

# Radio Resource Allocation in Heterogeneous Wireless Networks: A Spatial-Temporal Perspective

Wei Bao and Ben Liang

Department of Electrical and Computer Engineering, University of Toronto, Canada

Email: {wbao, liang}@comm.utoronto.ca

**Abstract**—We study optimal radio resource allocation across multiple tiers of a heterogeneous wireless network in order to maximize the downlink sum throughput. Different from prior works, we consider both the randomness of base stations in space and dynamic user traffic session arrivals in time, accounting for both elastic and inelastic user traffic. A new stochastic analysis framework, which accommodates both spatial and temporal dimensions, is proposed to quantify the throughput objective. The derived throughput function is not in closed form and is non-concave in terms of the radio resource allocation factors to be optimized, hindering the search for an efficient optimization solution. Therefore, we further develop closed-form concave bounds that envelop the throughput function, to form convex approximations of the original optimization problem that can be solved efficiently. We characterize the performance gap when these bounds are used instead of the original objective. Both analytical bounding and simulation experiments demonstrate that the proposed solution is nearly optimal.

## I. INTRODUCTION

Future cellular networks are expected to integrate a diverse set of both macrocells and small-cells (e.g., picocells and femtocells). Each small-cell is equipped with a shorter-range and lower-cost base station (BS) or access point (AP), to provide nearby User Equipments (UEs) with high bandwidth network access with low power usage, and to offload data traffic from macrocells. In such multi-tier Heterogeneous Wireless Networks (HWNs), the small-cell BSs are spread irregularly, sometimes in an “anywhere plug-and-play” manner, leading to a high level of spatial randomness. This results in complicated cell shapes and spatial interference patterns, which imposes substantial challenges to data rate analysis and radio resource management.

Further complicating the problem is the recent proliferation of smart mobile devices and the rapid growth of multimedia traffic. A defining characteristic of multimedia traffic is inelasticity. The required network throughput for a multimedia session is defined in discrete levels, and a minimum throughput must be satisfied. Meanwhile, elastic traffic, such as file transfer and instant messaging, remains important. In contrast to inelastic traffic, elastic traffic is more tolerant of packet delays, and its transmission rate can be decreased when the network is congested. The different timing requirements of inelastic and elastic traffic sessions add complicated temporal randomness to the spatial randomness of HWNs.

Fig. 1 illustrates an example of a three-tier HWN in two-dimensional space over time. The  $x$ - $y$  plane shows the locations of BSs and UEs, as well as cell boundaries. The  $t$ -axis indicates the temporal evolution of the HWN. We note

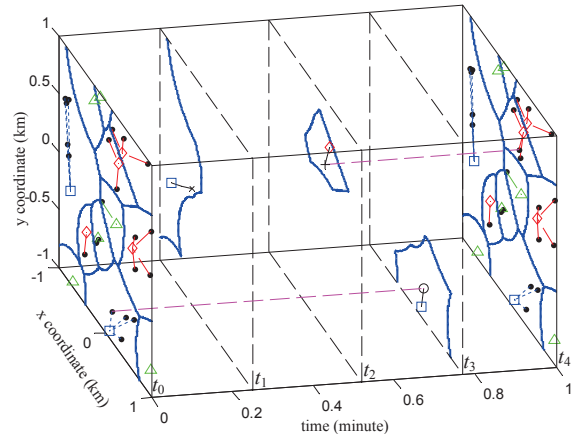


Fig. 1. An example of a three-tier HWN in both spatial and temporal dimensions. Macrocell BSs, picocell BSs, and femtocell BSs are represented by “□”, “◇”, and “△” respectively. UEs are represented by “.”. Blue solid lines show cell boundaries. At  $t_1$ , a UE rejection is shown as “×”. At  $t_2$ , a UE arrival is shown as “+”, and its session holding time is represented as the magenta dashed line between  $t_2$  and  $t_4$ . At  $t_3$ , a UE departure is shown as “o”, and its session holding time is represented as the magenta dashed line between  $t_0$  and  $t_3$ .

that complicated random spatial patterns are introduced by the randomly distributed BSs and different scales of cell sizes, and complicated random temporal patterns are introduced by the arrivals and departures of UE sessions. Furthermore, the mixture of elastic and inelastic traffic plays an important role in the radio resource management, influencing the temporal evolution of the HWN. At  $t_1$ , a new arrival is rejected because the cell is already occupied by many other active UEs, and the corresponding BS cannot satisfy the inelastic traffic requirement of the new arrival. At  $t_2$ , a new UE session arrival is admitted by the system, and the required radio resource at the corresponding BS is allocated to this UE at this instant. At  $t_3$ , a UE leaves the system, releasing its allocated radio resource to other potential users.

In the presence of the multi-tier structure in HWNs, the expensive radio spectrum licensed by the network operator must be shared by different tiers of BSs, each with widely different power and coverage areas. How to optimally allocate the radio resource among different tiers remains an 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 allocation mode is commonly advocated [1]–[4], where different tiers of BSs are allocated non-overlapping

portions of the radio resource. However, very few studies have tackled the problem of optimal radio resource apportioning among multiple tiers, especially when considering both spatial and temporal patterns together with mixed inelastic and elastic traffic.

In this work, our objective is to optimize the allocation of radio resource across multiple tiers in an HWN that supports mixed inelastic and elastic traffic, in order to maximize the downlink sum throughput. Our main contributions are as follows:

- In order to characterize the random spatial and temporal patterns of the network, we first present a multi-dimensional stochastic analysis model, accounting for both spatial and temporal dimensions, where each tier of BSs and UE session arrivals form homogeneous multi-dimensional Poisson point processes (PPPs). We allow the UE session holding time to be generally distributed. We then quantify the average downlink sum throughput of the system in terms of given resource allocation factors.
- Since the derived sum throughput is not in closed form and is non-concave, there is no efficient method to solve the original resource optimization problem. Instead, we further derive concave upper and lower bounds to the sum throughput in closed form, which are used to provide efficient approximate solutions to the original problem. We characterize the performance gap when these bounds are used instead of the original objective, to show that the performance gap is below 12.6%.
- Finally, simulation experiments are conducted to validate the correctness and usefulness of our theoretical conclusions. The simulation results suggest that the proposed solution often performs much closer to optimality than the analytical bound.

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 of the paper. In Sections IV and V, we present our contributions in the throughput performance derivation and our radio resource allocation solution respectively. In Section VI, we present simulation results. Finally, conclusions are given in Section VII.

## II. RELATED WORKS

Prior works on resource allocation in HWNs may be categorized based their modeling of randomness in the spatial and temporal dimensions.

Some previous studies on resource allocation in cellular networks, such as [5]–[7], considered only a fixed number of BSs and UEs, without accounting for their random spatial or temporal patterns. In other studies, such as [8], [9], the random UE arrivals and departures in time were taken into account. Błaszczyszyn *et al.* in [10] established a stochastic service model to evaluate the quality of real-time streaming in cellular networks, advocating a least-effort-served-first resource allocation policy. Joseph *et al.* in [11] studied optimal rate allocation and admission control for adaptive video streaming in wireless

networks. However, these works still did not consider spatial randomness in the network.

To address spatial randomness, the theory of stochastic geometry [12]–[14] has been applied in the development of tractable models for cellular networks and HWNs. Examples of downlink performance analysis include [15]–[19], for either the single-tier case or the multi-tier case. Toward resource allocation, Cheung *et al.* in [3] studied optimal spectrum sharing in two-tier cellular networks. Joint spectrum allocation and user association were studied in [20], where the authors analyzed the network coverage probability and data rate, but did not provide an analytical solution to the optimization problem. In [4], [21], [22], computationally efficient methods were proposed to jointly optimize spectrum allocation and user association. However, none of these works considered temporal randomness.

More recent works have drawn attention to both spatial and temporal randomness. Through stochastic geometric analysis, the authors of [23]–[25] studied the outage performance of mobile ad hoc networks considering both spatial and temporal random patterns. In these works, the temporal randomness arises from the temporal correlation of the interference signals, instead of the UE traffic sessions considered in our work. Furthermore, the derived performance metrics are generally in non-closed forms and cannot be further optimized.

As far as we are aware, the only existing study that jointly considers spatial and temporal randomness in resource optimization was presented in [26]. However, that work was limited in several aspects: (1) only inelastic traffic was considered; (2) each UE was assumed to occupy one unit spectrum bandwidth; and (3) session holding times were assumed to be exponentially distributed. The current work accommodates a significantly more general system model, with both inelastic and elastic traffic, multi-unit spectrum requirements, and general session holding times.

