Rate Maximization through Structured Spectrum Allocation and User Association in Heterogeneous Cellular Networks

Wei Bao, *Student Member, IEEE*, and Ben Liang, *Senior Member, IEEE*

Abstract—We study joint spectrum allocation and user association in heterogeneous cellular networks with multiple tiers of base stations. A stochastic geometric approach is applied as the basis to derive the average downlink user data rate in a closed-form expression. Then, the expression is employed as the objective function in jointly optimizing spectrum allocation and user association, which is of non-convex programming in nature. A computationally efficient Structured Spectrum Allocation and User Association (SSAUA) approach is proposed, solving the problem optimally and asymptotically optimally in two regions divided by a parameter specific threshold. A Surcharge Pricing Scheme (SPS) is also presented, such that the designed association bias values can be achieved in Nash equilibrium. Simulations and numerical studies are conducted to validate the accuracy and efficiency of the proposed SSAUA approach and SPS.

Index Terms—Heterogeneous cellular network, stochastic geometry, user association, spectrum allocation.

I. INTRODUCTION

Traditional single-tiered macro-cellular networks provide wide coverage for mobile user equipments (UEs), but they are insufficient to satisfy the exploding demand driven by modern mobile traffic, such as multimedia transmissions and cloud computing tasks. One efficient means to alleviate this problem is to install a diverse set of small-cells (e.g., picocells and femtocells), overlaying the macrocells, to form a multi-tiered heterogeneous cellular network [1]. Each small-cell is e-quipped with a shorter-range and lower-cost base station (BS), to provide nearby UEs with higher-quality communication links with lower power usage.

However, in the presence of multiple tiers of BSs in a cellular network, user association control becomes more challenging. A most direct approach is association by maximum received power, in which UEs are associated with the BS (in any tier) with the highest received power. However, in this case, because small-cell BSs transmit at lower power levels, only a small number of UEs close to them will connect with them, while most other UEs crowd in macrocells, leading to degraded performance. An example is shown in Fig. 1(a), in

The authors are with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, M5S 3G4, Canada (e-mail: wbao@ece.utoronto.ca; liang@ece.utoronto.ca).

(a) Association by maximum received power.
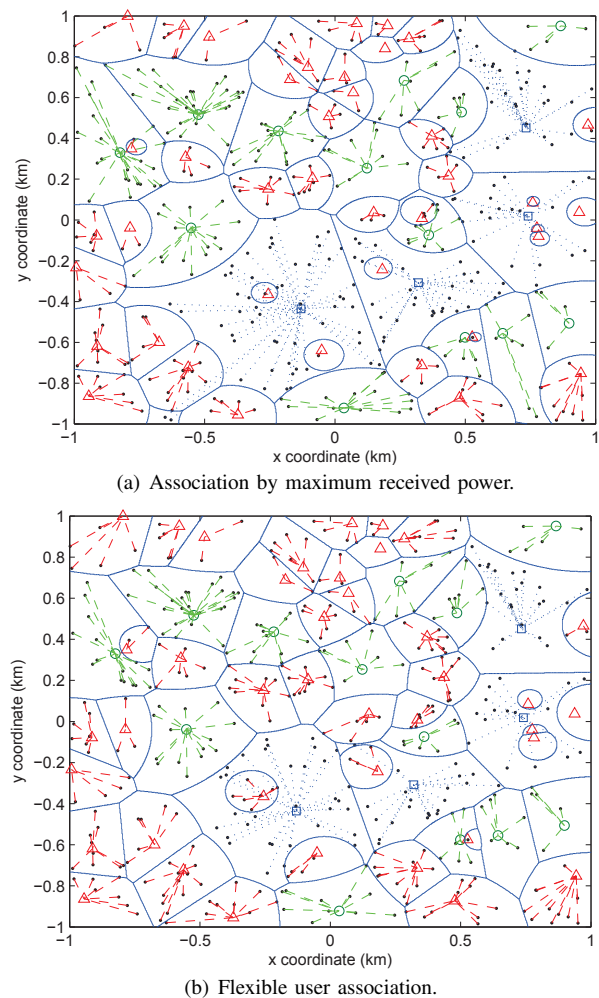

(b) Flexible user association.

Fig. 1. An example of a three-tier cellular network. Macrocell BSs, picocell BSs, and femtocell BSs are represented by squares, circles, and triangles respectively; UEs are represented by dots; blue solid lines show cell boundaries; dashed lines represent connections between UEs and BSs.

which many UEs occupy the macrocells, while some small-cells are nearly empty.

In order to resolve this issue, a *flexible user association* approach (also called ranged expansion in some literature) may be employed [2]–[6], in which each tier of BSs is assigned a *user association bias value*, and a UE is associated with a BS with the maximum received power multiplied by the bias value. If small-cell BSs are assigned larger association bias values, the small-cells are "expanded" accordingly. This can result in a more balanced mobile traffic pattern and thus better

network performance. Fig. 1(b) shows an example of flexible user association. However, if the association bias values for small-cell BSs are too large, it will cause improper expansion of small-cells such that UEs at their cell-edge may suffer from inadequate received power. As a consequence, the association bias values should be properly designed so that the overall network performance is optimized.

Further complicating the resource management problem in a multi-tier cellular network, the radio spectrum licensed by the network operator needs to be shared by BSs of widely different power and coverage areas. How to optimally allocate spectrum among different tiers is an important open problem. In order to avoid cross-tier interference, and the prohibitive complexity in tracking and provisioning for such interference especially with unplanned deployment of small cells, a disjoint spectrum mode is commonly advocated [7]–[9], where different tiers of BSs are allocated non-overlapping portions of the spectrum. Even so, it is still a challenging problem to properly divide the spectrum for optimal network performance.

In this work, our objective is to study jointly optimal spectrum allocation and user association in a heterogeneous celluar network with multiple tiers of BSs. First, we develop a stochastic geometric model to study the network performance analytically. A closed-form expression for the average downlink UE data rate is derived, which is then employed as the objective function for jointly optimizing the spectrum allocation among tiers and the user association bias values.

This resultant optimization problem is of non-convex programming in nature and cannot be solved with a standard method. Instead, we explore two important structures in solving the problem. Referred to as the *density thresholding* structure, we show that the problem can be studied separately over two regions, divided by a parameter specific UE density threshold. Referred to as the *priority ordering* structure, we show that a tier with higher BS density should have higher priority in spectrum allocation. Based on these observations, we propose a computationally efficient Structured Spectrum Allocation and User Association (SSAUA) approach to solve the problem optimally and asymptotically optimally in the two density regions respectively.

Finally, toward practical implementation of SSAUA, we propose a Surcharge Pricing Scheme (SPS), such that the designed association bias values can be achieved in Nash equilibrium. Hence, each UE is incentivized to adopt the proposed design with individual rationality.

The rest of the paper is organized as follows. In Section II, we discuss the relation between our work and prior works. In Section III, we describe the system model. In Sections IV, V, and VI, we present our contributions in UE data rate derivation, SSAUA design, and SPS, respectively. In Section VII, we present the extension of our model considering multiple modulation and coding schemes (MCSs). In Section VIII, we present numerical results. Finally, conclusions are given in Section IX.

II. RELATED WORKS

A. Stochastic Geometry as Analytical Basis

Stochastic geometry [10]–[14] is a powerful mathematical modeling tool to analyze the performance (e.g., outage probability and data rate) of cellular networks with random spatial patterns of UEs and BSs. In this work, we focus on the downlink user data rate as performance measure. Pioneering works on downlink performance analysis using stochastic geometry include [15], [16], and [17], for either the single-tier or the multi-tier case. None of them considered spectrum allocation or user association.

B. Resource Allocation and User Association

Assuming a fixed number of UEs and BSs and without considering their random spatial patterns, the authors of [18]–[21] studied optimal power allocation in cellular networks; the authors of [3], [4] investigated performance benefit introduced by flexible user association; and the authors of [22]–[27] studied optimal user association with deterministic utility optimization.

With a stochastic geometric approach, some prior studies considered either resource allocation or user association separately. For example, Cheung *et al.* [9] studied optimal spectrum allocation, in cellular networks limited to two tiers of BSs, without flexible user association. Bao and Liang [28], [29] compared the outage performance between open and closed small-cell access modes in a two-tier network with complete spectrum sharing. Jo *et al.* [5] presented the performance evaluation of flexible user association model with bias values. They derived the coverage probability and UE data rate, considering cross-tier interference, but in non-closed forms. They did not provide a means to optimize the derived performance metrics. This work was later extended in [6] to study optimal user association in a network with two tiers of BSs, without considering spectrum allocation.

Through stochastic geometric analysis, spectrum allocation and user association were jointly studied in [30]–[33]. Singh and Andrews [30] analyzed the network performance in terms of coverage probability and data rate under different spectrum allocation and user association settings. However, the optimal spectrum allocation and user association were not derived. A similar problem was also studied by Lin and Yu [31], with frequency reuse instead of tiered spectrum division as the approach for spectrum sharing. They provided conditionally optimal user association given frequency reuse factors or conditionally optimal frequency reuse factors given user association bias values. However, joint optimization remained an open problem.

Compared with the above studies, we consider multiple tiers of BSs with disjoint spectrum, and provide optimal and analytically bounded asymptotically optimal solutions for joint spectrum allocation and user association. The closest works to ours may be [32], [33], where the authors extended the analytical model in [31] to study joint spectrum allocation and user association. However, [32], [33] aimed to optimize the average log-utility per UE, such that the logarithm function cancels the exponential term corresponding to the Laplace

transform of interference [10], [11]. This alternate objective led to substantial reduction in analytical complexity. Different from [32], [33], our model aims to optimize the average UE data rate directly, leading to a much more challenging problem and requiring the proposed structured optimization solution. Furthermore, only the single-MCS case was considered in [32], [33], while we additionally study multiple MCSs in this work.

A preliminary version of this work was presented in [34], where only the single-MCS case was considered. The current version also contains substantial further analytical details, simulation results, and discussion.

III. System Model

A. Multi-tier Cellular Network

We consider a heterogeneous cellular network with random-ly spatially distributed $K \geq 2$ tiers of BSs. As in conventional stochastic geometric modeling of multi-tier cellular networks [5], [6], [16], [30], [31], each tier of BSs independently form a homogeneous Poisson point process (PPP) in two-dimensional Euclidean space $\mathbb{R}^2$. Let Φ_k denote the PPP corresponding to tier-k BSs, with intensity λ_k. Without loss of generality, we assume that $\lambda_1 < \lambda_2 \ldots < \lambda_K$. (If $\lambda_i = \lambda_j, i \neq j$ in reality, we may approximate by setting $\lambda_j = \lambda_i + \xi$, where ξ is arbitrarily close to 0.) UEs are also modeled as a homogeneous PPP Ψ with intensity μ, independent of all BSs. We assume each BS is connected to the core network by separate high-capacity wired or wireless links that have no influence on our performance analysis. In addition, because we focus on downlink analysis, we assume that the downlink and uplink of the system are operated in different spectra, so that uplink interference and capacity have no influence on downlink analysis.

B. Power and Path Loss Model

We define the tiers of BSs by their transmission power. Let P_k be the transmission power of tier-k BSs, which is a given parameter. Let $\mathbf{x}$ and $\mathbf{y}$ denote some two-dimensional coordinates throughout this paper. If $P_t(\mathbf{x})$, $P_t(\mathbf{x}) \in \{P_1, P_2, \ldots, P_K\}$, is the transmission power from a BS at $\mathbf{x}$ and $P_r(\mathbf{y})$ is the received power at $\mathbf{y}$, we have $P_r(\mathbf{y}) = \frac{P_t(\mathbf{x})h_{\mathbf{x},\mathbf{y}}}{\alpha |\mathbf{x}-\mathbf{y}|^\gamma}$, where $\alpha |\mathbf{x}-\mathbf{y}|^\gamma$ is the propagation loss function, α is a factor accounts for system losses, $\gamma > 2$ is the path loss exponent, and $h_{\mathbf{x},\mathbf{y}}$ is the fast fading term. We assume that α and γ are constant for all tiers. Corresponding to common Rayleigh fading with power normalization, $h_{\mathbf{x},\mathbf{y}}$ is independently exponentially distributed with unit mean. Let $\mathfrak{h}(\cdot)$ be the probability density function of $h_{\mathbf{x},\mathbf{y}}$.

C. Spectrum Allocation

In order to avoid cross-tier interference, different tiers of BSs are allocated separated spectrum. Assume the total spectrum bandwidth is W. The network operator allocates $\eta_k W$ to each tier-k BS, where η_k is the spectrum allocation factor and $\sum_{k=1}^{K} \eta_k = 1$. Let $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_K)$. Note that BSs in the same tier are operated on the same spectrum.

We additionally consider the possible constraints $\eta_{\min,k} \leq \eta_k \leq \eta_{\max,k}$, for $k = 1, 2, \ldots, K$. Clearly, we have $\sum_{k=1}^{K} \eta_{\min,k} \leq 1 \leq \sum_{k=1}^{K} \eta_{\max,k}$. Furthermore, we assume that $0 < \eta_{\min,1} \leq \eta_{\min,2} \ldots \leq \eta_{\min,K}$ and $0 < \eta_{\max,1} \leq \eta_{\max,2} \ldots \leq \eta_{\max,K}$, i.e., the network operator is likely (but not necessarily) to allocate more spectrum to a tier with higher BS density.

Given a specific tier-k BS, it is common to assume that all its associated UEs are equally allocated spectrum [5], [9], [31]. Hence, the per-UE assigned spectrum bandwidth is $\beta_k = \eta_k W/N_k$, where N_k is a random variable denoting the number of UEs associated with the BS.

D. Coverage Probability and UE Data Rate

Following conventional stochastic geometric modeling [5], [9], [31], in Section IV and V, we assume that UEs employ a single modulations coding scheme (MCS). In this case, let T denote the minimum required Signal-to-Interference Ratio (SIR) of UEs. The coverage probability of a UE is defined as the probability that its SIR is no lower than T [10]. Initially, we assume that if a UE experiences coverage probability $\mathbb{P}'$ and is allocated spectrum bandwidth β', its data rate is $\beta' \log(1+T)$ if the SIR is no lower than T, and its data rate is 0 if the SIR is lower than T (i.e., outage occurs). Thus, the overall data rate of the UE is $\beta' \log(1+T)\mathbb{P}'$. In Section VII, we present the extension of our model considering multiple MCSs, where N different threshold values $T_1, \ldots, T_N$ are accommodated.

