

A Tele-traffic-Aware Optimal Base-Station Deployment Strategy for Energy-Efficient Large-Scale Cellular Networks

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

Abstract—With the explosive proliferation of mobile devices and services, the user-density of large-scale cellular networks continues to increase, which results in further escalating traffic-generation. Yet, there is an urging demand for reducing the network’s energy dissipation. Hence we model the energy efficiency of large-scale cellular networks for characterizing its dependence on the base station (BS) density as well as for quantifying the impact of tele-traffic on the achievable energy efficiency under specific quality of service requirements. This allows us to match the BS deployment to the network’s tele-traffic, whilst conserving precious energy. More specifically, we formulate a practical tele-traffic-aware BS deployment problem for optimizing the network’s energy efficiency whilst satisfying the users’ maximum tolerable outage probability. This is achieved by analyzing the optimal BS-density under specific tele-traffic conditions. Furthermore, we study the energy saving potential of our optimal BS deployment strategy under diverse practical parameters and provide insights into the attainable energy savings in dense random cellular networks. Our simulation results confirm the accuracy of our analysis and verify the impact of the parameters considered on the network’s energy efficiency. Our results also demonstrate that the proposed tele-traffic-aware optimal BS deployment strategy significantly outperforms the existing approaches in terms of its energy efficiency.

Index Terms—Large-scale cellular networks, energy efficiency, tele-traffic-aware deployment, energy saving performance

I. INTRODUCTION

THE provision of high-speed wireless services for ubiquitous mobile devices has become a societal need in the context of next-generation cellular networks, because the predominant types of tele-traffic have been shifted from mobile voice to mobile data, such as mobile video and online gaming [1]. Hence mobile operators around the world are busily deploying soaring number of base stations (BSs) to meet the ever-increasing traffic demands [2].

G. Zhao and L. Zhao (G.Zhao@soton.ac.uk, lqzhao@mail.xidian.edu.cn) are with State Key Laboratory of Integrated Service Networks, Xidian University, Xian 710071, China.

S. Chen and L. Hanzo (sqc@ecs.soton.ac.uk, lh@ecs.soton.ac.uk) are with Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK. S. Chen is also with King Abdulaziz University, Jeddah 21589, Saudi Arabia.

This work was supported in part by National Natural Science Foundation of China (No. 61372070), Hong Kong, Macao and Taiwan Science and Technology Cooperation Special Project (Nos. 2014DFT10320), China Scholarship Council (No. 201506960042). The financial support of the EPSRC projects EP/N004558/1 and EP/L018659/1, as well as of the European Research Council’s Advanced Fellow Grant under the Beam-Me-Up project and of the Royal Society’s Wolfson Research Merit Award is gratefully acknowledged.

A. Motivation with Related Works

Increasing the spatial reuse of spectral resources by reducing the cell size has become the most influential factor in terms of increasing the attainable system capacity, which is an explicit benefit of reducing the pathloss [3]. Naturally, in the resultant predominantly line-of-sight (LOS) scenarios the fading-distribution also becomes more benevolent. However, the tremendous increase in the number of heterogeneous cells has become a major concern from both an environmental and economic perspective - especially in the light of the escalating energy-prices [4]–[6]. The fifth-generation (5G) mobile network of the near future is expected to increase the tele-traffic capacity, whilst simultaneously increasing both the end-user experience as well as the energy efficiency [7], [8].

Hence a substantial number of investigations have been focused on improving the network’s energy efficiency [9], bearing in mind that nearly 75 % of the energy consumption is imposed by the BSs. Furthermore, within a BS, about 70 % of the energy is consumed by the power amplifiers and the air conditioners [8], [10]. Therefore, network’s energy efficiency mainly depends on the BS density and how efficiently the deployed BSs operate. The BS allocation must be able to meet the maximum tele-traffic demand. However, during low tele-traffic periods at night, weekends or holidays, many BSs are under-utilized, but they still consume energy for the sake of maintaining adequate cellular coverage. Hence, it is desirable to exploit any energy saving potential arising from both the temporal and geographic fluctuations of tele-traffic.

Not surprisingly, a range of energy-efficient BS operation techniques have been conceived for exploiting the time-variant tele-traffic demands [11]–[14]. For instance, Guo and O’Farell [14] conceived a cell-expansion technique for carrying dynamically fluctuating traffic loads based on the innovative concept of cell-coordination, which switches off low-load BSs and compensates for the resultant coverage loss by expanding the coverage of the neighboring cells. By analyzing the BS power consumption profile, Holtkamp *et al.* [13] proposed a radio resource management algorithm for minimizing the BS’s power consumption in multi-user multiple-input multiple-output systems. However, we note that almost all the existing contributions considered the BS operations in isolation, in the context of a single cell or at most a few cells. This is because the optimization procedures behind these existing contributions are based on localized system performance measures, which cannot be readily extended to large-scale cellular networks. A major impediment in this context is the potentially

excessive computational complexity in evaluating the overall performance of a large-scale cellular network. Traditionally, researchers in this area have conducted system performance evaluations in relying on a circular or hexagonal grid based structure [15], [16]. These studies often ignored the large-scale tele-traffic variations, and the techniques proposed in these works cannot be readily applied for the analysis and optimization of the emerging 5G networks relying on realistic randomly deployed BSs. Hence the analysis and evaluation of randomly deployed large-scale cellular networks [17], [18] is of particular importance in practice. As a benefit of its mathematical tractability, the distribution of BSs is typically modeled by a Poisson point process (PPP) [17]–[24]. However, these contributions have mainly focused their attention on user-centric performance analysis, such as the coverage-quality and the users' average throughput, but they do not analyze network's energy efficiency.

Nonetheless, there are a few studies addressing the challenging problem of the network's energy efficiency in dense networks relying on randomly distributed BSs [25]–[29]. Specifically, Soh *et al.* [25] proposed a BS deployment strategy by minimizing the BS density subject to a coverage probability or outage constraint. Peng *et al.* [26] proposed to minimize a so-called network area power consumption (APC) metric subject to a coverage probability constraint. However, by assuming a constant power consumption for each BS, the APC defined in [26] remains independent of the network's tele-traffic load. Hence minimizing this APC metric is equivalent to minimizing the BS density. Cho and Choi [27] minimized the APC metric subject to an average user-rate requirement. Since a constant BS transmit power is assumed in [27] and the average user-rate requirement is equivalent to a coverage probability requirement, the BS placement strategy of [27] is also equivalent to minimizing the BS density subject to maintaining the minimum required coverage probability. The investigations of Deng *et al.* [28] were focused on the coverage performance analysis, but they also characterized the network's energy efficiency. However, their results are based on a constant power consumption for each BS, which implies that the network's APC is directly determined by the BS density. Similarly, Huang *et al.* [29] proposed a BS assignment strategy by maximizing a network energy efficiency metric, which is also defined based on the assumption of a constant BS power consumption and it is independent of the network's tele-traffic. In a nutshell, the optimization criteria used in these existing studies are all based on a constant power consumption assumption for each BS and hence do not take into account the dynamically fluctuation nature of the network's tele-traffic. Consequently, the deployment strategies proposed in [25]–[29] can all be formulated similarly to the one that minimizes the BS density subject to an outage probability constraint.

Indeed, accommodating the network's tele-traffic fluctuation, whilst maximizing the network's energy efficiency is challenging and has not been exploited in the open literature.

B. Our Contributions

Given the above motivation, our goal is to conceive a tele-traffic-aware and energy efficient BS allocation strategy

for large-scale cellular networks. Against this backdrop, our contributions are summarized as follows.

1) *Energy efficiency modeling for randomly deployed large-scale cellular networks:* We first embark on modeling the BS's downlink (DL) power consumption in a dense large-scale cellular network, where all the BSs are randomly distributed based on a PPP and the system's bandwidth is equally allocated by a round-robin resource scheduler to each BS **while ignoring the network's energy efficiency benefits provided by the resource scheduler**. In contrast to most of the existing investigations [17]–[29] which stipulate the idealized simplifying assumption of having a constant DL BS power consumption, the quantitative impact of having the network's time-varying tele-traffic load is characterized statistically. The network's energy efficiency is defined and derived, with the objective of providing insights into how the geographic tele-traffic intensity, spatially averaged data rate requirement, BS density and other salient network parameters influence the network's energy efficiency, which makes it possible to achieve energy savings with the aid of our tele-traffic-aware approach from a large-scale network optimization perspective. Our simulation and numerical results confirm the accuracy of our analytical expressions and verify the impact of various network parameters on the attainable energy efficiency.