## III. SYSTEM MODEL

### A. Multi-tier HWN

We consider an HWN with  $K$  tiers of BSs. In order to capture the random spatial patterns of the network, each tier of BSs is assumed to independently form a homogeneous Poisson point process (PPP) in two-dimensional Euclidean space  $\mathbb{R}^2$ . Let  $\Phi_k$  denote the PPP corresponding to tier- $k$  BSs, with intensity  $\lambda_k$ . The locations of BSs do not change over time. Two types of UEs are studied in this paper, UEs with inelastic traffic (IUEs) and UEs with elastic traffic (EUEs). We consider only UEs with an active traffic session. Therefore, for brevity, we refer to UE traffic session arrivals and departures synonymously as *UE arrivals and departures*. Due to dynamic session arrivals and departures, the UEs may be viewed as distributed randomly also in the temporal dimension. The arrivals of IUEs and EUEs are modeled as homogeneous PPPs  $\Psi_I$  and  $\Psi_E$  with intensities  $\mu_I$  and  $\mu_E$ , respectively, in  $\mathbb{R}^3$  (accounting for two spatial dimensions and one temporal dimension). Note that the spatial Poisson modeling of BSs and UEs follows conventions in stochastic geometric analysis in HWNs [15]–[22], and the temporal Poisson arrivals of traffic

sessions have been well supported by experimental data [27]–[30].

An arriving UE may be rejected or admitted by the system. If it is rejected, it leaves immediately. If it is admitted, it will stay in the system for a random duration, termed the *session holding time*, which has a general distribution with mean  $\frac{1}{\nu}$ . For different UEs, their session holding times are independent. In this work we consider the scenario where all UEs remain stationary. The additional consideration of user mobility in the spatial and temporal dimensions will introduce substantial difficulties to system modeling<sup>1</sup>, which is beyond the scope of this paper.

The HWN is operated on  $W$  orthogonal resource blocks (RBs)<sup>2</sup> in the downlink. Without loss of generality, we assume that the size of an RB is 1 unit time  $\times$  1 unit frequency bandwidth in our analytical study (i.e., Section III - V). However, we assign a practical value on the size of an RB in our numerical study in Section VI.

In order to avoid cross-tier interference, different tiers of BSs are operated on non-overlapping RBs [1]–[4]. The network operator allocates  $W_k = \eta_k W$  to each tier- $k$  BS, where  $\eta_k$  is the resource allocation factor, and  $\sum_{k=1}^K \eta_k = 1$ . We define  $\boldsymbol{\eta} \triangleq (\eta_1, \eta_2, \dots, \eta_K)$ . Note that BSs in the same tier are operated on the same RBs and will interfere at the co-tier UEs, but there is no cross-tier interference. In this work, we aim to optimize the resource allocation factors  $\boldsymbol{\eta}$ , in order to maximize the normalized average downlink sum throughput over all UEs.

### B. Power, Pathloss, and Cell Splitting

Let  $P_k$  be the transmission power of tier- $k$  BSs, which is a given parameter. If  $P_t(\mathbf{x}) \in \{P_1, \dots, P_K\}$  is the transmission power from a BS at location  $\mathbf{x}$  and  $P_r(\mathbf{y})$  is the received power at  $\mathbf{y}$ , we assume  $P_r(\mathbf{y}) = \frac{P_t(\mathbf{x})h_{\mathbf{x},\mathbf{y}}}{\beta|\mathbf{x}-\mathbf{y}|^\gamma}$ , where  $\beta|\mathbf{x}-\mathbf{y}|^\gamma$  is the propagation loss function with predetermined constants  $\beta$  and  $\gamma$  (where  $\gamma > 2$ ), and  $h_{\mathbf{x},\mathbf{y}}$  is the fast fading term. Corresponding to common Rayleigh fading with power normalization, we assume that  $h_{\mathbf{x},\mathbf{y}}$  is independently exponentially distributed with unit mean. Let  $h(\cdot)$  be the probability density function (pdf) of  $h_{\mathbf{x},\mathbf{y}}$ .

We assume that each BS is connected to the core network by a separate high-capacity wired or wireless link that has no influence on our performance analysis. Also, we assume that the uplink of the system is operated in different spectrum, so that it has no influence on the downlink analysis and optimization. In addition, we focus on the interference limited scenario, such that noise is negligible.

Each arriving UE requests a connection with the BS that provides the maximum average received power, such that  $\mathcal{BS}(\mathbf{y}) = \arg \max_{\mathbf{x} \in \Phi_k, \forall k} P_k |\mathbf{x} - \mathbf{y}|^{-\gamma}$ , where  $\mathcal{BS}(\mathbf{y})$  denotes the location of the BS that the UE attempts to connect, and  $P_k |\mathbf{x} - \mathbf{y}|^{-\gamma}$  is the average received power from a tier- $k$  BS

<sup>1</sup>For example, it becomes quite challenging to track boundary crossings (i.e., handoffs) made by UEs, due to the irregularly shaped network topologies.

<sup>2</sup>We borrow the term RB from OFDM and LTE terminology. It represents a minimum radio resource chunk in the time-frequency space. The proposed method is equally applicable to other means of dividing the radio resource.

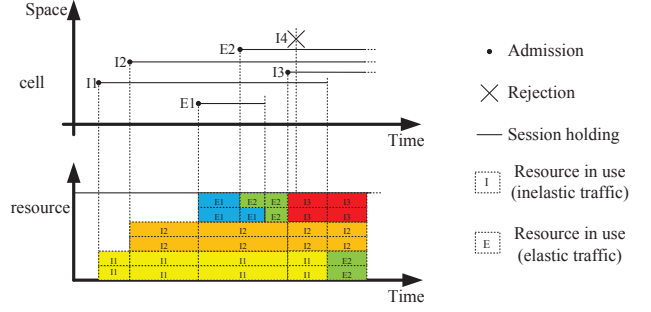


Fig. 2. An example of one cell in the HWN. The cell has 6 available RBs.  $w_I = 2$  and  $w_E = 2$ . Each admitted IUE's RB demand is satisfied. An EUE's demand may be partially satisfied. If there are 3 IUEs in the system, new IUE arrivals are rejected.

located at  $\mathbf{x}$ .<sup>3</sup> In this case, the resultant cell splitting on the spatial dimensions forms a generalized Dirichlet tessellation, or weighted Poisson Voronoi [31], and such cell splitting does not change over time. Let  $\text{Pr}_k = \frac{\lambda_k (P_k)^{\frac{2}{\gamma}}}{\sum_{j=1}^K \lambda_j (P_j)^{\frac{2}{\gamma}}}$  represent the proportion of area that tier- $k$  cells cover.

### C. Inelastic and Elastic Traffic

Each IUE demands  $w_I$  RBs for its inelastic traffic, and each EUE demands  $w_E$  RBs for its elastic traffic. We assume that the total number of available RBs at one BS is much larger than the number of RBs requested by one UE, i.e.,  $W_k \gg w_I$  and  $W_k \gg w_E$ . For admitted IUEs, their inelastic RB demands must be satisfied<sup>4</sup>, but the RB demands of EUEs are served in a best-effort fashion as follows. Suppose that there are  $M_k^I$  active IUEs and  $M_k^E$  active EUEs in some tier- $k$  cell. If  $w_I M_k^I + w_E M_k^E < W_k$ , the cell is *uncongested*. In this situation, each UE is fully allocated the number of RBs it requests. If  $w_I M_k^I + w_E M_k^E \geq W_k$  and  $w_I M_k^I < W_k$ , the cell is *mildly congested*. In this situation, each IUE is fully allocated  $w_I$  RBs, and all the EUEs equally share the remaining  $W_k - w_I M_k^I$  RBs not used by IUEs. In both the uncongested and the mildly congested cases, all new arrivals are admitted. However, in the mildly congested case, such admissions will cause RB reallocation from existing UEs to the new UEs. If  $w_I M_k^I = W_k$ , the cell is *heavily congested*. In this case, each IUE is allocated  $w_I$  RBs, and EUEs are temporarily not allocated any RBs. In addition, all new IUE arrivals are *rejected*. New EUE arrivals are admitted, but they are also temporarily not allocated any RBs. Fig. 2 shows an example with  $w_I = 2$  and  $w_E = 2$ .

We assume that  $W_k$  has a large value, and the numbers of IUEs and EUEs in a cell  $M_k^I$  and  $M_k^E$  are also large.

<sup>3</sup>This model can be easily extended to accommodate a flexible user association scheme [4], [15], [16], [21], [22], where tier- $k$  BSs are assigned an association bias value  $B_k$ , and a UE is associated with a BS that provides the largest biased received power, such that  $\mathcal{BS}(\mathbf{y}) = \arg \max_{\mathbf{x} \in \Phi_k, \forall k} B_k P_k |\mathbf{x} - \mathbf{y}|^{-\gamma}$ . In this case, we can simply replace  $P_k$  by  $P_k B_k$ , then the rest of the paper remains the same.

