Energy-Spectral-Efficiency Analysis and Optimization of Heterogeneous Cellular Networks: A Large-Scale User-Behavior Perspective

Guogang Zhao, Sheng Chen, *Fellow, IEEE*, Liqiang Zhao, *Member, IEEE*, and Lajos Hanzo, *Fellow, IEEE*

***Abstract*—Heterogeneous cellular networks (HCNs) are capable of meeting the explosive mobile-traffic demands. However, the conventional base station (BS) deployment strategy is unsuitable for supporting the often unpredictable nonuniform mobile-traffic demands, as governed by the large-scale user behavior (LUB). This results in the inefficient exploitation of the system's resources. In this paper, we develop an analytical framework for characterizing the achievable energy-spectral-efficiency (ESE) of HCNs, which explicitly quantifies the relationship between the network's ESE and the randomly time-varying LUBs as well as other network deployment parameters. Specifically, we model the quantitative impact of the geographical mobile-traffic intensity, the load migration factor, the users' required service rate and the per-tier BS densities on the achievable ESE of the network, while considering the area-spectral-efficiency requirements. Importantly, a closed-form ESE expression is derived, which enables us to explicitly analyze the properties of the network's ESE. Furthermore, the optimal LUB-aware BS deployment strategy is proposed for maximizing the ESE under specific outage constraints. Using numerical simulations, we verify the accuracy of the analytical ESE expression and quantify the impact of several relevant system parameters on the achievable ESE. Furthermore, we evaluate the achievable ESE performance of the network under diverse time-varying LUB scenarios. Our work, therefore, provides valuable insights for designing future ultradense HCNs.**

***Index Terms*—Heterogeneous cellular networks (HCNs), energy-spectral-efficiency (ESE), large-scale-user-behaviors (LUBs), base station (BS) deployment, mobile-traffic intensity, load migration factor, green communications and networks.**

Manuscript received April 7, 2017; revised September 16, 2017; accepted January 2, 2018. Date of publication January 8, 2018; date of current version May 14, 2018. This work was supported in parts by the National Natural Science Foundation of China (No. 61771358), by the Intergovernmental International Cooperation on Science and Technology Innovation (2016YFE0123200), by the Fundamental Research Funds for the Central Universities (JB170102), by the 111 Project (B08038), by the China Postdoctoral Science Foundation (2017M613074), and by the China Scholarship Council (No. 201506960042). L. Hanzo would like to acknowledge the financial support of the European Research Council's Advanced Fellow Grant Beam-Me-Up and of the Royal Society's Research Merit Award. The review of this paper was coordinated by Prof. Riku Jäntti. *(Corresponding author: Lajos Hanzo.)*

G. Zhao and L. Zhao are with the State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China (e-mail: ggzhao@s-an.org; lqzhao@mail.xidian.edu.cn).

S. Chen is with the School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K., and also with King Abdulaziz University, Jeddah 21589, Saudi Arabia (e-mail: sqc@ecs.soton.ac.uk).

L. Hanzo is with the School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: lh@ecs.soton.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2018.2789498

I. Introduction

A. Motivation

HETEROGENEOUS cellular networks (HCNs) are characterized by overlapping coverage areas supported both by the traditional macrocells and several different types of small-cells, whilst forming a large-scale irregularly and in many places dense network architecture [1], [2]. Naturally, the area-spectral-efficiency (ASE) of the system is expected to be significantly improved owing to spatially reusing the available spectrum in a denser manner. Thus, the explosively increasing mobile traffic can still be seamlessly conveyed, whilst satisfying high quality-of-service (QoS) requirements. On the other hand, driven both by economical and environmental concerns, network designers have to pay increasing attention to power-efficient 'green radio' (GR) techniques [3]–[12] in order to curb the rapidly escalating energy consumption of wireless communication networks. Both the 3GPP Long Term Evolution (LTE) and LTE-Advanced standards have specified energy efficiency (EE) as an important system performance indicator in addition to the enhanced QoS guarantees and to the dramatic boost of service data rates [13]. Therefore, technologies that are capable of efficiently exploiting the limited energy resources within the given licensed spectrum are of paramount importance.

The energy consumed by base stations (BSs) dominates the network's total energy consumption, and it is severely influenced by the fluctuating and geographically non-uniform mobile traffic. In the traditional worst-case design, the under-utilized BSs still consume substantial power simply for maintaining the cellular coverage [3]–[5]. This energy waste is serious, because again, the static power of a BS forms a predominant fraction of the total power consumption. Furthermore, the energy wastage may become more grave for larger-scale and ultra-dense networks [1], [2]. Therefore, it is critically important to appropriately capture and exploit the characteristics of the prevalent large-scale user behavior (LUB) for the sake of optimizing the network's spectral efficiency (SE) and/or EE subject to appropriate QoS guarantees in the context of large-scale irregularly and densely deployed multi-tier HCNs. Solving this challenging problem will result in energy-spectral efficient LUB-aware network design strategies. Unfortunately, it is inadequate to adopt the mobile-traffic-aware BS deployment strategy [14] developed for homogeneous cellular systems to the more challenging multi-tier HCN scenarios. This is because in an HCN, the LUBs

include not only the geographically non-uniform mobile traffic, but also the spatial and temporal distribution of the mobile traffic amongst the tiers. Hence, in this paper, we will investigate sophisticated LUB-aware network-level designs that have the potential to save substantially more energy than the conventional methods that are unaware of the LUB.

B. Related Works

A number of existing GR techniques focused on the energy-efficient physical layer transceiver design and radio resource allocation optimization [8]–[12], which however did not take into account the network's traffic characteristics. Diverse BS energy saving methods relying on dynamic operations and on BS deployment strategies were investigated in [5]–[7], [15]–[17]. More specifically, several optimization problems were formulated for minimizing the energy consumption of a single cell or that of a certain network area, while considering the users' data rate requirements [6], [7], [15]–[17], where it was shown that the dynamic operation and BS deployment as well as the associated algorithms conceived for finding the optimal solutions impose a high computational complexity. Nevertheless, these schemes constitute case-specific studies and they provide locally optimized solutions, which may be further improved for employment in large-scale network-wide optimization. Additionally, the energy saving performance of the network deploying a small number of BSs was evaluated in [18]–[22], where a local cell area based on the ideal circular or hexagonal cell model was assumed. However, it requires further work to use the results of [18]–[22] for guiding the design and optimization of large-scale irregularly and densely deployed HCNs.

In modeling the distribution of large-scale irregularly and densely deployed BSs, the BSs' spatial positions are typically characterized using Poisson point processes (PPPs) [23], [24]. Several contributions further extended this PPP model to multi-tier cellular networks [25]–[27], where the concept of cell association or cell load amongst tiers becomes important. Recently, the authors of [28] focused their attention on the user mobility with human tendency and on the clustering behaviors in 5G small-cell networks. Based on these works, several energy efficient network deployment strategies were proposed in [29]–[32], where the authors aimed either for minimizing the area-power-consumption (APC) of the entire network or for maximizing the network's EE performance by choosing a suitable BS density unilaterally under user-centric performance constraints. To be more specific, in [29], [30], the network's APC was minimized by switching off as many of the BSs as possible. More explicitly, the BS density was minimized, subject to the per-tier coverage/outage performance constraints. The network's EE was improved with the aid of dormant BSs in [31], where the network's EE maximization problem was formulated as a function of the sleeping factors. However, the studies of [29]–[32] all assumed that each BS consumes a constant power, which implies that the network's APC is completely determined by the density of BSs. Consequently, the quantitative impacts of LUBs, including the mobile traffic intensity and the geographical mobile traffic distribution amongst tiers both on the BS's power consumption as well as on the network's SE and EE performance were not taken into account. Compared to the aforementioned contributions, the study [33] analyzed both the SE and EE of a large-scale cellular network. Our previous treatise [14] developed a mobile-traffic-aware BS deployment strategy for homogeneous networks, but it is difficult to adopt it for the analysis and optimization of large-scale multi-tier HCNs.

It is widely recognized that beneficially exploiting the LUBs in a large-scale multi-tier HCN to optimize the energy-spectral-efficiency (ESE) is challenging, which has not been carried out in the existing literature.

C. Our Contributions

Against the above background, in this paper we propose an analytical ESE framework for large-scale irregularly and densely deployed M-tier HCNs, which facilitates the joint analysis of the impact of both the per-tier BS densities and the randomly time-varying LUBs, including the geographical mobile-traffic intensity and the load migration factor (LMF)[1], as well as the average required service rate per tier. Based on this framework, we develop the optimal LUB-aware design strategies to achieve the maximum ESE for large-scale M-tier HCNs, while taking into account the outage performance constraints. Hence this paper explicitly answers the following questions: 1) What is the quantitative impact of both the LUB and the per-tier BS densities on the achievable network ESE? 2) Is it possible to obtain the optimal ESE solutions for large-scale M-tier HCNs and how to exploit the optimal designs in practice? 3) How do the relevant system parameters influence the energy saving performance? Our main contributions are summarized as follows.

1) A new model for the BS's aggregate DL transmit power: We model the quantitative relationship between the LUB and the BS's aggregate downlink (DL) transmit power in large-scale irregularly and densely deployed M-tier HCNs, where the BS's aggregate DL transmit power is specified as the total DL power aggregated in each subband of a BS. In contrast to the related contributions [23]–[25], [29]–[32], where the BS's aggregate DL transmit power is assumed to be fixed and thus it is not influenced by the LUB, our model specifically characterizes the variations of the BS's aggregate DL transmit power caused by the geography-dependent mobile-traffic intensity and the LMF as well as the average required service rate per tier, the per-tier BS density and other network parameters. Furthermore, the modeling accuracy is verified with the aid of our simulation results.

2) A tractable network ESE analytical framework: An accurate and tractable network ESE expression is derived in closed-form based on the proposed BS aggregate DL transmit power model. This network ESE expression takes into account the per-tier BS densities, the LUB and several other relevant network parameters, and it is formulated as a function of the ratios of the BS densities to the geographical mobile-traffic intensities. The closed-form nature of this ESE expression makes the search of globally optimal solutions particularly tractable and efficient.

[1]The LMF reflects the mobile-traffic migration among the tiers, and it is similar to the bias factor of [23]–[27].

Additionally, a closed-form equivalent condition is established for the LUB to maximize the ESE, which allows us to numerically determine the LUB-aware optimal BS densities and makes the LUB-aware optimal network design strategy achievable.

3) System design insights and optimal strategy: According to the closed-form equivalent condition derived, we obtain several design guidelines for ultra-dense HCNs. In particular, we observe that the energy reduction achieved by our BS deployment strategy in a tier is partially offset by a denser BS deployment in other tiers. We propose an energy-efficient LUB-aware network design strategy for a M-tier HCN under outage probability constraints. Specifically, an optimal LUB-aware BS deployment strategy is developed by choosing a suitable BS density for the tier considered, based on the current LUBs and the BS densities in other tiers.

The rest of this paper is structured as follows. In Section II, the system model is introduced. The ESE of a large-scale M-tier HCN is modeled in Section III. In Section IV, the achievable performance of the energy-spectral-efficient LUB-aware network is analyzed, while the proposed optimal energy efficient LUB-aware network design strategy is provided in Section V. Our numerical simulation results are presented in Section VI, and our conclusions are offered in Section VII.

II. System Model

A. M-Tier HCN Characterization

We consider a large-scale dense DL M-tier HCN, whose BSs are deployed irregularly. Specifically, the M types of BS-locations are modeled by M independent PPPs in the Euclidean plane $\mathbb{R}^2$, and the BSs of the mth tier are collectively denoted by the set Ψ_m, which has the density λ_m, where $m \in \{1, 2, \ldots, M\}$. Similarly, all the user equipments (UEs), denoted by the set $\Psi_{\mathrm{u}} = \{u_i\}$, are also spatially scattered according to an independent PPP having the density λ_{u}. Fig. 2 depicts the topology of two-tier HCNs. Since the geographical mobile-traffic intensity is proportional to the users' spatial density λ_{u} given the average per-user data rate requirement, we will use λ_{u} to equivalently represent the mobile-traffic intensity. In contrast to a homogeneous network where the mobile-traffic intensity completely specifies the LUBs, the distribution of mobile traffic at each tier is also required for the M-tier HCN. We adopt the LMF to characterize the mobile-traffic distribution at every tier.

Specifically, let z represent the coordinate position of a typical user. Furthermore, let $b_m \in \Psi_m$ be the nearest BS in the mth tier to the typical user located at z. Then z is associated with the mth tier when the following condition is satisfied:

$$\|z - b_{\widetilde{m}}\|^{-\alpha} \leq \mu_{m \leftarrow \widetilde{m}} \|z - b_m\|^{-\alpha}, \forall \widetilde{m} \neq m, \qquad (1)$$

where $\|\cdot\|$ denotes the Euclidean distance, and α is the pathloss exponent, while $\mu_{m \leftarrow \widetilde{m}} > 0$ defines the LMF from tier $\widetilde{m}$ to tier m having the property of $\mu_{m \leftarrow \widetilde{m}} = \frac{1}{\mu_{\widetilde{m} \leftarrow m}}$. Physically, the LMF is related to the mobile-traffic distribution at each tier [26], [27]. In general, increasing $\mu_{m \leftarrow \widetilde{m}}$ can cause an increased mobile-traffic at the edge of tier $\widetilde{m}$ to migrate to tier m, which is equivalent to extending the BS's coverage area in tier m, where the weighted Voronoi tessellation model of [23]–[27] can be adopted and the cell of b_m is denoted by V_m. The effect of increasing $\mu_{m \leftarrow \widetilde{m}}$ to the topology of four-tier HCNs is illustrated in Fig. 1, where we have $m = \mathrm{micro}$, pico and femto tiers, while $\widetilde{m} = \mathrm{macro}$ tier. The LMF is a long-term averaged quantity, where the fading is averaged out, and it is proportional to the ratio of the power-level received from BS in tier m to the power level received from BS in tier $\widetilde{m}$ [26], [27]. For the two-tier HCN associated with $M = 2$, let m and $\widetilde{m}$ denote the macro tier and small tier, respectively. Since the power-level received from a macro BS is much higher than that from a small BS, the LMF $\mu_{m \leftarrow \widetilde{m}} \gg 1$ is true.