2) *Tractable energy efficiency analysis of large-scale cellular networks:* With the aid of this proposed energy efficiency modeling, we further derive an approximate closed-form energy efficiency expression. In particular, we find that both the BS-density and geographical tele-traffic intensity can be integrated into a unique independent variable in this energy efficiency expression. Based on this closed-form energy efficiency expression an optimum condition is established, which provides an effective analytical tool to study the relationship between the property of optimal solution and various network parameters. Specifically, we can study how the network parameters influence energy saving performance gain. For instance, we conclude that denser tele-traffic intensity and higher users' spatially averaged service rate requirement can lead to greater energy saving performance gain. The closed-form nature of this energy efficiency expression makes the search of network energy-efficiency-based optimal solutions particularly tractable and efficient, and this makes designing tele-traffic-aware optimal BS deployment strategy achievable for large-scale cellular networks.

3) *Design strategy for large-scale cellular networks:* With aid of the above-mentioned tractable energy efficiency analysis tool, we design a realistic tele-traffic-aware optimal BS deployment strategy for practical large-scale cellular networks, which carefully matches the geographical tele-traffic distribution to the BS density and thus maximizes the network's energy efficiency while satisfying the maximum tolerable outage probability constraint. We demonstrate that our proposed tele-traffic-aware optimal BS deployment strategy significantly outperforms the existing state-of-the-art energy efficient BS deployment strategies of [25]–[29]. An additional advantage of our design strategy is its computational efficiency, since our tele-traffic-aware optimal BS density solution under the given outage constraint is readily characterized in a closed-form.

The paper is structured as follows. In Section II, our system model is introduced. Then the energy efficiency of large-scale cellular networks is characterized in Section III. In Section IV, the optimal tele-traffic-aware BS-density is analyzed and a BS deployment strategy is derived. Our numerical and simulation results are presented in Section V, whilst our conclusions are summarized in Section VI.

II. SYSTEM MODEL

A. Network Deployment

A large number of BSs denoted by the set $\Psi_b = \{b_i, i = 0, 1, 2, \dots\}$ are randomly allocated in the Euclidean plane \mathbb{R}^2 according to a homogeneous PPP having a density of λ_b . The set of BSs Ψ_b forms the Voronoi tessellation model in \mathbb{R}^2 [17], [18], where the cell V_i is the coverage area of BS b_i . Similarly, the user equipments (UEs) hosted by the set $\Psi_u = \{u_k, k = 0, 1, 2, \dots\}$ are spatially distributed in \mathbb{R}^2 according to an independent homogeneous PPP with a density of λ_u . The serving BS of a UE u_k is the geographically nearest BS. Thus, in this network model, the distributions or locations of BSs and UEs follow a pair of PPPs having densities of λ_b and λ_u , respectively. Since the geographical tele-traffic intensity is proportional to the UEs' spatial density λ_u given the average user rate requirement, we will use λ_u for equivalently representing the tele-traffic intensity.

We may characterize a network scenario by choosing an appropriate BS density and a suitable tele-traffic intensity. For example, a BS density of $\lambda_b = 5 \times 10^{-6}$ BS/m² and a tele-traffic intensity of up to $\lambda_u = 4 \times 10^{-4}$ users/m² represent the network with an average cell radius of 250 m and average 80 UEs per cell. Generally, $\lambda_u/\lambda_b \leq 80$ is ensured by taking into account the total number of available resource blocks (RBs) in each BS. Further assuming $\lambda_b \geq 2 \times 10^{-6}$ BS/m², we have an average cell radius no larger than 400 m, **which corresponds to state-of-the-art small-cell scenarios**. Another advantage of employing a PPP network model is that we can switch BSs on/off statistically, since the superposition of independent PPPs and the independent thinning of a PPP still results in a PPP [27].

B. Resource Partitioning Modeling in BSs

Denote the set of UEs in an arbitrary cell V_i by $\Psi_{u,i}$ with the cardinality of $|\Psi_{u,i}| = N_i \leq N_{\max}$, where N_{\max} is the maximum number of users that can be served in a cell. Orthogonal frequency-division multiple access (OFDMA) is adopted in conjunction with unity frequency reuse (UFR). The total system bandwidth of B Hz is divided equally into N_i subbands, each having a bandwidth of B/N_i Hz. We invoke the classic round-robin scheduling for the sake of maintaining the maximum fairness. Any UE $u_{i,j} \in \Psi_{u,i}$ is associated with a specific subband B_n , where $n \in \{0, 1, \dots, N_i - 1\}$. Each UE $u_{i,j}$ has a service rate requirement denoted by $R_{i,j}$.

To avoid an excessive transmit power, power control is adopted in the BS scheduler by exploiting the long-term pathloss based channel state information (CSI). In other words, the power control does not consider the influence of the small-scale fading channel and, therefore, the instantaneous

CSI is not required. The subband B_n associated with the rate requirement $R_{i,j}$ of $u_{i,j}$ is assigned the right amount of transmit power $P_{i,j}$, specifically,

$$P_{i,j} = \min \{P_{i,j}^o, P_{\max}/N_i\}, \quad (1)$$

where $P_{i,j}^o$ is the transmit power necessitated by meeting the rate requirement $R_{i,j}$ and P_{\max} denotes the BS's maximum transmit power in total. The DL transmit power of BS b_i is given by the sum of the powers allocated to all the subbands

$$P_{cell}^i = \sum_{u_{i,j} \in \Psi_{u,i}} P_{i,j}. \quad (2)$$

In practice, the BS's transmit power is a random variable depending on many factors, but it is predominately influenced by the tele-traffic load. Our formulation (1) and (2) explicitly takes into account the impact of the tele-traffic intensity λ_u . This is in contrast to the existing contributions [27], [28], which assume that achieving energy saving is simply equivalent to minimizing the number of BSs and, therefore, they do not take the effects of tele-traffic into account.

Each BS supports the access of each user under its coverage area with the aid of its link-support decision unit. More specifically, the BS makes its access decisions based on the current CSI and on the quality of service (QoS) requirements of the requesting users.

C. Radio Channel Modeling

The DL is considered where both the large-scale path-loss and the small-scale fast fading are taken into account. Specifically, the path-loss is modeled by

$$L_\alpha(r) = cr^{-\alpha}, \quad (3)$$

where $c > 0$ is a constant, α is the pathloss exponent, and r is the distance between the UE and BS considered. Correspondingly, Rayleigh distributed fast-fading is assumed. The aggregate interference at UE $u_{i,j}$ is given by the total signal power received from the BSs of the set $\Psi_b \setminus b_i$. We focus our attention on the worst-case scenario, where the interfering BSs in the set $\Psi_b \setminus b_i$ all transmit at the maximum power P_{\max} . Unless stated otherwise, the effect of the channel noise is ignored, since the interference power experienced by a UE is far higher than the noise power in a practical interference-limited scenario with UFR. For any user $u_{i,j} \in \Psi_{u,i}$ in cell V_i having a bandwidth of B_n , the received desired signal power can be expressed as

$$P_{i,j}^{Rx} = L_\alpha(\|u_{i,j} - b_i\|) P_{i,j} h_{i,j}^i, \quad (4)$$

while its corresponding received interference power is given by

$$I_{\Psi_b \setminus b_i} = \sum_{b_k \in \Psi_b \setminus b_i} L_\alpha(\|u_{i,j} - b_k\|) \frac{P_{\max}}{N_i} h_{k,j}^i, \quad (5)$$

where $\|u_{i,j} - b_k\|$ denotes the distance between UE $u_{i,j}$ and BS b_k , and $h_{k,j}^i$ represents the gain of the fast-fading channel from BS b_k to UE $u_{i,j}$, which is assumed to follow the independent exponential distribution with unity mean, i.e., $h_{k,j}^i \sim \exp(1)$.

Table I: List of Notations

Notation	Definition
\mathbb{R}^2	Euclidean plane
Ψ_b, Ψ_u	Sets of BSs and UEs
λ_b, λ_u	Densities of BSs and UEs
V_i	Cellular cell covered by BS b_i
$\Psi_{u,i}, N_i$	Set of UEs in cell V_i and size of $\Psi_{u,i}$
N_{\max}	Maximum number of users that can be served in a cell
$u_{i,j}$	j -th UE in V_i
$R_{i,j}$	Required data rate of UE $u_{i,j}$
$P_{i,j}$	Transmit power of UE $u_{i,j}$
P_{\max}	Maximum BS transmit power
$P_{i,j}^o$	Required transmit power for ensuring $R_{i,j}$
B	Total bandwidth resource available to a BS
B_n	subband bandwidth
α	Pathloss exponent
η_{EE}	Network-level energy efficiency metric
$h_{k,j}^k$	Fast-fading channel gain between BS b_k and UE $u_{i,j}$
$Q_{i,j}^{out}$	Outage probability
A_i	Coverage area of V_i
$P_{cell}^{A_i}$	Aggregate DL transmit power of b_i with cell size A_i
P_{cell}^i	Aggregate DL transmit power of b_i
β	Power amplifier efficiency
P_{OM}	BS non-transmission power
ε_{out}	Outage threshold

D. Performance Metrics

We first define the network-level energy efficiency metric as

$$\eta_{EE} = \frac{\text{Area Spectral Efficiency}}{\text{Area Power Consumption}}, \quad (6)$$

where the area spectral efficiency (ASE) refers to the sum of the users' service rates per unit area per Hz, while the APC is the sum of the power consumption per unit area. When deriving the energy efficiency expression, we can adopt the general power consumption model for each BS specified by the standardization bodies [5]. Next we define the outage probability

$$Q_{i,j}^{out} = \Pr(P_{i,j}^o > P_{\max}/N_i), \forall u_{i,j} \in \Psi_{u,i}, b_i \in \Psi_b, \quad (7)$$

which is the probability that the service rate of user $u_{i,j}$ cannot be guaranteed even under the maximum BS power P_{\max}/N_i of the current subband. In this paper, our main attention is the network energy efficiency evaluation and, therefore, the users' outage probabilities are used as constraints.

