial \mathbf{r}_m^{\mathrm{BS}}} = \frac{1}{2}(\cos\theta_{m,n} \frac{\partial d_{m,n}}{\partial \mathbf{r}_m^{\mathrm{BS}}} - d_{m,n} \sin\theta_{m,n} \frac{\partial \theta_{m,n}}{\partial \mathbf{r}_m^{\mathrm{BS}}}) \\ \frac{\partial f_{\overline{CD}}(0,0)}{\partial \mathbf{r}_m^{\mathrm{BS}}} = \frac{\partial f_{\overline{FA}}(0,0)}{\partial \mathbf{r}_m^{\mathrm{BS}}} = \frac{E_{\mathrm{bui}}}{2}(\cos\theta_{m,n} - \sin\theta_{m,n}) \frac{\partial \theta_{m,n}}{\partial \mathbf{r}_m^{\mathrm{BS}}} \\ \frac{\partial f_{\overline{AB}}(x_{m,n,b}, y_{m,n,b})}{\partial \mathbf{r}_m^{\mathrm{BS}}} = (x_{m,n,b} \cos\theta_{m,n} + y_{m,n,b} \sin\theta_{m,n} \\ + \frac{d_{m,n} \cos\theta_{m,n}}{2}) \frac{\partial \theta_{m,n}}{\partial \mathbf{r}_m^{\mathrm{BS}}} + \sin\theta_{m,n} \frac{\partial x_{m,n,b}}{\partial \mathbf{r}_m^{\mathrm{BS}}} - \cos\theta_{m,n} \frac{\partial y_{m,n,b}}{\partial \mathbf{r}_m^{\mathrm{BS}}} \\ + \frac{\sin\theta_{m,n}}{2} \frac{\partial d_{m,n}}{\partial \mathbf{r}_m^{\mathrm{BS}}} \\ \frac{\partial f_{\overline{BC}}(x_{m,n,b}, y_{m,n,b})}{\partial \mathbf{r}_m^{\mathrm{BS}}} = (y_{m,n,b} \cos\theta_{m,n} - x_{m,n,b} \sin\theta_{m,n} \\ - \frac{d_{m,n} \sin\theta_{m,n}}{2}) \frac{\partial \theta_{m,n}}{\partial \mathbf{r}_m^{\mathrm{BS}}} + \cos\theta_{m,n} \frac{\partial x_{m,n,b}}{\partial \mathbf{r}_m^{\mathrm{BS}}} + \sin\theta_{m,n} \frac{\partial y_{m,n,b}}{\partial \mathbf{r}_m^{\mathrm{BS}}} \\ + \frac{\cos\theta_{m,n}}{2} \frac{\partial d_{m,n}}{\partial \mathbf{r}_m^{\mathrm{BS}}} \\ \frac{\partial f_{\overline{CD}}(x_{m,n,b}, y_{m,n,b})}{\partial \mathbf{r}_m^{\mathrm{BS}}} = \frac{E_{\mathrm{bui}}}{2}\big(\cos\theta_{m,n} - \sin\theta_{m,n}\big) \frac{\partial \theta_{m,n}}{\partial \mathbf{r}_m^{\mathrm{BS}}} + \frac{\partial y_{m,n,b}}{\partial \mathbf{r}_m^{\mathrm{BS}}} \\ \frac{\partial f_{\overline{DE}}(x_{m,n,b}, y_{m,n,b})}{\partial \mathbf{r}_m^{\mathrm{BS}}} = (-x_{m,n,b} \cos\theta_{m,n} - y_{m,n,b} \sin\theta_{m,n} \\ + \frac{d_{m,n} \cos\theta_{m,n}}{2}) \frac{\partial \theta_{m,n}}{\partial \mathbf{r}_m^{\mathrm{BS}}} - \sin\theta_{m,n} \frac{\partial x_{m,n,b}}{\partial \mathbf{r}_m^{\mathrm{BS}}} + \cos\theta_{m,n} \frac{\partial y_{m,n,b}}{\partial \mathbf{r}_m^{\mathrm{BS}}} \\ + \frac{\sin\theta_{m,n}}{2} \frac{\partial d_{m,n}}{\partial \mathbf{r}_m^{\mathrm{BS}}} \\ \frac{\partial f_{\overline{EF}}(x_{m,n,b}, y_{m,n,b})}{\partial \mathbf{r}_m^{\mathrm{BS}}} = (-y_{m,n,b} \cos\theta_{m,n} + x_{m,n,b} \sin\theta_{m,n} \\ - \frac{d_{m,n} \sin\theta_{m,n}}{2}) \frac{\partial \theta_{m,n}}{\partial \mathbf{r}_m^{\mathrm{BS}}} - \cos\theta_{m,n} \frac{\partial x_{m,n,b}}{\partial \mathbf{r}_m^{\mathrm{BS}}} - \sin\theta_{m,n} \frac{\partial y_{m,n,b}}{\partial \mathbf{r}_m^{\mathrm{BS}}} \\ + \frac{\cos\theta_{m,n}}{2} \frac{\partial d_{m,n}}{\partial \mathbf{r}_m^{\mathrm{BS}}} \\ \frac{\partial f_{\overline{FA}}(x_{m,n,b}, y_{m,n,b})}{\partial \mathbf{r}_m^{\mathrm{BS}}} = \frac{E_{\mathrm{bui}}}{2}\big(\cos\theta_{m,n} - \sin\theta_{m,n}\big) \frac{\partial \theta_{m,n}}{\partial \mathbf{r}_m^{\mathrm{BS}}} - \frac{\partial y_{m,n,b}}{\partial \mathbf{r}_m^{\mathrm{BS}}} \end{cases} \tag{63}$$