<sup>4</sup>In this work, we assume that the persistent scheduling mode [32], [33] is adopted, where BSs allocate a predefined number of RBs to IUEs with real-time sessions.

#### D. Coverage Probability and Data Rate

Let  $T$  denote the minimum required Signal-to-Interference Ratio (SIR). The coverage probability of a UE using an RB,  $P_{\text{cov}}$ , is defined as the probability that its SIR is no less than  $T$  [12]. Then, the data rate on the RB is  $\log(1+T)$  if the SIR is no less than  $T$ , and the data rate is 0 if the SIR is less than  $T$  (i.e., outage occurs). Thus, the overall data rate on the RB is  $\log(1+T)P_{\text{cov}}$ .  $\log$  is in base 2 throughout this paper.

#### IV. DOWNLINK THROUGHPUT DERIVATION

In this section, we present a joint spatial and temporal stochastic analysis framework to compute the average downlink throughput in an HWN with both IUEs and EUEs. This will serve as the optimization objective for radio resource allocation in the next section.

##### A. Interference Analysis and Coverage Probability

In this subsection, we derive the average coverage probability of UEs in the system. An approach similar to one studied in [22] is used. The main difference is that here not all RBs are active, and the inactive RBs do not generate interference. We present an outline of the derivation for completeness.

We consider a reference UE, termed the *typical UE*, communicating with its BS, termed the *typical BS*. The average coverage probability of UEs in the system is the same as the coverage probability of the typical UE. Furthermore, due to the statistical stationarity of UEs and BSs, we can re-define the coordinates so that the typical UE is located at  $\mathbf{0}$ .

We first derive the coverage probability given that the typical UE is at distance  $d$  from its associated tier- $k$  BS. In this case, the overall interference,  $I_k(d)$ , to the typical UE is the sum interference from all other tier- $k$  BSs that are actively using the same RB as the one being used by the typical UE-BS pair. Note that because we assume that different tiers of BSs are working on separated RBs, the interference from other tiers is zero.

Let  $\Phi'_k$  be the point process corresponding to all tier- $k$  BSs other than the typical BS using the same RB, given that the typical BS is located at distance  $d$  from the typical UE. By standard Palm theory,  $\Phi'_k$  is a PPP with intensity 0 in  $\mathcal{B}(\mathbf{0}, d)$  and intensity  $p_k \lambda_k$  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  $d$ , and  $p_k$  is the probability that the same RB is actively in use at a non-typical tier- $k$  BS. Note that  $p_k$  also depends non-trivially on other network parameters (e.g., BS intensities, UE arrival intensities, average session holding times, and resource allocation factors), which will be further studied in Section IV-B.

We have  $I_k(d) = \sum_{\mathbf{x} \in \Phi'_k} \frac{P_k h_{\mathbf{x}, \mathbf{0}}}{\beta |\mathbf{x}|^\gamma}$ , and its distribution is derived through its Laplace transform as follows:

$$\begin{aligned} \mathcal{L}_{I_k}(d, \omega) &= \mathbb{E} \left[ e^{-\omega I_k(d)} \right] = \mathbb{E} \left[ e^{-\sum_{\mathbf{x} \in \Phi'_k} \frac{\omega P_k h_{\mathbf{x}, \mathbf{0}}}{\beta |\mathbf{x}|^\gamma}} \right] \\ &= \exp \left( -p_k \lambda_k \int_{\mathbb{R}^2 \setminus \mathcal{B}(\mathbf{0}, d)} \left( 1 - \int_{\mathbb{R}^+} e^{-\frac{\omega P_k h}{\beta |\mathbf{x}|^\gamma}} \mathfrak{h}(h) dh \right) d\mathbf{x} \right) \end{aligned} \quad (1)$$

$$= \exp \left( -2\pi p_k \lambda_k \int_d^\infty \frac{\frac{\omega P_k r}{\beta}}{\frac{\omega P_k}{\beta} + r^\gamma} dr \right), \quad (2)$$

where (1) is obtained from the Laplace functional of PPP  $\Phi'_k$  [12]; and (2) is obtained by noting that the fading term  $h$  is exponentially distributed with unit mean, and by a transformation to polar coordinates.

Let  $P_{\text{cov},k}(d)$  denote the conditional coverage probability of the typical UE given  $k$  and  $d$ . We have

$$\begin{aligned} P_{\text{cov},k}(d) &= \mathbb{P} \left[ \frac{P_k h_{\mathbf{x}_B, \mathbf{0}}}{\beta d^\gamma} \geq T I_k(d) \right] = \mathcal{L}_{I_k}(d, \omega) \Big|_{\omega = \frac{T \beta d^\gamma}{P_k}}, \\ &= \exp \left( -\pi p_k \lambda_k T^{\frac{2}{\gamma}} d^2 \int_{(\frac{1}{T})^{\frac{2}{\gamma}}}^\infty \frac{1}{1+t^{\frac{\gamma}{2}}} dt \right), \end{aligned} \quad (3)$$

where  $\mathbf{x}_B$  is the coordinate of the typical BS and  $|\mathbf{x}_B| = d$ .

Furthermore, the probability density function (pdf) of the distance between the typical UE and the typical BS is known to be [15]

$$g_k(d) = \frac{2\pi \lambda_k}{\text{Pr}_k} d \exp \left( -\pi d^2 \frac{\lambda_k}{\text{Pr}_k} \right). \quad (4)$$

Based on (3) and (4), we can derive the coverage probability  $P_{\text{cov},k}$  of the typical UE deconditioning on  $d$ :

$$P_{\text{cov},k} = \int_0^\infty g_k(d) P_{\text{cov},k}(d) dd = \frac{1}{1 + C \text{Pr}_k p_k}, \quad (5)$$

where we define  $C \triangleq T^{\frac{2}{\gamma}} \int_{(\frac{1}{T})^{\frac{2}{\gamma}}}^\infty \frac{1}{1+t^{\frac{\gamma}{2}}} dt$ .

##### B. Normalized Downlink Throughput

Let  $N_k$  denote the number of active RBs in a tier- $k$  cell. Its mean is denoted by  $\mathbb{E}(N_k)$ . Then the average throughput of a tier- $k$  BS is<sup>5</sup>

$$\bar{R}_k = \mathbb{E}(N_k) \log(1+T) P_{\text{cov},k}, \quad (6)$$

and the average downlink sum throughput per unit area (normalized throughput) is computed as

$$\bar{R} = \sum_{k=1}^K \lambda_k \bar{R}_k = \sum_{k=1}^K \lambda_k \mathbb{E}(N_k) \frac{\log(1+T)}{1 + C \text{Pr}_k p_k}. \quad (7)$$

From (7), we observe that  $p_k$  and  $\mathbb{E}(N_k)$  are two key values to derive  $\bar{R}$ . Because  $\mathbb{E}(N_k) = p_k W_k$ , we focus on the derivation of  $p_k$ , which is formally stated as the following theorem.

**Theorem 1.** *The probability that an RB in a tier- $k$  cell is actively in use is*

$$\begin{aligned} p_k &= \frac{\text{Pr}_k \mu}{\lambda_k \nu W_k} - \frac{3.5^{3.5}}{\Gamma(3.5)} \\ &\quad \int_1^\infty (y-1) \left( \frac{\lambda_k \nu W_k}{\text{Pr}_k \mu} \right)^{3.5} y^{2.5} e^{-3.5 \frac{\lambda_k \nu W_k}{\text{Pr}_k \mu} y} dy, \end{aligned} \quad (8)$$

where  $\mu \triangleq \mu_I W_I + \mu_E W_E$ .

*Proof.* Let  $S_k$  denote the cell size of the tier- $k$  cell. **In the first step**, we study the probability that an RB in the tier- $k$  cell is actively in use given  $S_k = s$ , which is denoted by  $p_k(s)$ .

<sup>5</sup>We have used an approximation that the coverage event and  $N_k$  are independent for mathematical tractability. In reality, their dependency is negligible, which has been verified by simulations in [16], [19]. In Section VI, we also show via simulation that the throughput performance based on this approximation is close to that without it.