Note that $\log$ is in base 2 throughout this paper. Also, we have assumed the system is interference limited [6], [9], [18], [31]–[33], such that noise is negligible.

E. Flexible User Association

Given that a UE is located at $\mathbf{y}$, it associates itself with the BS that provides the maximum *biased received power* [5], [6], [31] as follows:

$$\mathcal{BS}(\mathbf{y}) = \arg \max_{\mathbf{x} \in \Phi_k, \forall k} \frac{1}{\alpha} B_k P_k |\mathbf{x}-\mathbf{y}|^{-\gamma}, \quad (1)$$

where $\mathcal{BS}(\mathbf{y})$ denotes the location of the BS associated with the UE, and $\frac{1}{\alpha} P_k |\mathbf{x}-\mathbf{y}|^{-\gamma}$ is the received power from a tier-k BS located at $\mathbf{x}$, and B_k is the association bias, indicating the connecting preference of a UE toward tier-k BSs. In this case, the resultant cell splitting forms a generalized Dirichlet tessellation, or weighted Poisson Voronoi [35], shown in Fig. 1(b). Note that for $B_1, B_2, \ldots, B_K$, their effects remain the same if we multiply all of them by the same positive constant. Thus, without loss of generality, in this paper, we normalize them such that $\sum_{k=1}^{K} B_k = 1$. Let $\mathbf{B} = (B_1, B_2, \ldots, B_K)$.

Let A_k denote the probability that a UE associates itself with a tier-k BS, and $\mathbf{A} = (A_1, A_2, \ldots, A_K)$. As derived in [5], we have

$$A_k = \frac{\lambda_k (P_k B_k)^{\frac{2}{\gamma}}}{\sum_{j=1}^{K} \lambda_j (P_j B_j)^{\frac{2}{\gamma}}}, \quad (2)$$

and thus

$$B_k = \frac{P_k^{-1}(A_k/\lambda_k)^{\frac{\gamma}{2}}}{\sum_{j=1}^{K} P_j^{-1}(A_j/\lambda_j)^{\frac{\gamma}{2}}}. \quad (3)$$

Hence, there is a one-to-one mapping between $\mathbf{A}$ and $\mathbf{B}$, so we can view them interchangeably.

F. Problem Statement

We first aim to derive a closed-form expression for the average UE data rate. Then, our objective is to maximize the average UE data rate by jointly optimizing the spectrum allocation factors $\boldsymbol{\eta}$ and the user association bias values $\mathbf{B}$ (or equivalently $\mathbf{A}$). Finally, we give a pricing scheme to incentivize each UE to adopt the designed $\mathbf{B}$.

IV. CLOSED-FORM AVERAGE UE DATA RATE

In this section, we derive the average UE data rate via stochastic geometric analysis. Consider a reference UE, termed the *typical UE*, communicating with its BS, termed the *typical BS*. Due to the stationarity of UEs and BSs, throughout this section we will re-define the coordinates so that the typical UE is located at $\mathbf{0}$. We are interested in the typical UE since the average UE performance in the system is the same as the performance of the typical UE [10].

First, we study the coverage probability given that the typical UE is associating with a tier-k BS and their distance is d. In this case, the overall interference to the typical UE is the sum interference from all tier-k BSs other than the typical BS. Let $I_k(d)$ denote such interference. Then

$$I_k(d) = \sum_{\mathbf{x} \in \Phi'_k} \frac{P_k h_{\mathbf{x},\mathbf{0}}}{\alpha |\mathbf{x}|^\gamma}. \tag{4}$$

where Φ'_k is the reduced Palm point process corresponding to all tier-k BSs other than the typical BS, given that the typical BS is located at a distance of d from the typical UE. It can be shown that Φ'_k is a PPP with intensity 0 in $\mathcal{B}(\mathbf{0},d)$ and intensity λ_k in $\mathbb{R}^2 \setminus \mathcal{B}(\mathbf{0},d)$, where $\mathcal{B}(\mathbf{0},d)$ denotes the disk region centered at $\mathbf{0}$ with radius d [10].

The distribution of $I_k(d)$ is derived through its Laplace transform as follows:

$$\mathcal{L}_{I_k}(d,s) = \mathbf{E}\left[\exp\left(-\sum_{\mathbf{x}\in\Phi'_k} \frac{sP_k h_{\mathbf{x},\mathbf{0}}}{\alpha|\mathbf{x}|^\gamma}\right)\right]$$

$$= \exp\left(-\lambda_k \int_{\mathbb{R}^2\setminus\mathcal{B}(\mathbf{0},d)} \left(1 - \int_{\mathbb{R}^+} e^{-\frac{sP_k h}{\alpha|\mathbf{x}|^\gamma}} \mathfrak{h}(h)\mathrm{d}h\right)\mathrm{d}\mathbf{x}\right) \tag{5}$$

$$= \exp\left(-\lambda_k \int_{\mathbb{R}^2\setminus\mathcal{B}(\mathbf{0},d)} \frac{\frac{sP_k}{\alpha|\mathbf{x}|^\gamma}}{\frac{sP_k}{\alpha|\mathbf{x}|^\gamma}+1}\mathrm{d}\mathbf{x}\right) \tag{6}$$

$$= \exp\left(-2\pi\lambda_k \int_d^\infty \frac{\frac{sP_k r}{\alpha}}{\frac{sP_k}{\alpha}+r^\gamma}\mathrm{d}r\right), \tag{7}$$

where (5) is obtained from the Laplace functional of PPP Φ'_k [10], (6) is because the fading term is exponentially distributed with unit mean, and (7) is through a transformation to polar coordinates.

Let $\mathbb{P}_{cover,k}(d)$ denote the conditional coverage probability of the typical UE (given k and d). Then

$$\mathbb{P}_{cover,k}(d) = \mathbb{P}\left(\frac{P_k h_{\mathbf{x}_B,\mathbf{0}}}{\alpha d^\gamma} \geq T I_k(d)\right)$$

$$= \mathcal{L}_{I_k}(d,s)\big|_{s=\frac{T\alpha d^\gamma}{P_k}}, \tag{8}$$

where $\mathbf{x}_B$ is the coordinate of the typical BS, and $|\mathbf{x}_B| = d$. Substituting (7) into (8), we have

$$\mathbb{P}_{cover,k}(d) = \exp\left(-2\pi\lambda_k \int_d^\infty \frac{Td^\gamma r}{Td^\gamma + r^\gamma}\mathrm{d}r\right)$$

$$\stackrel{t=\frac{r^2}{T^{2/\gamma}d^2}}{=} \exp\left(-\pi\lambda_k T^{\frac{2}{\gamma}} d^2 \int_{(\frac{1}{T})^{\frac{2}{\gamma}}}^\infty \frac{1}{1+t^{\frac{\gamma}{2}}}\mathrm{d}t\right). \tag{9}$$

Furthermore, the probability density function of the distance between the typical UE and its associated tier-k BS is

$$f_k(d) = \frac{2\pi\lambda_k}{A_k} d \exp\left(-\pi d^2 \sum_{j=1}^K \lambda_j \left(\frac{P_j B_j}{P_k B_k}\right)^{\frac{2}{\gamma}}\right) \tag{10}$$

$$= \frac{2\pi\lambda_k}{A_k} d \exp\left(-\pi d^2 \frac{\lambda_k}{A_k}\right), \tag{11}$$

where (10) is derived in [5], and (11) is by substituting (2) into (10).

Hence, the coverage probability $\mathbb{P}_{cover,k}$ of the typical UE associated with a tier-k BS can be computed as

$$\mathbb{P}_{cover,k} = \int_0^\infty f_k(d)\mathbb{P}_{cover,k}(d)\mathrm{d}d$$

$$= \int_0^\infty \frac{2\pi\lambda_k}{A_k} d \exp\left(-\pi d^2 \frac{\lambda_k}{A_k}\right)$$

$$\qquad \exp\left(-\pi\lambda_k (T)^{\frac{2}{\gamma}} d^2 \int_{(\frac{1}{T})^{\frac{2}{\gamma}}}^\infty \frac{1}{1+t^{\gamma/2}}\mathrm{d}t\right)\mathrm{d}d$$

$$= \frac{\pi\lambda_k}{A_k} \frac{1}{\pi\frac{\lambda_k}{A_k} + \pi\lambda_k (T)^{\frac{2}{\gamma}} \int_{(\frac{1}{T})^{\frac{2}{\gamma}}}^\infty \frac{1}{1+t^{\gamma/2}}\mathrm{d}t}$$

$$= \frac{1}{1+A_k C}, \tag{12}$$

where $C \triangleq (T)^{\frac{2}{\gamma}} \int_{(\frac{1}{T})^{\frac{2}{\gamma}}}^\infty \frac{1}{1+t^{\gamma/2}}\mathrm{d}t$ is a system-level constant only related to γ and T. Note that the coverage probability is given in non-closed form in [5] for a system where the spectrum is shared by all tiers. Here we are able to obtain a closed-form expression, mainly as a consequence of different tiers using separate spectrum.

Let $\mathbf{E}_0(\beta_k)$ denote the expected spectrum bandwidth allocated to the typical UE (connecting to a tier-k BS). Following the model in Section III-C, $\mathbf{E}_0(\beta_k)$ equals the spectrum bandwidth allocated to the typical tier-k BS divided by the average number of UEs associated with it conditioned on the typical UE, which is $A_k\mu/\lambda_k + 1$. Hence,

$$\mathbf{E}_0(\beta_k) = \frac{\eta_k W}{A_k\mu/\lambda_k + 1}. \tag{13}$$

Then, by Section III-D, the conditional expected data rate of the typical UE, given it is associated with a tier-k BS, can be computed as [5], [31]

$$\overline{R}_k = \mathbf{E}_0(\beta_k) \log(1+T)\mathbb{P}_{cover,k}. \tag{14}$$

Note that by doing so, we slightly underestimate the average data rate because the coverage event and β_k are not completely

independent. Although some efforts have been made to approximate their correlation [6], [36], all of them are inexact but result in tremendous mathematical complexity. In Section VIII, we show that the resultant analysis is close to actual performance via simulation.

Finally, the average data rate of the typical UE, and hence the average data rate per UE in the system, is

$$\mathbf{F} = \sum_{k=1}^{K} A_k \overline{R}_k = \sum_{k=1}^{K} A_k \mathbf{E}_0(\beta_k) \log(1+T) \mathbb{P}_{cover,k}$$
$$= \sum_{k=1}^{K} \frac{\eta_k W \log(1+T)}{(A_k \mu/\lambda_k + 1)(\frac{1}{A_k} + C)}. \tag{15}$$

Note that stochastic geometric analysis often leads to non-closed forms requiring numerical integrations (e.g., [5], [6], [15], [16]), due to the integral form of the Laplace functional or generating functional of PPPs applied in analysis [10], [14]. Fortunately, our derived closed-form expression for the average UE data rate facilitates the tractability of the resultant optimization problem.

V. Joint Optimization Problem and SSAUA

We aim to maximize the average UE data rate $\mathbf{F}$ with respect to $\boldsymbol{\eta}$ and $\mathbf{B}$. As there is a one-to-one mapping between $\mathbf{A}$ and $\mathbf{B}$, we study the optimization problem over $(\boldsymbol{\eta}, \mathbf{A})$ instead for analytical convenience. This is formally stated as optimization problem $\mathbf{P}$ as follows:

$$\underset{\boldsymbol{\eta}, \mathbf{A}}{\text{maximize}} \quad \mathbf{F}(\boldsymbol{\eta}, \mathbf{A}) = \sum_{k=1}^{K} \eta_k M_k(A_k)$$

$$\text{subject to} \quad \sum_{k=1}^{K} \eta_k = 1, \quad \eta_{\min,k} \leq \eta_k \leq \eta_{\max,k}, \forall k,$$

$$\sum_{k=1}^{K} A_k = 1, \quad A_k \geq 0, \forall k, \tag{16}$$

where $M_k(A_k)$ is defined as

$$M_k(A_k) = \frac{1}{(A_k \mu/\lambda_k + 1)\left(\frac{1}{A_k} + C\right)}. \tag{17}$$

Problem $\mathbf{P}$ is non-convex and cannot be solved through a standard method. Instead, we investigate into two important structures of the optimal solution, termed density thresholding and priority ordering, based on which we propose a computationally efficient Structured Spectrum Allocation and User Association (SSAUA) approach to solve the problem.

A. Density Thresholding Structure

First, we define an important parameter

$$a_k \triangleq \sqrt{\lambda_k/(\mu C)}. \tag{18}$$

Note that $M_k(A_k)$ is increasing on $[0, a_k]$ and decreasing on $[a_k, \infty)$. We further observe several useful properties of $M_k(A_k)$, which are presented in Appendix A. Based on these properties, we obtain the following lemma, whose proof is given in Appendix B.

Lemma 1. *Consider a potential solution* $(\boldsymbol{\eta}^{**}, \mathbf{A}^{**})$ *to Problem* $\mathbf{P}$. *If* $\exists i \neq j$, *such that* $A_i^{**} < a_i$ *and* $A_j^{**} > a_j$, *then* $(\boldsymbol{\eta}^{**}, \mathbf{A}^{**})$ *is not an optimal solution.*

Lemma 1 suggests that, in an optimal solution, every A_k must be on the same side of a_k. This directly leads to the following theorem, which is fundamental to our optimization solution.

Theorem 1. *(Density Thresholding) Let* $(\boldsymbol{\eta}^*, \mathbf{A}^*)$ *be an optimal solution to Problem* $\mathbf{P}$. *If* $\sum_{k=1}^{K} a_k > 1$, *then* $\forall k, A_k^* \leq a_k$; *if* $\sum_{k=1}^{K} a_k < 1$, *then* $\forall k, A_k^* \geq a_k$; *if* $\sum_{k=1}^{K} a_k = 1$, *then* $\forall k, A_k^* = a_k$.