with

$$\frac{\partial \theta_{m,n}}{\partial \mathbf{r}_m^{\mathrm{BS}}} = \Big(\frac{\pi}{2} S'\Big(\arctan \frac{y_n^{\mathrm{UE}} - y_m^{\mathrm{BS}}}{x_n^{\mathrm{UE}} - x_m^{\mathrm{BS}}}\Big) - 1\Big) \cdot \Big(1 + \Big(\frac{y_n^{\mathrm{UE}} - y_m^{\mathrm{BS}}}{x_n^{\mathrm{UE}} - x_m^{\mathrm{BS}}}\Big)^2\Big)^{-1} \Big[\frac{y_n^{\mathrm{UE}} - y_m^{\mathrm{BS}}}{(x_n^{\mathrm{UE}} - x_m^{\mathrm{BS}})^2}, \frac{1}{x_m^{\mathrm{BS}} - x_n^{\mathrm{UE}}}\Big] \tag{64}$$

based on (8) and (43). According to (6) and (7), the partial derivatives $\frac{\partial x_{m,n,b}}{\partial \mathbf{r}_m^{\mathrm{BS}}}$ and $\frac{\partial y_{m,n,b}}{\partial \mathbf{r}_m^{\mathrm{BS}}}$ are given by $\frac{\partial x_{m,n,b}}{\partial \mathbf{r}_m^{\mathrm{BS}}} = [\frac{\partial x_{m,n,b}}{\partial x_m^{\mathrm{BS}}}, \frac{\partial x_{m,n,b}}{\partial y_m^{\mathrm{BS}}}]$ and $\frac{\partial y_{m,n,b}}{\partial \mathbf{r}_m^{\mathrm{BS}}} = [\frac{\partial y_{m,n,b}}{\partial x_m^{\mathrm{BS}}}, \frac{\partial y_{m,n,b}}{\partial y_m^{\mathrm{BS}}}]$, respectively, where