B. Resource Assignment

There are N_m users located in V_m and they are collectively denoted by the set φ_m. Additionally, unity frequency reuse is employed for the cells, and the total bandwidth B available to a cell V_m is divided equally into N_m subbands, with subband $B_{j,m}$ allocated to user $u_{j,m} \in \varphi_m$ for $1 \leq j \leq N_m$. In LTE systems, $B_{j,m}$ is constituted by a subset of the total set of subcarriers available. The random scheduling strategy is employed at the BS for achieving maximum fairness amongst users. The subband $B_{j,m}$ associated with user $u_{j,m}$ is then assigned an appropriate transmit power $P_{j,m}$ for meeting the service rate requirement $R_{j,m}$ of $u_{j,m}$ based on the long-term pathloss information. Each BS has a link request decision unit to manage its users' access to service. More specifically, the BS makes decisions regarding the access based on the current CSI and on the QoS requirements of the requesting users, who are only granted access when their QoS requirements can be satisfied. This resource assignment model characterizes the quantitative impact imposed by the various network parameters on each BS's aggregate DL transmit power, expressed as $P_m = \sum_{j=1}^{N_m} P_{j,m}$, and P_m is mainly influenced by the LUBs.

We consider both the large-scale pathloss and small-scale Rayleigh distributed fast fading on all desired and interfering links. The aggregate interference imposed on the UE located at z is given by the total power received from all the BSs except its associated BS, i.e., from the BSs included in the set $(\bigcup_{\widetilde{m}=1}^{M} \Psi_{\widetilde{m}}) \backslash b_m$. Unless otherwise stated, the effect of noise is ignored, since the power of the aggregate interference at a given UE is far higher than the noise power in our interference-limited scenario. Without causing confusion, we also set $u_{j,m} = z$. In order to meet the rate requirement $R_{j,m}$, the DL transmit power $P_{j,m}$ of UE $u_{j,m}$ must satisfy

$$R_{j,m} = B_{j,m} \log_2 \left(1 + \frac{P_{j,m} g_j^m}{\|z - b_m\|^{\alpha} I_{j,m}}\right), \qquad (2)$$

where g_j^m is the power gain of the fast-fading channel spanning from BS b_m to z, which is assumed to follow the independent exponential distribution with unity mean, i.e., $g_j^m \sim \exp(1)$, and $I_{j,m}$ denotes the corresponding received interference power. We focus our attention on the worst-case scenario, where the interfering BSs are transmitting at their maximum power $P_{\max}^m$.

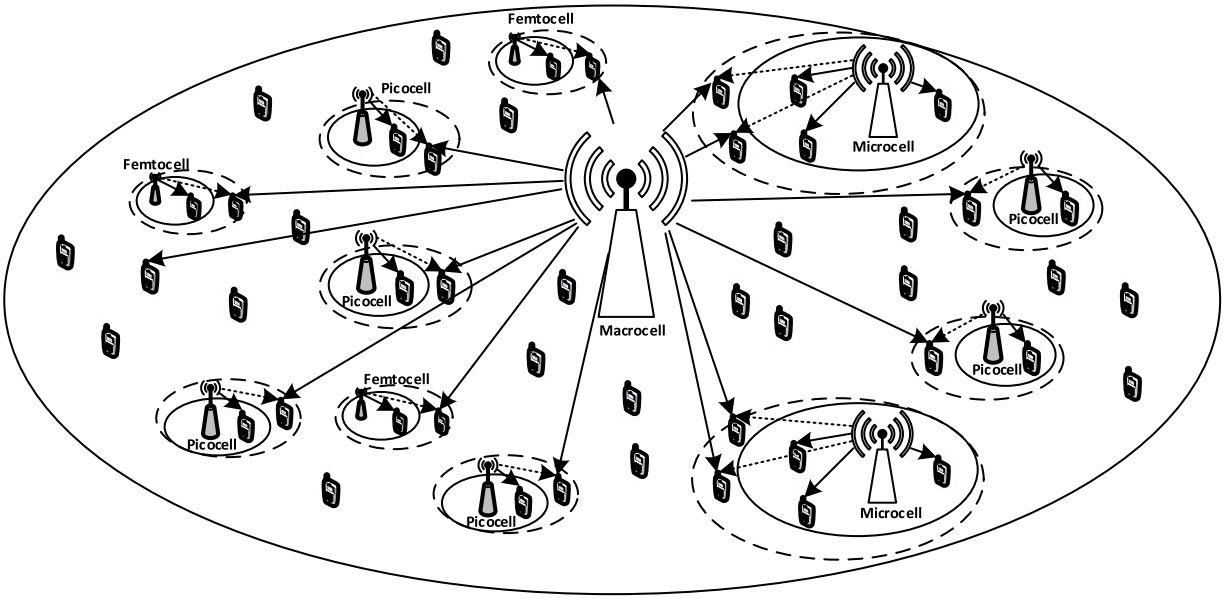


Fig. 1. A four-tier HCN, where the solid circles around the small-cell BSs are their original coverage areas with the solid arrows as their original transmit signals, while the dashed circles represent their enlarged coverage areas with the dashed arrows as their new transmit signals when the LMFs $\mu_{m\leftarrow\mathrm{mac}}$ are increased.

Hence,

$$I_{j,m} = \sum_{\widetilde{m}=1}^{M} \sum_{b_k\in\Psi_{\widetilde{m}}, b_k\neq b_m} \|z-b_k\|^{-\alpha}\, \frac{P_{\max}^{\widetilde{m}}}{N_m} g_j^k. \qquad (3)$$

Because of adopting this worst-case interference model, the results obtained in this paper represent a lower-bound for the network's achievable performance.

C. Performance Metric

We define the network-level ESE metric for our large-scale M-tier HCN as follows:

$$\eta_{\mathrm{ESE}} = \frac{\sum_{m=1}^{M} \mathrm{ASE}_m}{\sum_{m=1}^{M} \mathrm{APC}_m}\,[\mathrm{bit/Hz/Joule}], \qquad (4)$$

where ASE_m is the sum of the users' data rates [bit/s] per m^2 per Hz in the mth tier, and APC_m is the sum of the power consumption [Watt] per m^2 in the mth tier.

When deriving the ESE, we adopt the general power consumption model for each BS specified by the 3GPP standard [4] and use P_{tot}^m to denote the total BS power consumption in tier m. Additionally, a user is regarded to suffer from outage, if its service data rate cannot be guaranteed under the maximum BS transmit power available to the subband occupied by this user. Since our main goal is the network's ESE-level evaluation, the users' outage probabilities are treated as the constraints.

For clarity, the main notations used in this paper are summarized in Table I.

TABLE I
LIST OF MAIN NOTATIONS

Notation	Definition
Ψ_m	Set of BSs in tier m
λ_m	Density of BSs in tier m
Ψ_{u}	Set of all UEs
λ_{u}	Density of UEs
b_m	The nearest BS in tier m to position z
V_m	Cell centered at BS b_m
α	Pathloss exponent
$\mu_{m\leftarrow\widetilde{m}}$	LMF from tier $\widetilde{m}$ to tier m
φ_m	Set of UEs associated with BS b_m
N_m	Number of UEs in cell V_m
B	Total bandwidth available to a cell
$B_{j,m}$	Subband allocated to UE $u_{j,m}$
$R_{j,m}$	Required service data rate of UE $u_{j,m} \in \varphi_m$
$P_{j,m}$	DL transmit power of UE $u_{j,m}$
P_m	Aggregate DL transmit power of BS b_m in tier m
g_j^m	Fast-fading power gain from BS b_m to UE $u_{j,m}$
$P_{\max}^m$	maximum BS DL transmit power in tier m
η_{ESE}	Network-level energy-spectral efficiency
A_m	Coverage area of cell V_m
β_m	tier m BS power amplifier efficiency
P_{OM}^m	Each BS's static operation power in tier m
R_m	Average required service data rate (per UE) in tier m
Q_m	Outage probability in tier m
$\varepsilon_{\mathrm{out}}$	Outage threshold

III. NETWORK-LEVEL ESE FORMULATION IN LARGE-SCALE M-TIER HCNS

For the large-scale M-tier HCN, we characterize the average DL transmit power for a typical user and the average aggregate DL transmit power in a typical cell as well as derive a closed-form network-level ESE expression. In particular, we quantify the impact of LUB, including the geographical mobile-traffic intensity and LMFs, and the spatially averaged service data rate requirement.

A. Average DL Transmit Power for a Typical User

Proposition 1: In a large-scale M-tier HCN associated with irregularly and densely deployed BSs, given the BS densities λ_m, the mobile-traffic intensity λ_{u} and the LMFs $\mu_{m\leftarrow\widetilde{m}}$, the

$$\eta_{\mathrm{ESE}} = \frac{\sum_{m=1}^{M} \frac{R_m \lambda_m}{B \sum_{\widetilde{m}=1}^{M} \lambda_{\widetilde{m}} \mu_{m \leftarrow \widetilde{m}}^{-2/\alpha}}}{\sum_{m=1}^{M} \left(\frac{P_{\max}^{m} \beta_m \lambda_m}{(\alpha-2)\lambda_{\mathrm{u}} \left(1-\left(2^{R_m/B}-1\right)\lambda_{\mathrm{u}} \Big/ \left(K \sum_{\widetilde{m}=1}^{M} \lambda_{\widetilde{m}} \mu_{m \leftarrow \widetilde{m}}^{-2/\alpha}\right)\right)^{K}} - \frac{\lambda_m \beta_m P_{\max}^{m}}{\lambda_{\mathrm{u}}(\alpha-2)} + \frac{\lambda_m P_{\mathrm{OM}}^{m}}{\lambda_{\mathrm{u}}} \right)} \ [\mathrm{bits/Hz/Joule}]. \quad (11)$$

average DL transmit power required for user $u_{j,m}$ having the service data rate of $R_{j,m}$ is given by

$$E\left[P_{j,m}\right] = \frac{\lambda_m \left(2^{R_{j,m}/B_{j,m}} - 1\right)}{(\alpha-2)\lambda_{\mathrm{u}}} \sum_{\widetilde{m}=1}^{M} \frac{\lambda_{\widetilde{m}} P_{\max}^{\widetilde{m}} \mu_{m \leftarrow \widetilde{m}}^{(\alpha-2)/\alpha}}{\lambda_m}, \quad (5)$$

where $E[\cdot]$ denotes the expectation operator, and the pathloss exponent is $\alpha > 2$.

Proof: See Appendix A. ■

The impact of the mobile-traffic intensity λ_{u} and the BS densities λ_m on the average DL transmit power required by user $u_{j,m}$ is multifold. From the first fractional expression on the right-hand side of (5), we can see that increasing λ_m, which indicates having denser interfering BSs from tier m, is beneficial for increasing $E[P_{j,m}]$. However, a larger λ_m increases $B_{j,m}$ since each cell has less users, which in turn decreases $E[P_{j,m}]$. Since this reduction is exponential, $E[P_{j,m}]$ decreases as λ_m increases. On the other hand, $E[P_{j,m}]$ includes the factor of $\frac{1}{\lambda_{\mathrm{u}}}$, and hence it decreases as λ_u increases. But increasing λ_{u} will results in less subbands $B_{j,m}$ being allocated for $u_{j,m}$, which in turn increases $E[P_{j,m}]$ exponentially. Therefore, increasing λ_{u} will increase $E[P_{j,m}]$ in general. Furthermore, increasing the LMF $\mu_{m \leftarrow \widetilde{m}}$ implicitly also reduces $B_{j,m}$, because more users will migrate from tier $\widetilde{m}$ to m, which results in a higher mobile traffic in tier m. Note that from (5), by setting $M = 1$ and $\mu_{m \leftarrow \widetilde{m}} = 1$, the average DL transmit power required for a typical UE in the homogeneous cellular network can be obtained [14].

B. Average Aggregate DL Transmit Power in a Typical Cell

Proposition 2: In a large-scale M-tier HCN with irregularly and densely deployed BSs, given the BS densities λ_m, the mobile-traffic intensity λ_{u} and the LMFs $\mu_{m \leftarrow \widetilde{m}}$, upon assuming that all the users in tier m have an identical required service data rate of R_m, the averaged aggregate DL transmit power of BS b_m is given by

$$E\left[P_m\right] = \left(K \left(\sum_{\widetilde{m}=1}^{M} \lambda_{\widetilde{m}} \mu_{m \leftarrow \widetilde{m}}^{-2/\alpha} \right) - \left(2^{R_m/B} - 1\right) \lambda_{\mathrm{u}} \right)^{-K} \times \frac{P_{\max}^{m}}{(\alpha-2)} \left(\sum_{\widetilde{m}=1}^{M} K \lambda_{\widetilde{m}} \mu_{m \leftarrow \widetilde{m}}^{-2/\alpha} \right)^{K} - \frac{P_{\max}^{m}}{(\alpha-2)}, \quad (6)$$

where $K = 3.57$.