Proof. If $\sum_{k=1}^{K} a_k > 1$, because $\sum_{k=1}^{K} A_k^* = 1$, $\exists l$ such that $A_l^* < a_l$. This leads to $A_k^* \leq a_k, \forall k$, according to Lemma 1. The cases where $\sum_{k=1}^{K} a_k < 1$ and $\sum_{k=1}^{K} a_k = 1$ are similar. $\square$

Note that, the condition $\sum_{i=1}^{K} a_i > 1$ (i.e., $\sqrt{\frac{1}{C}}\left(\sum_{i=1}^{K} \sqrt{\lambda_i}\right) > \sqrt{\mu}$) is referred to as the *optimality region* throughout the paper, since the optimization problem $\mathbf{P}$ can be solved optimally in this case, as shown later in Section V-B. The condition $\sum_{i=1}^{K} a_i < 1$ (i.e., $\sqrt{\frac{1}{C}}\left(\sum_{i=1}^{K} \sqrt{\lambda_i}\right) < \sqrt{\mu}$) is referred to as the *asymptotic-optimality region* throughout the paper, since the optimization problem $\mathbf{P}$ can be solved asymptotically optimally in this case, as shown later in Section V-C. If $\sum_{i=1}^{K} a_i = 1$, Problem $\mathbf{P}$ can be trivially solved and is ignored in the rest of our discussion. Note that because a_k can be computed directly from the given parameters, one can judge in which region Problem $\mathbf{P}$ falls before solving the problem. Next, the solution to $\mathbf{P}$ will be investigated separately under $\sum_{i=1}^{K} a_i > 1$ and $\sum_{i=1}^{K} a_i < 1$.

B. SSAUA under $\sum_{i=1}^{K} a_i > 1$

In this case, the original Problem $\mathbf{P}$ becomes Problem $\mathbf{P1}$ as follows:

$$\underset{\boldsymbol{\eta}, \mathbf{A}}{\text{maximize}} \quad \mathbf{F}(\boldsymbol{\eta}, \mathbf{A}) = \sum_{k=1}^{K} \eta_k M_k(A_k)$$

$$\text{subject to} \quad \sum_{k=1}^{K} \eta_k = 1, \quad \eta_{\min,k} \leq \eta_k \leq \eta_{\max,k}, \forall k,$$

$$\sum_{k=1}^{K} A_k = 1, \quad 0 \leq A_k \leq a_k, \forall k. \tag{19}$$

We first observe an important ordering property of the optimal solution to $\mathbf{P1}$, as shown in the following lemma, whose proof is given in Appendix C.

Lemma 2. *(Ordering Property) If* $\mathbf{A}^*$ *is optimal for Problem* $\mathbf{P1}$, *then* $M_1(A_1^*) \leq M_2(A_2^*) \leq \ldots \leq M_K(A_K^*)$.

Next, by sequentially computing $\boldsymbol{\eta}^*$ as follows:

$$
\begin{cases}
\eta_K^* = \min(1 - \sum_{k=1}^{K-1} \eta_{\min,k}, \eta_{\max,K}), \\
\eta_{K-1}^* = \min(1 - \eta_K^* - \sum_{k=1}^{K-2} \eta_{\min,k}, \eta_{\max,K-1}), \\
\cdots, \\
\eta_l^* = \min(1 - \sum_{k=l+1}^{K} \eta_k^* - \sum_{k=1}^{l-1} \eta_{\min,k}, \eta_{\max,l}), \\
\cdots, \\
\eta_1^* = \min(1 - \sum_{k=2}^{K} \eta_k^*, \eta_{\max,1}),
\end{cases}
\tag{20}
$$

we have the following theorem:

Theorem 2. *(Priority Ordering) If $\mathbf{A}^*$ is optimal for Problem* **P1***, then $(\boldsymbol{\eta}^*, \mathbf{A}^*)$, where $\boldsymbol{\eta}^*$ is computed in (20), is an optimal solution to* **P1***.*

Proof. Consider Problem **P1A** as follows:

$$
\begin{aligned}
\underset{\boldsymbol{\eta}}{\text{maximize}} \quad & \sum_{k=1}^{K} \eta_k M_k(A_k^*) \\
\text{subject to} \quad & \sum_{k=1}^{K} \eta_k = 1, \quad \eta_{\min,k} \le \eta_k \le \eta_{\max,k}, \forall k.
\end{aligned}
\tag{21}
$$

It is a simple linear programming problem with ordered linear coefficients in the objective, since $M_1(A_1^*) \le M_2(A_2^*) \le \ldots \le M_K(A_K^*)$ due to Lemma 2. Note that $\boldsymbol{\eta}^*$ does not depend on the exact values of $\mathbf{A}^*$; it only requires the ordering property as shown in Lemma 2. Also, $\boldsymbol{\eta}^*$ is in the feasible region due to $\sum_{k=1}^{K} \eta_{\min,k} \le 1 \le \sum_{k=1}^{K} \eta_{\max,k}$. It is easy to verify that $(\boldsymbol{\eta}^*, \mathbf{A}^*)$ is an optimal solution to **P1**. $\square$

In Theorem 2, we note that (20) indicates a priority ordering structure in spectrum allocation. Tier-K has the highest priority in spectrum allocation, followed by tier-$(K-1)$, and so forth. Furthermore, Theorem 2 provides a means to derive an optimal $\boldsymbol{\eta}^*$ regardless of the $\mathbf{A}^*$ values. We need one further step to derive the corresponding optimal $\mathbf{A}^*$ by solving the following Problem **P1B**:

$$
\begin{aligned}
\underset{\mathbf{A}}{\text{maximize}} \quad & \sum_{k=1}^{K} \eta_k^* M_k(A_k) \\
\text{subject to} \quad & \sum_{k=1}^{K} A_k = 1, \quad 0 \le A_k \le a_k, \forall k.
\end{aligned}
\tag{22}
$$

Note that **P1B** is a convex programming problem, since $M_k(A_k)$ is concave on $[0, a_k]$. Thus, $\mathbf{A}^*$ can be computed by a computationally efficient algorithm, such as the interior-point method. Hence both steps to compute the jointly optimal solution $(\boldsymbol{\eta}^*, \mathbf{A}^*)$ have low computational complexity.

In summary, under $\sum_{i=1}^{K} a_i > 1$, the original optimization problem can be solved optimally, and thus $\sum_{i=1}^{K} a_i > 1$ is referred to as the optimality region.

C. SSAUA under $\sum_{i=1}^{K} a_i < 1$

In this case, the original Problem **P** becomes Problem **P2** as follows:

$$
\underset{\boldsymbol{\eta}, \mathbf{A}}{\text{maximize}} \quad \mathbf{F}(\boldsymbol{\eta}, \mathbf{A}) = \sum_{k=1}^{K} \eta_k M_k(A_k)
$$

$$
\begin{aligned}
\text{subject to} \quad & \sum_{k=1}^{K} \eta_k = 1, \quad \eta_{\min} \le \eta_k \le \eta_{\max}, \forall k, \\
& \sum_{k=1}^{K} A_k = 1, \quad A_k \ge a_k, \forall k.
\end{aligned}
\tag{23}
$$

Problem **P2** is more complicated than Problem **P1**, as $M_k(A_k)$ is not concave, but an S-shaped function, in the feasible region. Hence, **P2** generally incurs high computational complexity even if an optimal $\boldsymbol{\eta}^*$ is given [37], [38].

Therefore, instead of directly solving **P2**, we first approximate $M_k(A_k)$ by $\widetilde{M}_k(A_k)$ defined as follows:

$$
\widetilde{M}_k(A_k) = \frac{1}{(A_k \mu / \lambda_k)\left(\frac{1}{A_k} + C\right)}.
\tag{24}
$$

Note that this approximation is reasonable when $A_k \mu / \lambda_k$ is much larger than 1, e.g., when μ is large. This observation is also supported by the performance gap as derived in Section V-C2. Some useful properties of $\widetilde{M}_k(A_k)$ are shown in Appendix D.

The approximated problem is referred to as Problem **P2A**, where we simply replace the objective function of **P2** by the following:

$$
\mathbf{F}'(\boldsymbol{\eta}, \mathbf{A}) = \sum_{k=1}^{K} \eta_k \widetilde{M}_k(A_k).
\tag{25}
$$

1) Solution to **P2A***:* The important ordering property still holds for Problem **P2A**, as formalized in the following lemma, whose proof is given in Appendix E.

Lemma 3. *(Ordering Property) If $\widetilde{\mathbf{A}}^*$ is optimal for Problem* **P2A***, then $\widetilde{M}_1(\widetilde{A}_1^*) \le \widetilde{M}_2(\widetilde{A}_2^*) \le \ldots \le \widetilde{M}_K(\widetilde{A}_K^*)$.*

We observe that with the same ordering property, (20) can again be adopted as an optimal solution to **P2A**, leading to the following theorem:

Theorem 3. *(Priority Ordering) If $\widetilde{\mathbf{A}}^*$ is optimal for Problem* **P2A***, then $(\widetilde{\boldsymbol{\eta}}^*, \widetilde{\mathbf{A}}^*)$, where $\widetilde{\boldsymbol{\eta}}^*$ is computed the same way as $\boldsymbol{\eta}^*$ in (20), is an optimal solution to Problem* **P2A***.*

Proof. The proof is similar to that of Theorem 2. $\square$

Given an optimal $\widetilde{\boldsymbol{\eta}}^*$ for **P2A**, we find the corresponding optimal $\widetilde{\mathbf{A}}^*$ for **P2A** by solving the following Problem **P2B**:

$$
\begin{aligned}
\underset{\mathbf{A}}{\text{maximize}} \quad & \sum_{k=1}^{K} \widetilde{\eta}_k^* \widetilde{M}_k(A_k) \\
\text{subject to} \quad & \sum_{k=1}^{K} A_k = 1, \quad A_k \ge a_k, \forall k.
\end{aligned}
\tag{26}
$$

Unlike in the optimality region, here we have an explicit solution, as stated in the following theorem:

Theorem 4. *Given an optimal $\widetilde{\boldsymbol{\eta}}^*$ for* **P2A** *(computed the same way as $\boldsymbol{\eta}^*$ in (20)), the corresponding optimal $\widetilde{\mathbf{A}}^*$ can be expressed as follows:*

$$
\begin{cases}
\widetilde{A}_k^* = a_k, & k \ge 2 \\
\widetilde{A}_1^* = 1 - \sum_{l=2}^{K} \widetilde{A}_l^*.
\end{cases}
\tag{27}
$$

Proof. See Appendix F. ∎

Note that both (20) and (27) can be computed with low computational complexity.

2) Bounding the Performance Gap: Since $(\widetilde{\boldsymbol{\eta}}^*, \widetilde{\mathbf{A}}^*)$ is optimal for **P2A** rather than **P2**, we next quantify the performance gap between $(\widetilde{\boldsymbol{\eta}}^*, \widetilde{\mathbf{A}}^*)$ and an optimal solution $(\boldsymbol{\eta}^*, \mathbf{A}^*)$ to **P2**.

The performance gap is defined as

$$E = \mathbf{F}(\boldsymbol{\eta}^*, \mathbf{A}^*) - \mathbf{F}(\widetilde{\boldsymbol{\eta}}^*, \widetilde{\mathbf{A}}^*). \tag{28}$$

Because $\mathbf{F}(\widetilde{\boldsymbol{\eta}}^*, \widetilde{\mathbf{A}}^*) \leq \mathbf{F}(\boldsymbol{\eta}^*, \mathbf{A}^*) \leq \mathbf{F}'(\boldsymbol{\eta}^*, \mathbf{A}^*) \leq \mathbf{F}'(\widetilde{\boldsymbol{\eta}}^*, \widetilde{\mathbf{A}}^*)$, we have

$$E \leq \mathbf{F}'(\widetilde{\boldsymbol{\eta}}^*, \widetilde{\mathbf{A}}^*) - \mathbf{F}(\widetilde{\boldsymbol{\eta}}^*, \widetilde{\mathbf{A}}^*) \triangleq E'. \tag{29}$$

Substituting $\widetilde{\boldsymbol{\eta}}^*$ and $\widetilde{\mathbf{A}}^*$ into E', we have

$$E' = \sum_{k=1}^{K} \frac{\widetilde{\eta}_k^*}{\frac{\widetilde{A}_k^*\mu}{\lambda_k}\left(\frac{\widetilde{A}_k^*\mu}{\lambda_k}+1\right)\left(\frac{1}{\widetilde{A}_k^*}+C\right)}. \tag{30}$$

Therefore, the relative performance gap is bounded:

$$
\begin{aligned}
\epsilon &\triangleq \frac{E}{\mathbf{F}(\boldsymbol{\eta}^*, \mathbf{A}^*)} \leq \frac{E'}{\mathbf{F}(\widetilde{\boldsymbol{\eta}}^*, \widetilde{\mathbf{A}}^*)} \\
&= \frac{\sum_{k=1}^{K} \frac{\widetilde{\eta}_k^*}{\frac{\widetilde{A}_k^*\mu}{\lambda_k}\left(\frac{\widetilde{A}_k^*\mu}{\lambda_k}+1\right)\left(\frac{1}{\widetilde{A}_k^*}+C\right)}}{\sum_{k=1}^{K} \frac{\widetilde{\eta}_k^*}{\left(\frac{\widetilde{A}_k^*\mu}{\lambda_k}+1\right)\left(\frac{1}{\widetilde{A}_k^*}+C\right)}} \qquad (31) \\
&\overset{(a)}{\leq} \max_k \frac{\lambda_k}{\widetilde{A}_k^*\mu} \leq \max_k \frac{\lambda_k}{a_k\mu} \\
&= \sqrt{\lambda_K C/\mu}, \tag{32}
\end{aligned}
$$

where inequality (a) is obtained by observing the common factor in the summations in the numerator and denominator of (31). This result implies that ϵ scales as $O\left(\sqrt{\lambda_K/\mu}\right)$. Note that when $\mu \gg \lambda_K$, we have $\epsilon \simeq 0$ and the performance of SSAUA is asymptotically optimal. Thus $\sum_{i=1}^{K} a_i < 1$ is referred to as the asymptotic-optimality region.