Let  $N_k^I = w_I M_k^I$ ,  $N_k^E = w_E M_k^E$ , and  $N_k^{IE} = N_k^I + N_k^E$ . Note that  $N_k^{IE}$  and  $N_k$  are with different definitions. We have  $N_k \leq W_k$ , but  $N_k^{IE}$  could be arbitrarily large. We define  $\rho_I \triangleq \frac{\mu_I s}{\nu}$  and  $\rho_E \triangleq \frac{\mu_E s}{\nu}$ . Let  $M_k^{\max} \triangleq \frac{W_k}{w_I}$  denote the maximum possible number of admitted IUEs in the cell. IUE arrivals are accepted if  $M_k^I < M_k^{\max}$ , and are rejected if  $M_k^I = M_k^{\max}$ . Thus,  $M_k^I$  can be equivalently regarded as the number of units in an  $M/G/M_k^{\max}/M_k^{\max}$  queue, with arrival rate  $\mu_I s$ , and average service time  $\frac{1}{\nu}$ . Hence, we have

$$\mathbb{P}_1(m_1) \triangleq \mathbb{P}(M_k^I = m_1) = \frac{\frac{\rho_I^{m_1}}{m_1!}}{\sum_{i=0}^{M_k^{\max}} \frac{\rho_I^i}{i!}}, \quad 0 \leq m_1 \leq M_k^{\max}. \quad (9)$$

EUE arrivals are always accepted. Thus,  $M_k^E$  can be equivalently regarded as the number of units in an  $M/G/\infty$  queue, with arrival rate  $\mu_E s$ , and average service time  $\frac{1}{\nu}$ . Hence,

$$\mathbb{P}_2(m_2) \triangleq \mathbb{P}(M_k^E = m_2) = \frac{\rho_E^{m_2}}{m_2!} e^{-\rho_E}, \quad m_2 \geq 0. \quad (10)$$

Since  $M_k^I$  and  $M_k^E$  do not depend on each other, their joint distribution is

$$\mathbb{P}_A(m_1, m_2) \triangleq \mathbb{P}(M_k^I = m_1, M_k^E = m_2) = \mathbb{P}_1(m_1) \mathbb{P}_2(m_2). \quad (11)$$

If  $N_k^{IE} = w_I M_k^I + w_E M_k^E < W_k$ , only a fraction  $\frac{N_k^{IE}}{W_k}$  of the  $W_k$  RBs are active. Otherwise, if  $N_k^{IE} \geq W_k$ , all RBs are active. Thus, we have

$$p_k(s) = \sum_{m_1=0}^{M_k^{\max}} \sum_{m_2=0}^{\infty} \min\left(1, \frac{w_I m_1 + w_E m_2}{W_k}\right) \mathbb{P}_A(m_1, m_2). \quad (12)$$

**Case 1,  $\rho_I > M_k^{\max}$ .**

Because  $\rho_I$  and  $M_k^{\max}$  are large values,  $\mathbb{P}_1(m_1)$  is equivalent to a geometric distribution [34] as follows

$$\mathbb{P}_1(m_1) \simeq \left(1 - \frac{M_k^{\max}}{\rho_I}\right) \left(\frac{M_k^{\max}}{\rho_I}\right)^{M_k^{\max} - m_1}. \quad (13)$$

Let

$$p'_k(s) = \sum_{m_1=0}^{M_k^{\max}} \frac{w_I m_1}{W_k} \mathbb{P}_1(m_1) \quad (14)$$

be the probability that an RB is actively used by some IUE. Obviously, we have  $p'_k(s) \leq p_k(s) \leq 1$ .

Plugging (13) into (14), and after some manipulations, we have  $p'_k(s) \simeq 1$ , which implies  $p_k(s) \simeq 1$ .

**Case 2,  $\rho_I < M_k^{\max}$ .**

In this case, the distribution of  $M_k^I$  (resp.  $M_k^E$ ) is equivalent to the normal distribution [34], with mean  $\rho_I$  and standard deviation  $\sqrt{\rho_I}$  (resp. with mean  $\rho_E$  and standard deviation  $\sqrt{\rho_E}$ ). Therefore,  $N_k^{IE} = w_I M_k^I + w_E M_k^E$  is normal distributed with mean  $\rho \triangleq \frac{\mu_I w_I + \mu_E w_E}{\nu} s = \frac{\mu s}{\nu}$  and standard deviation  $\sigma = \sqrt{w_I^2 \rho_I + w_E^2 \rho_E}$ . Note that  $\sigma \ll \rho$  as  $\rho_I$  and  $\rho_E$  are large values.

In this case, (12) becomes

$$p_k(s) = \sum_{w=0}^{W_k} \frac{w}{W_k} \mathbb{P}(N_k^{IE} = w) + \sum_{w=W_k+1}^{\infty} \mathbb{P}(N_k^{IE} = w) \quad (15)$$

$$\simeq \int_{-\infty}^{W_k} \frac{x}{W_k} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\rho)^2}{2\sigma^2}} dx + \int_{W_k}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\rho)^2}{2\sigma^2}} dx. \quad (16)$$

**Case 2.1,  $\rho \geq W_k$ .**

$$\begin{aligned} \text{In this case, we have } \int_{-\infty}^{W_k} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\rho)^2}{2\sigma^2}} dx &\simeq \\ \int_{-\infty}^{W_k} \frac{x}{W_k} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\rho)^2}{2\sigma^2}} dx, \text{ leading to } p_k(s) &\simeq \\ \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\rho)^2}{2\sigma^2}} dx = 1. \end{aligned}$$

**Case 2.2,  $\rho < W_k$ .**

$$\begin{aligned} \text{In this case, we have } \int_{W_k}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\rho)^2}{2\sigma^2}} dx &\simeq \\ \int_{W_k}^{\infty} \frac{x}{W_k} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\rho)^2}{2\sigma^2}} dx, \text{ leading to } p_k(s) &\simeq \\ \int_{-\infty}^{\infty} \frac{x}{W_k} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\rho)^2}{2\sigma^2}} dx = \frac{\rho}{W_k}. \end{aligned}$$

**Case 3,  $\rho_I = M_k^{\max}$ .**

In this case, the distribution of  $M_k^I$  is equivalent to a truncated normal distribution. Because  $\rho_I$  is a large value, we have  $p'_k(s) \simeq \int_{-\infty}^{\rho_I} \frac{x}{\rho_I} \frac{2}{\sqrt{2\pi}\rho_I} e^{-\frac{(x-\rho_I)^2}{2\rho_I}} dx \simeq 1$ , which implies  $p_k(s) \simeq 1$ .

As a consequence, by combining Cases 1-3, we have

$$p_k(s) = \min\left(\frac{\mu s}{\nu W_k}, 1\right). \quad (17)$$

**In the second step**, the cell size of a tier- $k$  cell,  $S_k$ , is known to have the following pdf [16], [35]:

$$f_{S_k}(s) = \frac{3.5^{3.5}}{\Gamma(3.5)} \frac{\lambda_k}{\text{Pr}_k} \left(\frac{\lambda_k}{\text{Pr}_k} s\right)^{2.5} e^{-3.5s \frac{\lambda_k}{\text{Pr}_k}}. \quad (18)$$

Consequently, we can derive  $p_k$  by deconditioning on  $s$  as follows

$$\begin{aligned} p_k &= \int_0^{\infty} p_k(s) f_{S_k}(s) ds \\ &= \int_0^{\infty} \frac{\mu s}{\nu W_k} \frac{3.5^{3.5}}{\Gamma(3.5)} \frac{\lambda_k}{\text{Pr}_k} \left(\frac{\lambda_k}{\text{Pr}_k} s\right)^{2.5} e^{-3.5s \frac{\lambda_k}{\text{Pr}_k}} ds \\ &\quad - \int_{\frac{\nu W_k}{\lambda_k}}^{\infty} \left(\frac{\mu s}{\nu W_k} - 1\right) \frac{3.5^{3.5}}{\Gamma(3.5)} \frac{\lambda_k}{\text{Pr}_k} \left(\frac{\lambda_k}{\text{Pr}_k} s\right)^{2.5} e^{-3.5s \frac{\lambda_k}{\text{Pr}_k}} ds \\ &= \frac{\text{Pr}_k \mu}{\lambda_k \nu W_k} - \frac{3.5^{3.5}}{\Gamma(3.5)} \\ &\quad \int_1^{\infty} (y-1) \left(\frac{\lambda_k \nu W_k}{\text{Pr}_k \mu}\right)^{3.5} y^{2.5} e^{-3.5 \frac{\lambda_k \nu W_k}{\text{Pr}_k \mu} y} dy, \quad (19) \end{aligned}$$

which completes the proof.  $\square$

For presentation convenience, we define  $\alpha_k \triangleq \frac{\lambda_k \nu W_k}{\text{Pr}_k \mu}$  and  $f(\alpha_k) \triangleq \frac{3.5^{3.5}}{\Gamma(3.5)} \int_1^{\infty} (y-1) \alpha_k^{4.5} y^{2.5} e^{-3.5\alpha_k y} dy$ , such that  $p_k$  can be rewritten as  $p_k = \frac{1}{\alpha_k} - \frac{1}{\alpha_k} f(\alpha_k)$ . Correspondingly,  $\mathbb{E}(N_k) = W_k p_k = \frac{\text{Pr}_k \mu}{\lambda_k \nu} - \frac{\text{Pr}_k \mu}{\lambda_k \nu} f(\alpha_k)$ .