$$\begin{cases} \frac{\partial x_{m,n,b}}{\partial x_m^{\mathrm{BS}}} = -\frac{1}{2} \cos\Big(\arctan \frac{y_n^{\mathrm{UE}} - y_m^{\mathrm{BS}}}{x_n^{\mathrm{UE}} - x_m^{\mathrm{BS}}}\Big) \\ \quad + \Big(1 + \Big(\frac{y_n^{\mathrm{UE}} - y_m^{\mathrm{BS}}}{x_n^{\mathrm{UE}} - x_m^{\mathrm{BS}}}\Big)^2\Big)^{-1} \frac{y_{m,n,b}(y_n^{\mathrm{UE}} - y_m^{\mathrm{BS}})}{(x_n^{\mathrm{UE}} - x_m^{\mathrm{BS}})^2} \\ \frac{\partial x_{m,n,b}}{\partial y_m^{\mathrm{BS}}} = -\frac{1}{2} \sin\Big(\arctan \frac{y_n^{\mathrm{UE}} - y_m^{\mathrm{BS}}}{x_n^{\mathrm{UE}} - x_m^{\mathrm{BS}}}\Big) \\ \quad + \Big(1 + \Big(\frac{y_n^{\mathrm{UE}} - y_m^{\mathrm{BS}}}{x_n^{\mathrm{UE}} - x_m^{\mathrm{BS}}}\Big)^2\Big)^{-1} \frac{y_{m,n,b}}{x_m^{\mathrm{BS}} - x_n^{\mathrm{UE}}} \\ \frac{\partial y_{m,n,b}}{\partial x_m^{\mathrm{BS}}} = \frac{1}{2} \sin\Big(\arctan \frac{y_n^{\mathrm{UE}} - y_m^{\mathrm{BS}}}{x_n^{\mathrm{UE}} - x_m^{\mathrm{BS}}}\Big) \\ \quad - \Big(1 + \Big(\frac{y_n^{\mathrm{UE}} - y_m^{\mathrm{BS}}}{x_n^{\mathrm{UE}} - x_m^{\mathrm{BS}}}\Big)^2\Big)^{-1} \frac{x_{m,n,b}(y_n^{\mathrm{UE}} - y_m^{\mathrm{BS}})}{(x_n^{\mathrm{UE}} - x_m^{\mathrm{BS}})^2} \\ \frac{\partial y_{m,n,b}}{\partial y_m^{\mathrm{BS}}} = -\frac{1}{2} \cos\Big(\arctan \frac{y_n^{\mathrm{UE}} - y_m^{\mathrm{BS}}}{x_n^{\mathrm{UE}} - x_m^{\mathrm{BS}}}\Big) \\ \quad - \Big(1 + \Big(\frac{y_n^{\mathrm{UE}} - y_m^{\mathrm{BS}}}{x_n^{\mathrm{UE}} - x_m^{\mathrm{BS}}}\Big)^2\Big)^{-1} \frac{x_{m,n,b}}{x_m^{\mathrm{BS}} - x_n^{\mathrm{UE}}}. \end{cases} \tag{65}$$

Finally, $\frac{\partial \mathcal{I}_{m,n}^{\mathrm{block}}}{\partial \mathbf{r}^{\mathrm{BS}}}$ can be obtained by combining (58)–(65).

APPENDIX D
DERIVATION OF THE APPROXIMATION GUARANTEE OF ALGORITHM 2

The derivation of the approximation guarantee of Algorithm 2 is mainly based on the following performance guarantee

result in [41]: For an optimization problem with a monotone submodular objective and K matroid constraints, the ratio of the solution obtained by a greedy algorithm to the optimal solution is lower bounded by $\frac{1}{K+1}$.

For Algorithm 2, it associates one UE with one BS in a greedy manner in each iteration to improve the objective function of Problem P2. In the following, we will further show that Problem P2 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 [42] as follows.