D. Computational Complexity Comparison

In this subsection, we discuss the computational complexity of SSAUA and that of an exhaustive search approach to solve the original joint optimization problem.

1) Computational Complexity of SSAUA: In the optimality region, the optimal $\boldsymbol{\eta}^*$ can be derived with computational complexity $O(K)$ through (20). Given the optimal $\boldsymbol{\eta}^*$, the remaining problem (22) is a convex optimization problem, which can be solved using the interior-point method. Let ε denote the error bound between the output and the optimal solution. According to Section 11.5 of [39], the interior-point method involves $O(\log(\frac{K}{\varepsilon})K)$ Newton iterations; each Newton iteration involves a matrix inversion operation, which has computational complexity $O(K^3)$. Therefore, the overall computational complexity is $O(\log(\frac{K}{\varepsilon})K^4)$. In the asymptotic-optimality region, the asymptomatically optimal $\widetilde{\boldsymbol{\eta}}^*$ and $\widetilde{\mathbf{A}}^*$ can be derived through (20) and (27) respectively, with an overall computational complexity $O(K)$.

2) Computational Complexity of Exhaustive Search: As explained in detail in Appendix G, we observe that at least one of the optimal solutions to **P**, $(\boldsymbol{\eta}^*, \mathbf{A}^*)$, has the following property: there is at most one $k \in \{1, 2, \dots K\}$ such that $\eta_{\min,k} < \eta_k^* < \eta_{\max,k}$; $\forall j \neq k$, either $\eta_j^* = \eta_{\min,j}$ or $\eta_j^* = \eta_{\max,j}$ (i.e., at the boundary). Thus, the search for $\boldsymbol{\eta}^*$ needs to be performed at these boundary cases only, leading to a complexity of $O(2^K)$.

In the optimality region, if $\boldsymbol{\eta}$ is fixed, the remaining problem is a convex optimization problem. The interior-point method takes another fold of computational complexity of $O(\log(\frac{K}{\varepsilon})K^4)$. Consequently, the overall computational complexity is $O(2^K \log(\frac{K}{\varepsilon})K^4)$. In the asymptomatic-optimality region, if $\boldsymbol{\eta}$ is fixed, the remaining problem is still a non-convex optimization problem. A numerical search over all solutions to the KKT condition is required, leading to another fold of computational complexity of $O(\frac{2^K}{\varepsilon})$. Consequently, the overall computational complexity is $O(\frac{4^K}{\varepsilon})$.

VI. Nash Equilibrium for SSAUA

Individual UEs may behave selfishly to derive unfair advantage despite our design of $\mathbf{B}^*$ (or equivalently $\mathbf{A}^*$). Thus, in this section, we propose a Surcharge Pricing Scheme (SPS), such that the designed $\mathbf{B}^*$ is the natural outcome of a Nash equilibrium. We assume that the designed spectrum allocation factors $\boldsymbol{\eta}^*$ are centrally maintained by the network operator.

We consider a reference individual UE, whose association bias values are $\mathbf{B}' = (B_1', B_2', \dots, B_K')$. Let $\mathbf{A}' = (A_1', A_2', \dots, A_K')$ be its corresponding association probabilities. For the other UEs, suppose they all obey the association bias values $\mathbf{B}^*$ assigned by the network operator. Similar to the discussions in Section III and IV, the average data rate of the reference UE is

$$\mathbb{F} = \sum_{k=1}^{K} \frac{\eta_k^* W \log(1+T)}{(A_k^*\mu/\lambda_k + 1)(\frac{1}{A_k'}+C)}. \tag{33}$$

If the reference UE performs an optimization on $\mathbb{F}$ with respect to $\mathbf{A}'$, the resultant optimal $\mathbf{A}'^* = (A_1'^*, A_2'^*, \dots, A_K'^*)$ is unlikely to be the same as $\mathbf{A}^*$. Therefore, we add the following Surcharge Pricing Scheme: the network operator applies a surcharge c_k to each UE associated with a tier-k BS. Let $\mathbf{c} = (c_1, c_2, \dots, c_K)$. In this case, the average surcharge for the reference UE is $\sum_{k=1}^{K} c_k A_k'$. Accordingly, the reference UE will perform the following optimization Problem **P3**:

$$\underset{\mathbf{A}'}{\text{maximize}} \quad \mathbb{F}' = \sum_{k=1}^{K} \left(\frac{\eta_k^* W \log(1+T)}{(A_k^*\mu/\lambda_k + 1)\left(\frac{1}{A_k'}+C\right)} - c_k A_k' \right)$$

$$\text{subject to} \quad \sum_{k=1}^{K} A_k' = 1, \quad A_k' \geq 0. \tag{34}$$

Different from **P**, it can be shown that **P3** is a standard convex optimization problem. By the KKT conditions, its optimal solution $\mathbf{A}'^*$ satisfies

$$\frac{H_k}{(1+CA_k'^*)^2} - c_k - \nu + \theta_k = 0, \tag{35}$$

$$\theta_k A_k'^* = 0, \quad \theta_k \geq 0, \qquad (36)$$

where $H_k = \frac{\eta_k^* W \log(1+T)}{A_k^* \mu / \lambda_k + 1}$, θ_k is a Lagrange multiplier corresponding to the inequality constraint $A_k' \geq 0$, and ν is a Lagrange multiplier corresponding to the equality constraint $\sum_{k=1}^K A_k' = 1$.

Setting $A_k'^* = A_k^*$, we have

$$c_k = \begin{cases} \infty, & \text{if } A_k^* = 0, \\ \frac{H_k}{(1+CA_k^*)^2} - \nu, & \text{otherwise.} \end{cases} \qquad (37)$$

Note that ν could be set arbitrarily due to the equality constraint. Without loss of generality, we set $\nu = \min_k \frac{H_k}{(1+CA_k^*)^2}$ so that the minimum surcharge among tiers is 0. As a consequence, a Nash Equilibrium is achieved where every UE adopts the assigned $\mathbf{B}^*$.

VII. THE MULTIPLE-MCS CASE

In this section, we discuss the usefulness of our proposed SSAUA in systems with multiple modulation and coding schemes (i.e., the multiple-MCS case). Instead of considering only one SIR threshold T (see Section III-D), N SIR threshold values, $T_1, T_2, \ldots, T_N$, where $T_1 < T_2 < \ldots < T_N$, corresponding to N MCSs, are accommodated. In this case, if a UE is allocated spectrum bandwidth β', its data rate is $\beta' \log(1 + T_N)$, $\beta' \log(1 + T_{N-1})$, $\ldots$, $\beta' \log(1 + T_1)$, and 0, respectively, if its SIR is in $[T_N, \infty)$, $[T_{N-1}, T_N)$, $\ldots$, $[T_1, T_2)$, and $[0, T_1)$.

A. Average UE Data Rate

First, similar to the derivations of (4)-(12), given that the typical UE is associated with a tier-k BS, we can find its coverage probabilities under $T_1, T_2, \ldots, T_N$ to be

$$\mathbb{P}(SIR_k \geq T_1) = \mathbb{P}_{cover,k,1} = \frac{1}{1 + A_k C_1}, \qquad (38)$$

$$\ldots$$

$$\mathbb{P}(SIR_k \geq T_n) = \mathbb{P}_{cover,k,n} = \frac{1}{1 + A_k C_n}, \qquad (39)$$

$$\ldots$$

$$\mathbb{P}(SIR_k \geq T_N) = \mathbb{P}_{cover,k,N} = \frac{1}{1 + A_k C_N}, \qquad (40)$$

where $C_n \triangleq (T_n)^{\frac{2}{\gamma}} \int_{(\frac{1}{T_n})^{\frac{2}{\gamma}}}^{\infty} \frac{1}{1 + t^{\gamma/2}} dt$. Note that because $T_1 < T_2 < \ldots < T_N$, we have $C_1 < C_2 < \ldots < C_N$ and $\mathbb{P}_{cover,k,1} > \mathbb{P}_{cover,k,2} > \ldots > \mathbb{P}_{cover,k,N}$.

Then, the conditional expected data rate of the typical UE, given it is associated with a tier-k BS, is recomputed as

$$\overline{R}_k = \mathbf{E}_0(\beta_k) \Bigg[\log(1 + T_N) \mathbb{P}_{cover,k,N} \qquad (41)$$

$$+ \sum_{n=1}^{N-1} \log(1 + T_n) \big(\mathbb{P}_{cover,k,n} - \mathbb{P}_{cover,k,n+1} \big) \Bigg]$$

$$= \mathbf{E}_0(\beta_k) \Bigg[\log(1 + T_1) \mathbb{P}_{cover,k,1} \qquad (42)$$

$$+ \sum_{n=2}^{N} \big(\log(1 + T_n) - \log(1 + T_{n-1}) \big) \mathbb{P}_{cover,k,n} \Bigg]$$

$$= \frac{\eta_k W}{(A_k \mu / \lambda_k + 1)} \Bigg[\frac{\log(1 + T_1)}{1 + A_k C_1} \qquad (43)$$

$$+ \sum_{n=2}^{N} \frac{\log(1 + T_n) - \log(1 + T_{n-1})}{1 + A_k C_n} \Bigg]$$

$$= \frac{\eta_k W}{(A_k \mu / \lambda_k + 1)} \left(\sum_{n=1}^{N} \frac{b_n}{1 + A_k C_n} \right), \qquad (44)$$

where $b_1 \triangleq \log(1 + T_1), b_2 \triangleq \log(1 + T_2) - \log(1 + T_1), \ldots, b_N \triangleq \log(1 + T_N) - \log(1 + T_{N-1})$.

Similar to (15), the average data rate of the typical UE (unconditioned on k), and hence the average data rate per UE in the system, is recomputed as

$$\mathbf{F} = \sum_{k=1}^{K} A_k \overline{R}_k$$

$$= \sum_{k=1}^{K} \frac{\eta_k W}{(A_k \mu / \lambda_k + 1)} \left(\sum_{n=1}^{N} \frac{b_n}{\frac{1}{A_k} + C_n} \right). \qquad (45)$$

The optimization problem (16) is updated correspondingly, such that $M_k(A_k)$ is redefined as

$$M_k(A_k) = \frac{1}{(A_k \mu / \lambda_k + 1)} \left(\sum_{n=1}^{N} \frac{b_n}{\frac{1}{A_k} + C_n} \right). \qquad (46)$$

B. SSAUA in the Multiple-MCS Case

In this subsection, we discuss the usefulness of our proposed SSAUA in solving the modified optimization problem considering multiple-MCS.

1) Density Thresholding: We redefine a_k as the unique positive solution to $M_k'(A_k) = 0$. The existence and uniqueness of a_k is shown in Appendix H. Properties (M-1) to (M-4) presented in Appendix A still hold with the redefined $M_k(A_k)$ and a_k. The proofs of these properties are also shown in Appendix H.

As a result, Lemma 1 and Theorem 1 still hold in the multiple-MCS case. Note that different from the single-MCS case (18), a_k cannot be represented in an explicit expression. As shown in Appendix H, a_k is the unique solution to $f_k(A) = 0$, where $f_k(A) = \sum_{n=1}^{N} \frac{b_n(1 - A^2 C_n \mu / \lambda_k)}{(AC_n + 1)^2}$ is a decreasing function. A simple binary search method can be applied to compute a_k, which has low computational complexity.

2) Optimality Region: When $\sum_{i=1}^{K} a_i > 1$, since Properties (M-1) to (M-4) still hold, Lemma 2 and Theorem 2 still hold in the multiple-MCS case. Then, the same method presented in Section V-B can be applied to solve the optimization problem.

3) Asymptotic-optimality Region: When $\sum_{i=1}^{K} a_i < 1$, $M_k(A_k)$ is approximated by $\widetilde{M}_k(A_k)$ redefined as follows:

$$\widetilde{M}_k(A_k) = \frac{1}{A_k \mu / \lambda_k} \left[\sum_{n=1}^{N} \frac{b_n}{\frac{1}{A_k} + C_n} \right]. \qquad (47)$$

If Properties (M-1')-(M-6') presented in Appendix D still hold for the above redefined $\widetilde{M}_k(A_k)$, then Lemma 3, Theorem 3, and Theorem 4 still hold in the multiple-MCS case, and the same method employed in Section V-C can be applied to solve the optimization problem with performance gap bounded by (32).
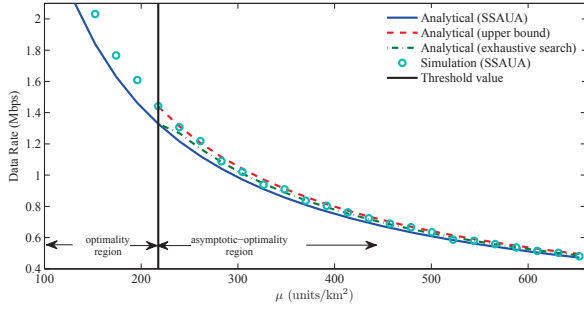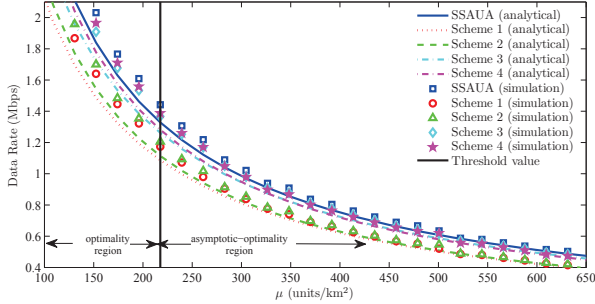
Fig. 2. Average UE data rate under different UE density μ.

(a) Optimality region. (b) Asymptotic-optimality region.

Fig. 4. Average UE data rate under different path loss exponent γ.

Fig. 3. Comparison of different schemes under different UE density μ.

(a) Asymptotic-optimality region. (b) Asymptotic-optimality region.

Fig. 5. Comparison of different schemes under different path loss exponent γ, analytical case.

It is straightforward to verify Properties (M-1'), (M-3'), and (M-4'). Property (M-6') is implied by (M-4') and (M-5'). However, Properties (M-2') and (M-5') are difficult to verify analytically, as a_k is no longer expressed in closed form. Therefore, we conduct numerical validation to check Properties (M-2') and (M-5') under a wide range of parameter settings.