Table I summarized the notations used throughout our discussions.

III. ENERGY EFFICIENCY OF LARGE-SCALE CELLULAR NETWORKS

In this section, we characterize the energy efficiency of large-scale cellular networks as a function of the average transmit power of a typical user and that of the aggregate transmit power in a typical cell. This will allow us to establish the theoretical foundations for the optimal tele-traffic-aware BS allocation derived in the next section.

Proposition 1. *In a large-scale cellular network having a BS density of λ_b and a tele-traffic intensity of λ_u , the average DL transmit power of user $u_{i,j}$ requiring a service rate of $R_{i,j}$ is*

given by

$$E[P_{i,j}^o] = \frac{p_{act} P_{\max} \lambda_b}{(\alpha - 2) \lambda_u} \left(2^{\frac{R_{i,j}}{B_n}} - 1 \right), \quad (8)$$

where $E[\cdot]$ denotes the expectation operator, $\alpha > 2$, and $p_{act} = 1 - (1 + \frac{\lambda_u}{K\lambda_b})^{-K}$ with $K = 3.57$ is the average probability of a cell having one or more users to serve.

Proof. See Appendix A. \square

As expected, the average DL transmit power required for user $u_{i,j}$ is directly linked to its required service rate $R_{i,j}$. Increasing the BS density also increases the required $E[P_{i,j}^o]$ proportionally as a direct result of the increased interference. However, the influence of the tele-traffic intensity λ_u on $E[P_{i,j}^o]$ is complex. On one hand, the formula of $E[P_{i,j}^o]$ contains a factor of $\frac{1}{\lambda_u}$, which decreases as λ_u increases. But increasing λ_u will reduce B_n implicitly, which in turn increases the required $E[P_{i,j}^o]$. Since this increase is exponential, in general increasing λ_u will lead to increasing the required average DL transmit power.

Proposition 2. *In a large-scale cellular network having a BS density of λ_b and a tele-traffic intensity of λ_u , the averaged aggregate DL transmit power of BS b_i having a cell coverage area of A_i , where all the users $\forall u_{i,j} \in \Psi_{u,i}$ have an identical rate requirement of $R_{i,j} = R$, is given by*

$$E[P_{cell}^{A_i}] = \frac{p_{act} P_{\max}}{(\alpha - 2)} \left(\exp \left(\left(2^{\frac{R}{B}} - 1 \right) \lambda_u A_i \right) - 1 \right). \quad (9)$$

Proof. See Appendix B. \square

Observe from (9) that $E[P_{cell}^{A_i}]$ increases exponentially as the tele-traffic intensity λ_u or as the BS's coverage area A_i increases, and $E[P_{cell}^{A_i}]$ increases even faster with the user's service data rate R . However, λ_u and R are not independent and, therefore, their relationship with $E[P_{cell}^{A_i}]$ is in fact far more complicated.

Proposition 3. *In a large-scale cellular network having a BS density of λ_b and a tele-traffic intensity of λ_u , assuming that all the users have an identical rate requirement of R , the averaged aggregate DL transmit power of BS b_i is obtained as*

$$E[P_{cell}^i] = \frac{p_{act} P_{\max} (K\lambda_b)^K}{(\alpha - 2) (K\lambda_b - (2^{R/B} - 1) \lambda_u)^K} - \frac{p_{act} P_{\max}}{(\alpha - 2)}. \quad (10)$$

Proof. See Appendix C. \square

We are now ready to consider the network-level energy efficiency metric (6). The APC can be expressed as

$$\text{APC} = \beta \lambda_b E[P_{cell}^i] + \lambda_b P_{OM}, \quad (11)$$

where β is the power amplifier efficiency, and P_{OM} is the BS's circuit-dissipation based power consumption, including the baseband signal processing, battery backup, BS cooling, etc. By using R to denote the spatially averaged users' required rate over \mathbb{R}^2 , the ASE can be formulated as

$$\text{ASE} = R\lambda_u/B. \quad (12)$$

Therefore, the network-level energy efficiency is given by

$$\eta_{EE} = \frac{R\lambda_u/B}{\beta\lambda_b E [P_{cell}^i] + \lambda_b P_{OM}}. \quad (13)$$

Corollary 1. *The energy efficiency of a large-scale cellular network having a BS density of λ_b , a tele-traffic intensity of λ_u , and a spatially averaged user rate of R can be expressed as*

$$\eta_{EE} = \frac{R/B}{\beta \frac{\lambda_b}{\lambda_u} \frac{P_{act} P_{max}}{(\alpha-2)} \left(\frac{K^K}{(K-(2^{R/B}-1)\frac{\lambda_u}{\lambda_b})^K} - 1 \right) + \frac{\lambda_b}{\lambda_u} P_{OM}}. \quad (14)$$

Corollary 1 is obvious. Observe in (14) that the energy efficiency is not a simple monotonic function of λ_b/λ_u .

IV. ENERGY EFFICIENT TELE-TRAFFIC-AWARE NETWORK DEPLOYMENT STRATEGY

We first invoke our analytical results from the previous section to derive and analyze the optimal tele-traffic-aware BS density solution in large-scale cellular networks, providing some useful guidelines for the overall energy savings. Then, we formulate and solve the targeted optimization problem. More specifically, given that a sufficient number of BSs are controllable by network operator and they can be switched on/off when needed, the energy efficiency of the network can always be optimized by finding the optimal tele-traffic-aware BS-density, while maintaining all the users' QoS constraints.

A. Tele-Traffic-Aware Optimal BS-Density Analysis

Proposition 4. *There exists a unique tele-traffic-aware optimal BS-density, denoted by λ_b^* , which maximizes the network's energy efficiency metric η_{EE} and can be obtained numerically from*

$$\lambda_b^* = \frac{(2^{R/B}-1)\lambda_u}{K - \kappa^{+1} \sqrt{\frac{X((K+1)(2^{R/B}-1)\lambda_u - K\lambda_b^*)}{\lambda_b^*(P_{OM}-X/K^K)}}}, \quad (15)$$

where $X = \frac{\beta P_{max} K^K}{(\alpha-2)} > 0$ is a constant.

Proof. By taking the derivative of η_{EE} with respect to λ_b/λ_u , we have

$$\frac{\partial \eta_{EE}}{\partial \left(\frac{\lambda_b}{\lambda_u}\right)} = M_{P_{max}}^{\alpha, \beta, P_{OM}} \left(\frac{\lambda_b}{\lambda_u}\right) \times \underbrace{\left(\frac{X \left((K+1)(2^{R/B}-1) \left(\frac{\lambda_u}{\lambda_b}\right) - K \right)}{\left(K - (2^{R/B}-1) \left(\frac{\lambda_u}{\lambda_b}\right) \right)^{K+1}} + \frac{X}{K^K} - P_{OM} \right)}_{\Theta_{B, P_{OM}}^R \left(\frac{\lambda_b}{\lambda_u}\right)}, \quad (16)$$

where the first part $M_{P_{max}}^{\alpha, \beta, P_{OM}} \left(\frac{\lambda_b}{\lambda_u}\right) > 0$ always holds. Hence the sign of $\partial \eta_{EE} / \partial \left(\frac{\lambda_b}{\lambda_u}\right)$ is equal to the sign of

$\Theta_{B, P_{OM}}^R \left(\frac{\lambda_b}{\lambda_u}\right)$ in (16). Furthermore, we have

$$\begin{cases} \lim_{\frac{\lambda_b}{\lambda_u} \rightarrow +\infty} \Theta_{B, P_{OM}}^R \left(\frac{\lambda_b}{\lambda_u}\right) = \frac{-KX}{K^{K+1}} + \frac{X}{K^K} - P_{OM} \\ \quad \quad \quad = 0^+ - P_{OM} < 0, \\ \lim_{\frac{\lambda_b}{\lambda_u} \rightarrow \frac{(2^{R/B}-1)^+}{K}} \Theta_{B, P_{OM}}^R \left(\frac{\lambda_b}{\lambda_u}\right) = +\infty + \frac{X}{K^K} - P_{OM} \\ \quad \quad \quad > 0. \end{cases} \quad (17)$$