Definition 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 $f : 2^{\mathcal{V}} \to \mathbb{R}$, is submodular if*

$$f(\mathcal{S} \cup \{v\}) - f(\mathcal{S}) \geq f(\mathcal{T} \cup \{v\}) - f(\mathcal{T}), \tag{66}$$

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

$$f(\mathcal{S}) \leq f(\mathcal{T}), \tag{67}$$

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

Definition 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

$$Z = \{\mathcal{X} \subseteq \mathcal{V} : |\mathcal{X} \cap \mathcal{V}_l| \leq \gamma_l, \forall l = 1, 2, \cdots, L\}, \tag{68}$$

for some given parameters $\gamma_1, \gamma_2, \cdots, \gamma_L$.

In the user association problem P2, the constraint (17) indicates that UE n can associate with BS m only if BS m is physically accessible to UE n, i.e., $I\left(\mathcal{I}_{m,n}^{\text{out}} + \mathcal{I}_{m,n}^{\text{block}}\right) = 0$. In order to reformulate the objective function (16) as a set function, with constraint (17) taken into account, we define the ground set $\mathcal{V}$ as

$$\mathcal{V}=\{v_{m,n} : m=1, \cdots, M; n=1, \cdots, N; I\left(\mathcal{I}_{m,n}^{\text{out}}+\mathcal{I}_{m,n}^{\text{block}}\right)=0\}, \tag{69}$$

and the user association set $\mathcal{S}_{\text{a}}$ as a subset of $\mathcal{V}$ such that $v_{m,n} \in \mathcal{S}_{\text{a}}$ if and only if UE n associates with BS m, i.e., $\mathcal{I}_{m,n}^{\text{associate}} = 1$. Therefore, minimizing the objective (16) of Problem P2 becomes equivalent to maximizing the set function $f(\mathcal{S}_{\text{a}}) = |\mathcal{S}_{\text{a}}|$, which is equal to the number of UEs that can associate with BSs. As the set function $f(\cdot)$ satisfies

$$f(\mathcal{S} \cup \{v\}) - f(\mathcal{S}) = f(\mathcal{T} \cup \{v\}) - f(\mathcal{T}) \text{ and } f(\mathcal{S}) \leq f(\mathcal{T}), \tag{70}$$

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 1.

With the ground set $\mathcal{V}$ given in (69), the constraint (18) can be written as

$$\mathcal{S} \in Z_B, \tag{71}$$

where $Z_B = \{\mathcal{X} \subseteq \mathcal{V} : |\mathcal{X} \cap \mathcal{V}_m^B| \leq C, \forall m = 1, 2, \cdots, M\}$. $\mathcal{V}_m^B = \{v_{m,n} : n = 1, \cdots, N; I(\mathcal{I}_{m,n}^{\text{out}} + \mathcal{I}_{m,n}^{\text{block}}) = 0\}$ is the set containing all the possible user associations between BS m and all the UEs, which satisfies $\bigcup_{m=1,\cdots,M} \mathcal{V}_m^B = \mathcal{V}$ and $\mathcal{V}_i^B \cap \mathcal{V}_j^B = \emptyset, \forall i \neq j$. Based on Definition 2, we can see that $(\mathcal{V}, Z_B)$ is a partition matroid, and hence (71) is a matroid constraint [41].

Similarly, the constraint (19) can also be written as

$$\mathcal{S} \in Z_U, \tag{72}$$

where $Z_U = \{\mathcal{X} \subseteq \mathcal{V} : |\mathcal{X} \cap \mathcal{V}_n^U| \leq 1, \forall n = 1, 2, \cdots, N\}$. $\mathcal{V}_n^U = \{v_{m,n} : m = 1, \cdots, M; I(\mathcal{I}_{m,n}^{\text{out}} + \mathcal{I}_{m,n}^{\text{block}}) = 0\}$ is the set containing all the possible user associations between UE n and all the BSs, which satisfies $\bigcup_{n=1,\cdots,N} \mathcal{V}_n^U = \mathcal{V}$ and $\mathcal{V}_i^U \cap \mathcal{V}_j^U = \emptyset, \forall i \neq j$. It is clear from Definition 2 that $(\mathcal{V}, Z_U)$ is also a partition matroid. Therefore, (72) is also a matroid constraint [41].