Substituting  $p_k$  and  $\mathbb{E}(N_k)$  into (7), we can compute the normalized throughput of the HWN. Note that this normalized

throughput is a function of  $\mu$ , which remains the same as long as  $\mu \triangleq \mu_I w_I + \mu_E w_E$  does not change (even if the portion of inelastic traffic demands  $\frac{\mu_I w_I}{\mu_I w_I + \mu_E w_E}$  changes).

## V. RADIO RESOURCE ALLOCATION

In this section, we present an efficient approximate solution to optimal radio resource allocation in order to maximize the average downlink throughput derived in the previous section.

### A. Optimization Problem Formulation

We aim to maximize  $\bar{R}$  with respect to  $\boldsymbol{\eta}$ , with the constraint  $\sum_{k=1}^K \eta_k = 1$ . We additionally consider a *basic tier-level RB allocation requirement*, where the average number of RB demands per tier- $k$  cell,  $\frac{\text{Pr}_k \mu}{\lambda_k \nu}$ , is no greater than  $W_k$ , such that the  $k$ th tier should at least support its average traffic demands to avoid too many congestions in this tier. By doing so, we are able to avoid unfair resource allocation when some tiers are allocated too few RBs to support their average traffic demands while some other tiers are allocated much more RBs than their average traffic demands.

The optimization problem is formally stated as problem **P**:

$$\underset{\boldsymbol{\eta}}{\text{maximize}} \quad \bar{R}(\boldsymbol{\eta}) \quad (20)$$

$$\text{subject to} \quad \sum_{k=1}^K \eta_k = 1, \quad (21)$$

$$\frac{\text{Pr}_k \mu}{\lambda_k \nu} \leq W_k, \forall k. \quad (22)$$

Note that we only need to consider the scenario where  $\sum_{k=1}^K \frac{\text{Pr}_k \mu}{\lambda_k \nu} \leq W$ , i.e., the overall system has enough RBs to support its average traffic demands, such that the optimization problem is feasible.

Let  $\boldsymbol{\alpha} \triangleq (\alpha_1, \alpha_2, \dots, \alpha_K)$ , where  $\alpha_k$  is as defined in Section IV-B. We observe that there is a one-to-one mapping between  $\boldsymbol{\eta}$  and  $\boldsymbol{\alpha}$ . In this case, we can instead study the optimization over  $\boldsymbol{\alpha}$  for analytical convenience. As a consequence, the objective function (20) can be re-written as

$$\bar{R}(\boldsymbol{\alpha}) = \sum_{k=1}^K \frac{\text{Pr}_k (1 - f(\alpha_k))}{1 + \text{Pr}_k C \frac{1}{\alpha_k} (1 - f(\alpha_k))}, \quad (23)$$

where  $\log(1 + T)$ ,  $\mu$ , and  $\nu$  are omitted as they are common factors of all the summation terms. Furthermore, constraints (21) and (22) can be re-written as

$$\sum_{k=1}^K \frac{\alpha_k \text{Pr}_k \mu}{\lambda_k \nu} = W, \quad (24)$$

$$\alpha_k \geq 1, \forall k. \quad (25)$$

In either (20) or (23), the objective function is in non-closed form and is non-convex. Hence Problem **P** cannot be solved using a standard method. In the rest of this section, we propose our approximate solution to Problem **P**, and quantify its performance bound.

### B. Bounds of the Function $f(\cdot)$

In order to remove the non-closed-form expression in the objective function, we first derive closed-form lower and upper bounds of  $f(\cdot)$ . Let  $f_{lb}(\alpha) \triangleq \frac{3.5^4}{3!} e^{-3.5\alpha} \left[ \frac{\alpha^2}{3.5^2} + \frac{4\alpha}{3.5^3} + \frac{6}{3.5^4} \right]$  and  $f_{ub}(\alpha) \triangleq \frac{3.5^5}{4!} e^{-3.5\alpha} \left[ \frac{\alpha^3}{3.5^2} + \frac{6\alpha^2}{3.5^3} + \frac{18\alpha}{3.5^4} + \frac{24}{3.5^5} \right]$ . We have the following theorem:

**Theorem 2.**  $f_{lb}(\alpha) < f(\alpha) < f_{ub}(\alpha)$ .

*Proof.* Let  $g(y) = \begin{cases} 0 & \text{if } y \leq 1 \\ 1 - \frac{1}{y} & \text{if } y > 1 \end{cases}$ , and let  $\text{pdf}_Y(y)$  be the pdf of a gamma distributed random variable  $Y$ ,  $Y \sim \Gamma(4.5, \frac{1}{3.5\alpha})$ . Then, according to the definition of  $f(\cdot)$ , we have

$$\begin{aligned} f(\alpha) &= \frac{3.5^{3.5}}{\Gamma(3.5)} \int_1^\infty \left(1 - \frac{1}{y}\right) \alpha^{4.5} y^{3.5} e^{-3.5\alpha y} dy \\ &= \int_0^\infty g(y) \text{pdf}_Y(y) dy = \mathbb{E}(g(Y)). \end{aligned} \quad (26)$$

Let  $Y_u \sim \Gamma(5, \frac{1}{3.5\alpha})$  and  $Y_l \sim \Gamma(4, \frac{1}{3.5\alpha})$  be two gamma distributed random variables. According to the properties of gamma distributions,  $\text{ccdf}_{Y_u}(y) > \text{ccdf}_Y(y) > \text{ccdf}_{Y_l}(y)$  on  $[0, \infty)$ , where  $\text{ccdf}_{Y_u}(y)$ ,  $\text{ccdf}_Y(y)$ , and  $\text{ccdf}_{Y_l}(y)$  are the complementary cumulative distribution function (ccdf) of  $Y_u$ ,  $Y$ , and  $Y_l$  respectively.

Note that  $g(\cdot)$  is a non decreasing function. We define a non-negative function  $g_0(x) \triangleq \begin{cases} 0 & \text{if } x \leq 1 \\ \frac{1}{x^2} & \text{if } x > 1 \end{cases}$ , such that  $g(y) = \int_0^y g_0(x) dx$ . Then we can show that

$$\begin{aligned} \mathbb{E}(g(Y)) &= \int_0^\infty g(y) \text{pdf}_Y(y) dy = \int_0^\infty g_0(x) \text{ccdf}_Y(x) dx \\ &< \int_0^\infty g_0(x) \text{ccdf}_{Y_u}(x) dx = \mathbb{E}(g(Y_u)). \end{aligned}$$

where  $\mathbb{E}(g(Y_u))$  can be expressed in closed form as follows

$$\begin{aligned} \mathbb{E}(g(Y_u)) &= \frac{3.5^5}{4!} \int_1^\infty \left(1 - \frac{1}{y}\right) \alpha^5 y^4 e^{-3.5\alpha y} dy \\ &= \frac{3.5^5}{4!} e^{-3.5\alpha} \left[ \frac{\alpha^3}{3.5^2} + \frac{6\alpha^2}{3.5^3} + \frac{18\alpha}{3.5^4} + \frac{24}{3.5^5} \right], \end{aligned} \quad (27)$$

Similarly, we can prove  $\mathbb{E}(g(Y)) > \mathbb{E}(g(Y_l))$ , and  $\mathbb{E}(g(Y_l))$  can be expressed as

$$\mathbb{E}(g(Y_l)) = \frac{3.5^4}{3!} e^{-3.5\alpha} \left[ \frac{\alpha^2}{3.5^2} + \frac{4\alpha}{3.5^3} + \frac{6}{3.5^4} \right]. \quad (28)$$

Finally, we have  $f_{lb}(\alpha) < f(\alpha) < f_{ub}(\alpha)$  due to (27), (28), and  $\mathbb{E}(g(Y_l)) < \mathbb{E}(g(Y)) < \mathbb{E}(g(Y_u))$ .  $\square$

### C. Approximate Convex Optimization

Since the normalized throughput  $\bar{R}$  decreases if  $f$  increases, we have the following upper and lower bounds that envelop the original objective function  $\bar{R}(\boldsymbol{\alpha})$ :

$$\bar{R}_{ub}(\boldsymbol{\alpha}) = \sum_{k=1}^K \frac{\text{Pr}_k (1 - f_{lb}(\alpha_k))}{1 + \text{Pr}_k C \frac{1}{\alpha_k} (1 - f_{lb}(\alpha_k))}, \quad (29)$$

$$\bar{R}_{lb}(\boldsymbol{\alpha}) = \sum_{k=1}^K \frac{\text{Pr}_k (1 - f_{ub}(\alpha_k))}{1 + \text{Pr}_k C \frac{1}{\alpha_k} (1 - f_{ub}(\alpha_k))}. \quad (30)$$

More importantly, the following theorem shows that  $\bar{R}_{ub}(\alpha)$  and  $\bar{R}_{lb}(\alpha)$  have the desirable property of being concave:

**Theorem 3.**  $\bar{R}_{ub}(\alpha)$  and  $\bar{R}_{lb}(\alpha)$  are concave functions in  $\alpha$ .

*Proof.* For presentation convenience, we use the letter  $b$  to represent 3.5. Let  $A$  and  $C$  be arbitrary positive numbers.