We set $\mathcal{T} = \{10^{-1}, 10^{-0.9}, \ldots, 10^{0.9}, 10^1\}$, $\mathcal{U} = \{10^1, 10^{1.1}, \ldots, 10^3\}$, $\Gamma = \{3, 3.5, 4, 4.5, 5\}$. It is shown that Properties (M-2') and (M-5') are true under $N = 2$, 3, and 4; $\forall T_1, T_2, \ldots, T_N \in \mathcal{T}$ ($T_1 < T_2 < \ldots < T_N$); $\forall \mu/\lambda_i, \mu/\lambda_j \in \mathcal{U}$ ($\mu/\lambda_i > \mu/\lambda_j$); and $\forall \gamma \in \Gamma$. Therefore, at least for the wide range of parameter settings that are tested, the proposed SSAUA is still useful in the asymptotic-optimality region. See Appendix I for visualized verification results.

VIII. NUMERICAL STUDY

In this section, we present numerical studies on the performance of SSAUA. Unless otherwise stated, we label the S-SAUA solution as $(\widehat{\boldsymbol{\eta}}^*, \widehat{\mathbf{B}}^*)$ and $(\widetilde{\boldsymbol{\eta}}^*, \widetilde{\mathbf{B}}^*)$ in the optimality and asymptotic-optimality regions, respectively. Note that $(\widehat{\boldsymbol{\eta}}^*, \widehat{\mathbf{B}}^*)$ is optimal in the optimality region. We use $(\boldsymbol{\eta}^*, \mathbf{B}^*)$ to label an optimal solution obtained from exhaustive search in the asymptotic-optimality region. In this section, each simulation point is derived as follows: in each round of simulation, UEs and BSs are generated on a 10 km × 10 km square, and the UEs in the central 5 km × 5 km square are sampled for performance evaluation (in order to remove the edge effect). Each simulation data point is averaged over all sampled UEs during 100 rounds of simulations.
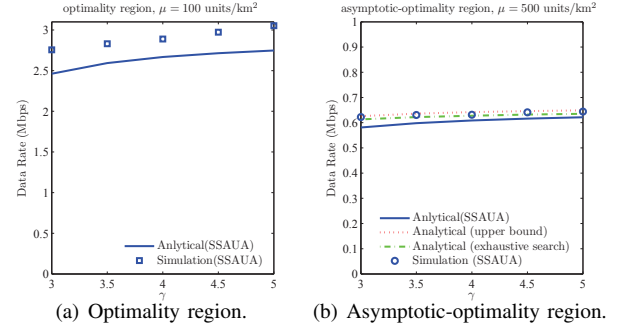
A. Average UE Data Rate under Different UE Densities

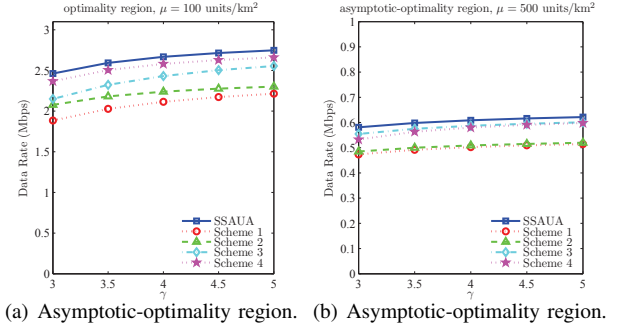In this subsection, we study the average UE data rate of SSAUA under different values of UE density μ, and compare it with four reference schemes. The network parameters are as follows: $K = 3$, $\lambda_1 = 1$ units/km^2, $\lambda_2 = 5$ units/km^2, $\lambda_3 = 10$ units/km^2, $P_1 = 56$ dBm, $P_2 = 46$ dBm, $P_3 = 36$ dBm, $\eta_{\min,1} = 0.2$, $\eta_{\min,2} = 0.25$, $\eta_{\min,3} = 0.3$, $\eta_{\max,1} = 0.35$, $\eta_{\max,2} = 0.4$, $\eta_{\max,3} = 0.45$, $\gamma = 4$, $W = 200$ MHz, and $T = 0.2$.

The performance of SSAUA is shown in Fig. 2. A vertical line indicates the threshold value of μ, as given in Theorem 1, separating the optimality and asymptotic-optimality regions. For both regions, we show results of the analytical and simulated performance of SSAUA. Since SSAUA is not optimal in the asymptotic-optimality region, we also add two sets of results accordingly: the optimal performance $\mathbf{F}(\boldsymbol{\eta}^*, \mathbf{B}^*)$ through exhaustive search and its analytical upper bound $\mathbf{F}'(\widetilde{\boldsymbol{\eta}}^*, \widetilde{\mathbf{B}}^*)$. Fig. 2 illustrates that the performance of SSAUA is very close to the optimal solution in the asymptotic-optimality region.

Next, we compare the performance of SSAUA with four reference schemes listed as follows:

- Scheme 1 employs equal spectrum allocation, and user association based on the maximum received power.
- Scheme 2 employs equal spectrum allocation, and optimal user association as in SSAUA.
- Scheme 3 employs optimal spectrum allocation as in SSAUA, and user association based on the maximum received power.
- Scheme 4 employs optimal spectrum allocation as in SSAUA, and user association based on a simple range
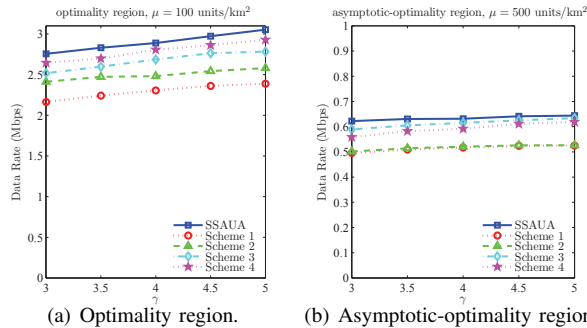
Fig. 6. Comparison of different schemes under different path loss exponent γ, simulation case.
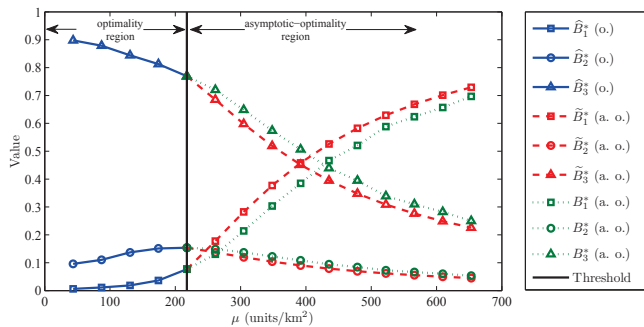

Fig. 7. Designed association bias values under different UE density μ. "o" represents the optimality region, and "a. o." represents the asymptotic-optimality region.

Fig. 8. Pricing values under different UE density μ. "o" represents the optimality region, and "a. o." represents the asymptotic-optimality region.

Fig. 9. Comparison of run time.

expansion scheme by setting $B_k = \frac{1}{P_k}, \forall k$.

Fig. 3 shows that SSAUA outperforms these schemes. This illustrates that both the spectrum allocation component and the user association component of SSAUA bring performance benefits. Note that due to the approximation made in (14), the analytical data rate is slightly lower than the simulated one, matching our discussion in Section IV.

B. Average UE Data Rate under Different Path Loss Exponents

Fig. 4 shows the optimal network performance under different values of the path loss exponent γ. The network parameters are the same as those used in Fig. 2 except μ is fixed at 100 (i.e., optimality region) and 500 (i.e., asymptotic-optimality region) units/km^2 in Figs. 4 (a) and (b) respectively. This figure further confirms the observations from Fig. 2. Furthermore, it shows that SSAUA is effective for a wide range of path loss conditions.

Fig. 5 shows further analytical performance comparison of SSAUA and the four reference schemes stated in Section VIII-A. The results illustrate that in both the optimality and asymptotic-optimality regions, SSAUA outperforms Scheme 2, Scheme 3, and Scheme 4, and both Scheme 2 and Scheme 3 outperform Scheme 1, confirming the observations from Fig. 3. Fig. 6 shows the simulated performance comparison of SSAUA and the four reference schemes. The results show that the performance orders of the five schemes agree with those in Fig. 5 in both the optimality region and the asymptotic-optimality region.
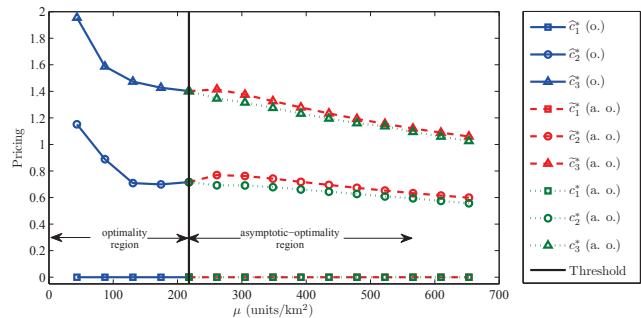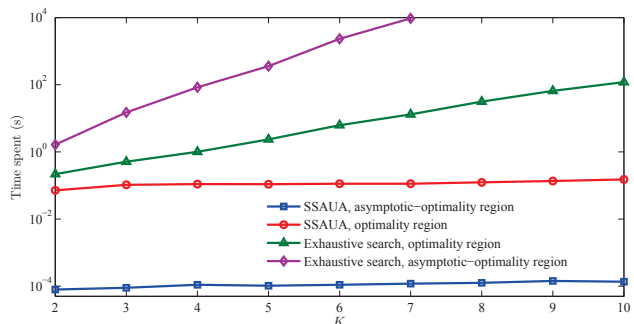
C. Association Bias Values and Prices

Fig. 7 shows $\widehat{\mathbf{B}}^*$, $\widetilde{\mathbf{B}}^*$, and $\mathbf{B}^*$; and Fig. 8 shows their corresponding prices $\widehat{\mathbf{c}}^*$, $\widetilde{\mathbf{c}}^*$, and $\mathbf{c}^*$, under different μ values. The other network parameters are the same as those used in Fig. 2. We observe that the $\widetilde{\mathbf{B}}^*$ and $\widetilde{\mathbf{c}}^*$ computed based on SSAUA are close to their counterparts $\mathbf{B}^*$ and $\mathbf{c}^*$. Similar results have been observed under different path loss exponents and are omitted for brevity.

D. Run Time Experiment

In this subsection, a run time experiment is presented to compare the computational complexity of SSAUA with that of exhaustive search. The experiment is executed by Matlab R2011a on an ASUS PC with Intel i7-3610QM 2.3GHz processor and 4GB RAM. The results are averaged over 1000 runs for SSAUA and 10 runs for exhaustive search (both with randomly generated parameters). Fig. 9 shows that the run time of SSAUA is almost negligible compared with exhaustive search. Note that the y-axis is in log scale. When K increases, the run time of exhaustive search exhibits an exponential growth tendency, while SSAUA remains computationally efficient.

E. Performance of SSAUA in the Multiple-MCS Case

In this subsection, a numerical experiment is presented to validate the usefulness of SSAUA in the multiple-MCS case. The network parameters are as follows: $N = 3$, $T_1 = 0.1$, $T_2 = 0.3$, $T_3 = 0.5$, $K = 3$, $\lambda_1 = 1$ units/km^2, $\lambda_2 = 5$ units/km^2, $\lambda_3 = 10$ units/km^2, $P_1 = 56$ dBm, $P_2 = 46$ dBm, $P_3 = 36$ dBm, $\eta_{\min,1} = 0.2$, $\eta_{\min,2} = 0.25$, $\eta_{\min,3} = 0.3$,
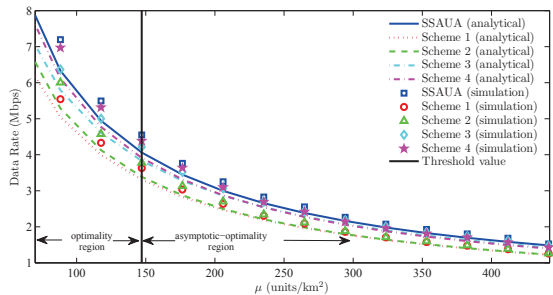
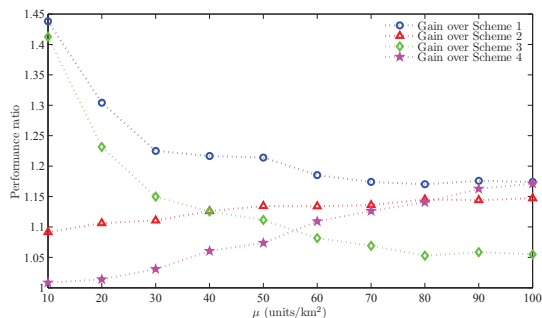Fig. 10. Comparison of different schemes in the multiple-MCS case.

Fig. 11. Comparison of different schemes under more realistic network settings, $(\lambda'_1, \lambda'_2, \lambda'_3) = (1, 2, 3)$ unit/km^2.
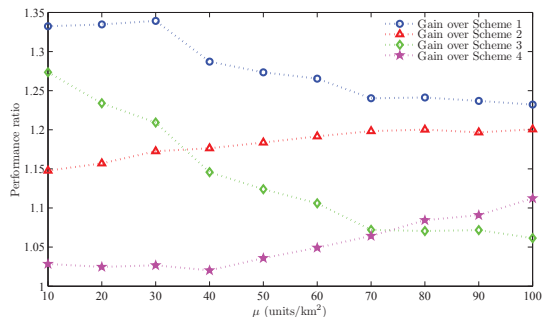

Fig. 12. Comparison of different schemes under more realistic network settings, $(\lambda'_1, \lambda'_2, \lambda'_3) = (1, 5, 10)$ unit/km^2.

$\eta_{\max,1} = 0.35$, $\eta_{\max,2} = 0.4$, $\eta_{\max,3} = 0.45$, $\gamma = 4$, and $W = 200$ MHz.