Proof: See Appendix B. ■

C. Network-Level ESE Formulation

We are now ready to complete the network-level ESE metric (4) for a large-scale M-tier HCN. By using R_m to represent the users' spatially averaged required service data rate in tier m, the ASE can be expressed as

$$\mathrm{ASE} = \sum_{m=1}^{M} \frac{1}{B} \frac{\lambda_{\mathrm{u}} R_m}{\sum_{\widetilde{m}=1}^{M} \frac{\lambda_{\widetilde{m}} \mu_{m \leftarrow \widetilde{m}}^{-2/\alpha}}{\lambda_m}} \ [\mathrm{bit/s/Hz/m^2}], \quad (7)$$

where λ_{u} has the unit of $\mathrm{UE/m^2}$, and

$$\frac{1}{\sum_{\widetilde{m}=1}^{M} \frac{\lambda_{\widetilde{m}}}{\lambda_m} \mu_{m \leftarrow \widetilde{m}}^{-2/\alpha}} = \frac{1}{\chi_m} \quad (8)$$

represents the number of BSs associated with tier m [26], [27]. The network's APC on the other hand can be expressed as

$$\mathrm{APC} = \sum_{m=1}^{M} \lambda_m \left(\beta_m E\left[P_m\right] + P_{\mathrm{OM}}^{m}\right) \ [\mathrm{Joule/s/m^2}], \quad (9)$$

where $\lambda_m (\beta_m E[P_m] + P_{\mathrm{OM}}^{m})$ represents the BS power consumption per $\mathrm{m^2}$ in tier m, β_m is the mth tier BS power amplifier efficiency, and $P_{\mathrm{OM}}^{\mathrm{m}}$ represents each tier-m BS's static operational power consumption, including the power consumed by baseband processing, battery backup, BS cooling, and so on. Thus,

$$\eta_{\mathrm{ESE}} = \frac{\sum_{m=1}^{M} \frac{\lambda_{\mathrm{u}} \lambda_m R_m}{B \sum_{\widetilde{m}=1}^{M} \lambda_{\widetilde{m}} \mu_{m \leftarrow \widetilde{m}}^{-2/\alpha}}}{\sum_{m=1}^{M} \lambda_m \left(\beta_m E\left[P_m\right] + P_{\mathrm{OM}}^{m}\right)} \ [\mathrm{bits/Hz/Joule}]. \quad (10)$$

Noting Proposition 2, the following proposition is obvious.

Proposition 3: In a large-scale M-tier HCN with irregularly and densely deployed BSs having the BS density λ_m in tier m, the mobile-traffic intensity λ_{u} and the LMF $\mu_{m \leftarrow \widetilde{m}}$, upon assuming that all the users in tier m have an identical required service data rate of R_m, the closed-form network-level ESE expression is given by (11) at the top of this page.

Additionally, in the special case of $M = 2$, we have the following corollary.

Corollary 1: Consider the large-scale two-tier HCN, where the density of macro BSs is λ_{mac}, the density of small BSs is λ_{mic}, the mobile-traffic intensity is λ_{u}, and the LMF is $\mu_{\mathrm{mac} \leftarrow \mathrm{mic}} = \mu$, while the average required service data rates in the macro-BS tier and small-BS tier are R_{mac} and R_{mic}, respectively. The closed-form network-level ESE expression for this two-tier HCN is given by

$$\eta_{\mathrm{ESE}} = \frac{R_{\mathrm{mac}}/\chi_{\mathrm{mac}} + R_{\mathrm{mic}}/\chi_{\mathrm{mic}}}{B\left(U_{\mathrm{mac}} + U_{\mathrm{mic}}\right)} \ [\mathrm{bit/Hz/Joule}], \quad (12)$$

where

$$\chi_{\rm mac} = 1 + \frac{\lambda_{\rm mic}}{\lambda_{\rm mac}} \mu^{-\frac{2}{\alpha}}, \tag{13}$$

$$\chi_{\rm mic} = 1 + \frac{\lambda_{\rm mac}}{\lambda_{\rm mic}} \mu^{\frac{2}{\alpha}}, \tag{14}$$

$$U_{\rm mac} = \frac{P_{\max}^{\rm mac} \beta_{\rm mac} \lambda_{\rm mac}}{(\alpha-2)\lambda_{\rm u} \left(1 - \frac{\left(2^{R_{\rm mac}/B}-1\right)\lambda_{\rm u}}{K \lambda_{\rm mac} \chi_{\rm mac}}\right)^K} - \frac{\lambda_{\rm mac} \beta_{\rm mac} P_{\max}^{\rm mac}}{\lambda_{\rm u}(\alpha-2)} + \frac{\lambda_{\rm mac} P_{\rm OM}^{\rm mac}}{\lambda_{\rm u}}, \tag{15}$$

$$U_{\rm mic} = \frac{P_{\max}^{\rm mic} \beta_{\rm mic} \lambda_{\rm mic}}{(\alpha-2)\lambda_{\rm u} \left(1 - \frac{\left(2^{R_{\rm mic}/B}-1\right)\lambda_{\rm u}}{K \lambda_{\rm mic} \chi_{\rm mic}}\right)^K} - \frac{\lambda_{\rm mic} \beta_{\rm mic} P_{\max}^{\rm mic}}{\lambda_{\rm u}(\alpha-2)} + \frac{\lambda_{\rm mic} P_{\rm OM}^{\rm mic}}{\lambda_{\rm u}}. \tag{16}$$

IV. LUBs-Aware Performance Analysis and Optimization in Large-Scale M-Tier HCNs

In this section, the properties of the network-level ESE for a large-scale M-tier HCN are explicitly analyzed. First, the optimal LUB-aware BS densities maximizing the network's ESE are derived in the following proposition.

Proposition 4: In a large-scale M-tier HCN having the BS density λ_m in tier m, the mobile-traffic intensity $\lambda_{\rm u}$ and the LMF $\mu_{m \leftarrow \widetilde{m}}$, where all the users in tier m have an identical required service data rate of R_m, there exists a unique optimal LUB-aware BS density in tier m, denoted by $\lambda_m^\star$, which maximizes the network's ESE $\eta_{\rm ESE}$. Furthermore, $\lambda_m^\star$ can be numerically obtained from

$$\frac{1-(1+K)\Upsilon_m}{(1-\Upsilon_m)^{K+1}} + \sum_{\widetilde{m} \neq m} \mu_{m \leftarrow \widetilde{m}}^{\frac{2-\alpha}{\alpha}} \frac{1-(1+K)\Upsilon_{\widetilde{m}}}{(1-\Upsilon_{\widetilde{m}})^{K+1}} = \gamma_m^{(M)}, \tag{17}$$

where

$$\Upsilon_m = \frac{\left(2^{R_m/B}-1\right)\lambda_{\rm u}}{K\left(\lambda_m^\star + \sum\limits_{\widetilde{m} \neq m} \lambda_{\widetilde{m}} \mu_{m \leftarrow \widetilde{m}}^{-\frac{2}{\alpha}}\right)}, \tag{18}$$

$$\Upsilon_{\widetilde{m}} = \frac{\left(2^{R_{\widetilde{m}}/B}-1\right)\lambda_{\rm u}}{K\left(\lambda_{\widetilde{m}} + \sum\limits_{m' \neq m, m' \neq \widetilde{m}} \lambda_{m'} \mu_{\widetilde{m} \leftarrow m'}^{-\frac{2}{\alpha}} + \lambda_m^\star \mu_{\widetilde{m} \leftarrow m}^{-\frac{2}{\alpha}}\right)}, \tag{19}$$

$$\gamma_m^{(M)} = \left(1 + \sum_{\widetilde{m} \neq m} \mu_{m \leftarrow \widetilde{m}}^{\frac{2-\alpha}{\alpha}}\right) - \frac{P_{\rm OM}^m}{P_{\max}^m}(\alpha-2). \tag{20}$$

Proof: See Appendix C. ■

In the special case of $M = 2$, i.e., a large-scale macro-small two-tier HCN, we have the following corollary.

Corollary 2: In a large-scale two-tier HCN with irregularly and densely deployed macro BSs in tier $\rm mac$ and small BSs in tier $\rm mic$, the unique optimal LUBs-aware BS density $\lambda_m^\star$ can be numerically obtained from

$$\frac{1-(1+K)\bar{\Upsilon}_m}{(1-\bar{\Upsilon}_m)^{K+1}} + \mu_{m \leftarrow \widetilde{m}}^{\frac{2-\alpha}{\alpha}} \frac{1-(1+K)\bar{\Upsilon}_{\widetilde{m}}}{(1-\bar{\Upsilon}_{\widetilde{m}})^{K+1}} = \gamma_m, \tag{21}$$

where $m = \rm mac$ and $\widetilde{m} = \rm mic$ or $m = \rm mic$ and $\widetilde{m} = \rm mac$,

$$\bar{\Upsilon}_m = \frac{\left(2^{R_m/B}-1\right)\lambda_{\rm u}}{K\left(\lambda_m^\star + \lambda_{\widetilde{m}} \mu_{m \leftarrow \widetilde{m}}^{-\frac{2}{\alpha}}\right)}, \tag{22}$$

$$\bar{\Upsilon}_{\widetilde{m}} = \frac{\left(2^{R_{\widetilde{m}}/B}-1\right)\lambda_{\rm u}}{K\left(\lambda_{\widetilde{m}} + \lambda_m^\star \mu_{m \leftarrow \widetilde{m}}^{\frac{2}{\alpha}}\right)}, \tag{23}$$

$$\gamma_m = \left(1 + \mu_{m \leftarrow \widetilde{m}}^{\frac{2-\alpha}{\alpha}}\right) - \frac{P_{\rm OM}^m}{P_{\max}^m}(\alpha-2). \tag{24}$$

From Corollary 2, we have the following corollary.

Corollary 3: Consider the large-scale two-tier HCN with irregularly and densely deployed macro BSs in tier $\rm mac$ and small BSs in tier $\rm mic$. Given two BS densities $\lambda_{\widetilde{m},1} \neq \lambda_{\widetilde{m},2}$ in tier $\widetilde{m}$, we have a pair of corresponding optimal BS densities $\lambda_{m,1}^\star$ and $\lambda_{m,2}^\star$ in tier m, which satisfy

$$\left(\lambda_{m,1}^\star - \lambda_{m,2}^\star\right)\left(\lambda_{\widetilde{m},1} - \lambda_{\widetilde{m},2}\right) < 0, \tag{25}$$

where $m = \rm mac$ and $\widetilde{m} = \rm mic$ or $m = \rm mic$ and $\widetilde{m} = \rm mac$.

Proof: See Appendix D. ■

Corollary 3 implies that when more BSs are deployed in tier $\widetilde{m}$, i.e., when increasing $\lambda_{\widetilde{m}}$, the optimal BS density $\lambda_m^\star$ in tier m has to be reduced, if the network is expected to operate at a near-optimal ESE. This is consistent with the generally accepted insight that deploying denser macro BSs will degrade the performance of small BSs due to the increased cross-tier interference imposed on small cells, and vice versa.

Proposition 5: Consider the large-scale M-tier HCN with irregularly and densely deployed BSs having the BS density λ_m in tier m and the LMF $\mu_{m \leftarrow \widetilde{m}}$, where all the users in tier m have an identical required service rate of R_m. There exists a unique optimal mobile-traffic intensity for maximizing the network ESE, denoted as $\lambda_{\rm u}^\star$, which can be numerically obtained from

$$\sum_{m=1}^{M} \left(\frac{K P_{\max}^m \beta_m \lambda_m \lambda_{\rm u}^\star S_m}{(1-S_m\lambda_{\rm u}^\star)^{K+1}} - \frac{P_{\max}^m \beta_m \lambda_m}{(1-S_m\lambda_{\rm u}^\star)^K} + \lambda_m \beta_m P_{\max}^m - \lambda_m P_{\rm OM}^m(\alpha-2) \right) = 0, \tag{26}$$

where

$$S_m = \frac{\left(2^{R_m/B}-1\right)}{\left(K\sum_{\widetilde{m}=1}^{M} \lambda_{\widetilde{m}} \mu_{m \leftarrow \widetilde{m}}^{-2/\alpha}\right)}. \tag{27}$$

Proof: See Appendix E. ■

V. Optimal Energy Efficient LUBs-Aware Network Design Strategy

In this section, we commence by presenting our framework conceived for designing the optimal energy efficient LUB-aware

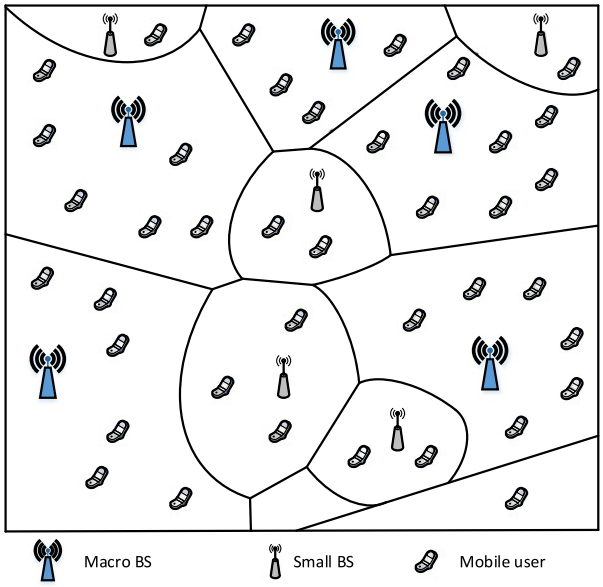


Fig. 2. Voronoi topology of a large-scale two-tier heterogeneous cellular network where macro and small BSs as well as mobile users are located according to their respective PPP distributions.

M-tier HCN. However, our main goal is to derive an energy efficient LUBs-aware design strategy for a large-scale two-tier HCN with irregularly and densely deployed macro BSs in tier mac and small BSs in tier mic.

A. Optimal Energy Efficient LUBs-Aware Network Design Strategy for M-Tier HCNs

The following proposition presents the outage probability of a typical user associated with a tier-m BS, which is defined as the probability that a typical user's data rate requirement cannot be met even at the maximum DL transmit power.