(a) Let  $v_1(x) = \frac{A(1-f_{lb}(x))}{1+AC\frac{1}{x}(1-f_{lb}(x))}$ . Its second derivative is

$$v_1''(x) = -\frac{6Ae^{bx}(ACB_1(x) + 6b^4x^5e^{bx})}{(AC(-b^2x^2 - 4bx + 6e^{bx} - 6) + 6xe^{bx})^3}, \quad (31)$$

where  $B_1(x) = [b^6x^6 + 8b^5x^5 + 36b^4x^4 + 96b^3x^3 + 144b^2x^2 + 6e^{bx}(b^4x^4 - 4b^3x^3 - 12b^2x^2 - 24bx - 24) + 144bx + 72e^{2bx} + 72]$ . It can be shown that  $-b^2x^2 - 4bx + 6e^{bx} - 6$  is positive and  $B_1(x)$  is positive on  $[1, \infty)$ . Thus,  $v_1''(x)$  is negative on  $[1, \infty)$ .

(b) Let  $v_2(x) = \frac{A(1-f_{ub}(x))}{1+AC\frac{1}{x}(1-f_{ub}(x))}$ . Its second derivative is

$$v_2''(x) = \frac{24Ae^{bx}(ACB_2(x) + 24b^5x^6e^{bx})}{(AC(-b^3x^3 - 6b^2x^2 - 18bx + 24e^{bx} - 24) + 24xe^{bx})^3}, \quad (32)$$

where  $B_2(x) = [b^8x^8 + 10b^7x^7 + 62b^6x^6 + 264b^5x^5 + 768b^4x^4 + 1536b^3x^3 + 2304b^2x^2 + 24e^{bx}(b^5x^5 - 4b^4x^4 - 16b^3x^3 - 48b^2x^2 - 96bx - 96) + 2304bx + 1152e^{2bx} + 1152]$ . It can be shown that  $-b^3x^3 - 6b^2x^2 - 18bx + 24e^{bx} - 24$  is positive and  $B_2(x)$  is positive on  $[1, \infty)$ . Thus,  $v_2''(x)$  is negative on  $[1, \infty)$ .  $\square$

Based on Theorems 2 and 3, instead of solving the original problem  $\mathbf{P}$ , we propose to solve two approximated problems, by replacing the original objective function by its upper and lower bounds:

$$\begin{aligned} \mathbf{P}_U & \text{ maximize } \bar{R}_{ub}(\alpha) \quad \text{subject to (24) and (25),} \\ \mathbf{P}_L & \text{ maximize } \bar{R}_{lb}(\alpha) \quad \text{subject to (24) and (25).} \end{aligned}$$

Since  $\bar{R}_{ub}(\alpha)$  and  $\bar{R}_{lb}(\alpha)$  are concave, and (24) and (25) are linear constraints,  $\mathbf{P}_U$  and  $\mathbf{P}_L$  are convex programming problems and can be efficiently solved by standard convex optimization methods.

#### D. Quantifying the Performance Gap

In this subsection, we aim to quantify the performance gap between the approximate solutions (of  $\mathbf{P}_U$  and  $\mathbf{P}_L$ ) and the original optimal solution. Let  $\alpha^*$ ,  $\hat{\alpha}^*$ , and  $\tilde{\alpha}^*$  denote the optimal solutions to problems  $\mathbf{P}$ ,  $\mathbf{P}_U$ , and  $\mathbf{P}_L$  respectively. We are interested to quantify  $\delta_1 \triangleq \frac{|\bar{R}_{ub}(\hat{\alpha}^*) - \bar{R}(\alpha^*)|}{\bar{R}(\alpha^*)}$  and  $\delta_2 \triangleq \frac{|\bar{R}_{lb}(\tilde{\alpha}^*) - \bar{R}(\alpha^*)|}{\bar{R}(\alpha^*)}$  as the relative performance gap corresponding to  $\hat{\alpha}^*$  and  $\tilde{\alpha}^*$  respectively.

Let  $\delta \triangleq \frac{\bar{R}_{ub}(\hat{\alpha}^*) - \bar{R}_{lb}(\tilde{\alpha}^*)}{\bar{R}(\alpha^*)}$ . We have  $\delta \geq \max(\delta_1, \delta_2)$ . Then, we can bound the worst case performance gap  $\delta$  in the following theorem:

**Theorem 4.**  $\delta \leq \max_{\alpha \in [1, \infty)} \frac{f_{ub}(\alpha) - f_{lb}(\alpha)}{1 - f_{ub}(\alpha)}$ .

*Proof.* We have

$$\delta = \frac{\bar{R}_{ub}(\hat{\alpha}^*) - \bar{R}_{lb}(\tilde{\alpha}^*)}{\bar{R}(\alpha^*)} \leq \frac{\bar{R}_{ub}(\hat{\alpha}^*) - \bar{R}_{lb}(\hat{\alpha}^*)}{\bar{R}_{lb}(\hat{\alpha}^*)} \quad (33)$$

$$\begin{aligned} &= \frac{\sum_{k=1}^K \left( \frac{\text{Pr}_k(1-f_{lb}(\hat{\alpha}_k^*))}{1+\text{Pr}_k C \frac{1}{\hat{\alpha}_k^*}(1-f_{lb}(\hat{\alpha}_k^*))} - \frac{\text{Pr}_k(1-f_{ub}(\hat{\alpha}_k^*))}{1+\text{Pr}_k C \frac{1}{\hat{\alpha}_k^*}(1-f_{ub}(\hat{\alpha}_k^*))} \right)}{\sum_{k=1}^K \frac{\text{Pr}_k(1-f_{ub}(\hat{\alpha}_k^*))}{1+\text{Pr}_k C \frac{1}{\hat{\alpha}_k^*}(1-f_{ub}(\hat{\alpha}_k^*))}} \\ &\leq \max_k \left( \frac{\left( \frac{\text{Pr}_k(1-f_{lb}(\hat{\alpha}_k^*))}{1+\text{Pr}_k C \frac{1}{\hat{\alpha}_k^*}(1-f_{lb}(\hat{\alpha}_k^*))} - \frac{\text{Pr}_k(1-f_{ub}(\hat{\alpha}_k^*))}{1+\text{Pr}_k C \frac{1}{\hat{\alpha}_k^*}(1-f_{ub}(\hat{\alpha}_k^*))} \right)}{\frac{\text{Pr}_k(1-f_{ub}(\hat{\alpha}_k^*))}{1+\text{Pr}_k C \frac{1}{\hat{\alpha}_k^*}(1-f_{ub}(\hat{\alpha}_k^*))}} \right) \quad (34) \end{aligned}$$

$$\begin{aligned} &= \max_k \left( \frac{\left( \frac{f_{ub}(\hat{\alpha}_k^*) - f_{lb}(\hat{\alpha}_k^*)}{(1-f_{ub}(\hat{\alpha}_k^*))(1-f_{lb}(\hat{\alpha}_k^*))} \right)}{\frac{1}{1-f_{lb}(\hat{\alpha}_k^*)} + \frac{\text{Pr}_k C}{\hat{\alpha}_k^*}} \right) \\ &< \max_{\alpha \in [1, \infty)} \frac{f_{ub}(\alpha) - f_{lb}(\alpha)}{1 - f_{ub}(\alpha)}, \quad (35) \end{aligned}$$

where (33) is because  $\bar{R}_{lb}(\hat{\alpha}^*) \geq \bar{R}_{lb}(\tilde{\alpha}^*)$  and  $\bar{R}(\alpha^*) \geq \bar{R}_{lb}(\hat{\alpha}^*)$ , (34) is because  $\frac{\sum_{k=1}^K x_k}{\sum_{k=1}^K y_k} \leq \max_k \frac{x_k}{y_k}$  when  $y_k > 0, \forall k$ , and (35) is because  $\frac{\text{Pr}_k C}{\hat{\alpha}_k^*}$  is greater than 0.  $\square$

Note that (35) is an unconstrained optimization over a one-dimensional function. Simple numerical search finds that its maximum value is 12.6%. As a consequence, we have shown that the proposed solutions  $\hat{\alpha}^*$  and  $\tilde{\alpha}^*$  are near-optimal with a worst case optimization performance gap of 12.6%.