Fig. 10 shows the comparison of SSAUA with the four reference schemes. Similar to the results in VIII-A, in the multiple-MCS case, we can still observe that SSAUA outperforms Scheme 2, Scheme 3, and Scheme 4, and both Scheme 2 and Scheme 3 outperform Scheme 1. Both the spectrum allocation part and the user association part used in SSAUA bring performance benefits, and the joint spectrum allocation and user association of SSAUA brings the greatest performance benefits in the Multiple-MCS case.

F. Performance Comparison under More Realistic Network Topologies

In this section, we present simulation of SSAUA and the four reference schemes in more realistic network settings. First, the UE and BS density settings follow the 3GPP simulation guidelines from Annex A of [40], which recommend that there are between 1 to 10 small cell BSs, and 10 to 100 UEs per macrocell. Second, BSs are non-Poisson, and UEs are inhomogeneously distributed.

The Matern hard core (MHC) point process is an alternative point process adopted in the literature to counter the drawback of PPP modeling of cellular networks [41], [42]. MHC point processes can additionally capture the effect that two BSs are unlikely to be located very close to each other. In our simulation, tier-k BSs are generated as an MHC point process as follows: First, we generate a PPP with intensity λ'_k. Each point in the PPP is associated with a "mark", which is independently uniformly distributed on $[0, 1]$. A point is retained in the point process if its mark is the largest among all the points within a distance D_k from it (or there are no other points within this range); otherwise, the point is removed from the point process. The remaining points form an MHC point process. Note that the distance between any two points in the point process is no less than D_k. In our simulation, each tier of BSs is independently generated as an MHC point process. Then, the spectrum allocation factors and user association bias values are determined, by either SSAUA or one of the reference schemes, based on PPP BSs with equivalent densities $\lambda_k = \frac{1-e^{-\pi D_k^2 \lambda'_k}}{\pi D_k^2 \lambda'_k}$.

We also consider the scenario where UEs are likely to crowd near BSs. At the beginning, UEs follow a PPP with intensity μ'. Then, each tier-k BS brings increment of μ_0 to the UE density in the disk region centered at the BS with a radius of R_k. In our simulation, the spectrum allocation factors and user association bias values are determined, by either SSAUA or one of the reference schemes, based on an equivalent UE density $\mu = \mu' + \mu_0 \left(\sum_{k=1}^{K} \lambda_k \pi R_k^2 \right)$.

In Figs. 11 and 12, we set $\gamma = 4$; $(P_1, P_2, P_3) = (56, 43, 33)$ dBm; $\eta_{\min,1} = 0.2$, $\eta_{\min,2} = 0.25$, $\eta_{\min,3} = 0.3$, $\eta_{\max,1} = 0.35$, $\eta_{\max,2} = 0.4$, $\eta_{\max,3} = 0.45$; $W = 50$ MHz; $N = 3$, $T_1 = 1$, $T_2 = 2$, $T_3 = 3$. In Fig. 11, $(\lambda'_1, \lambda'_2, \lambda'_3) = (1, 2, 3)$ unit/km^2, $(D_1, D_2, D_3) = (200, 80, 40)$ m, $(R_1, R_2, R_3) = (200, 80, 40)$ m, and $\mu_0 = \mu'$. The equivalent BS densities are $(\lambda_1, \lambda_2, \lambda_3) = (0.9397, 1.9603, 2.9775)$ unit/km^2, and the equivalent UE density is $\mu = 1.1725\mu'$. In Fig. 12, we set $(\lambda'_1, \lambda'_2, \lambda'_3) = (1, 5, 10)$ unit/km^2, $(D_1, D_2, D_3) = (200, 80, 40)$ m, $(R_1, R_2, R_3) = (200, 80, 40)$ m, and $\mu_0 = \mu'$. The equivalent BS densities are $(\lambda_1, \lambda_2, \lambda_3) = (0.9397, 4.7569, 9.7528)$ unit/km^2, and the equivalent UE density is $\mu = 1.2628\mu'$. In both figures, μ ranges from 10 to 100. The performance ratios of SSAUA to all four reference schemes are shown in the figures. The result suggests that, under more realistic network topologies, SSAUA still brings useful performance gain compared with the reference schemes.

IX. CONCLUSIONS AND FUTURE WORKS

In this work, we provide a theoretical framework to study the joint optimization of spectrum allocation and user association in heterogeneous cellular networks. We establish a stochastic geometric model that captures the random spatial patterns of BSs and UEs, and a closed-form expression of the analytical average UE data rate is derived. We then consider the problem of maximizing the average UE data rate by jointly optimizing the spectrum allocation factors and user association bias values, which is non-convex programming in nature. We propose the SSAUA approach to solve this problem with low

computational complexity. We show that SSAUA is either optimal or asymptotically optimal with a quantified bound scaling as $O(\sqrt{\lambda_K/\mu})$. We also propose a pricing scheme such that the designed association bias values can be achieved in Nash equilibrium.

So far we have focused on maximizing the user data rate. However, in multimedia applications, a defining characteristic is inelasticity, where some predetermined minimum throughput must be satisfied, or the application session is rejected or dropped. By taking inelastic traffic into consideration, the user data rate is no longer the only design goal [43]. For future work, it will be crucial to design user association and spectrum allocation schemes to benefit inelastic traffic through decreasing traffic delay, reducing data rate fluctuation, and improving quality of experience.

APPENDIX

A. Useful Properties of $M_k(A_k)$

(M-1) $M_k(A_k)$ is increasing on $[0, a_k)$ and decreasing on $[a_k, \infty)$.

(M-2) $M_k(A_k)$ is concave on $[0, a_k)$.

(M-3) If $\lambda_i < \lambda_j$, then $M_i(A) < M_j(A), \forall A > 0$.

(M-4) If $\lambda_i < \lambda_j$, then $M_j(A) - M_i(A)$ is strictly increasing on $[0, a_j]$.

B. Proof of Lemma 1

Proof. Suppose $\mathbf{A}^{**}$ is optimal, $A_i^{**} < a_i$, and $A_j^{**} > a_j$. Consider that we increase A_i^{**} by a small value $\Delta > 0$ and decrease A_j^{**} by Δ. According to property (M-1), $\eta_i^{**} M_i(A_i^{**}) + \eta_j^{**} M_j(A_j^{**}) < \eta_i^{**} M_i(A_i^{**} + \Delta) + \eta_j^{**} M_j(A_j^{**} - \Delta)$. Thus, through replacing A_i^{**} and A_j^{**} by $A_i^{**} + \Delta$ and $A_j^{**} - \Delta$ respectively, we find a better solution to $\mathbf{P}$, leading to a contradiction. $\square$

C. Proof of Lemma 2

Proof. Suppose $\exists i < j$ such that $M_i(A_i^*) > M_j(A_j^*)$. This implies that $A_i^* > A_j^*$. (Otherwise, if $A_i^* \leq A_j^*$, then we have $M_i(A_i^*) \leq M_i(A_j^*) < M_j(A_j^*)$, leading to a contradiction.) A corresponding diagram is shown in Fig. 13(a).

Case 1: $\eta_i^* \leq \eta_j^*$.

Let $\widehat{A}_j^* = A_i^*$ and $\widehat{A}_i^* = A_j^*$. Then we have

$$[\eta_i^* M_i(\widehat{A}_i^*) + \eta_j^* M_j(\widehat{A}_j^*)] - [\eta_i^* M_i(A_i^*) + \eta_j^* M_j(A_j^*)]$$
$$= [\eta_i^* M_i(A_j^*) + \eta_j^* M_j(A_i^*)] - [\eta_i^* M_i(A_i^*) + \eta_j^* M_j(A_j^*)]$$
$$= \eta_j^* [M_j(A_i^*) - M_j(A_j^*)] + \eta_i^* [M_i(A_j^*) - M_i(A_i^*)]$$
$$\geq \eta_i^* [M_j(A_i^*) - M_j(A_j^*) + M_i(A_j^*) - M_i(A_i^*)] > 0, \quad (48)$$

where (48) is due to property (M-4).

As a consequence, if A_i^* and A_j^* are replaced by $\widehat{A}_i^*$ and $\widehat{A}_j^*$ respectively, we obtain a larger $\mathbf{F}$, leading to a contradiction.

Case 2: $\eta_i^* > \eta_j^*$.

Let $\widehat{A}_j^* = A_i^*$, $\widehat{A}_i^* = A_j^*$, $\widehat{\eta}_j^* = \eta_i^*$, and $\widehat{\eta}_i^* = \eta_j^*$. (Note that because $\eta_{\min,i} \leq \eta_{\min,j}$ and $\eta_{\max,i} \leq \eta_{\max,j}$, $\widehat{\eta}_j^*$ and $\widehat{\eta}_i^*$ are guaranteed to be in the feasible region.)

$$[\widehat{\eta}_i^* M_i(\widehat{A}_i^*) + \widehat{\eta}_j^* M_j(\widehat{A}_j^*)] - [\eta_i^* M_i(A_i^*) + \eta_j^* M_j(A_j^*)]$$

(a) Diagram of $M_i(\cdot)$ and $M_j(\cdot)$.

(b) Diagram of $\widetilde{M}_i(\cdot)$ and $\widetilde{M}_j(\cdot)$, Case 1.1 and 2.1.

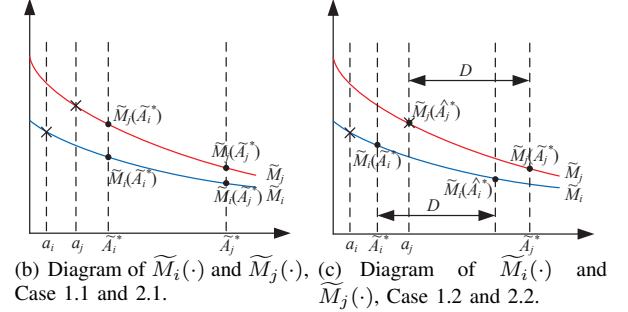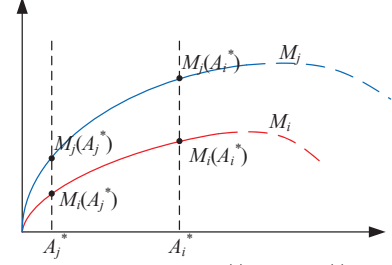(c) Diagram of $\widetilde{M}_i(\cdot)$ and $\widetilde{M}_j(\cdot)$, Case 1.2 and 2.2.

Fig. 13. Diagrams of $M_i(\cdot)$, $M_j(\cdot)$, $\widetilde{M}_i(\cdot)$, and $\widetilde{M}_j(\cdot)$.

$$= [\eta_j^* M_i(A_j^*) + \eta_i^* M_j(A_i^*)] - [\eta_i^* M_i(A_i^*) + \eta_j^* M_j(A_j^*)]$$
$$= \eta_i^* [M_j(A_i^*) - M_i(A_i^*)] + \eta_j^* [M_i(A_j^*) - M_j(A_j^*)]$$
$$> \eta_j^* [M_j(A_i^*) - M_i(A_i^*) + M_i(A_j^*) - M_j(A_j^*)] > 0. \quad (49)$$

Thus, if A_i^*, A_j^*, η_i^*, and η_j^* are replaced by $\widehat{A}_i^*$, $\widehat{A}_j^*$, $\widehat{\eta}_i^*$, and $\widehat{\eta}_j^*$ respectively, we can find a larger $\mathbf{F}$, leading to a contradiction. $\square$

D. Useful Properties of $\widetilde{M}_k(A_k)$

(M-1') $\widetilde{M}_k(A_k)$ is a decreasing convex function.

(M-2') If $\lambda_i < \lambda_j$, then $\widetilde{M}_i(a_i) < \widetilde{M}_j(a_j)$.

(M-3') If $\lambda_i < \lambda_j$, then $\widetilde{M}_j(A) - \widetilde{M}_i(A)$ is a strictly decreasing function.

(M-4') $\widetilde{M}_k(A) - \widetilde{M}_k(A+D) > \widetilde{M}_k(A') - \widetilde{M}_k(A'+D)$, for any $A' > A \geq a_k$ and $D > 0$.

(M-5') If $\lambda_i < \lambda_j$, then $\widetilde{M}_j(a_j) - \widetilde{M}_j(a_j + D) > \widetilde{M}_i(a_i) - \widetilde{M}_i(a_i + D)$, for any $D > 0$.

(M-6') If $\lambda_i < \lambda_j$, then $\widetilde{M}_j(a_j) - \widetilde{M}_j(a_j + D) > \widetilde{M}_i(A') - \widetilde{M}_i(A'+D)$, for any $D > 0$ and $A' > a_i$ (combining (M-4') and (M-5')).

E. Proof of Lemma 3

Proof. Suppose that $\exists i < j$ (i.e., $\lambda_i < \lambda_j$) such that $\widetilde{M}_i(\widetilde{A}_i^*) > \widetilde{M}_j(\widetilde{A}_j^*)$, which also implies that $a_i \leq \widetilde{A}_i^* < \widetilde{A}_j^*$. The corresponding diagrams are shown in Figs. 13(b) and 13(c).

Case 1: $\widetilde{\eta}_i^* \leq \widetilde{\eta}_j^*$.

Case 1.1: $\widetilde{A}_i^* \geq a_j$.

Let $\widehat{A}_j^* = \widetilde{A}_i^*$ and $\widehat{A}_i^* = \widetilde{A}_j^*$. We have

$$[\widetilde{\eta}_i^* \widetilde{M}_i(\widehat{A}_i^*) + \widetilde{\eta}_j^* \widetilde{M}_j(\widehat{A}_j^*)] - [\widetilde{\eta}_i^* \widetilde{M}_i(\widetilde{A}_i^*) + \widetilde{\eta}_j^* \widetilde{M}_j(\widetilde{A}_j^*)]$$
$$= [\widetilde{\eta}_i^* \widetilde{M}_i(\widetilde{A}_j^*) + \widetilde{\eta}_j^* \widetilde{M}_j(\widetilde{A}_i^*)] - [\widetilde{\eta}_i^* \widetilde{M}_i(\widetilde{A}_i^*) + \widetilde{\eta}_j^* \widetilde{M}_j(\widetilde{A}_j^*)]$$

$$=\widetilde{\eta}_j^*[\widetilde{M}_j(\widehat{A}_i^*) - \widetilde{M}_j(\widetilde{A}_j^*)] + \widetilde{\eta}_i^*[\widetilde{M}_i(\widetilde{A}_j^*) - \widetilde{M}_i(\widetilde{A}_i^*)]$$
$$\geq \widetilde{\eta}_i^*[\widetilde{M}_j(\widehat{A}_i^*) - \widetilde{M}_j(\widetilde{A}_j^*) + \widetilde{M}_i(\widetilde{A}_j^*) - \widetilde{M}_i(\widetilde{A}_i^*)] > 0, \quad (50)$$

where (50) is due to property (M-3').