Proposition 6: In a large-scale M-tier HCN with irregularly and densely deployed BSs where the BS-density is λ_m in tier m with the LMF $\mu_{m\leftarrow\widetilde{m}}$ and the mobile-traffic intensity is λ_{u}, the outage probability of a typical user $u_{j,m}$ associated with a tier-m BS b_m that requires a service data rate of $R_{j,m}$ is given by

$$Q_{j,m} = \frac{\lambda_{\mathrm{u}}}{N_m \sum_{\widetilde{m}=1}^{M} \mu_{m\leftarrow\widetilde{m}}^{-\frac{2}{\alpha}} \lambda_{\widetilde{m}} \rho_{j,m} \exp\left(\frac{N_m}{\lambda_{\mathrm{u}}} \sum_{\widetilde{m}=1}^{M} \mu_{m\leftarrow\widetilde{m}}^{-\frac{2}{\alpha}} \lambda_{\widetilde{m}} \rho_{j,m}\right)} + 1 - \frac{\lambda_{\mathrm{u}}}{N_m \sum_{\widetilde{m}=1}^{M} \mu_{m\leftarrow\widetilde{m}}^{-\frac{2}{\alpha}} \lambda_{\widetilde{m}} \rho_{j,m}}, \tag{28}$$

where N_m is the number of UEs associated with BS b_m, whose probability mass function (PMF) can be found in Appendix B, and

$$\rho_{j,m} = T_{j,m}^{2/\alpha} \int_{T_{j,m}^{-2/\alpha}}^{+\infty} \frac{1}{1+u^{\alpha/2}} du, \tag{29}$$

with $T_{j,m} = 2^{R_{j,m}/B_{j,m}} - 1$.

Proof: See Appendix F. ■

In $Q_{j,m}$, by replacing N_m with $E[N_m]$ of (52) in Appendix B and by replacing $R_{j,m}$ with the average service data rate R_m, we obtain the outage probability of the mth tier as

$$Q_m = 1 + \frac{1}{\rho_m \exp(\rho_m)} - \frac{1}{\rho_m}, \tag{30}$$

in which

$$\rho_m = T_m^{2/\alpha} \int_{T_m^{-2/\alpha}}^{+\infty} \frac{1}{1+u^{\alpha/2}} du, \tag{31}$$

with

$$T_m = 2^{(R_m \lambda_{\mathrm{u}})\Big/\left(B \sum_{\widetilde{m}=1}^{M} \mu_{m\leftarrow\widetilde{m}}^{-2/\alpha} \lambda_{\widetilde{m}}\right)} - 1. \tag{32}$$

Denote $\boldsymbol{\lambda} = [\lambda_1\, \lambda_2 \cdots \lambda_M]^{\mathrm{T}}$. Clearly, Q_m is a function of $\boldsymbol{\lambda}$ and λ_{u}, and is therefore denoted as $Q_m(\boldsymbol{\lambda}, \lambda_{\mathrm{u}})$.

The problem of finding the optimal LUB-aware energy efficient BS deployment for the M-tier HCN under the outage constraint in each tier can be formulated as

$$\begin{aligned} \max_{\boldsymbol{\lambda}} \quad & \eta_{\mathrm{ESE}}, \\ \text{s.t.} \quad & Q_m\left(\boldsymbol{\lambda}, \lambda_{\mathrm{u}}\right) \le \varepsilon_{\mathrm{out}}^m,\ 1 \le m \le M, \end{aligned} \tag{33}$$

where $0 < \varepsilon_{\mathrm{out}}^m < 1$ is the outage probability threshold in the mth tier and η_{ESE} is given by (11).

The optimization problem (33) is complex. Moreover, its practical application is somewhat limited, since it is unlikely in practice that a M-tier HCN is designed and deployed from the beginning. A most likely practical senario is that when a homogeneous macro network can no longer meet the service requirements, a micro tier is introduced. Therefore, the task becomes one of optimizing the small BS deployment to achieve the optimal ESE operating status of the two-tier HCN. We will study this more tractable and practical two-tier HCN senario, whose topology is illustrated in Fig. 2.

B. Optimal Energy Efficient LUBs-Aware Network Design Strategy for Two-Tier HCNs

We now focus our attention on large-scale two-tier HCNs and our task is to find the optimal LUB-aware small BS density under the relevant outage constraint in the macro tier associated with the given macro BS density λ_{mac} and user intensity λ_{u}, which will be denoted as $\lambda_{\mathrm{mic}}^{out\star}$. The optimization problem associated with this optimal LUB-aware small BS deployment design for the two-tier HCN is formally formulated as

$$\begin{aligned} \max_{\lambda_{\mathrm{min}}} \quad & \eta_{\mathrm{ESE}}, \\ \text{s.t.} \quad & Q_{\mathrm{mac}}\left(\lambda_{\mathrm{mic}}; \lambda_{\mathrm{mac}}, \lambda_{\mathrm{u}}\right) \le \varepsilon_{\mathrm{out}}, \end{aligned} \tag{34}$$

where $0 < \varepsilon_{\mathrm{out}} < 1$ is the outage probability threshold in tier mac and η_{ESE} is given by (12). First, we explain why only the macro-tier outage constraint has to be considered. A small-cell user not only gets more bandwidth, since the number of UEs in a small cell is relatively small in comparison with a macro cell, but also requires less transmit power to meet its rate requirement, since the distance involved is much smaller. Therefore, the outage performance of the mic tier is inherently much better than that of the mac tier. Consequently, we only

have to consider the outage probability of the mac tier as the outage constraint.

The optimal solution $\lambda_{\rm mic}^{\star}$ to the unconstrained optimization problem $\max_{\lambda_{\rm min}} \eta_{\rm ESE}$ is given in Corollary 2. In order to solve the outage-constrained optimization problem (34), we examine the macro-tier outage constraint

$$Q_{\rm mac}\left(\lambda_{\rm mic};\lambda_{\rm mac},\lambda_{\rm u}\right) = 1 + \frac{1}{\rho_{\rm mac}\exp(\rho_{\rm mac})} - \frac{1}{\rho_{\rm mac}}, \quad (35)$$

where

$$\rho_{\rm mac} = T_{\rm mac}^{2/\alpha}\int_{T_{\rm mac}^{-2/\alpha}}^{+\infty}\frac{1}{1+u^{\alpha/2}}du, \quad (36)$$

$$T_{\rm mac} = 2^{\frac{R_{\rm mac}\lambda_{\rm u}}{B\left(\lambda_{\rm mac}+\mu^{-2/\alpha}_{\rm mac\leftarrow mic}\lambda_{\rm mic}\right)}} - 1. \quad (37)$$

The derivative of $Q_{\rm mac}$ with respect to $\rho_{\rm mac}$ satisfies

$$\frac{\partial Q_{\rm mac}}{\partial \rho_{\rm mac}} = \frac{1}{\rho_{\rm mac}^2}\left(1-\exp\left(-\rho_{\rm mac}\right)\right) - \frac{1}{\rho_{\rm mac}}\exp\left(-\rho_{\rm mac}\right)$$
$$> \frac{\partial Q_{\rm mac}}{\partial \rho_{\rm mac}}\Big|_{\exp\left(\rho_{\rm mac}\right)=\rho_{\rm mac}+1} = 0, \quad (38)$$

Since $\exp\left(\rho_{\rm mac}\right) > \rho_{\rm mac}+1$ always holds for $\rho_{\rm mac} > 0$. This indicates that $Q_{\rm mac}$ is a monotonically increasing function of $\rho_{\rm mac}$. From (36) and (37), it is easy to see that $\rho_{\rm mac}$ increases as $\lambda_{\rm mic}$ decreases, since $\rho_{\rm mac}$ increases with $T_{\rm mac}$ and $T_{\rm mac}$ decreases as $\lambda_{\rm mic}$ increases. Therefore, we have

$$\frac{\partial Q_{\rm mac}}{\partial \lambda_{\rm mic}} = \frac{\partial Q_{\rm mac}}{\partial \rho_{\rm mac}}\frac{\partial \rho_{\rm mac}}{\partial \lambda_{\rm mic}} < 0, \quad (39)$$

which implies that the outage-constrained feasible region of $\lambda_{\rm mic}$ that satisfies $Q_{\rm mac}\left(\lambda_{\rm mic};\lambda_{\rm mac},\lambda_{\rm u}\right) \le \varepsilon_{\rm out}$ is given by

$$\lambda_{\rm mic} \in \left[Q_{\rm mac}^{-1}\left(\varepsilon_{\rm out};\lambda_{\rm mac},\lambda_{\rm u}\right), +\infty\right), \quad (40)$$

where $Q_{\rm mac}^{-1}\left(\cdot;\lambda_{\rm mac},\lambda_{\rm u}\right)$ denotes the inverse function of $Q_{\rm mac}\left(\lambda_{\rm mic};\lambda_{\rm mac},\lambda_{\rm u}\right)$.

Moreover, it can be seen from Appendix C that for $\lambda_{\rm mic} \in \left[\lambda_{\rm mic}^{(1)}, \lambda_{\rm mic}^{\star}\right)$, $\eta_{\rm ESE}$ is a monotonically increasing function of $\lambda_{\rm mic}$, while for $\lambda_{\rm mic} \in \left[\lambda_{\rm mic}^{\star}, +\infty\right)$, $\eta_{\rm ESE}$ is a monotonically decreasing function of $\lambda_{\rm mic}$. Based on these properties of $\eta_{\rm ESE}$ and the outage-constrained feasible region of (40) as well as by defining $\varepsilon_{\rm out}^{\star}$ as the solution of

$$Q_{\rm mac}^{-1}\left(\varepsilon_{\rm out}^{\star};\lambda_{\rm mac},\lambda_{\rm u}\right) = \lambda_{\rm mic}^{\star}, \quad (41)$$

we readily obtain the outage-constrained optimal solution $\lambda_{\rm mic}^{out\star}$ to the outage-constrained optimization problem (34), as formulated in the following theorem.

Theorem 1: For $\varepsilon_{\rm out} < \varepsilon_{\rm out}^{\star}$, i.e., for $Q_{\rm mac}^{-1}(\varepsilon_{\rm out};\lambda_{\rm mac},\lambda_{\rm u}) > \lambda_{\rm mic}^{\star}$, the outage-constrained optimal micro-tier BS density is given by $\lambda_{\rm mic}^{out\star} = Q_{\rm mac}^{-1}(\varepsilon_{\rm out};\lambda_{\rm mac},\lambda_{\rm u})$, while for $\varepsilon_{\rm out} > \varepsilon_{\rm out}^{\star}$, i.e., for $Q_{\rm mac}^{-1}(\varepsilon_{\rm out};\lambda_{\rm mac},\lambda_{\rm u}) < \lambda_{\rm mic}^{\star}$, the outage-constrained optimal micro-tier BS density is given by $\lambda_{\rm mic}^{out\star} = \lambda_{\rm mic}^{\star}$.

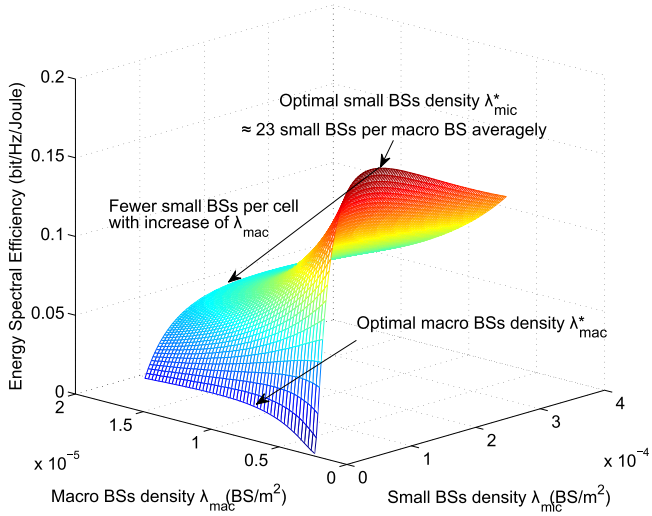


Fig. 3. The network ESE $\eta_{\rm ESE}$ as the function of $\lambda_{\rm mac}$ and $\lambda_{\rm mic}$, given $\lambda_{\rm u} = 4\times10^{-4}$ user/m^2, $R_{\rm mac} = 0.3$ Mbit/s, $R_{\rm mic} = 1$ Mbit/s and $\mu = 8$.

VI. Numerical and Simulation Results

Both numerical and simulation results are provided to characterize the ESE of large-scale two-tier HCNs and to verify the accuracy of our ESE performance metrics. Moreover, the APC required by our optimal-EE LUB-aware network design strategy is compared to that of the practical baseline cellular design [39]–[41], for the scenario of an approximate macro BS density of $2 \sim 10\,{\rm BS/km^2}$. Specifically, we use the energy saving gain to quantify the gain of our optimized design over the baseline design of [39]–[41], which is defined as

$$\Delta{\rm APC} = \frac{{\rm APC}_{\rm base} - {\rm APC}_{\rm opt}}{{\rm APC}_{\rm base}}, \quad (42)$$

where ${\rm APC}_{\rm base}$ and ${\rm APC}_{\rm opt}$ denote the APCs required by the baseline design and our optimal design, respectively. Unless otherwise stated, we choose the network parameters as follows: $P_{\rm OM}^{\rm mac} = 20$ W, $P_{\rm OM}^{\rm mic} = 0.5$ W, $P_{\rm max}^{\rm mac} = 40$ W, $P_{\rm max}^{\rm mic} = 2$ W, $\alpha = 4$, $\beta = 1.2$ and $\varepsilon_{\rm out} = 0.3$. Note that according to [42], [43], the power consumption of the power amplifier - including the feeder - accounts for approximately 50–80% of the total power consumption of a BS. Based on this result, we opted for $P_{\rm OM}^{\rm mac} = 20$ W and $P_{\rm max}^{\rm mac} = 40$ W.

A. Characteristics of Key Performance Metrics

In Fig. 3, the network's ESE quantified in terms of bit/Hz/Joule, as the function of $\lambda_{\rm mac}$ and $\lambda_{\rm mic}$ is illustrated. It can be observed from Fig. 3 that for a given $\lambda_{\rm mac}$, there always exists a unique optimal $\lambda_{\rm mic}^{\star}$ to maximize $\eta_{\rm ESE}$, while for a given $\lambda_{\rm mic}$, there always exists a unique optimal $\lambda_{\rm mac}^{\star}$ which maximizes $\eta_{\rm ESE}$. This confirms Corollary 2. It is also worth noting that when the BS density λ_m in tier m increases or decreases, the optimal density $\lambda_{\widetilde{m}}^{\star}$ in tier $\widetilde{m}$ varies in an inverse direction, where $m = {\rm mac}$ or mic, while $\widetilde{m} = {\rm mic}$ or mac. This agrees with Corollary 3.