The derivative of $\Theta_{B, P_{OM}}^R \left(\frac{\lambda_b}{\lambda_u}\right)$ with respect to $\frac{\lambda_b}{\lambda_u}$ is

$$\frac{\partial \Theta_{B, P_{OM}}^R \left(\frac{\lambda_b}{\lambda_u}\right)}{\partial \left(\frac{\lambda_b}{\lambda_u}\right)} = \underbrace{\Xi_B^R \left(\frac{\lambda_b}{\lambda_u}\right)}_{f_1 > 0} \underbrace{\left(K \left(1 - 2^{R/B} \right) \frac{\lambda_u}{\lambda_b} \right)}_{f_2 < 0}, \quad (18)$$

where f_1 is always positive and f_2 is always negative, implying that $\Theta_{B, P_{OM}}^R \left(\frac{\lambda_b}{\lambda_u}\right)$ is a monotonically decreasing function of $\frac{\lambda_b}{\lambda_u}$. From (17) and (18), there must exist a unique $\frac{\lambda_b}{\lambda_u}$ to enable $\Theta_{B, P_{OM}}^R \left(\frac{\lambda_b}{\lambda_u}\right) = 0$. Therefore, there always exists an optimal $\frac{\lambda_b}{\lambda_u}$ for maximizing η_{EE} in the range of $[(2^{R/B}-1)/K, +\infty)$. Setting (16) to zero yields (15). \square

Corollary 2. *When the network is operating near the optimal energy efficiency state of $\eta_{EE}(\lambda_b^*)$, increasing the BS's static operational power P_{OM} has to be compensated by reducing the BS-density λ_b and vice versa.*

Proof. Since $\Theta_{B, P_{OM}}^R \left(\frac{\lambda_b^*}{\lambda_u}\right) = 0$ and the network is operating near this optimal energy efficiency state, we have $\Theta_{B, P_{OM}}^R \left(\frac{\lambda_b}{\lambda_u}\right) \approx 0$, that is,

$$\underbrace{\frac{X \left((K+1)(2^{R/B}-1) \left(\frac{\lambda_u}{\lambda_b}\right) - K \right)}{\left(K - (2^{R/B}-1) \left(\frac{\lambda_u}{\lambda_b}\right) \right)^{K+1}}}_{\delta(\lambda_b, \lambda_u)} \approx \underbrace{P_{OM} - \frac{X}{K^K}}_{\kappa(R, B, P_{OM})}. \quad (19)$$

The derivative of $\delta(\lambda_b, \lambda_u)$ with respect to λ_b satisfies

$$\frac{\partial \delta(\lambda_b, \lambda_u)}{\partial \lambda_b} = - \frac{X(K+1)K(2^{R/B}-1)^2 \left(\frac{\lambda_u}{\lambda_b}\right)^2}{\lambda_b \left(K - (2^{R/B}-1) \left(\frac{\lambda_u}{\lambda_b}\right) \right)^{K+2}} < 0. \quad (20)$$

Thus $\delta(\lambda_b, \lambda_u)$ is a monotonically decreasing function of λ_b . On the other hand, $\kappa(R, B, P_{OM})$ is a monotonically increasing function of P_{OM} . Therefore, in order to operate or to maintain the network near its optimal energy efficiency state, i.e., at the state of (19), increasing P_{OM} has to be compensated by reducing λ_b , and vice versa. \square

The implication of Corollary 2 is as follows. When more BSs are deployed, i.e., when increasing λ_b , the BS's static operational power has to be reduced, if the network is to operate in the vicinity of its optimal energy efficiency state. It is generally believed in the literature that installing more BSs is capable of achieving energy savings because of the reduced BS-user distance and the increased achievable user capacity. However, Corollary 2 tells us that this can only be attained by employing higher energy-efficiency BSs.

Corollary 3. When the network is operating near the optimal energy efficiency state of $\eta_{EE}(\lambda_b^*)$, increasing the BS-density λ_b must be compensated by increasing average service rate R , and vice versa.

Proof. Since the network is operating near its optimal energy efficiency state, we have

$$\lambda_b \approx \frac{\tau_1}{K - \tau_2}, \quad (21)$$

where

$$\tau_1 = \left(2^{R/B} - 1\right) \lambda_u, \quad (22)$$

$$\tau_2 = \sqrt[\kappa+1]{\frac{X \left((2^{R/B} - 1) (K + 1) \lambda_u - K \lambda_b \right)}{\lambda_b (P_{OM} - X/K^K)}}. \quad (23)$$

It can readily be seen that $\frac{\tau_1}{K - \tau_2}$ is a monotonically decreasing function of λ_b and a monotonically increasing function of R . Therefore, again when the network operates near its optimal energy efficiency state defined in (21), increasing λ_b must be compensated by increasing R , and vice versa. \square

The physical interpretation of Corollary 3 is plausible. Increasing R or equivalently increasing λ_u can only be accomplished by increasing λ_b , i.e., by employing more BS, if the network is to be operated at a high energy efficiency state.

Proposition 5. Given BS density λ_b , there exists a unique tele-traffic intensity, denoted by λ_u^* , which maximizes the network's energy efficiency metric η_{EE} and can be obtained numerically from

$$\lambda_u^* = \frac{\lambda_b \left(K - \sqrt[\kappa+1]{\frac{X \left((K+1) (2^{R/B} - 1) \lambda_u^* - K \lambda_b \right)}{\lambda_b (P_{OM} - X/K^K)}} \right)}{\left(2^{\frac{R}{B}} - 1 \right)}. \quad (24)$$

Proof. See Appendix D. \square

Combining Propositions 4 and 5, the following corollary can be inferred.

Corollary 4. There exists a unique optimal ratio of λ_b to λ_u , denoted by $\left. \frac{\lambda_b}{\lambda_u} \right|_{\text{opt}}$, which maximizes η_{EE} and it specifies the unique optimal network energy efficiency state.

B. Outage-Constrained Tele-Traffic-Aware Optimal BS Density Analysis

Proposition 6. In a large-scale cellular network having a BS density of λ_b and a tele-traffic intensity of λ_u , the outage probability associated with user $u_{i,j}$ having a required service rate of $R_{i,j}$ is given by

$$Q_{out}^{i,j} = 1 + \frac{\lambda_u}{\lambda_b \rho N_i \exp\left(\frac{N_i \lambda_b \rho}{\lambda_u}\right)} - \frac{\lambda_u}{\lambda_b \rho N_i}, \quad (25)$$

where

$$\rho = T^{2/\alpha} \int_{T^{-2/\alpha}}^{+\infty} \frac{1}{1 + u^{\alpha/2}} du, \quad (26)$$

$$T = \left(2^{R_{i,j}/B_n} - 1 \right). \quad (27)$$

Proof. See Appendix E. \square

In general, a numerical solution has to be applied to calculate the value of ρ . However, for some specific values of α , a closed-form solution can be obtained. For example, for the pathloss exponent of $\alpha = 4$, we have

$$\rho = \sqrt{T} \left(\frac{\pi}{2} - \arctan \left(\frac{1}{\sqrt{T}} \right) \right). \quad (28)$$

The outage probability $Q_{out}^{i,j}$ is a function of N_i and $R_{i,j}$. Utilizing the result of $E[N_i] = \frac{\lambda_u}{\lambda_b}$ and substituting $R_{i,j}$ by the average service rate R , we define the network-level average outage probability as

$$Q_{out} = Q_{out}^{i,j} \Big|_{N_i = \frac{\lambda_u}{\lambda_b}, R_{i,j} = R} = 1 + \frac{1}{\rho \exp(\rho)} - \frac{1}{\rho}, \quad (29)$$

in which the parameter T of (27) becomes

$$T = 2^{\frac{\lambda_u}{\lambda_b} \frac{R}{B}} - 1. \quad (30)$$

Since Q_{out} is a function of λ_b and λ_u , it can be denoted as $Q_{out}(\lambda_b, \lambda_u)$.