Thus, if $\widetilde{A}_i^*$ and $\widetilde{A}_j^*$ are replaced by $\widehat{A}_i^*$ and $\widehat{A}_j^*$ respectively, we obtain a larger $\mathbf{F}'$, leading to a contradiction.

Case 1.2: $\widetilde{A}_i^* < a_j$.

Let $\widehat{A}_j^* = a_j$, $D = \widetilde{A}_j^* - a_j$ and $\widehat{A}_i^* = \widetilde{A}_i^* + D$. We have

$$[\widetilde{\eta}_i^*\widetilde{M}_i(\widehat{A}_i^*) + \widetilde{\eta}_j^*\widetilde{M}_j(\widehat{A}_j^*)] - [\widetilde{\eta}_i^*\widetilde{M}_i(\widetilde{A}_i^*) + \widetilde{\eta}_j^*\widetilde{M}_j(\widetilde{A}_j^*)]$$
$$=[\widetilde{\eta}_i^*\widetilde{M}_i(\widetilde{A}_i^* + D) + \widetilde{\eta}_j^*\widetilde{M}_j(a_j)] - [\widetilde{\eta}_i^*\widetilde{M}_i(\widetilde{A}_i^*) + \widetilde{\eta}_j^*\widetilde{M}_j(\widetilde{A}_j^*)]$$
$$=\widetilde{\eta}_j^*[\widetilde{M}_j(a_j) - \widetilde{M}_j(\widetilde{A}_j^*)] + \widetilde{\eta}_i^*[\widetilde{M}_i(\widetilde{A}_i^* + D) - \widetilde{M}_i(\widetilde{A}_i^*)]$$
$$\geq \widetilde{\eta}_i^*[\widetilde{M}_j(a_j) - \widetilde{M}_j(\widetilde{A}_j^*) + \widetilde{M}_i(\widetilde{A}_i^* + D) - \widetilde{M}_i(\widetilde{A}_i^*)] > 0, \quad (51)$$

where (51) is due to property (M-6').

Thus, if $\widetilde{A}_i^*$ and $\widetilde{A}_j^*$ are replaced by $\widehat{A}_i^*$ and $\widehat{A}_j^*$ respectively, we obtain a larger $\mathbf{F}'$, leading to a contradiction.

Case 2: $\widetilde{\eta}_i^* > \widetilde{\eta}_j^*$.

Case 2.1: $\widetilde{A}_i^* \geq a_j$.

Let $\widehat{A}_j^* = \widetilde{A}_i^*$, $\widehat{A}_i^* = \widetilde{A}_j^*$, $\widehat{\eta}_j^* = \widetilde{\eta}_i^*$, and $\widehat{\eta}_i^* = \widetilde{\eta}_j^*$. (Note that because $\eta_{\min,i} \leq \eta_{\min,j}$ and $\eta_{\max,i} \leq \eta_{\max,j}$, $\widehat{\eta}_j^*$ and $\widehat{\eta}_i^*$ are guaranteed to be in the feasible region.)

$$[\widehat{\eta}_i^*\widetilde{M}_i(\widehat{A}_i^*) + \widehat{\eta}_j^*\widetilde{M}_j(\widehat{A}_j^*)] - [\widetilde{\eta}_i^*\widetilde{M}_i(\widetilde{A}_i^*) + \widetilde{\eta}_j^*\widetilde{M}_j(\widetilde{A}_j^*)]$$
$$=[\widehat{\eta}_i^*\widetilde{M}_i(\widetilde{A}_j^*) + \widehat{\eta}_j^*\widetilde{M}_j(\widetilde{A}_i^*)] - [\widetilde{\eta}_i^*\widetilde{M}_i(\widetilde{A}_i^*) + \widetilde{\eta}_j^*\widetilde{M}_j(\widetilde{A}_j^*)]$$
$$=\widetilde{\eta}_i^*[\widetilde{M}_j(\widetilde{A}_i^*) - \widetilde{M}_i(\widetilde{A}_i^*)] + \widetilde{\eta}_j^*[\widetilde{M}_i(\widetilde{A}_j^*) - \widetilde{M}_j(\widetilde{A}_j^*)]$$
$$>\widetilde{\eta}_j^*[\widetilde{M}_j(\widetilde{A}_i^*) - \widetilde{M}_i(\widetilde{A}_i^*) + \widetilde{M}_i(\widetilde{A}_j^*) - \widetilde{M}_j(\widetilde{A}_j^*)] > 0. \quad (52)$$

Thus, if $\widetilde{A}_i^*$, $\widetilde{A}_j^*$, $\widetilde{\eta}_i^*$, and $\widetilde{\eta}_j^*$ are replaced by $\widehat{A}_i^*$, $\widehat{A}_j^*$, $\widehat{\eta}_i^*$, and $\widehat{\eta}_j^*$ respectively, we obtain a larger $\mathbf{F}'$, leading to a contradiction.

Case 2.2: $\widetilde{A}_i^* < a_j$.

Let $\widehat{A}_j^* = a_j$, $D = \widetilde{A}_j^* - a_j$, $\widehat{A}_i^* = \widetilde{A}_i^* + D$, $\widehat{\eta}_j^* = \widetilde{\eta}_i^*$, and $\widehat{\eta}_i^* = \widetilde{\eta}_j^*$. We have

$$[\widehat{\eta}_i^*\widetilde{M}_i(\widehat{A}_i^*) + \widehat{\eta}_j^*\widetilde{M}_j(\widehat{A}_j^*)] - [\widetilde{\eta}_i^*\widetilde{M}_i(\widetilde{A}_i^*) + \widetilde{\eta}_j^*\widetilde{M}_j(\widetilde{A}_j^*)]$$
$$=[\widetilde{\eta}_j^*\widetilde{M}_i(\widetilde{A}_i^* + D) + \widetilde{\eta}_i^*\widetilde{M}_j(a_j)] - [\widetilde{\eta}_i^*\widetilde{M}_i(\widetilde{A}_i^*) + \widetilde{\eta}_j^*\widetilde{M}_j(\widetilde{A}_j^*)]$$
$$=\widetilde{\eta}_i^*[\widetilde{M}_j(a_j) - \widetilde{M}_i(\widetilde{A}_i^*)] + \widetilde{\eta}_j^*[\widetilde{M}_i(\widetilde{A}_i^* + D) - \widetilde{M}_j(\widetilde{A}_j^*)]$$
$$>\widetilde{\eta}_j^*[\widetilde{M}_j(a_j) - \widetilde{M}_i(\widetilde{A}_i^*) + \widetilde{M}_i(\widetilde{A}_i^* + D) - \widetilde{M}_j(\widetilde{A}_j^*)] > 0. \quad (53)$$

Thus, if $\widetilde{A}_i^*$, $\widetilde{A}_j^*$, $\widetilde{\eta}_i^*$, and $\widetilde{\eta}_j^*$ are replaced by $\widehat{A}_i^*$, $\widehat{A}_j^*$, $\widehat{\eta}_i^*$, and $\widehat{\eta}_j^*$ respectively, we obtain a larger $\mathbf{F}'$, leading to a contradiction. $\square$

F. Proof of Theorem 4

Proof. Suppose $\exists k \geq 2$ such that $\widetilde{A}_k^* > a_k$. Let $l = 1$, $\widehat{A}_k^* = a_k$, $D = \widetilde{A}_k^* - a_k$, and $\widehat{A}_l^* = \widetilde{A}_l^* + D$. Note that we have $\widetilde{\eta}_k^* \geq \widetilde{\eta}_l^*$ through Lemma 3. Thus,

$$[\widetilde{\eta}_k^*\widetilde{M}_k(\widehat{A}_k^*) + \widetilde{\eta}_l^*\widetilde{M}_l(\widehat{A}_l^*)] - [\widetilde{\eta}_k^*\widetilde{M}_k(\widetilde{A}_k^*) + \widetilde{\eta}_l^*\widetilde{M}_l(\widetilde{A}_l^*)]$$
$$=[\widetilde{\eta}_k^*\widetilde{M}_k(a_k) + \widetilde{\eta}_l^*\widetilde{M}_l(\widetilde{A}_l^* + D)] - [\widetilde{\eta}_k^*\widetilde{M}_k(\widetilde{A}_k^*) + \widetilde{\eta}_l^*\widetilde{M}_l(\widetilde{A}_l^*)]$$
$$=\widetilde{\eta}_k^*[\widetilde{M}_k(a_k) - \widetilde{M}_k(\widetilde{A}_k^*)] + \widetilde{\eta}_l^*[\widetilde{M}_l(\widetilde{A}_l^* + D) - \widetilde{M}_l(\widetilde{A}_l^*)]$$
$$\geq \widetilde{\eta}_l^*[\widetilde{M}_k(a_k) - \widetilde{M}_k(\widetilde{A}_k^*) + \widetilde{M}_l(\widetilde{A}_l^* + D) - \widetilde{M}_l(\widetilde{A}_l^*)] > 0. \quad (54)$$

As a consequence, if we replace $\widetilde{A}_l^*$ and $\widetilde{A}_k^*$ by $\widehat{A}_l^*$ and $\widehat{A}_k^*$ respectively, we find a better solution to Problem **P2A**, which leads to a contradiction. $\square$

G. Some Properties Used for Exhaustive Search

Lemma 4. *Let* $(\boldsymbol{\eta}^*, \mathbf{A}^*)$ *be an optimal solution to* **P***. Suppose* $\exists i \neq j$, *such that* $\eta_{\min,i} < \eta_i^* < \eta_{\max,i}$ *and* $\eta_{\min,j} < \eta_j^* < \eta_{\max,j}$. *Then, let* $\widehat{\eta}_i^* = \eta_{\max,i}$ *and* $\widehat{\eta}_j^* = \eta_i^* + \eta_j^* - \eta_{\max,i}$ *if* $\eta_{\max,i} - \eta_i^* \leq \eta_j^* - \eta_{\min,j}$; *let* $\widehat{\eta}_i^* = \eta_i^* + \eta_j^* - \eta_{\min,j}$ *and* $\widehat{\eta}_j^* = \eta_{\min,j}$ *otherwise. Let* $\widehat{\boldsymbol{\eta}}^* = (\eta_1^*, \ldots, \widehat{\eta}_i^*, \ldots, \widehat{\eta}_j^*, \ldots, \eta_K^*)$. *Then* $(\widehat{\boldsymbol{\eta}}^*, \mathbf{A}^*)$ *is still an optimal solution to Problem* **P***.*

Proof. First, we have $M_i(A_i^*) = M_j(A_j^*)$. Otherwise, suppose $M_i(A_i^*) > M_j(A_j^*)$ (without loss of generality); then we can find a better solution by replacing η_i^* and η_j^* by $\eta_i^* + \delta$ and $\eta_j^* - \delta$ respectively, where $\delta > 0$.

Given that $M_i(A_i^*) = M_j(A_j^*)$, the same $\mathbf{F}$ can be obtained when we replace η_i^* and η_j^* by $\widehat{\eta}_i^*$ and $\widehat{\eta}_j^*$. $\square$

Lemma 4 demonstrates that if there are η_i^* and η_j^* not at the boundary, we can "push" one of them to the boundary and maintain the optimization in **P**. If there are more than two terms in $\boldsymbol{\eta}^*$ not at the boundary, we can "push" them to the boundary one by one, until there is at most one term in $\boldsymbol{\eta}^*$ not at the boundary. Thus, Lemma 4 directly leads to the following Theorem:

Theorem 5. *At least one of the optimal solutions to* **P***,* $(\boldsymbol{\eta}^*, \mathbf{A}^*)$, *has the following property: There is at most one* $k \in \{1, 2, \ldots K\}$ *such that* $\eta_{\min,k} < \eta_k^* < \eta_{\max,k}$, *and* $\forall j \neq k$, *either* $\eta_j^* = \eta_{\min,j}$ *or* $\eta_j^* = \eta_{\max,j}$.

H. Proof of Properties of $M_k(\cdot)$ in the Multiple-MCS Case

In this subsection, we prove Properties (M-1) to (M-4) when we redefine $M_k(A)$ as in (46). In the proof, we set $u_k = \mu/\lambda_k$.

1) Property (M-1): The first derivative of $M_k(A)$ is

$$M_k'(A) = \frac{1}{(Au_k + 1)^2}\left[\sum_{n=1}^N \frac{b_n(1 - A^2 C_n u_k)}{(AC_n + 1)^2}\right]. \quad (55)$$

$M_k'(A)$ is positive at $A = 0$, and is negative when A is sufficient large. Thus, there exists positive a_k such that it is a solution to $M_k'(A) = 0$. In addition, we have $\left[\sum_{n=1}^N \frac{b_n(1 - A^2 C_n u_k)}{(AC_n + 1)^2}\right]' = -\sum_{n=1}^N \frac{2b_n C_n(Au_k + 1)}{(AC_n + 1)^3} < 0$, demonstrating that $\sum_{n=1}^N \frac{b_n(1 - A^2 C_n u_k)}{(AC_n + 1)^2}$ is a decreasing function. Thus, a_k is the unique solution to $M_k'(A) = 0$. $M_k'(A)$ is positive on $[0, a_k)$, and negative on (a_k, ∞). Note that a simple binary search method can be applied to numerically search for a_k.

Note that

$$\sum_{n=1}^N \frac{b_n(1 - A^2 C_n u_k)}{(AC_n + 1)^2} > 0 \quad (56)$$

on $[0, a_k)$, which will be used in the subsequent steps.