## VI. NUMERICAL STUDY

In this section, we present numerical studies on the performance of the proposed resource allocation approach. We label the solutions to problems  $\mathbf{P}_U$  and  $\mathbf{P}_L$  as  $\hat{\eta}^*$  and  $\tilde{\eta}^*$ , respectively. For reference, we also obtain  $\eta^*$  as the optimal solution to the original problem  $\mathbf{P}$  by numerical exhaustive search. The ‘‘analytical optimal’’, ‘‘analytical upper bound’’, and ‘‘analytical lower bound’’ labels in Fig. 3 indicate the throughput computed from  $\bar{R}(\eta^*)$ ,  $\bar{R}_{ub}(\hat{\eta}^*)$ , and  $\bar{R}_{lb}(\tilde{\eta}^*)$  respectively.

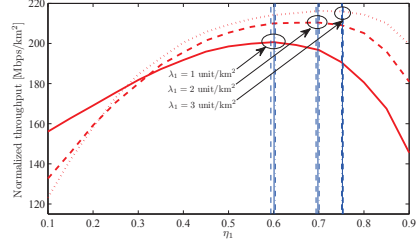
In addition to the analytical throughput performance, we also present the simulated throughput under  $\eta^*$ ,  $\hat{\eta}^*$ , and  $\tilde{\eta}^*$ . In each round of simulation, multiple tiers of BSs are generated on a 10 km  $\times$  10 km square, according to their intensities. UE session arrivals are generated in 10 km  $\times$  10 km  $\times$  10 minute spatial-temporal space. UEs in the central 5 km  $\times$  5 km square are selected for performance evaluation in order to remove the edge effect. Each data point is averaged over 100 rounds of simulation. In all figures, each RB covers 10 ms time duration and 1 MHz frequency bandwidth, and we set  $T = 1$ ,  $\gamma = 4$ ,  $W = 20000$  unit,  $w_I = 10$  unit,  $w_E = 10$  unit, and  $\frac{1}{\nu} = 1$  minute. All other network parameters used in the figures are listed in Table I.

First, we present the simulated throughput performance under different resource allocation factors in a two-tier scenario (i.e.,  $K = 2$ ) in Fig. 3(a). The simulated results are plotted for  $\lambda_1 = 1, 2$ , and 3 unit/km<sup>2</sup> respectively. The solid, dashed, and dashed dotted vertical lines correspond to the solutions  $\eta_1^*$ ,  $\hat{\eta}_1^*$ , and  $\tilde{\eta}_1^*$  respectively. The results validate the optimality of the solution  $\eta_1^*$ . In addition, the optimal solution is close to our proposed solutions, illustrating the effectiveness of our analysis. The figure also suggests that when the intensity of

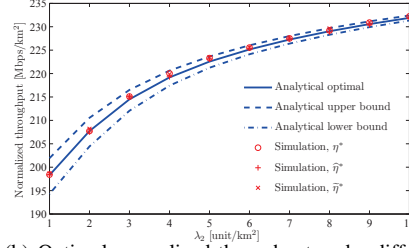


TABLE I  
LIST OF PARAMETERS USED IN NUMERICAL STUDIES.

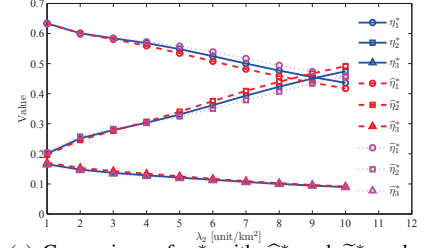
Figure	Tiers $K$	Power (dBm)	$\lambda$ (unit/km <sup>2</sup> )	$\mu_I$ and $\mu_E$ (unit/km <sup>2</sup> ·min)
Fig. 3(a)	2	$(P_1, P_2) = (50, 40)$	$\lambda_2 = 5, \lambda_1$ various	$\mu_I = 500, \mu_E = 2000$
Fig. 3(b)-3(c)	3	$(P_1, P_2, P_3) = (50, 40, 30)$	$(\lambda_1, \lambda_3) = (1, 10), \lambda_2$ various	$\mu_I = 500, \mu_E = 2000$
Fig. 3(d)-3(e)	3	$(P_1, P_3) = (53, 30), P_2$ various	$(\lambda_1, \lambda_2, \lambda_3) = (1, 4, 10)$	$\mu_I = 500, \mu_E = 2000$
Fig. 3(f)-3(g)	3	$(P_1, P_2, P_3) = (50, 40, 30)$	$(\lambda_1, \lambda_2, \lambda_3) = (1, 4, 10)$	Various with $\mu_E = 4\mu_I$
Fig. 3(h)	3	$(P_1, P_2, P_3) = (50, 40, 30)$	$(\lambda_1, \lambda_2, \lambda_3) = (1, 4, 10)$	Various



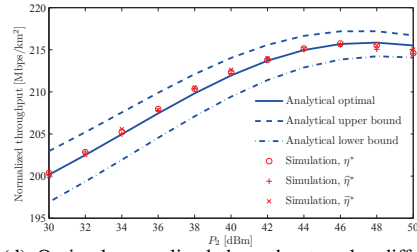
(a) Normalized simulated throughput under different  $\eta_1$ .



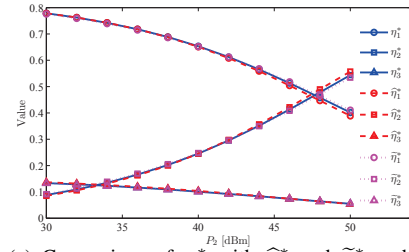
(b) Optimal normalized throughput under different  $\lambda_2$ .



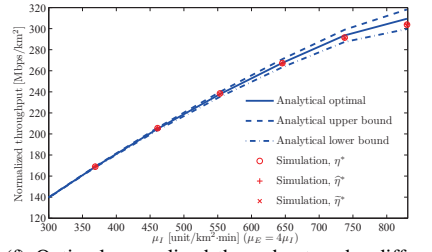
(c) Comparison of  $\eta^*$  with  $\hat{\eta}^*$  and  $\tilde{\eta}^*$  under different  $\lambda_2$ .



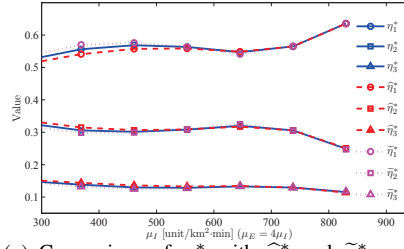
(d) Optimal normalized throughput under different  $P_2$ .



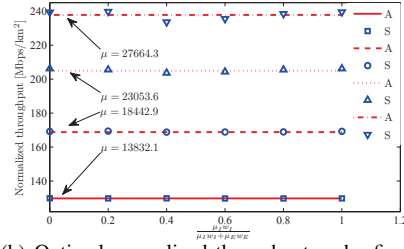
(e) Comparison of  $\eta^*$  with  $\hat{\eta}^*$  and  $\tilde{\eta}^*$  under different  $P_2$ .



(f) Optimal normalized throughput under different UE arrival intensities.



(g) Comparison of  $\eta^*$  with  $\hat{\eta}^*$  and  $\tilde{\eta}^*$  under different UE arrival intensities.



(h) Optimal normalized throughput under fixed  $\mu$ . "A" and "S" indicate analytical and simulation results respectively.

Fig. 3. Numerical studies.

tier-1 BSs increases, more RBs should be allocated to this tier, and the optimal throughput performance also increases accordingly.

In Fig. 3(b) and Fig. 3(c), we study the optimal throughput performance under different  $\lambda_2$  values, and in Fig. 3(d) and Fig. 3(e), we study the optimal throughput performance under different  $P_2$  values. The figures validate that the bounding approximations provide nearly optimal throughput for a wide range of BS intensities and BS powers. The results show that when the intensity or the power of tier-2 BSs increases, more portion of RBs should be allocated to this tier. In addition, Fig. 3(b) suggests that the optimal throughput performance keeps increasing, but the increase rate slows down, when the intensity of tier-2 BSs increases. Fig. 3(d) suggests that increasing the BS power in a single tier may not always bring throughput gain. Two opposing effects arise in this situation: (a) increasing the BS power of one tier could expand the cell

ranges in this tier, and thus more UEs are likely to locate at cell edges in this tier, decreasing the throughput performance; and (b) the cell ranges in the other tiers are reduced, and thus less UEs are likely to locate at cell edges in the other tiers, improving the throughput performance. The relative effect between (a) and (b) varies, leading to either throughput loss or gain.