Therefore, the generic problem of optimal tele-traffic-aware energy efficient BS deployment (EE-TTAD) satisfying a specific outage constraint can be formulated as: $\max_{\lambda_b} \eta_{EE}$ subject to the constraint of $Q_{out}(\lambda_b, \lambda_u) \leq \varepsilon_{out}$, where $0 < \varepsilon_{out} < 1$ is the given outage threshold, that is,

$$\begin{aligned} \max_{\lambda_b} & \frac{R/B}{\beta \frac{\lambda_b}{\lambda_u} \frac{P_{act} P_{max}}{(\alpha-2)} \left(\frac{K^K}{(K - (2^{R/B} - 1) \frac{\lambda_u}{\lambda_b})^K} - 1 \right) + \frac{\lambda_b}{\lambda_u} P_{OM}}, \\ \text{s.t.} & 1 + \frac{1}{\rho \exp(\rho)} - \frac{1}{\rho} \leq \varepsilon_{out}. \end{aligned} \quad (31)$$

We now derive the optimal solution of the EE-TTAD (31). The derivative of Q_{out} with respect to ρ is

$$\frac{\partial Q_{out}}{\partial \rho} = \frac{1}{\rho^2} (1 - \exp(-\rho)) - \frac{1}{\rho} \exp(-\rho). \quad (32)$$

Since $\exp(\rho) > \rho + 1$ always holds when $\rho > 0$, we have

$$\begin{aligned} \frac{\partial Q_{out}}{\partial \rho} & > \frac{\partial Q_{out}}{\partial \rho} \Big|_{\exp(\rho) = \rho + 1} \\ & = \frac{1}{\rho^2} \left(1 - \frac{1}{\rho + 1} \right) - \frac{1}{\rho} \left(\frac{1}{\rho + 1} \right) = 0, \end{aligned} \quad (33)$$

which implies that Q_{out} is a monotonically increasing function of ρ . Moreover, it is readily seen from (26) and (30) that ρ increases as λ_b decreases, and therefore

$$\frac{\partial Q_{out}}{\partial \lambda_b} = \frac{\partial Q_{out}}{\partial \rho} \frac{\partial \rho}{\partial \lambda_b} < 0, \quad (34)$$

which implies that the outage-constrained feasible region of λ_b is $\lambda_b \in [Q_{out}^{-1}(\varepsilon_{out}; \lambda_u), +\infty)$, where $Q_{out}^{-1}(\cdot; \lambda_u)$ denotes the inverse function of $Q_{out}(\cdot, \lambda_u)$.

Let λ_b^* be the optimal solution of the unconstrained optimization $\max_{\lambda_b} \eta_{EE}$, as given in (15). From the proof of Proposition 4, the unconstrained feasible region of λ_b is $[(2^{R/B} - 1) \lambda_u / K, +\infty)$. Moreover, in the region of $[(2^{R/B} - 1) \lambda_u / K, \lambda_b^*)$, η_{EE} is a monotonically increasing function of λ_b , while in the region of $[\lambda_b^*, +\infty)$, η_{EE} is a monotonically decreasing function of λ_b .

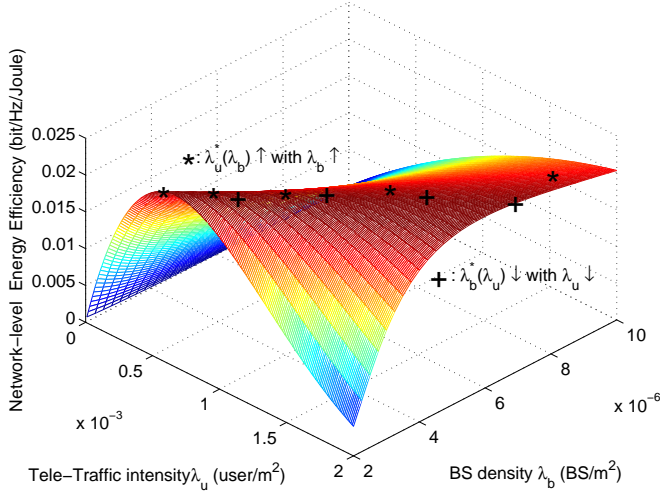


Figure 1: Networks energy efficiency η_{EE} as a function of BS density λ_b and tele-traffic intensity λ_u , given bandwidth resource $B = 30$ MHz and average user rate $R = 0.1$ Mbits/s.

Define $Q_{out}^{-1}(\varepsilon_{out}^*; \lambda_u) = \lambda_b^*$. Then the outage-constrained optimal solution of the optimization problem (31), denoted by λ_b^{out} , is readily given as follows.

1) **Case 1:** $\varepsilon_{out} < \varepsilon_{out}^*$, i.e., $Q_{out}^{-1}(\varepsilon_{out}; \lambda_u) > Q_{out}^{-1}(\varepsilon_{out}^*; \lambda_u) = \lambda_b^*$. The outage-constrained optimal BS density is given by $\lambda_b^{out} = Q_{out}^{-1}(\varepsilon_{out}; \lambda_u)$.

2) **Case 2:** $\varepsilon_{out} > \varepsilon_{out}^*$, i.e., $Q_{out}^{-1}(\varepsilon_{out}; \lambda_u) < Q_{out}^{-1}(\varepsilon_{out}^*; \lambda_u) = \lambda_b^*$. The outage-constrained optimal BS density is given by $\lambda_b^{out} = \lambda_b^*$.

V. NUMERICAL AND SIMULATION RESULTS

In this section our numerical and simulation results are provided for characterizing the network's energy efficiency and to verify the accuracy of our formulation. Moreover, energy savings achieved under various practical scenarios by our design strategy are compared to those of the existing state-of-the-art BS deployment strategies proposed in [25]–[29]. Unless otherwise stated, we set $\alpha = 4$, $P_{OM} = 20$ W, $P_{max} = 20$ W, $\beta = 1.2$, and $\varepsilon_{out} = 0.2$.

A. Characteristics of Key Performance Indicators

Fig. 1 portrays the network's energy efficiency η_{EE} as a function of λ_b and λ_u , given $B = 30$ MHz and $R = 0.1$ Mbits/s. Observe from Fig. 1 that given λ_u , there exists a unique maximum energy efficiency point $\lambda_b^*(\lambda_u)$. Moreover, in the range of $[(2^{R/B} - 1)\lambda_u/K, \lambda_b^*]$, η_{EE} is a monotonically increasing function of λ_b , while in the range of $[\lambda_b^*, +\infty)$, η_{EE} is a monotonically decreasing function of λ_b , as proved in Proposition 4. Similarly, given λ_b , there exists a unique maximum energy efficiency point $\lambda_u^*(\lambda_b)$, and additionally for $\lambda_u < \lambda_u^*$, η_{EE} is a monotonically increasing function of λ_u , while for $\lambda_u \geq \lambda_u^*$, η_{EE} is a monotonically decreasing function of λ_u , as seen in the proof of Proposition 5. It can also be observed from Fig. 1 that $\lambda_b^*(\lambda_u)$ is an increasing function of λ_u , and $\lambda_u^*(\lambda_b)$ is an increasing function of λ_b . The effects of system's bandwidth on the network's energy efficiency is illustrated in Fig. 2, where the system's bandwidth is reduced to $B = 10$ MHz. Upon comparing Fig. 2 to Fig. 1, it can be

seen that increasing B generally reduces the network's energy efficiency.

In Fig. 3 (a), the average aggregate DL transmit power $E[P_{cell}^{A_i}]$ of typical BS b_i is plotted as a function of λ_u , given the cell size $A_i = \pi 250^2$ m² and $B = 20$ MHz with three different values of R , while Fig. 3 (b) portrays the average aggregate BS DL transmit power, $E[P_{cell}^i]$, as a function of λ_b given $\lambda_u = 10^{-4}$ user/m² and $B = 20$ MHz for three different values of R . Observe from Fig. 3 that $E[P_{cell}^{A_i}]$ increases exponentially upon increasing λ_u , while $E[P_{cell}^i]$ decreases exponentially upon increasing λ_b , as indicated in Propositions 2 and 3, respectively. The effect of the average user rate R is clearly illustrated in Fig. 3, where it is seen that increasing R leads to the increase of the average aggregate BS DL transmit power. Moreover, the accuracy of our models (9) and (10) for $E[P_{cell}^{A_i}]$ and $E[P_{cell}^i]$ is verified in Fig. 3, where the theoretical values calculated according to (9) and (10) are shown to agree well with the simulation results.

Fig. 4 depicts the network's energy efficiency η_{EE} as a function of λ_u/λ_b , given the system's bandwidth of $B = 10$ MHz for different average user rates R . The results of Fig. 4 confirm Corollary 4, namely that there exists a unique optimal network energy efficiency state, defined by $\frac{\lambda_b}{\lambda_u}|_{opt}$ or $\frac{\lambda_u}{\lambda_b}|_{opt}$, which maximizes the network's energy efficiency metric η_{EE} . The accuracy of our theoretical formulation is also verified in Fig. 4, where the theoretical values are seen to agree well with the simulation results. Observe from Fig. 4 that the higher R is, the fewer users can be accommodated in the network, when it operates at its optimal energy efficiency. This is because an increase in R must be compensated by a corresponding increase in λ_b , see Corollary 3, or equivalently by a corresponding decrease in λ_u , if the network is to be operated at its optimal energy efficiency.