So far, we have shown that Problem P2 can be reformulated as a submodular optimization problem with two matroid constraints. Therefore, based on the performance guarantee result in [41], P2 can be solved by a greedy algorithm, e.g., Algorithm 2, with a constant-factor $\frac{1}{3}$ approximation guarantee.

ACKNOWLEDGEMENT

The authors would like to thank the editor, Dr. Narayan Prasad, for his insightful comments on the maximum weighted bipartite matching formulation of the user association problem, and the optimality gap of the proposed low-complexity user association algorithm.

REFERENCES

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

[2] *Special Issue: Millimeter wave communications for future mobile networks, Part I, IEEE J. Sel. Areas Commun.,* vol. 35, no. 7, Jul. 2017.

[3] X. Wang, L. Kong, F. Kong, F. Qiu, M. Xia, S. Arnon, and G. Chen, "Millimeter wave communication: a comprehensive survey," *IEEE Commun. Surveys Tuts.,* vol. 20, no. 3, pp. 1616–1653, 3rd Quart. 2018.

[4] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: potentials and challenges," *Proc. IEEE,* vol. 102, no. 3, pp. 366–385, Mar. 2014.

[5] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.,* vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[6] D. Liu, L. Wang, Y. Chen, M. Elkashlan, K.-K. Wong, R. Schober, and L. Hanzo, "User association in 5G networks: a survey and an outlook," *IEEE Commun. Surveys Tuts.,* vol. 18, no. 2, pp. 1018–1044, 2nd Quart. 2016.

[7] H. S. Ghadikolaei, C. Fischione, G. Fodor, P. Popovski, and M. Zorzi, "Millimeter wave cellular networks: a MAC layer perspective," *IEEE Trans. Commun.,* vol. 63, no. 10, pp. 3437–3458, Oct. 2015.

[8] 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.

[9] 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.

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

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[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] Y. Zhang and L. Dai, "Joint optimization of placement and coverage of access points for IEEE 802.11 networks," in *Proc. IEEE ICC 2020*, pp. 1–7.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

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

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

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

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

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

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

[35] 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.

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

[37] 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.

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

[39] 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.

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

[41] 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.

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

[43] 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.

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

[45] 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.

Yue Zhang (Member, IEEE) received his B.S. and M.S. degrees from South China University of Technology, Guangzhou, China, in 2012 and 2015, respectively. He is currently pursuing the Ph.D. degree with the Department of Electrical Engineering, City University of Hong Kong, Hong Kong. His research interests include capacity analysis of distributed antenna systems and access point placement optimization of wireless communication networks.

Lin Dai (Senior Member, IEEE) received the B.S. degree from the Huazhong University of Science and Technology, Wuhan, China, in 1998, and the M.S. and Ph.D. degrees from Tsinghua University, Beijing, China, in 2003, all in electronic engineering. She was a Post-Doctoral Fellow with the Hong Kong University of Science and Technology and the University of Delaware. Since 2007, she has been with the City University of Hong Kong, where she is currently a Full Professor. She has broad interests in communications and networking theory, with special interests in wireless communications. She was a co-recipient of the Best Paper Award at the IEEE Wireless Communications and Networking Conference (WCNC) 2007 and the IEEE Marconi Prize Paper Award in 2009.

Eric W. M. Wong (Senior Member, IEEE) received the B.Sc. and M.Phil. degrees in electronic engineering from the Chinese University of Hong Kong, Hong Kong, in 1988 and 1990, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Massachusetts, Amherst, MA, USA, in 1994. He is currently an Associate Professor with the Department of Electrical Engineering, City University of Hong Kong, Hong Kong. His research interests include analysis and design of telecommunications and computer networks, energy-efficient data center design, green cellular networks, 5G millimeter wave communication networks and optical networking.