Fig. 4 illustrates the impact of the required service rates in the two tiers on the network's ESE under varying $\lambda_{\rm mic}$ as well

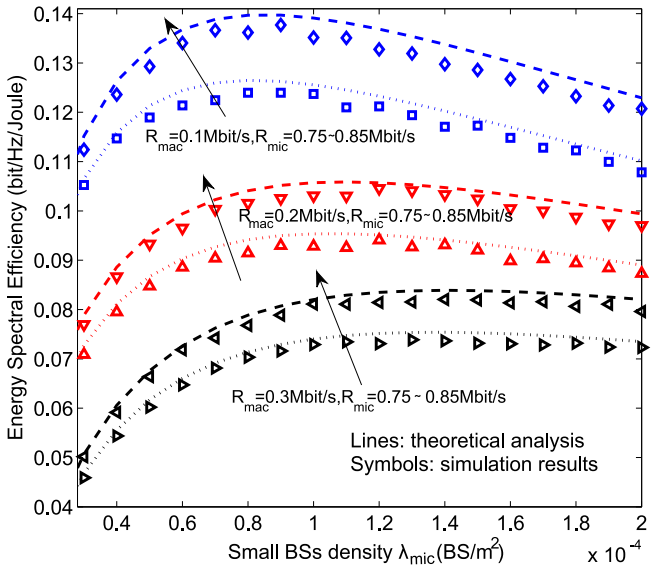


Fig. 4. The impacts of R_{mac} and R_{mic} on the network ESE η_{ESE} for various λ_{mic}, as well as given $\lambda_{\mathrm{mac}} = 5 \times 10^{-6}\,\mathrm{BS/m^2}$, $\lambda_{\mathrm{u}} = 4 \times 10^{-4}\,\mathrm{user/m^2}$, and $\mu = 10$. The arrow indicates increasing R_{mic}.

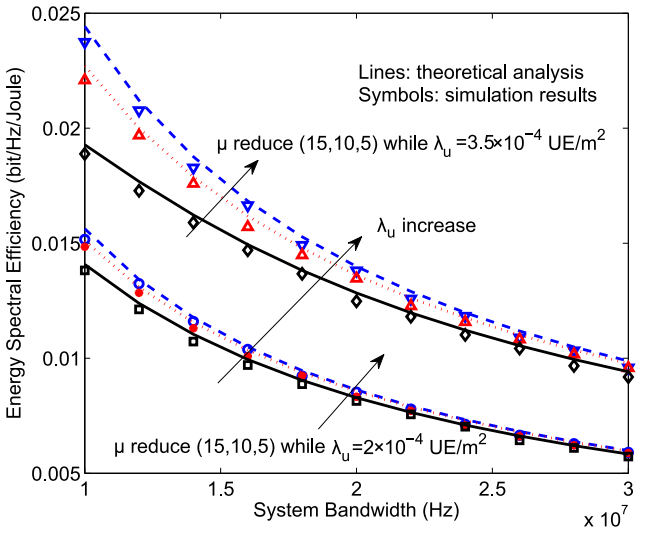


Fig. 5. The impacts of λ_{u} and μ on the network ESE η_{ESE} for various B, under the condition of $\lambda^{\star}_{\mathrm{mic}}$ as well as given $R_{\mathrm{mac}} = R_{\mathrm{mic}} = 0.1$ Mbit/s and $\lambda_{\mathrm{mac}} = 5 \times 10^{-6}\,\mathrm{BS/m^2}$.

as given $\lambda_{\mathrm{mac}} = 5 \times 10^{-6}\,\mathrm{BS/m^2}$, $\lambda_{\mathrm{u}} = 4 \times 10^{-4}\,\mathrm{user/m^2}$, and $\mu = 10$. It is clear that increasing R_{mac} causes a degradation of the network's ESE. This makes sense, since the higher required service rate of macrocell users imposes a higher power consumption due to having a large user-to-macro-BS distance. By contrast, increasing R_{mic} improves the network's ESE. In other words, increasing the service rate of small cell users consumes less power. Furthermore, the analytical values and simulation results of the network's ESE agree with each other well. This verifies the accuracy of our ESE modeling.

Under the optimal small BS density $\lambda^{\star}_{\mathrm{mic}}$ as well as given $R_{\mathrm{mac}} = R_{\mathrm{mic}} = 0.1$ Mbit/s and $\lambda_{\mathrm{mac}} = 5 \times 10^{-6}\,\mathrm{BS/m^2}$, the relationships between the network ESE and the load migration factor as well as the mobile-traffic intensity are shown in

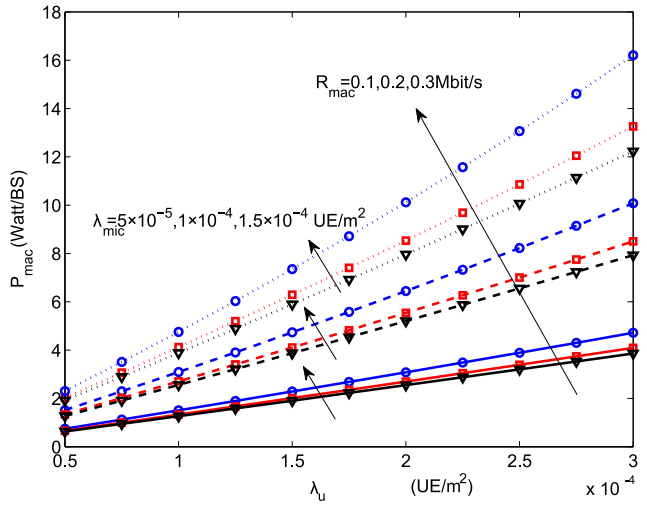


Fig. 6. The impacts of λ_{mic} and R_{mac} on $E\left[P_{\mathrm{mac}}\right]$ for various λ_{u}, given $\mu = 8$, and $\lambda_{\mathrm{mac}} = 5 \times 10^{-6}\,\mathrm{BS/m^2}$.

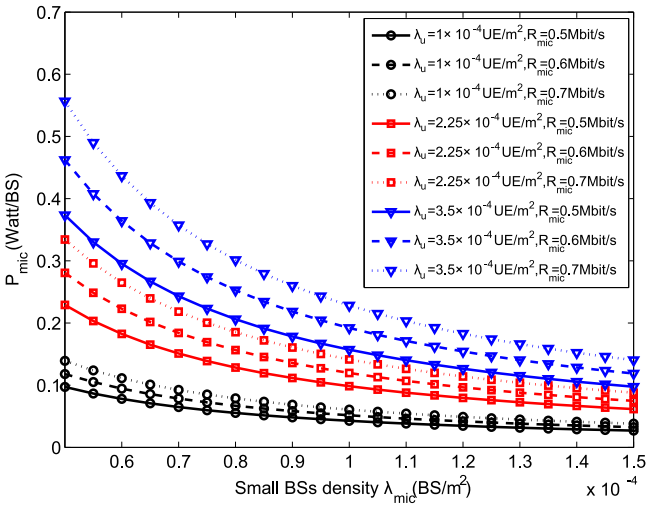


Fig. 7. The impacts of λ_{u} and R_{mic} on $E\left[P_{\mathrm{mic}}\right]$ for various λ_{mic}, given $\mu = 8$ and $\lambda_{\mathrm{mac}} = 5 \times 10^{-6}\,\mathrm{BS/m^2}$.

Fig. 5 for various system bandwidths. Obviously, increasing μ decreases the network's ESE, since more users are associated with macro BSs, which consume more transmit power, while a denser λ_{u} improves the ESE to some extent. Furthermore, the network's ESE decreases as B increases. This is because although the network's APC is reduced for an increased B, the network's ASE is reduced at a faster rate.

The average aggregate DL transmit power of a macro BS $E[P_{\mathrm{mac}}]$ versus the mobile traffic intensity λ_{u}, λ_{mic} and R_{mac} is illustrated in Fig. 6, given $\mu = 8$ and $\lambda_{\mathrm{mac}} = 5 \times 10^{-6}\,\mathrm{BS/m^2}$. As expected, $E[P_{\mathrm{mac}}]$ increases as either λ_{u} or R_{mac} increases. Furthermore, increasing λ_{mic} causes a larger $E[P_{\mathrm{mac}}]$ as a result of the increased interference.

Next the average aggregate DL transmit power of small BS $E[P_{\mathrm{mic}}]$ versus λ_{mic}, λ_{u} and R_{mic} is shown in Fig. 7, given $\mu = 8$ and $\lambda_{\mathrm{mac}} = 5 \times 10^{-6}\,\mathrm{BS/m^2}$. Note that $E[P_{\mathrm{mic}}]$ decreases upon increasing λ_{mic}, which is a result of the reduced small cell size. Furthermore, as expected, increasing R_{mic} also

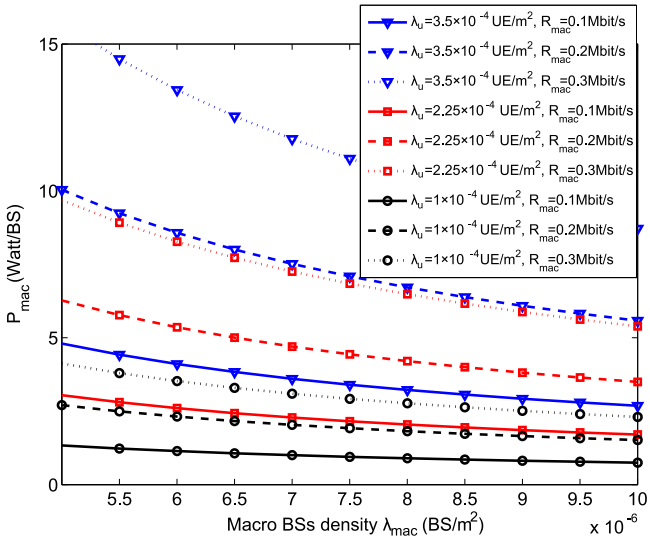


Fig. 8. The impacts of λ_{u} and R_{mac} on $E\left[P_{\mathrm{mac}}\right]$ for various λ_{mac}, given $\mu = 8$ and $\lambda_{\mathrm{mic}} = 1 \times 10^{-4}\,\mathrm{BS/m^2}$.

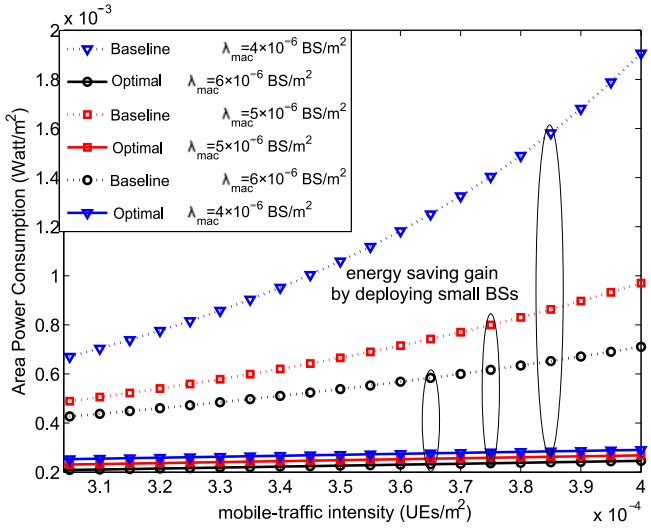


Fig. 9. The network APCs as the functions of λ_{u} and λ_{mac} for the optimal and non-optimal small-BS deployment designs, given $R_{\mathrm{mac}} = R_{\mathrm{mic}} = 0.3$ Mbit/s.

increases $E[P_{\mathrm{mic}}]$. Similarly, the average aggregate DL transmit power of macro BS $E[P_{\mathrm{mac}}]$ versus λ_{mac}, λ_{u} and R_{mac} is depicted in Fig. 8, given $\mu = 8$ and $\lambda_{\mathrm{mic}} = 1 \times 10^{-4}\,\mathrm{BS/m^2}$. Observe that R_{mac} has a very significant influence on $E[P_{\mathrm{mac}}]$.

B. Energy Saving of Optimal Small BS Deployment Strategy

We now evaluate the energy saving gain of the proposed optimal LUB-aware small-BS deployment strategy over the non-optimal baseline design of [39]–[41]. First, the APC performance of the optimal LUB-aware small-BS deployment design is compared to that of the baseline design in Fig. 9. Observe that the proposed optimal LUB-aware small-BS deployment design significantly outperforms the non-optimal baseline design in terms of the network's APC. Moreover, when the mobile traffic intensity λ_{u} is fixed, for the sparser macro BS scenario an increased energy saving gain can be obtained by our optimal

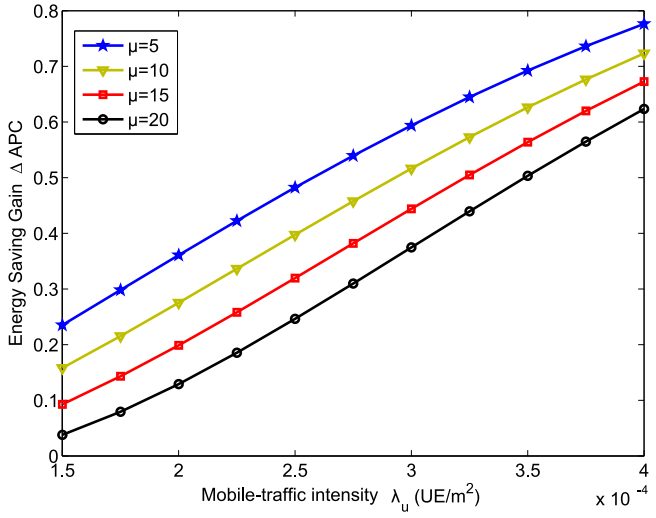


Fig. 10. Energy saving gain attained by the optimal LUBs-aware small-BS deployment strategy over the baseline design as the function of λ_{u} for different μ, given $B = 10$ Mbit/s, $\lambda_{\mathrm{mac}} = 5 \times 10^{-6}\,\mathrm{BS/m^2}$, and $R_{\mathrm{mac}} = R_{\mathrm{mic}} = 0.3$ Mbit/s.