B. Evaluation of the Attainable Energy Savings

As discussed in the introduction section, the existing state-of-the-art BS deployment strategies [25]–[29] can all be for-

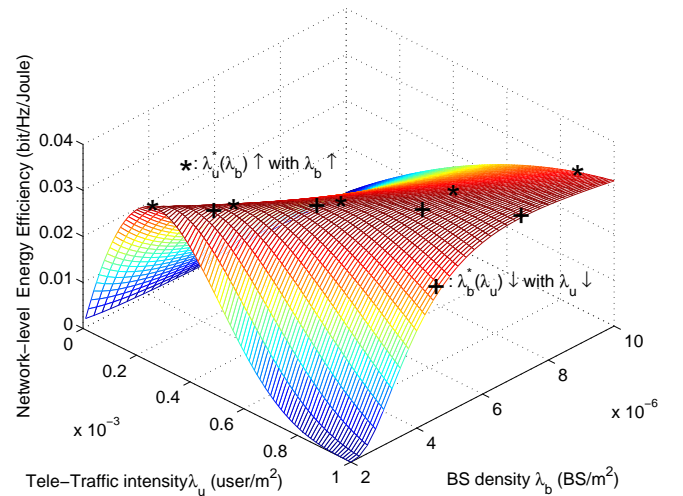


Figure 2: Networks energy efficiency η_{EE} as a function of BS density λ_b and tele-traffic intensity λ_u , given bandwidth resource $B = 10$ MHz and average user rate $R = 0.1$ Mbits/s.

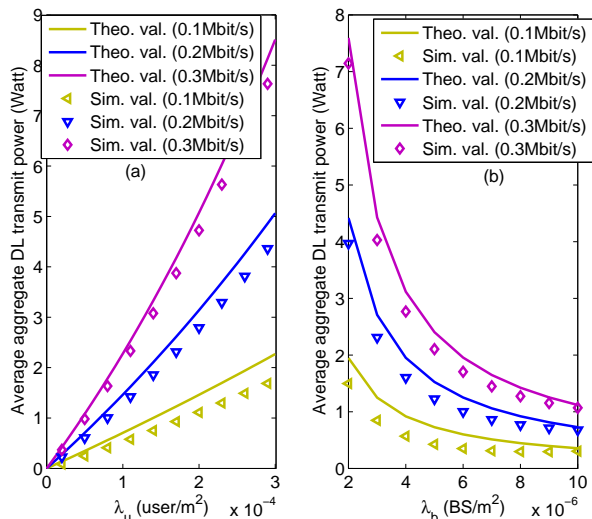


Figure 3: (a) Theoretical and simulated average aggregate DL transmit powers of typical BS b_i , $E[P_{cell}^{A_i}]$, as a functions of λ_u given $A_i = \pi 250^2$ m² and $B = 20$ MHz with different average user rates R , and (b) theoretical and simulated average aggregate BS DL transmit powers, $E[P_{cell}^i]$, as a functions of λ_b given $\lambda_u = 10^{-4}$ user/m² and bandwidth resource $B = 20$ MHz with different average user rates R .

mulated as the problem of minimizing the BS density subject to a specific outage probability constraint, which is denoted as the existing BM-OutC strategy. Fig. 5 compares the APC achieved as a function of the network's tele-traffic intensity λ_u by our proposed EE-TTAD to that of the existing BM-OutC, given the BS static power consumption of $P_{OM} = 40$ Watts, the system's bandwidth of $B = 18$ MHz and three different values of the average user rate R . The range of λ_u from 0 to 4×10^{-4} users/m² corresponds to 0 to 80 users/per cell. Observe from Fig. 5 that the APC consumed by our proposed EE-TTAD is significantly lower than that of the existing state-of-the-art solutions. Moreover, the energy savings achieved by our proposed EE-TTAD over the existing BM-OutC becomes

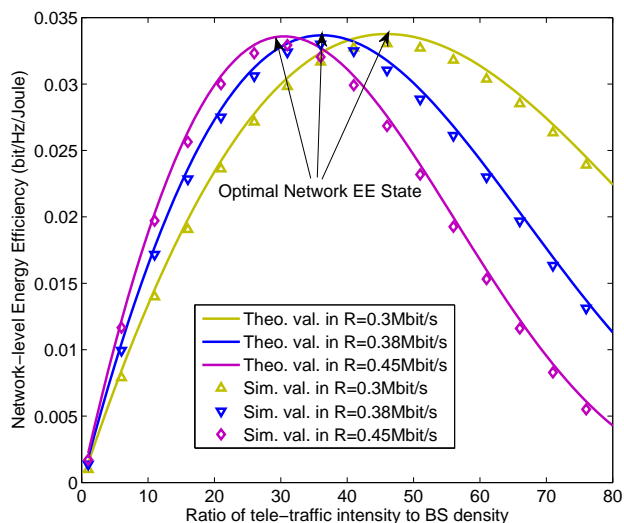


Figure 4: Theoretical and simulated network energy efficiency η_{EE} as a function of λ_u/λ_b given bandwidth resource $B = 10$ MHz with different average user rates R .

higher for the network supporting a higher average user rate. Similarly, Fig. 6 depicts the APCs achieved by our proposed EE-TTAD and the existing BM-OutC as the functions of network's tele-traffic intensity, given $R = 0.4$ Mb/s, $B = 18$ MHz and three different values of P_{OM} . Again, the results of Fig. 6 confirm that our proposed EE-TTAD considerably outperforms the existing state-of-the-art solutions in terms of its achievable network energy efficiency. It can be seen from Fig. 6 that energy savings achieved by our EE-TTAD over the existing BM-OutC are higher for the network associated with a higher BS static power consumption related to power-dissipation other than the transmission power.

VI. CONCLUSIONS

In this paper, we have explicitly modelled the energy efficiency of large-scale cellular networks and derived an accurate energy efficiency expression as a function of the BS-density, tele-traffic intensity, average data rate requirement, and other network parameters. Most crucially, we have quantitatively characterized the relationship between the network's energy efficiency and tele-traffic intensity. With the aid of our network's energy efficiency model, we have derived its optimal energy efficiency state and have studied how the various network parameters influence its energy efficiency. Moreover, we have designed an optimal tele-traffic-aware energy efficient BS deployment strategy, which maximizes the energy efficiency while satisfying the outage probability constraint. Our simulation results have confirmed that the proposed EE-TTAD strategy significantly outperforms the existing state-of-the-art energy efficient BS deployment strategies. Our study has therefore offered valuable insights and guidelines to mobile operators for operating their energy efficient cellular networks.

APPENDIX

Appendix A. Proof of Proposition 1

Proof. From (4) and (5), it can be seen that for typical UE $u_{i,j}$ allocated to B_n , its required DL transmit power $P_{i,j}^o$ must

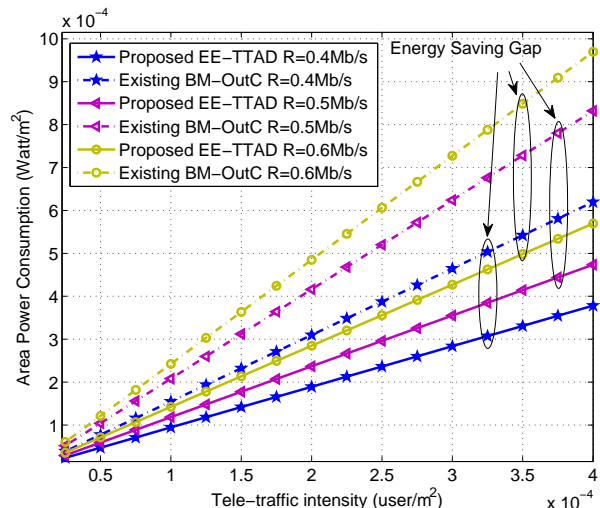


Figure 5: APCs achieved by the proposed EE-TTAD and the existing BM-OutC as the functions of network tele-traffic intensity λ_u , given $P_{OM} = 40$ W, $B = 18$ MHz and three different values of R .