In Fig. 3(f) and Fig. 3(g), we study the optimal throughput performance under different UE arrival intensities. We set  $\mu_E = 4\mu_I$  for these figures. The figures show (a) the simulated optimal results agree with analytical results; (b) the bounding approximations provide nearly optimal throughput for a wide range of UE arrival intensities; and (c) the simulated throughput of our proposed solutions almost overlaps that of optimal solutions, illustrating that the proposed solutions often perform much closer to optimality than the analytical bounds derived in Section V-D.



In Fig. 3(h), we study the optimal throughput performance under different  $\frac{\mu_I w_I}{\mu_I w_I + \mu_E w_E}$  (which is equivalent to  $\frac{\mu_I}{\mu_I + \mu_E}$  as we set  $w_I = w_E = 10$  unit), by fixing the value  $\mu = \mu_I w_I + \mu_E w_E$  (in unit/km<sup>2</sup>·min). This figure illustrates that the throughput remains almost the same when  $\frac{\mu_I w_I}{\mu_I w_I + \mu_E w_E}$  changes, validating our discussions in Section IV-B. Note that under different  $\frac{\mu_I w_I}{\mu_I w_I + \mu_E w_E}$ , there may be small fluctuation in the throughput performance in Fig. 3(h). This is because of two reasons. (a) We set a finite number for the total available RBs in the simulation study, so (17) is slightly inexact. (b) An approximation is made in (6), as explained in Footnote 5. However, both of the above factors bring only slight performance deviations.

## VII. CONCLUSION

In this paper, a new radio resource allocation method has been presented for multi-tier HWNs with inelastic and elastic traffic. The downlink sum throughput is derived based on a stochastic geometric framework, with consideration for random spatial and temporal patterns of the networks. Concave upper and lower bounds in closed-form expressions are derived for the throughput performance, to allow efficient approximate solutions to the resource allocation problem, with quantified performance bound.

## REFERENCES

- [1] 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.
- [2] V. Chandrasekhar and J. G. Andrews, "Spectrum allocation in tiered cellular networks," *IEEE Trans. on Communications*, vol. 57, no. 10, pp. 3059–3068, Oct. 2009.
- [3] W. C. Cheung, T. Q. S. Quek, and M. Kountouris, "Throughput optimization, spectrum allocation, and access control in two-tier femtocell networks," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 3, pp. 561–574, Apr. 2012.
- [4] Y. Lin and W. Yu, "Joint spectrum partition and user association in multi-tier heterogeneous networks," in *Proc. of Conference on Information Sciences and Systems (CISS)*, Princeton, NJ, Mar. 2014.
- [5] V. Chandrasekhar and J. Andrews, "Spectrum allocation in tiered cellular networks," *IEEE Trans. on Communications*, vol. 57, no. 10, pp. 3059–3068, Oct. 2009.
- [6] I. Guvenc, M.-R. Jeong, F. Watanabe, and H. Inamura, "A hybrid frequency assignment for femtocells and coverage area analysis for co-channel operation," *IEEE Communications Letters*, vol. 12, no. 12, pp. 880–882, Dec. 2008.
- [7] M. H. Cheung, A.-H. Mohsenian-Rad, V. Wong, and R. Schober, "Random access for elastic and inelastic traffic in WLANs," *IEEE Trans. on Wireless Communications*, vol. 9, no. 6, pp. 1861–1866, Jun. 2010.
- [8] L. Cimini, G. Foschini, and C.-L. I, "Call blocking performance of distributed algorithms for dynamic channel allocation in microcells," in *Proc. of IEEE ICC*, Chicago, IL, Jun. 1992.
- [9] S. Anand, A. Sridharan, and K. Sivarajan, "Performance analysis of channelized cellular systems with dynamic channel allocation," *IEEE Trans. on Vehicular Technology*, vol. 52, no. 4, pp. 847–859, Jul. 2003.
- [10] B. Błaszczyszyn, M. Jovanovic, and M. K. Karray, "Quality of real-time streaming in wireless cellular networks - stochastic modeling and analysis," *IEEE Trans. on Wireless Communications*, vol. 13, no. 6, pp. 3124–3136, Jun. 2014.
- [11] V. Joseph, S. Borst, and M. Reiman, "Optimal rate allocation for adaptive wireless video streaming in networks with user dynamics," in *Proc. of IEEE INFOCOM*, Toronto, Canada, Apr. 2014.
- [12] F. Baccelli and B. Błaszczyszyn, "Stochastic geometry and wireless networks, volume 1: Theory," *Foundations and Trends in Networking*, vol. 3, no. 3-4, pp. 249 – 449, 2009.
- [13] 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 Journal on Selected Areas in Communications*, vol. 27, no. 7, pp. 1029 – 1046, Sep. 2009.
- [14] D. Stoyan, W. Kendall, and J. Mecke, *Stochastic Geometry and Its Applications*, 2nd ed. Wiley, 1995.
- [15] 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. on Wireless Communications*, vol. 11, no. 10, pp. 3484–3495, Oct. 2012.
- [16] S. Singh, H. Dhillon, and J. G. Andrews, "Offloading in heterogeneous networks: Modeling, analysis, and design insights," *IEEE Trans. on Wireless Communications*, vol. 12, no. 5, pp. 2484–2497, May 2013.
- [17] J. G. Andrews, F. Baccelli, and R. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Trans. on Communications*, vol. 59, no. 11, pp. 3122–3134, Nov. 2011.
- [18] H. Dhillon, R. Ganti, F. Baccelli, and J. G. Andrews, "Modeling and analysis of K-tier downlink heterogeneous cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 3, pp. 550–560, Apr. 2012.
- [19] S. M. Yu and S.-L. Kim, "Downlink capacity and base station density in cellular networks," in *Proc. of IEEE International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, Tsukuba Science City, Japan, May 2013.
- [20] S. Singh and J. G. Andrews, "Joint resource partitioning and offloading in heterogeneous cellular networks," *IEEE Trans. on Wireless Communications*, vol. 13, no. 2, pp. 888–901, Feb. 2014.
- [21] 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.
- [22] W. Bao and B. Liang, "Structured spectrum allocation and user association in heterogeneous cellular networks," in *Proc. of IEEE INFOCOM*, Toronto, Canada, Apr. 2014.
- [23] R. Ganti and M. Haenggi, "Spatial and temporal correlation of the interference in ALOHA ad hoc networks," *IEEE Communications Letters*, vol. 13, no. 9, pp. 631–633, Sep. 2009.
- [24] Z. Gong and M. Haenggi, "Interference and outage in mobile random networks: Expectation, distribution, and correlation," *IEEE Trans. on Mobile Computing*, vol. 13, no. 2, pp. 337–349, Feb. 2014.
- [25] U. Schilcher, C. Bettstetter, and G. Brandner, "Temporal correlation of interference in wireless networks with Rayleigh block fading," *IEEE Trans. on Mobile Computing*, vol. 11, no. 12, pp. 2109–2120, Dec. 2012.
- [26] W. Bao and B. Liang, "Near-optimal spectrum allocation in multi-tier cellular networks with random inelastic traffic," in *Proc. of IEEE ICASSP*, Florence, Italy, May 2014.
- [27] V. B. Iversen, *Teletraffic Engineering Handbook*. ITU-D SG 2/16 & ITC Draft 2001-06-20, 2001.
- [28] P. Heegaard, "Empirical observations of traffic patterns in mobile and IP telephony," in *Next Generation Teletraffic and Wired/Wireless Advanced Networking*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2007, vol. 4712, pp. 26–37.
- [29] Y. Chen, J. Kurose, and D. Towsley, "A mixed queueing network model of mobility in a campus wireless network," in *Proc. of IEEE INFOCOM Mini-Conference*, Orlando, FL, Mar. 2012.
- [30] W. Bao and B. Liang, "Insensitivity of user distribution in multicell networks under general mobility and session patterns," *IEEE Trans. on Wireless Communications*, vol. 12, no. 12, pp. 6244–6254, Dec. 2013.
- [31] P. F. Ash and E. D. Bolker, "Generalized Dirichlet tessellations," *Geometriae Dedicata*, vol. 20, no. 2, pp. 209–243, Apr. 1986.
- [32] F. Capozzi, G. Piro, L. A. Grieco, G. Boggia, and P. Camarda, "Downlink packet scheduling in LTE cellular networks: Key design issues and a survey," *IEEE Communications Surveys and Tutorials*, vol. 2, no. 15, pp. 678–700, second quarter, 2013.
- [33] D. Jiang, H. Wang, E. Malkamaki, and E. Tuomaala, "Principle and performance of semi-persistent scheduling for VoIP in LTE system," in *Proc. of IEEE Wireless Communications, Networking and Mobile Computing (WiCom)*, Shanghai, China, Sep. 2007.
- [34] G. F. Newell, *The M/M/Infinity service system with ranked servers in heavy traffic*. Springer-Verlag, 1984.
- [35] 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.