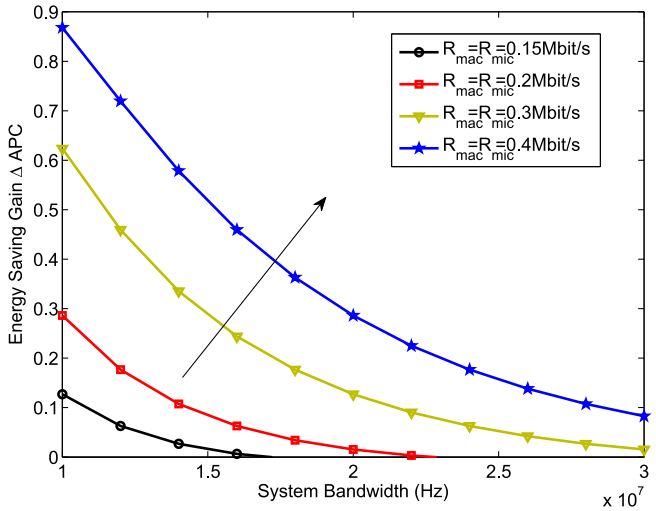


Fig. 11. Energy saving gain attained by the optimal LUBs-aware small-BS deployment strategy over the baseline design as the function of B for various $R_{\mathrm{mac}} = R_{\mathrm{mic}}$ values, given $\lambda_{\mathrm{mac}} = 5 \times 10^{-6}\,\mathrm{BS/m^2}$, $\mu = 20$ and $\lambda_{\mathrm{u}} = 4 \times 10^{-4}\,\mathrm{UE/m^2}$.

LUB-aware small-BS deployment strategy over the baseline design, as clearly seen from Fig. 9. Note that given the fixed λ_{u}, for our optimal design, a reduction in λ_{mac} results in an increase in $\lambda^{\star}_{\mathrm{mic}}$. See Corollary 3. Also when λ_{mac} is fixed, our optimal deployment strategy attains a higher energy saving gain over the baseline design upon increasing λ_{u}.

As illustrated in Fig. 10, increasing the coverage range of small BSs by decreasing the LMF μ can provide increased energy saving gains. This is because decreasing μ is capable of increasing the efficiency of small BSs deployment by associating more users with the micro cell tier. In Fig. 11, we observe the impact of the service data rate on the energy saving gain, when $R_{\mathrm{mac}} = R_{\mathrm{mic}}$. Clearly, the higher the required service rate, the larger the energy saving gain becomes. This implies

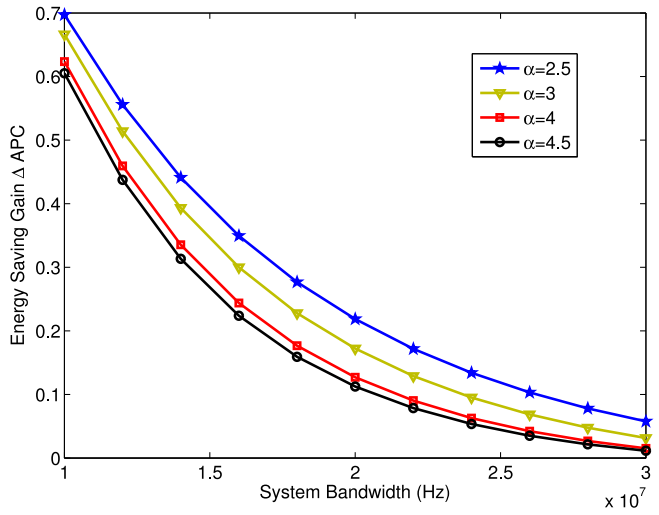


Fig. 12. Energy saving gain attained by the optimal LUBs-aware small-BS deployment strategy over the baseline design as the function of B for various α values, given $\lambda_{\rm mac} = 5 \times 10^{-6}$ BS/m^2, $\mu = 20$, $\lambda_{\rm u} = 4 \times 10^{-4}$ UE/m^2, and $R_{\rm mac} = R_{\rm mic} = 0.3$ Mbit/s.

that a higher service rate requirement will necessitate denser small BS deployment.

Lastly, in Fig. 12 we investigate how the path-loss exponent α impacts on the achievable energy saving gain. It is widely recognized that a smaller α on one hand is capable of decreasing the transmit power required for users, while on the other hand it can also result in a more serious interference. Observe from Fig. 12 that a smaller α results in a higher energy saving gain for our optimal LUB-aware small-BS deployment strategy over the baseline design.

VII. Conclusion

We have proposed an analytical ESE framework for large-scale densely deployed multi-tier HCNs. Specifically, we have integrated the key parameters of HCNs, including the per-tier BS densities, the LMFs and the mobile-traffic intensity, into the network's ESE modeling. Significantly, a closed-form ESE expression has been derived, which enables us to analyze the properties of multi-tier HCNs. Moreover, an optimal LUB-aware ESE BS deployment strategy was conceived for two-tier macro-small HCNs, which maximizes the network's ESE under outage constraints. The accuracy of our ESE modeling has been verified by our simulations results. This study has qualitatively and quantitatively shown the significant energy saving gains achievable by our proposed optimal LUBs-aware BS deployment design conceived for large-scale HCNs, and our results have provided valuable design insights for the ultra-dense HCN of the near future.

Appendix

A. Proof of Proposition 1

Proof: Given typical UE $u_{j,m}$ in position z and associated to BS b_m in tier m, which requests rate $R_{j,m}$ and is allocated with subband $B_{j,m}$, its required DL transmit power conditioning on $r_m = \|z - b_m\|$ and interference topology $(\bigcup_{\widetilde{m}} \Psi_{\widetilde{m}}) \backslash b_m$ is given by

$$E\left[P_{j,m}\middle| r_m, \left(\bigcup\nolimits_{\widetilde{m}} \Psi_{\widetilde{m}}\right)\backslash b_m\right] = \frac{\left(2^{R_{j,m}/B_{j,m}} - 1\right)}{r_m^{-\alpha}} E\left[I_{j,m}\right]$$
$$= \frac{\left(2^{R_{j,m}/B_{j,m}} - 1\right)}{r_m^{-\alpha}} E\left[\sum_{\widetilde{m}=1}^{M} \sum_{b_k \in \Psi_{\widetilde{m}}, b_k \neq b_m} \frac{g_j^k P_{\max}^{\widetilde{m}}}{N_m r_k^{\alpha}}\right]. \quad (43)$$

In deriving (43) from (2) and (3), we have utilized $E[g_j^m] = 1$. The moment generating function (MGF) of $I_{j,m}$ is [34]

$$\mathcal{M}_I(s) = E\left[\exp\left(-\sum_{\widetilde{m}=1}^{M} \sum_{b_k \in \Psi_{\widetilde{m}}, b_k \neq b_m} s \frac{g_j^k P_{\max}^{\widetilde{m}}}{N_m r_k^{\alpha}}\right)\right]$$
$$= E\left[\prod_{\widetilde{m}=1}^{M} \prod_{b_k \in \Psi_{\widetilde{m}}, b_k \neq b_m} \exp\left(-s \frac{g_j^k P_{\max}^{\widetilde{m}}}{N_m r_k^{\alpha}}\right)\right]. \quad (44)$$

Now consider the aggregate interference power I' from the BSs in all the tiers where each tier has a finite region with the inner radius $r_{\widetilde{m}}$ and the outer radius $r_{\rm bn}^{\widetilde{m}} = r_{\rm bn}$ for tier $\widetilde{m}$. We have $I' = \sum_{\widetilde{m}=1}^{M} I'_{\widetilde{m}}$, where $I'_{\widetilde{m}}$ is the interference from tier $\widetilde{m}$. The MGF of $I'_{\widetilde{m}}$ is given by [14], [35]

$$\mathcal{M}_{I'_{\widetilde{m}}}(s) = \exp\left(2\pi\lambda_{\widetilde{m}} \int_{r_{\widetilde{m}}}^{r_{\rm bn}} \left(1 - \exp\left(-\frac{s P_{\max}^{\widetilde{m}}}{N_m t^{\alpha}}\right)\right) t dt\right). \quad (45)$$

Again $E[g_j^k] = 1$ is utilized in deriving (45). Hence, the MGF of I' is given by

$$\mathcal{M}_{I'}(s) = \prod_{\widetilde{m}} \mathcal{M}_{I'_{\widetilde{m}}}(s). \quad (46)$$

According to the well-known property of MGF, we have

$$E\left[I'\right] = -\frac{\partial}{\partial s} \ln \mathcal{M}_{I'}(s)\Big|_{s=0}$$
$$= \sum_{\widetilde{m}} \frac{2\pi\lambda_{\widetilde{m}} P_{\max}^{\widetilde{m}}}{N_m} \frac{r_{\rm bn}^{2-\alpha} - r_{\widetilde{m}}^{2-\alpha}}{(2-\alpha)}, \quad (47)$$

assuming $\alpha > 2$. Clearly, $E[I_{j,m}] = \lim_{r_{\rm bn} \to \infty} E[I']$. By letting $r_{\rm bn} \to \infty$ in (47) and substituting the result into (43) as well as utilizing $\mu_{m \leftarrow \widetilde{m}} = (\frac{r_{\widetilde{m}}}{r_m})^{-\alpha}$ [26], [27], we obtain

$$E\left[P_{j,m}|r_m\right] = \frac{2\pi\lambda_m P_{\max}^m r_m^2 \left(2^{R_{j,m}/B_{j,m}} - 1\right)}{N_m(\alpha - 2)}$$
$$\times \left(\sum_{\widetilde{m}=1}^{M} \frac{\lambda_{\widetilde{m}} P_{\max}^{\widetilde{m}}}{\lambda_m P_{\max}^m} \mu_{m \leftarrow \widetilde{m}}^{(\alpha-2)/\alpha}\right). \quad (48)$$

The user distribution within V_m can be accurately approximated by a uniform distribution[2]. Further approximate cell V_m with area A_m by the circle having area $A_m = \pi R_{\rm ave}^2$, where $R_{\rm ave}$ is the radius of the cell. Since $N_m = A_m \lambda_{\rm u}$,

[2] Although the user distribution in the whole network is modelled by a nonuniform PPP process, the distribution of users in a single cell can accurately be modelled by a uniform distribution.

$R_{\rm ave}=\sqrt{\frac{N_m}{\pi\lambda_{\rm u}}}$. Thus the probability density function (PDF) of the distance x between user $u_{j,m}$ and BS b_m is $f(x)=\frac{2x}{R_{\rm ave}^2}$, and therefore we have

$$E\left[P_{j,m}\right] = \int_0^{R_{\rm ave}} \frac{2r_m E\left[P_{j,m}|r_m\right]}{R_{\rm ave}^2} dr_m = \frac{\lambda_m\left(2^{R_{j,m}/B_{j,m}}-1\right)}{(\alpha-2)\lambda_{\rm u}} \sum_{\widetilde{m}=1}^{M} \frac{\lambda_{\widetilde{m}} P_{\max}^{\widetilde{m}} \mu_{m\leftarrow\widetilde{m}}^{(\alpha-2)/\alpha}}{\lambda_m}. \tag{49}$$

This completes the proof. ■

B. Proof of Proposition 2

Proof: Given the typical cell V_m centered by b_m, the average aggregate DL transmit power conditioned on the cell area of $|V_m|=A_m$ can be expressed as

$$E\left[P_m|A_m\right] = \sum_{u_i\in\Psi^{\rm u}} \mathcal{P}_{u_i}\mathcal{I}\left\{u_i\in\varphi_m\right\} = \sum_{N_m=1}^{\infty}\left(\sum_{j=1}^{N_m} E\left[P_{j,m}\right]\right) g\left(N_m\right), \tag{50}$$

where $\mathcal{P}_{u_i}$ denotes the DL transmit power for user u_i and $\mathcal{I}\{\,\}$ is the indicator function, which is equal to 1 when the condition inside the bracket is satisfied and 0 otherwise, while $g(N_m)$ denotes the PMF of the number of users in V_m, which is given by $g(N_m)=(\lambda_{\rm u}A_m)^{N_m}\exp(-\lambda_{\rm u}A_m)/N_m!$ for $N_m\geq 1$ [36].