2) Property (M-2): We have

$$M_k''(A) = \frac{\left[\sum_{n=1}^{N} z_n'(A)\right] v(A) - \left[\sum_{n=1}^{N} z_n(A)\right] v'(A)}{(v(A))^2},$$

(57)

where $z_n(A) = \frac{b_n(1-A^2C_nu_k)}{(AC_n+1)^2}$, and $v(A) = (Au_k+1)^2$. Then, we have

$$z_n'(A) = -\frac{2b_nC_n(Au_k+1)}{(AC_n+1)^3},$$

(58)

which is negative.

Consequently, on $[0, a_k)$, $\sum_{n=1}^{N} z_n'(A)$ is negative, $v(A)$ is positive, $\sum_{n=1}^{N} z_n(A)$ is positive (due to (56)), and $v'(A)$ is positive. Thus, $M_k''(A)$ is negative. Property (M-2) is proved.

3) Property (M-3): Trivially true.

4) Property (M-4): Let $\lambda_k > \lambda_j$. Then we have $u_k < u_j$. Let $\Delta M_{kj}(A) = M_k(A) - M_j(A)$. Then

$$\Delta M_{kj}'(A) = M_k'(A) - M_j'(A)$$

(59)

$$= \frac{u_k - u_j}{(Au_k+1)^2(Au_j+1)^2} \cdot$$

(60)

$$\left[\sum_{n=1}^{N} \frac{Ab_n(A^3C_nu_ku_j - Au_k - Au_j - AC_n - 2)}{(AC_n+1)^2}\right]$$

$$= \frac{u_k - u_j}{(Au_k+1)^2(Au_j+1)^2} \cdot$$

(61)

$$\left[\sum_{n=1}^{N} \frac{A^2b_nu_j(A^2C_nu_k - 1)}{(AC_n+1)^2} - \sum_{n=1}^{N} \frac{Ab_n(Au_k + AC_n + 2)}{(AC_n+1)^2}\right].$$

Note that $\sum_{n=1}^{N} \frac{A^2b_nu_j(A^2C_nu_k-1)}{(AC_n+1)^2} = A^2u_j\sum_{n=1}^{N} \frac{b_n(A^2C_nu_k-1)}{(AC_n+1)^2} < 0$ on $[0, a_k)$ (see (56)), and $u_k < u_j$. Consequently, $\Delta M_{kj}'(A) > 0$. Property (M-4) is proved.

I. Numerical Verification of Properties (M-2') and (M-5') of $\widetilde{M}_k(A_k)$ in the Multiple-MCS Case

Fig. 14 visualizes a typical example in our numerical verification of Properties (M-2') and (M-5'). In this figure, we narrow the parameter sets to $\mathcal{T} = \{10^{-1}, 10^{-0.5}, 1, 10^{0.5}, 10^1\}$, $\mathcal{U} = \{10^1, 10^{1.5}, \ldots, 10^3\}$, and $\Gamma = \{4\}$. For all $T_1, T_2, T_3 \in \mathcal{T}$ ($T_1 < T_2 < T_3$), $\forall \mu/\lambda_i, \mu/\lambda_j \in \mathcal{U}$ ($\mu/\lambda_i > \mu/\lambda_j$), and $\forall \gamma \in \Gamma$, we plot $[\widetilde{M}_j(a_j) - \widetilde{M}_j(a_j+D)] - [\widetilde{M}_i(a_i) - \widetilde{M}_i(a_i+D)]$ versus D, and values of $\widetilde{M}_j(a_j) - \widetilde{M}_i(a_i)$. Consequently, 100 curves and points are plotted in Fig. 14. The curves and points are all above zero, illustrating the correctness of Properties (M-2') and (M-5') under the range of parameter settings that are tested. The results for other parameter setting are similar and are omitted to avoid redundancy.
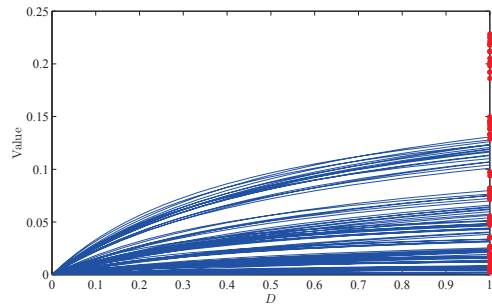

Fig. 14. $[\widetilde{M}_j(a_j) - \widetilde{M}_j(a_j + D)] - [\widetilde{M}_i(a_i) - \widetilde{M}_i(a_i + D)]$ versus D (curves) and $\widetilde{M}_j(a_j) - \widetilde{M}_i(a_i)$ (dots).

REFERENCES

[1] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3GPP heterogeneous networks," *IEEE Wireless Commun.*, vol. 18, no. 3, pp. 10–21, Jun. 2011.

[2] "Performance of eICIC with control channel coverage limitation," 3GPP Standard Contribution, R1-103264, NTT DOCOMO, May 2010.

[3] R. Madan, J. Borran, A. Sampath, N. Bhushan, A. Khandekar, and T. Ji, "Cell association and interference coordination in heterogeneous LTE-A cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1479–1489, Dec. 2010.

[4] D. Lopez-Perez, X. Chu, and I. Guvenc, "On the expanded region of picocells in heterogeneous networks," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 3, pp. 281–294, Jun. 2012.

[5] H.-S. Jo, Y. J. Sang, P. Xia, and J. G. Andrews, "Heterogeneous cellular networks with flexible cell association: A comprehensive downlink SINR analysis," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3484–3495, Oct. 2012.

[6] S. Singh, H. Dhillon, and J. G. Andrews, "Offloading in heterogeneous networks: Modeling, analysis, and design insights," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2484–2497, May 2013.

[7] W. C. Cheung, T. Q. S. Quek, and M. Kountouris, "Stochastic analysis of two-tier networks: Effect of spectrum allocation," in *Proc. of IEEE ICASSP*, Prague, Czech Republic, May 2011, pp. 2964–2967.

[8] V. Chandrasekhar and J. G. Andrews, "Spectrum allocation in tiered cellular networks," *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 3059–3068, Oct. 2009.

[9] W. C. Cheung, T. Q. S. Quek, and M. Kountouris, "Throughput optimization, spectrum allocation, and access control in two-tier femtocell networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 561–574, Apr. 2012.

[10] F. Baccelli and B. Blaszczyszyn, "Stochastic geometry and wireless networks, volume 1: Theory," *Foundations and Trends in Networking*, vol. 3, no. 3-4, pp. 249 – 449, 2009.

[11] ——, "Stochastic geometry and wireless networks, volume 2: Applications," *Foundations and Trends in Networking*, vol. 4, no. 1-2, pp. 1–312, 2009.

[12] M. Haenggi, J. G. Andrews, F. Baccelli, O. Dousse, and M. Franceschetti, "Stochastic geometry and random graphs for the analysis and design of wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 7, pp. 1029 – 1046, Sep. 2009.

[13] H. ElSawy, E. Hossain, and M. Haenggi, "Stochastic geometry for modeling, analysis, and design of multi-tier and cognitive cellular wireless networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 996–1019, Third Quarter, 2013.

[14] D. Stoyan, W. Kendall, and J. Mecke, *Stochastic Geometry and Its Applications*, 2nd ed. Wiley, 1995.

[15] J. G. Andrews, F. Baccelli, and R. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 3122–3134, Nov. 2011.

[16] H. Dhillon, R. Ganti, F. Baccelli, and J. G. Andrews, "Modeling and analysis of K-tier downlink heterogeneous cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 550–560, Apr. 2012.

[17] H.-S. Jo, P. Xia, and J. G. Andrews, "Open, closed, and shared access femtocells in the downlink," *EURASIP J. on Wireless Commun. and Netw.*, vol. 2012, no. 363, Dec. 2012.

[18] S. Kandukuri and S. Boyd, "Optimal power control in interference-limited fading wireless channels with outage-probability specifications," *IEEE Trans. Wireless Commun.*, vol. 1, no. 1, pp. 46–55, Jan. 2012.

[19] C. W. Tan, "Optimal power control in Rayleigh-fading heterogeneous networks," in *Proc. of IEEE INFOCOM*, Shanghai, China, Apr. 2011, pp. 2552–2560.

[20] ——, "Optimal power control in Rayleigh-fading heterogeneous networks," *accepted for publication in IEEE/ACM Trans. Netw.*, 2015.

[21] L. Zheng and C. W. Tan, "Optimal algorithms in wireless utility maximization: Proportional fairness decomposition and nonlinear Perron-

Frobenius theory framework," *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 2086–2095, Feb. 2014.

[22] J.-W. Lee, R. Mazumdar, and N. Shroff, "Joint resource allocation and base-station assignment for the downlink in CDMA networks," *IEEE/ACM Trans. Netw.*, vol. 14, no. 1, pp. 1–14, Feb. 2006.

[23] 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.

[24] M. Hong and A. Garcia, "Mechanism design for base station association and resource allocation in downlink OFDMA network," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 11, pp. 2238–2250, Dec. 2012.

[25] C. S. Chen, F. Baccelli, and L. Roullet, "Joint optimization of radio resources in small and macro cell networks," in *Proc. of VTC (Spring)*, Budapest, Hungary, May 2011.

[26] K. Shen and W. Yu, "Downlink cell association optimization for het- eregeneous networks via dual coordinate descent," in *Proc. of IEEE ICASSP*, Vancouver, Canada, May 2013, pp. 4779–4783.

[27] D. Fooladivanda, A. A. Daoud, and C. Rosenberg, "Joint channel alloca- tion and user association for heterogeneous wireless cellular networks," in *Proc. of IEEE PIMRC*, Toronto, Canada, Sep. 2011, pp. 384–390.

[28] W. Bao and B. Liang, "Understanding the benefits of open access in femtocell networks: Stochastic geometric analysis in the uplink," in *Proc. of ACM/IEEE MSWiM*, Barcelona, Spain, Nov. 2013, pp. 237–246.

[29] ——, "Stochastic analysis of uplink interference in two-tier femtocell networks: Open versus closed access," *accepted for publication in IEEE Trans. Wireless Commun.*, 2015.

[30] S. Singh and J. G. Andrews, "Joint resource partitioning and offloading in heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 888–901, Feb. 2014.

[31] Y. Lin and W. Yu, "Optimizing user association and frequency reuse for heterogeneous network under stochastic model," in *Proc. of IEEE Globecom*, Atlanta, GA, Dec. 2013, pp. 2045–2050.

[32] ——, "Joint spectrum partition and user association in multi-tier het- erogeneous networks," in *Proc. of CISS*, Princeton, NJ, Mar. 2014.

[33] Y. Lin, W. Bao, W. Yu, and B. Liang, "Optimizing user association and spectrum allocation in HetNets: a utility perspective," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 6, pp. 1025–1039, Jun. 2015.

[34] W. Bao and B. Liang, "Structured spectrum allocation and user associa- tion in heterogeneous cellular networks," in *Proc. of IEEE INFOCOM*, Toronto, Canada, Apr.-May 2014, pp. 1069–1077.

[35] P. F. Ash and E. D. Bolker, "Generalized Dirichlet tessellations," *Geometriae Dedicata*, vol. 20, no. 2, pp. 209–243, Apr. 1986.

[36] J.-S. Ferenc and Z. Neda, "On the size distribution of Poisson Voronoi cells," *Physica A: Statistical Mechanics and its Applications*, vol. 385, no. 2, pp. 518–526, Nov. 2007.

[37] M. Udell and S. Boyd, "Maximizing a sum of sigmoids." [Online]. Available: http://www.stanford.edu/~udell/doc/max_sum_sigmoids.pdf

[38] M. Chiang, "Nonconvex optimization of communication systems," *Ad- vances in Mechanics and Mathematics, Special Volumn on Strang's 70th Birthday*, pp. 137–196, Oct. 2007.

[39] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[40] 3GPP-TR-36.814, "Evolved universal terrestrial radio access (E-UTRA); further advancements for E-UTRA physical layer aspects (release 9)," Mar. 2010.

[41] A. M. Ibrahim, T. ElBatt, and A. El-Keyi, "Coverage probability analysis for wireless networks using repulsive point processes," in *Proc. of IEEE PIMRC*, London, UK, Sep. 2013, pp. 1002–1007.

[42] A. Guo and M. Haenggi, "Spatial stochastic models and metrics for the structure of base stations in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 11, pp. 5800–5812, Nov. 2013.

[43] W. Bao and B. Liang, "Radio resource allocation in heterogeneous wireless networks: a spatial-temporal perspective," in *Proc. of IEEE INFOCOM*, Hong Kong, China, Apr. 2015, pp. 334–342.

Wei Bao (S'10) received the B.E. degree in Commu- nications Engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2009, and the M.A.Sc. degree in Electrical and Computer Engineering from the University of British Columbia, Vancouver, BC, Canada, in 2011. He is currently a Ph.D. student at the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada. His research interests are stochastic and queueing networks, stochastic ge- ometry, and performance analysis of heterogeneous wireless networks. He received the Ontario Trillium Scholarship in 2011, and the Best Paper Award in ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM) in 2013.
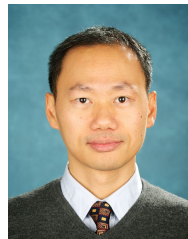
Ben Liang (S'94-M'01-SM'06) received honors- simultaneous B.Sc. (valedictorian) and M.Sc. de- grees in Electrical Engineering from Polytechnic University in Brooklyn, New York, in 1997 and the Ph.D. degree in Electrical Engineering with Computer Science minor from Cornell University in Ithaca, New York, in 2001. In the 2001 - 2002 academic year, he was a visiting lecturer and post- doctoral research associate at Cornell University. He joined the Department of Electrical and Computer Engineering at the University of Toronto in 2002, where he is now a Professor. His current research interests are in mobile communications and networked systems. He has served as an Editor for IEEE TRANSACTIONS ON COMMUNICATIONS, as an Editor for IEEE TRANSAC- TIONS ON WIRELESS COMMUNICATIONS, and as an Associate Editor for the Wiley SECURITY AND COMMUNICATION NETWORKS journal, in addition to regularly serving on the organizational and technical committees of a number of conferences. He is a senior member of IEEE and a member of ACM and Tau Beta Pi.