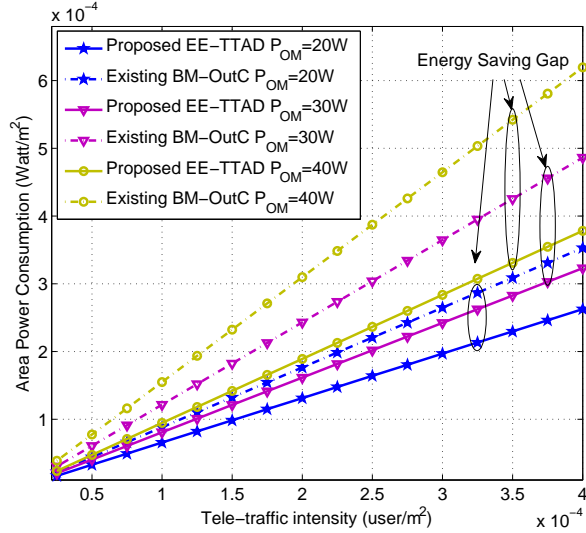


Figure 6: APCs achieved by the proposed EE-TTAD and the existing BM-OutC as the functions of network tele-traffic intensity λ_u , given $R = 0.4$ Mb/s, $B = 18$ MHz and three different values of P_{OM} .

satisfy

$$R_{i,j} = B_n \log_2 \left(1 + \frac{cP_{i,j}^o h_{i,j}^i}{\|u_{i,j} - b_i\|^\alpha I_{\Psi_b \setminus b_i}} \right), \quad (35)$$

in order to meet the service rate $R_{i,j}$. The averaged required DL transmit power conditioned on $\|u_{i,j} - b_i\| = r$ and the interfer layout $\Psi_b \setminus b_i$ is then given by

$$\begin{aligned} E [P_{i,j}^o | r, \Psi_b \setminus b_i] &= E \left[\left(2^{\frac{R_{i,j}}{B_n}} - 1 \right) \frac{I_{\Psi_b \setminus b_i} r^\alpha}{h_{i,j}^i c} \right] \\ &= E \left[\left(2^{\frac{R_{i,j}}{B_n}} - 1 \right) \frac{r^\alpha}{h_{i,j}^i} \sum_{b_k \in \Psi_b \setminus b_i} \frac{h_{k,j}^i P_{\max}}{N_i r_k^\alpha} \right], \quad (36) \end{aligned}$$

where we have introduced $r_k = \|u_{i,j} - b_k\|$. The moment generating function (MGF) of $\frac{1}{c} I_{\Psi_b \setminus b_i}$ is given by [30]

$$\begin{aligned} M_I(s) &= E \left[\exp \left(-s \sum_{b_k \in \Psi_b \setminus b_i} \frac{h_{k,j}^i P_{\max}}{N_i r_k^\alpha} \right) \right] \\ &= E \left[\prod_{b_k \in \Psi_b \setminus b_i} \exp \left(-s \frac{h_{k,j}^i P_{\max}}{N_i r_k^\alpha} \right) \right]. \quad (37) \end{aligned}$$

Note that in calculating the interference, we must consider only the active interfering BSs. Let the BS sleeping probability be $p_s = (1 + \frac{\lambda_u}{K\lambda_b})^{-K}$ with $K = 3.57$ [31]. Then the density of active BSs is $p_{act}\lambda_b$ where $p_{act} = 1 - p_s$. Now let us specifically consider the interference power I' of the PPP distributed BSs in the finite region with the inner radius r and the outer radius r_{bn} . Clearly, $\lim_{r_{bn} \rightarrow \infty} I' = I_{\Psi_b \setminus b_i}$. The MGF of $\frac{1}{c} I'$ can be calculated according to [32]

$$M_{I'}(s) = \exp \left(2\pi p_{act} \lambda_b \int_r^{r_{bn}} \left(1 - \exp \left(-s \frac{P_{\max}}{N_i x^\alpha} \right) \right) x dx \right), \quad (38)$$

where we have utilized $E[h_{k,j}^i] = 1$. According to the well-known property of MGF, we have

$$\begin{aligned} E \left[\frac{1}{c} I' \right] &= - \frac{\partial}{\partial s} \ln M_{I'}(s) \Big|_{s=0} \\ &= \frac{2\pi p_{act} \lambda_b P_{\max}}{N_i} \frac{r_{bn}^{2-\alpha} - r^{2-\alpha}}{(2-\alpha)}, \quad (39) \end{aligned}$$

assuming $\alpha > 2$. By taking $r_{bn} \rightarrow \infty$ in (39) and plugging the result into (36), we obtain

$$E [P_{i,j}^o | r] = \frac{2\pi p_{act} \lambda_b P_{\max}}{(\alpha - 2) N_i} \left(2^{\frac{R_{i,j}}{B_n}} - 1 \right) r^2, \quad (40)$$

where again we have utilized $E[h_{i,j}^i] = 1$. Although the UE distribution in the network follows a PPP with density λ_u , the user distribution within a cell V_i can be accurately approximated by a uniform distribution. Further consider cell V_i as a circle with area $A_i = \pi R_{ave}^2$, where R_{ave} is the radius of the cell. Since $N_i = A_i \lambda_u$, $R_{ave} = \sqrt{\frac{N_i}{\pi \lambda_u}}$. Thus the probability density function (PDF) of the distance r between user $u_{i,j}$ and BS b_i is $f(r) = \frac{2r}{R_{ave}^2}$, and therefore we have

$$\begin{aligned} E [P_{i,j}^o] &= \int_0^{R_{ave}} E [P_{i,j}^o | r] \frac{2r}{R_{ave}^2} dr \\ &= \int_0^{R_{ave}} \frac{\pi^2 p_{act} \lambda_b \lambda_u P_{\max}}{(\alpha - 2) N_i^2} \left(2^{\frac{R_{i,j}}{B_n}} - 1 \right) 4r^3 dr \\ &= \frac{p_{act} \lambda_b P_{\max}}{(\alpha - 2) \lambda_u} \left(2^{\frac{R_{i,j}}{B_n}} - 1 \right). \quad (41) \end{aligned}$$

This completes the proof. \square

Appendix B. Proof of Proposition 2

Proof. In practice, the maximum number of users that a BS can serve, denoted as N_{\max} , is limited by the total RBs available. If the number of users N_i in cell V_i is larger than N_{\max} , obviously some of the users cannot be served. For BS b_i with cell size A_i and the number of serving users $N_i = |\Psi_{u,i}|$, given that all the users have the identical rate requirement of $R_{i,j} = R$, the average aggregate DL transmit power of b_i conditioned on N_i is given by

$$\begin{aligned} E [P_{cell}^{A_i} | N_i] &= E [N_i E [P_{i,j}^o]] \\ &= \begin{cases} \frac{p_{act} P_{\max}}{(\alpha - 2)} \left(2^{\frac{N_i R}{B}} - 1 \right), & N_i \leq N_{\max}, \\ \frac{p_{act} P_{\max}}{(\alpha - 2)} \left(2^{\frac{N_{\max} R}{B}} - 1 \right), & N_i > N_{\max}, \end{cases} \quad (42) \end{aligned}$$

where we have utilized the result that the average of N_i is $E[N_i] = \frac{\lambda_u}{\lambda_b}$. Under an idealized condition of $N_{\max} \rightarrow \infty$, we have the probability mass function (PMF) of N_i given by [33]

$$Q_{A_i}(N_i) = \frac{(\lambda_u A_i)^{N_i}}{N_i!} \exp(-\lambda_u A_i), \quad N_i \geq 1. \quad (43)$$

Therefore, the average aggregate DL transmit power of BS b_i with cell size A_i is given by

$$\begin{aligned} E[P_{cell}^{A_i}] &= \sum_{N_i=1}^{\infty} \frac{p_{act} P_{max}}{(\alpha-2)} \left(2^{\frac{N_i R}{B}} - 1\right) \exp(-\lambda_u A_i) \frac{(\lambda_u A_i)^{N_i}}{N_i!} \\ &= \frac{p_{act} P_{max}}{(\alpha-2)} \left(\sum_{N_i=1}^{\infty} \frac{\left(2^{\frac{R}{B}} \lambda_u A_i\right)^{N_i}}{N_i! \exp(\lambda_u A_i)} - \sum_{N_i=1}^{\infty} \frac{(\lambda_u A_i)^{N_i}}{N_i! \exp(\lambda_u A_i)} \right). \end{aligned} \quad (44)$$

With the aid of the well-known property of infinite series for sufficiently small $\frac{R}{B}$, i.e., $\frac{R}{B} \rightarrow 0$, the first infinite series in (44) becomes

$$\begin{aligned} \sum_{N_i=1}^{\infty} \frac{\left(2^{\frac{R}{B}} \lambda_u A_i\right)^{N_i}}{N_i! \exp(\lambda_u A_i)} &= \sum_{N_i=0}^{\infty} \frac{\left(2^{\frac{R}{B}} \lambda_u A_i\right)^{N_i}}{N_i! \exp(\lambda_u A_i)} - \exp(-\lambda_u A_i) \\ &= \exp\left(\left(2^{\frac{R}{B}} - 1\right) \lambda_u A_i\right) - \exp(-\lambda_u A_i). \end{aligned} \quad (45)$$