Since all the UEs in V_m have the identical rate of $R_{j,m}=R_m$, from (5), we have

$$E\left[P_{j,m}\right]=\frac{P_{\max}^m\left(2^{N_m R_m/B}-1\right)}{(\alpha-2)\lambda_{\rm u}}\sum_{\widetilde{m}=1}^{M}\lambda_{\widetilde{m}}\mu_{m\leftarrow\widetilde{m}}^{-2/\alpha}, \tag{51}$$

where $\mu_{m\leftarrow\widetilde{m}}=P_{\max}^m/P_{\max}^{\widetilde{m}}$ [26], [27] is utilized. Further note that the average value of N_m is

$$E\left[N_m\right]=\frac{\lambda_{\rm u}}{\sum_{\widetilde{m}=1}^{M}\lambda_{\widetilde{m}}\mu_{m\leftarrow\widetilde{m}}^{-2/\alpha}}. \tag{52}$$

Note that $1/\sum_{\widetilde{m}=1}^{M}\lambda_{\widetilde{m}}\mu_{m\leftarrow\widetilde{m}}^{-2/\alpha}$ in (52) is approximately the average coverage area of b_m for the M-tier HCN [26], [27]. Thus the average aggregate DL transmit power conditioned on A_m can be expressed as

$$E\left[P_m|A_m\right]=E\left[N_m E\left[P_{j,m}\right]\right] = \sum_{N_m=1}^{\infty}\frac{P_{\max}^m\left(2^{N_m R_m/B}-1\right)}{(\alpha-2)}\frac{\left(\lambda_{\rm u}A_m\right)^{N_m}}{N_m!\exp\left(\lambda_{\rm u}A_m\right)}. \tag{53}$$

Consequently,

$$E\left[P_m\right]=\int_{A_m>0}E\left[P_m|A_m\right]\mathcal{G}(A_m)\,dA_m, \tag{54}$$

where $\mathcal{G}(A_m)$ is the PDF of A_m. According to [37],

$$\mathcal{G}(A_m)=\frac{K^K\left(\lambda_m+\sum\limits_{\widetilde{m}\neq m}\lambda_{\widetilde{m}}\mu_{m\leftarrow\widetilde{m}}^{-2/\alpha}\right)^K A_m^{K-1}}{\Gamma(k)\exp\left(K\left(\lambda_m+\sum\limits_{\widetilde{m}\neq m}\lambda_{\widetilde{m}}\mu_{m\leftarrow\widetilde{m}}^{-2/\alpha}\right)A_m\right)}, \tag{55}$$

where $K=3.57$ [38], and the Gamma function $\Gamma(x)=\int_0^{+\infty}t^{x-1}\exp(-t)\,dt$. Substituting (53) and (55) into (54) and completing the resulting integration yields (6). The detailed final part of this proof is similar to Appendix C of [14]. ■

C. Proof of Proposition 4

Proof: By taking the derivative of the network-level ESE with respect to λ_m, we obtain

$$\frac{\partial\eta_{\rm ESE}}{\partial\lambda_m}=-\frac{N_\eta}{(D_\eta)^2}\Omega, \tag{56}$$

where N_η and D_η are the numerator and denominator of $\eta_{\rm ESE}$, respectively,

$$\Omega=\frac{1-(1+K)\Upsilon_m}{(1-\Upsilon_m)^{K+1}}+\sum_{\widetilde{m}\neq m}\mu_{m\leftarrow\widetilde{m}}^{\frac{2-\alpha}{\alpha}}\frac{1-(1+K)\Upsilon_{\widetilde{m}}}{(1-\Upsilon_{\widetilde{m}})^{K+1}}-\gamma_m^{(M)}, \tag{57}$$

and $\gamma_m^{(M)}$ is given in (20), while Υ_m and $\Upsilon_{\widetilde{m}}$ are given in (18) and (19), respectively, but with $\lambda_m^\star$ replaced by λ_m. Since $N_\eta>0$ and $D_\eta>0$, the sign of $\frac{\partial\eta_{\rm ESE}}{\partial\lambda_m}$ is equal to the sign of $-\Omega$.

Clearly, from (18), there exists a $\lambda_m^{(1)}$ such that $\Upsilon_m\to 1^-$ as $\lambda_m\to\lambda_m^{(1)+}$. Therefore,

$$\lim_{\lambda_m\to\lambda_m^{(1)+}}\Omega=-\infty\Rightarrow\lim_{\lambda_m\to\lambda_m^{(1)+}}\frac{\partial\eta_{\rm ESE}}{\partial\lambda_m}>0. \tag{58}$$

In fact, from (18), we have

$$\lambda_m^{(1)}=\left(2^{R_m/B}-1\right)\frac{\lambda_{\rm u}}{K}-\sum_{\widetilde{m}\neq m}\lambda_{\widetilde{m}}\mu_{m\leftarrow\widetilde{m}}^{-2/\alpha}. \tag{59}$$

Also $\Upsilon_m\to 0$ and $\Upsilon_{\widetilde{m}}\to 0$ as $\lambda_m\to+\infty$. Therefore, we have

$$\lim_{\lambda_m\to+\infty}\frac{1-(1+K)\Upsilon_m}{(1-\Upsilon_m)^{K+1}}=1, \tag{60}$$

$$\lim_{\lambda_m\to+\infty}\sum_{\widetilde{m}\neq m}\mu_{m\leftarrow\widetilde{m}}^{\frac{2-\alpha}{\alpha}}\frac{1-(1+K)\Upsilon_{\widetilde{m}}}{(1-\Upsilon_{\widetilde{m}})^{K+1}}=\sum_{\widetilde{m}\neq m}\mu_{m\leftarrow\widetilde{m}}^{\frac{2-\alpha}{\alpha}}, \tag{61}$$

which imply that

$$\lim_{\lambda_m\to+\infty}\Omega=\frac{P_{\rm OM}^m}{P_{\max}^m}(\alpha-2)>0\Rightarrow\lim_{\lambda_m\to+\infty}\frac{\partial\eta_{\rm ESE}}{\partial\lambda_m}<0. \tag{62}$$

Next, it can be shown that $\frac{\partial\Omega}{\partial\lambda_m}>0$. This implies that Ω is a monotonically increasing function of λ_m. Therefore, there exists an unique $\lambda_m^\star\in[\lambda_m^{(1)},+\infty)$ which maximizes $\eta_{\rm ESE}$. Setting (57) to zero completes the proof. ■

D. Proof of Corollary 3

Proof: First, we introduce

$$\Theta\left(\lambda_m^{\star},\lambda_{\widetilde{m}}\right)=\underbrace{\frac{1-(1+K)\bar{\Upsilon}_m}{\left(1-\bar{\Upsilon}_m\right)^{K+1}}}_{\delta_1}+\underbrace{\mu_{m\leftarrow\widetilde{m}}^{\frac{2-\alpha}{\alpha}}\frac{1-(1+K)\bar{\Upsilon}_{\widetilde{m}}}{\left(1-\bar{\Upsilon}_{\widetilde{m}}\right)^{K+1}}}_{\delta_2}. \tag{63}$$

The derivative of Θ with respect to $\lambda_m^{\star}$ satisfies

$$\begin{aligned}\frac{\partial\Theta\left(\lambda_m^{\star},\lambda_{\widetilde{m}}\right)}{\partial\lambda_m^{\star}}&=\frac{\partial\delta_1}{\partial\bar{\Upsilon}_m}\frac{\partial\bar{\Upsilon}_m}{\partial\lambda_m^{\star}}+\frac{\partial\delta_2}{\partial\bar{\Upsilon}_{\widetilde{m}}}\frac{\partial\bar{\Upsilon}_{\widetilde{m}}}{\partial\lambda_m^{\star}}\\&=\left(\frac{-K(K+1)\bar{\Upsilon}_m}{\left(1-\bar{\Upsilon}_m\right)^{K+2}}\right)\left(\frac{-\left(2^{R_m/B}-1\right)\lambda_{\mathrm{u}}}{K\left(\lambda_m^{\star}+\lambda_{\widetilde{m}}\mu_{m\leftarrow\widetilde{m}}^{-\frac{2}{\alpha}}\right)^2}\right)\\&\quad+\left(\frac{-K(K+1)\bar{\Upsilon}_{\widetilde{m}}\mu_{m\leftarrow\widetilde{m}}^{\frac{2-\alpha}{\alpha}}}{\left(1-\bar{\Upsilon}_{\widetilde{m}}\right)^{K+2}}\right)\left(\frac{-\mu_{\widetilde{m}\leftarrow m}^{-\frac{2}{\alpha}}\left(2^{R_m/B}-1\right)\lambda_{\mathrm{u}}}{K\left(\lambda_{\widetilde{m}}+\lambda_m^{\star}\mu_{\widetilde{m}\leftarrow m}^{-\frac{2}{\alpha}}\right)^2}\right)>0.\end{aligned} \tag{64}$$

Similarly, we have

$$\begin{aligned}\frac{\partial\Theta\left(\lambda_m^{\star},\lambda_{\widetilde{m}}\right)}{\partial\lambda_{\widetilde{m}}}&=\frac{\partial\delta_1}{\partial\bar{\Upsilon}_m}\frac{\partial\bar{\Upsilon}_m}{\partial\lambda_{\widetilde{m}}}+\frac{\partial\delta_2}{\partial\bar{\Upsilon}_{\widetilde{m}}}\frac{\partial\bar{\Upsilon}_{\widetilde{m}}}{\partial\lambda_{\widetilde{m}}}\\&=\left(\frac{-K(K+1)\bar{\Upsilon}_m}{\left(1-\bar{\Upsilon}_m\right)^{K+2}}\right)\left(\frac{-\left(2^{R_m/B}-1\right)\lambda_{\mathrm{u}}\mu_{m\leftarrow\widetilde{m}}^{-\frac{2}{\alpha}}}{K\left(\lambda_m^{\star}+\lambda_{\widetilde{m}}\mu_{m\leftarrow\widetilde{m}}^{-\frac{2}{\alpha}}\right)^2}\right)\\&\quad+\left(\frac{-K(K+1)\bar{\Upsilon}_{\widetilde{m}}\mu_{m\leftarrow\widetilde{m}}^{\frac{2-\alpha}{\alpha}}}{\left(1-\bar{\Upsilon}_{\widetilde{m}}\right)^{K+2}}\right)\left(\frac{-\left(2^{R_m/B}-1\right)\lambda_{\mathrm{u}}}{K\left(\lambda_{\widetilde{m}}+\lambda_m^{\star}\mu_{\widetilde{m}\leftarrow m}^{-\frac{2}{\alpha}}\right)^2}\right)>0.\end{aligned} \tag{65}$$

The conditions (64) and (65) indicate that $\Theta\left(\lambda_m^{\star},\lambda_{\widetilde{m}}\right)$ is a monotonically increasing function of $\lambda_m^{\star}$ and $\lambda_{\widetilde{m}}$, when the network operates near the optimal ESE status of $\lambda_m^{\star}$. Therefore, to maintain the network at a near optimal ESE status, increasing $\lambda_{\widetilde{m}}$ has to be compensated by reducing $\lambda_m^{\star}$, and vice versa. This completes the proof. ■

E. Proof of Proposition 5

Proof: Firstly, express the network's η_{ESE} (11) as

$$\eta_{\mathrm{ESE}}=\frac{N_\eta}{D_\eta}=\left(\sum_{m=1}^{M}\frac{1}{B}\frac{R_m\lambda_m}{\left(\sum_{\widetilde{m}=1}^{M}\lambda_{\widetilde{m}}\mu_{m\leftarrow\widetilde{m}}^{-2/\alpha}\right)}\right)\frac{1}{D_\eta}. \tag{66}$$

The derivative of η_{ESE} with respect to λ_{u} is

$$\frac{\partial\eta_{\mathrm{ESE}}}{\partial\lambda_{\mathrm{u}}}=-\frac{N_\eta}{D_\eta^2}\frac{\partial D_\eta}{\partial\lambda_{\mathrm{u}}}, \tag{67}$$

where

$$\begin{aligned}\frac{\partial D_\eta}{\partial\lambda_{\mathrm{u}}}&=\frac{1}{\left(\lambda_{\mathrm{u}}\right)^2}\sum\nolimits_{m=1}^{M}\left(\frac{KP_{\max}^m\beta_m\lambda_m\lambda_{\mathrm{u}}S_m}{(\alpha-2)\left(1-S_m\lambda_{\mathrm{u}}\right)^{K+1}}\right.\\&\quad\left.-\frac{P_{\max}^m\beta_m\lambda_m}{(\alpha-2)\left(1-S_m\lambda_{\mathrm{u}}\right)^{K}}+\frac{\lambda_m\beta_mP_{\max}^m}{(\alpha-2)}-\lambda_mP_{\mathrm{OM}}^m\right)\\&=\frac{1}{\left(\lambda_{\mathrm{u}}\right)^2}D^{'}.\end{aligned} \tag{68}$$

Clearly, the sign of $\frac{\partial D_\eta}{\partial\lambda_{\mathrm{u}}}$ is determined by the sign of $D^{'}$, and furthermore

$$\lim_{\lambda_{\mathrm{u}}\to0^+}D^{'}=-\sum_{m=1}^{M}\lambda_mP_{\mathrm{OM}}^m<0, \tag{69}$$

$$\lim_{\lambda_{\mathrm{u}}\to\frac{1}{S_m}^-}D^{\prime}=+\infty+\sum_{m=1}^{M}\left(\frac{\lambda_m\beta_mP_{\max}^m}{(\alpha-2)}-\lambda_mP_{\mathrm{OM}}^m\right)>0. \tag{70}$$

Since the derivative of $D^{'}$ with respect to λ_{u}

$$\frac{\partial D^{'}}{\partial\lambda_{\mathrm{u}}}=\sum_{m=1}^{M}\frac{KP_{\max}^m\beta_m\lambda_m\lambda_{\mathrm{u}}(K+1)S_m^2}{(\alpha-2)\left(1-S_m\lambda_{\mathrm{u}}\right)^{K+2}}>0, \tag{71}$$

$D^{'}$ is a monotonically increasing function of λ_u.

Based on (69) to (71), there must exist a unique $\lambda_{\mathrm{u}\star}\in[0,\frac{1}{S_m})$ to enable $\frac{\partial D_\eta}{\partial\lambda_{\mathrm{u}}}|_{\lambda_{\mathrm{u}\star}}=0$. That is, there always exists the unique optimal $\lambda_{\mathrm{u}\star}\in[0,\frac{1}{S_m})$ as the solution of (26), which maximizes η_{ESE}. ■

F. Proof of Proposition 6

Proof: The outage probability of user $u_{j,m}$ located at z, i.e., conditioned on $r_m=\|z-b_m\|$, is given by

$$Q_{j,m}[r_m]=1-\Pr\left(P_{j,m}\le P_{\max}^m/N_m\,\middle|\,r_m\right). \tag{72}$$

From (2), $P_{j,m}=(2^{R_{j,m}/B_{j,m}}-1)I_{j,m}r_m^{\alpha}/g_j^m$, and hence according to [23] we have

$$Q_{j,m}[r_m]=1-\left.\mathcal{L}_{I_{j,m}}(s)\right|_{s=\frac{N_m}{P_{\max}^m}T_{j,m}r_m^{\alpha}}, \tag{73}$$

where $\mathcal{L}_{I_{j,m}}(s)$ is the Laplace transform of $I_{j,m}$, and $T_{j,m}=\left(2^{R_{j,m}/B_{j,m}}-1\right)$. From (3),

$$I_{j,m}=\sum_{\widetilde{m}}\sum_{b_k\in\Psi_{\widetilde{m}},b_k\neq b_m}P_{\max}^{\widetilde{m}}g_j^kr_k^{\alpha}/N_m, \tag{74}$$

in which $\frac{\widetilde{P_{\max}^m}}{N_m}g_j^k\sim\exp\left(N_m/P_{\max}^{\widetilde{m}}\right)$. Therefore, according to [23, Appendix B], we have

$$\mathcal{L}_{I_{j,m}}\left(\frac{N_m}{P_{\max}^m}T_{j,m}r_m^{\alpha}\right)=\prod_{\widetilde{m}}\exp\left(-\pi\mu_{m\leftarrow\widetilde{m}}^{-2/\alpha}r_m^2\lambda_{\widetilde{m}}\rho_{j,m}\right), \tag{75}$$

where $\rho_{j,m}$ is defined in (29). Noting that the PDF of r is $f(r) = \frac{2r}{R_{\text{ave}}^2}$, $0 \le r \le R_{\text{ave}}$, with $R_{\text{ave}}^2 = \frac{N_m}{\pi \lambda_{\text{u}}}$, we have

$$Q_{j,m} = 1 - \int_0^{R_{\text{ave}}} \frac{2x}{R_{\text{ave}}^2} \exp\left(-\pi \sum_{\widetilde{m}=1}^{M} \mu_{m \leftarrow \widetilde{m}}^{-2/\alpha} \lambda_{\widetilde{m}} \rho_{j,m} x^2\right) dx. \tag{76}$$

Completing the integration (76) leads to (28). ■

REFERENCES

[1] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[2] J. G. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. C. Reed, "Femtocells: Past, present, and future," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 497–508, Apr. 2012.

[3] A. P. Bianzino, C. Chaudet, D. Rossi, and J.-L. Rougier, "A survey of green networking research," *IEEE Commun. Surveys Tuts.*, vol. 14, no. 1, pp. 3–20, 2012.

[4] C. Han *et al.*, "Green radio: Radio techniques to enable energy-efficient wireless networks," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 46–54, Jun. 2011.

[5] Z. Niu, Y. Wu, J. Gong, and Z. Yang, "Cell zooming for cost-efficient green cellular networks," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 74–79, Nov. 2010.

[6] N. Saxena, B. J. R. Sahu, and Y. S. Han, "Traffic-aware energy optimization in green LTE cellular systems," *IEEE Commun. Lett.*, vol. 18, no. 1, pp. 38–41, Jan. 2014.

[7] E. Oh, K. Son, and B. Krishnamachari, "Dynamic base station switching-on/off strategies for green cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2126–2136, May 2013.

[8] K. T. K. Cheung, S. Yang, and L. Hanzo, "Achieving maximum energy-efficiency in multi-relay OFDMA cellular networks: A fractional programming approach," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 2746–2757, Jul. 2013.

[9] K. T. K. Cheung, S. Yang, and L. Hanzo, "Spectral and energy spectral efficiency optimization of joint transmit and receive beamforming based multi-relay MIMO-OFDMA cellular networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 11, pp. 6147–6165, Nov. 2014.

[10] W. Jing, Z. Lu, X. Wen, Z. Hu, and S. Yang, "Flexible resource allocation for joint optimization of energy and spectral efficiency in OFDMA multi-cell networks," *IEEE Commun. Lett.*, vol. 19, no. 3, pp. 451–454, Mar. 2015.

[11] K. T. K. Cheung, S. Yang, and L. Hanzo, "Distributed energy spectral efficiency optimization for partial/full interference alignment in multi-user multi-relay multi-cell MIMO systems," *IEEE Trans. Signal Process.*, vol. 64, no. 4, pp. 882–896, Feb. 2016.

[12] T. Abräo *et al.*, "Energy efficient OFDMA networks maintaining statistical QoS guarantees for delay-sensitive traffic," *IEEE Access*, vol. 4, pp. 774–791, 2016.

[13] D. Astely, E. Dahlman, G. Fodor, S. Parkvall, and J. Sachs, "LTE release 12 and beyond," *IEEE Commun. Mag.*, vol. 51, no. 7, pp. 154–160, Jul. 2013.

[14] G. Zhao, S. Chen, L. Zhao, and L. Hanzo, "A tele-traffic-aware optimal base-station deployment strategy for energy-efficient large-scale cellular networks," *IEEE Access*, vol. 4, pp. 2083–2095, 2016.

[15] K. Son, H. Kim, Y. Yi, and B. Krishnamachari, "Base station operation and user association mechanisms for energy-delay tradeoffs in green cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1525–1536, Sep. 2011.

[16] L. Saker, S. E. Elayoubi, R. Combes, and T. Chahed, "Optimal control of wake up mechanisms of femtocells in heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 664–672, Apr. 2012.

[17] Z. Yang and Z. Niu, "Energy saving in cellular networks by dynamic RS-BS association and BS switching," *IEEE Trans. Veh. Technol.*, vol. 62, no. 9, pp. 4602–4614, Nov. 2013.

[18] L. Zhao, G. Zhao, and T. O'Farrell, "Efficiency metrics for wireless communications," in *Proc. Int. Symp. Pers., Indoor, Mobile Radio Commun.*, London, U.K., Sep. 8–11, 2013, pp. 2825–2829.

[19] M. M. A. Hossain, K. Koufos, and R. Jäntti, "Energy efficient deployment of HetNets: Impact of power amplifier and delay," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Shanghai, China, Apr. 7–10, 2013, pp. 778–782.

[20] W. Guo and T. O'Farrell, "Dynamic cell expansion with self-organizing cooperation," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 5, pp. 851–860, May 2013.

[21] H. Claussen, L. T. W. Ho, and F. Pivit, "Effects of joint macrocell and residential picocell deployment on the network energy efficiency," in *Proc. Int. Symp. Pers., Indoor, Mobile Radio Commun.*, Cannes, France, Sep. 15–18, 2008, pp. 1–6.

[22] F. Richter, A. J. Fehske, and G. P. Fettweis, "Energy efficiency aspects of base station deployment strategies for cellular networks," in *Proc. IEEE 70th Veh. Technol. Conf.*, Anchorage, AK, USA, Sep. 20–23, 2009, pp. 1–5.

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

[24] H. Zhang, S. Chen, L. Feng, Y. Xie, and L. Hanzo, "A universal approach to coverage probability and throughput analysis for cellular networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 9, pp. 4245–4256, Sep. 2015.

[25] H. S. Dhillon, R. K. 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.

[26] 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.

[27] S. Singh, H. S. 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.

[28] X. Ge, J. Ye, Y. Yang, and Q. Li, "User mobility evaluation for 5G small cell networks based on individual mobility model," *IEEE J. Sel. Area Commun.*, vol. 34, no. 3, pp. 528–541, Mar. 2016.

[29] J. Peng, P. Hong, and K. Xue, "Energy-aware cellular deployment strategy under coverage performance constraints," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 69–80, Jan. 2015.

[30] S.-R. Cho and W. Choi, "Energy-efficient repulsive cell activation for heterogeneous cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 5, pp. 870–882, May 2013.

[31] Y. Huang *et al.*, "Energy-efficient design in heterogeneous cellular networks based on large-scale user behavior constraints," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 4746–4757, Sep. 2014.

[32] Y. S. Soh, T. Q. S. Quek, M. Kountouris, and H. Shin, "Energy efficient heterogeneous cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 5, pp. 840–850, May 2013.

[33] X. Ge *et al.*, "Spatial spectrum and energy efficiency of random cellular networks," *IEEE Trans. Commun.*, vol. 63, no. 3, pp. 1019–1030, Mar. 2015.

[34] G. Grimmett and D. Stirzaker, *Probability and Random Processes*, 3rd ed. Oxford, U.K.: Cambridge Univ. Press, 2004.

[35] S. Srinivasa, "Modeling interference in uniformly random wireless networks: Theory and applications," M.Sc. thesis, University of Notre Dame, Notre Dame, IN, USA, 2007.

[36] R. L. Streit, *Poisson Point Processes: Imaging, Tracking and Sensing*. London, U.K.: Springer, 2010.

[37] J. Ferenc and Z. Néda, "On the size distribution of Poisson Voronoi cells," *Physica A: Stat. Mech. Appl.*, vol. 385, no. 2, pp. 518–526, 2007.

[38] S. Lee and K. Huang, "Coverage and economy of cellular networks with many base stations," *IEEE Commun. Lett.*, vol. 16, no. 7, pp. 1038–1040, Jul. 2012.

[39] W. Guo *et al.*, "Automated small-cell deployment for heterogeneous cellular networks," *IEEE Commun. Mag.*, vol. 51, no. 5, pp. 46–53, May 2013.

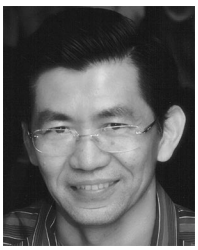
[40] Y. Zhou *et al.*, "Large-scale spatial distribution identification of base stations in cellular networks," *IEEE Access*, vol. 3, pp. 2987–2999, 2015.

[41] S. Tombaz, A. Vastberg, and J. Zander, "Energy- and cost-efficient ultra-high-capacity wireless access," *IEEE Wireless Commun.*, vol. 18, no. 5, pp. 18–24, Oct. 2011.

[42] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, "Green cellular networks: A survey, some research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 4, pp. 524–540, 4th Quarter 2011.

[43] V. Mancuso and S. Alouf, "Reducing costs and pollution in cellular networks," *IEEE Commun. Mag.*, vol. 49, no. 8, pp. 63–71, Aug. 2011.

Guogang Zhao received the B.S. degree in communications engineering from People's Liberation Army Information Engineering University, Zhengzhou, China. He is currently working toward the Ph.D. degree at Department of Telecommunication Engineering, Xidian University, Xi'an, China. Since 2015, he has been a visiting Ph.D. student with Prof. Hanzo and Prof. Chen at the School of Electronics and Computer Science, University of Southampton, Southampton, U.K. His research interests focus on 5G wireless communication, network planning, and ultra-dense cellular network.



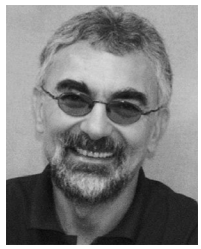
Sheng Chen (M'90–SM'97–F'08) received the B.Eng. degree in control engineering from the East China Petroleum Institute, Dongying, China, in 1982, the Ph.D. degree in control engineering from the City University, London, U.K., in 1986, and the higher doctoral degree, Doctor of Sciences (D.Sc.), from the University of Southampton, Southampton, U.K., in 2005.

From 1986 to 1999, he held research and academic appointments at the University of Sheffield, University of Edinburgh, and University of Portsmouth, all in U.K. Since 1999, he has been with the School of Electronics and Computer Science, University of Southampton, where he holds the post of Professor in Intelligent Systems and Signal Processing. His research interests include adaptive signal processing, wireless communications, modeling and identification of nonlinear systems, neural network and machine learning, intelligent control system design, evolutionary computation methods and optimisation. He has published over 550 research papers.

Dr. Chen is a Fellow of the United Kingdom Royal Academy of Engineering, a Fellow of IET, a Distinguished Adjunct Professor at King Abdulaziz University, Jeddah, Saudi Arabia, and an ISI highly cited researcher in engineering (March 2004).

Liqiang Zhao received the B.Eng. degree in electrical engineering from Shanghai Jiaotong University, Shanghai, China, in 1992, the M.Sc. degree in communications and information systems, and the Ph.D. degree in information and communications engineering from Xidian University, Xi'an, China, in 2000 and 2003, respectively. In 2008, he was awarded by the Program for New Century Excellent Talents in University, Ministry of Education, China.

From 1992 to 2005, he was a Senior Research Engineer with the 20th Research Institute, Chinese Electronics Technology Group Corporation, China, where his research focused on mobile communication systems and spread spectrum communications. From 2005 to 2007, he was an Associate Professor with State Key Laboratory of Integrated Service Networks, Xidian University, where his research focused on WiMAX, WLAN, and wireless sensor network. He was appointed as a Marie Curie Research Fellow at the Centre for Wireless Network Design, University of Bedfordshire in June 2007 to conduct research in the GAWIND project funded under EU FP6 HRM programme, where his activities focused on the area of automatic wireless broadband access network planning and optimization. Since June 2008, he has been with Xidian University. His current research focuses on broadband wireless communications and space communications. He has more than 70 publications in authorized academic periodicals both in China and abroad and in international science conferences, wherein 20 of which are retrieved in SCI, and more than 50 of them are EI indexed. He has also applied for 5 national invention patents. He has hosted/participated many vertical research projects, such as the National Natural Science Foundation, 863 Program and the National Science and Technology Major Projects, China, as well as many horizontal scientific research projects, including the EU FP6, FP7 Plans for International Cooperation and Exchange Projects, and the Huawei Fund.



Lajos Hanzo (F'04) received the Master degree in electronics in 1976, the Ph.D. degree in 1983. In 2009, he was awarded the honorary doctorate, Doctor Honoris Causa, by the Technical University of Budapest. He is a Fellow of the Royal Academy of Engineering, the IEEE, the Institution of Engineering and Technology, and the EURASIP.

During his 38-year career in telecommunications, he has held various research and academic posts in Hungary, Germany and the U.K. Since 1986, he has been with the School of Electronics and Computer Science, University of Southampton, U.K., where he holds the chair in telecommunications. He has successfully supervised about 112 Ph.D. students, coauthored 18 John Wiley/IEEE Press books on mobile radio communications totalling in excess of 10 000 pages, published 1703 research contributions at IEEE Xplore, acted both as TPC and General Chair of IEEE conferences, presented keynote lectures and has been awarded a number of distinctions. He is currently directing a 50-strong academic research team, working on a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council, U.K., the European Research Council's Advanced Fellow Grant, and the Royal Society's Wolfson Research Merit Award. He is an enthusiastic supporter of industrial and academic liaison and he offers a range of industrial courses. He is also a Governor of the IEEE VTS. From 2008 to 2012, he was the Editor-in-Chief of the IEEE Press and a Chaired Professor at Tsinghua University, China.