In a similar way, for the second infinite series in (44), we have

$$\begin{aligned} \sum_{N_i=1}^{\infty} \frac{(\lambda_u A_i)^{N_i}}{N_i! \exp(\lambda_u A_i)} &= \sum_{N_i=0}^{\infty} \frac{(\lambda_u A_i)^{N_i}}{N_i! \exp(\lambda_u A_i)} - \exp(-\lambda_u A_i) \\ &= 1 - \exp(-\lambda_u A_i). \end{aligned} \quad (46)$$

Substituting (45) and (46) into (44) completes the proof. \square

Appendix C. Proof of Proposition 3

Proof. The PDF of the cell size can be approximated by the Gamma distribution as follow [34]

$$f(A_i) = \lambda_b^K \frac{K^K}{\Gamma(K)} A_i^{K-1} \exp(-K\lambda_b A_i), \quad (47)$$

where the gamma function $\Gamma(x) = \int_0^{+\infty} t^{x-1} \exp(-t) dt$. According to Proposition 2, we have

$$\begin{aligned} E[P_{cell}^i] &= \int_{A>0} \frac{p_{act} P_{max} \lambda_b^K K^K}{(\alpha-2)\Gamma(K)} \\ &\quad \times \frac{A^{K-1} \left(\exp\left(\left(2^{\frac{R}{B}} - 1\right) \lambda_u A\right) - 1\right)}{\exp(K\lambda_b A)} dA \\ &= \int_{A>0} \frac{p_{act} P_{max} K^K \lambda_b^K}{(\alpha-2)\Gamma(K)} \frac{A^{K-1} \exp\left(\left(2^{\frac{R}{B}} - 1\right) \lambda_u A\right)}{\exp(K\lambda_b A)} dA \\ &\quad - \int_{A>0} \frac{p_{act} P_{max} K^K \lambda_b^K}{(\alpha-2)\Gamma(K)} \frac{A^{K-1}}{\exp(K\lambda_b A)} dA = Y_1 - Y_2. \end{aligned} \quad (48)$$

According to [35], the first integration can be evaluated as

$$\begin{aligned} Y_1 &= \frac{p_{act} P_{max} K^K \lambda_b^K}{\Gamma(K)(\alpha-2)} \frac{\Gamma(K)}{(K\lambda_b - (2^{R/B} - 1)\lambda_u)^K} \\ &= \frac{p_{act} P_{max} (K\lambda_b)^K}{(\alpha-2)(K\lambda_b - (2^{R/B} - 1)\lambda_u)^K}, \end{aligned} \quad (49)$$

while the second integration yields

$$Y_2 = \frac{p_{act} P_{max} K^K \lambda_b^K}{\Gamma(K)(\alpha-2)} \frac{\Gamma(K)}{(K\lambda_b)^K} = \frac{p_{act} P_{max}}{(\alpha-2)}. \quad (50)$$

Substituting (49) and (50) back to (48) completes the proof. \square

Appendix D. Proof of Proposition 5

Proof. By taking the derivative of η_{EE} with respect to λ_u , we have

$$\begin{aligned} \frac{\partial \eta_{EE}}{\partial \lambda_u} &= \frac{\lambda_b}{\lambda_u^2} M_{P_{max}}^{\alpha, \beta, P_{OM}} \left(\frac{\lambda_b}{\lambda_u}\right) \\ &\quad \times \underbrace{\left(\frac{X \left((K+1) \left(2^{R/B} - 1\right) \left(\frac{\lambda_u}{\lambda_b}\right) - K\right)}{\left(K - \left(2^{R/B} - 1\right) \left(\frac{\lambda_u}{\lambda_b}\right)\right)^{K+1}} - \frac{X}{K^K} + P_{OM} \right)}_{U_{B, P_{OM}}^R \left(\frac{\lambda_b}{\lambda_u}\right)}. \end{aligned} \quad (51)$$

Hence the sign of $\partial \eta_{EE} / \partial \lambda_u$ is equal to the sign of $U_{B, P_{OM}}^R \left(\frac{\lambda_b}{\lambda_u}\right)$ in (51). Furthermore, we have

$$\begin{cases} \lim_{\lambda_u \rightarrow +0} U_{B, P_{OM}}^R \left(\frac{\lambda_b}{\lambda_u}\right) = \frac{KX}{K^{K+1}} - \frac{X}{K^K} + P_{OM} = P_{OM} > 0, \\ \lim_{\lambda_u \rightarrow \left(\frac{R\lambda_b}{2^{R/B} - 1}\right)^+} U_{B, P_{OM}}^R \left(\frac{\lambda_b}{\lambda_u}\right) = -\infty - \frac{X}{K^K} + P_{OM} < 0. \end{cases} \quad (52)$$

The derivative of $U_{B, P_{OM}}^R \left(\frac{\lambda_b}{\lambda_u}\right)$ with respect to λ_u is

$$\frac{\partial U_{B, P_{OM}}^R \left(\frac{\lambda_b}{\lambda_u}\right)}{\partial \lambda_u} = \left(\frac{\lambda_b}{\lambda_u^2}\right) \underbrace{\Xi_B^R \left(\frac{\lambda_b}{\lambda_u}\right)}_{f_1 > 0} \underbrace{\left(K \left(1 - 2^{R/B}\right) \frac{\lambda_u}{\lambda_b}\right)}_{f_2 < 0}, \quad (53)$$

which implies that $U_{B, P_{OM}}^R \left(\frac{\lambda_b}{\lambda_u}\right)$ is a monotonically decreasing function of λ_u .

From (51) and (53), there must exist a unique λ_u to enable $U_{B, P_{OM}}^R \left(\frac{\lambda_b}{\lambda_u}\right) = 0$. Therefore, there always exists an optimal λ_u for maximizing η_{EE} in the range of $\lambda_u \in \left[0, \frac{K\lambda_b}{2^{R/B} - 1}\right)$. Setting (51) to zero yields (24). \square

Appendix E. Proof of Proposition 6

Proof. According to the definition (7), the outage probability of user $u_{i,j}$ conditioned on r is given by

$$Q_{i,j}^{out}[r] = 1 - \Pr(P_{i,j}^o \leq P_{max}/N_i | r). \quad (54)$$

From (35), we have $P_{i,j}^o = \left(2^{R_{i,j}/B_n} - 1\right) \frac{I_{\Psi_b \setminus b_i}}{h_{i,j}^{c\alpha} r^{-\alpha}}$, and hence according to [17]

$$\begin{aligned} Q_{i,j}^{out}[r] &= 1 - \Pr\left(h_{i,j}^i \geq \frac{N_i}{P_{max}} \left(2^{R_{i,j}/B_n} - 1\right) r^\alpha \frac{1}{c} I_{\Psi_b \setminus b_i}\right) \\ &= 1 - \mathcal{L}_{\frac{1}{c} I_{\Psi_b \setminus b_i}}(s) \Big|_{s = \frac{N_i}{P_{max}} T r^\alpha}, \end{aligned} \quad (55)$$

where $\mathcal{L}_{\frac{1}{c} I_{\Psi_b \setminus b_i}}(s)$ is the Laplace transform of $\frac{1}{c} I_{\Psi_b \setminus b_i}$, and $T = \left(2^{R_{i,j}/B_n} - 1\right)$. From (5),

$$\frac{1}{c} I_{\Psi_b \setminus b_i} = \sum_{b_k \in \Psi_b \setminus b_i} r_k^{-\alpha} \frac{P_{max}}{N_i} h_{k,j}^i, \quad (56)$$

in which $\frac{P_{\max}}{N_i} h_{k,j}^i \sim \exp\left(\frac{N_i}{P_{\max}}\right)$. Therefore, according to Appendix B of [17], we have

$$\begin{aligned} \mathcal{L}_{\frac{1}{c} I_{\Psi_b \setminus b_i}} & \left(\frac{N_i}{P_{\max}} T r^\alpha \right) \\ &= \exp \left(-2\pi\lambda_b \int_r^\infty \left(1 - \frac{\frac{N_i}{P_{\max}}}{\frac{N_i}{P_{\max}} + \frac{N_i}{P_{\max}} T r^\alpha v^{-\alpha}} \right) v dv \right) \\ &= \exp \left(-2\pi\lambda_b \int_r^\infty \left(\frac{T}{T + \left(\frac{v}{r}\right)^\alpha} \right) v dv \right) \\ &= \exp \left(-\pi r^2 \lambda_b \rho \right), \end{aligned} \quad (57)$$

where $\rho = T^{2/\alpha} \int_{T^{-2/\alpha}}^\infty \frac{1}{1+u^{\alpha/2}} du$.

Noting that the PDF of r is $f(r) = \frac{2r}{R_{\text{ave}}^2}$, $0 \leq r \leq R_{\text{ave}}$, with $R_{\text{ave}}^2 = \frac{N_i}{\pi\lambda_u}$, we have

$$\begin{aligned} Q_{i,j}^{\text{out}} &= 1 - \int_0^{R_{\text{ave}}} \frac{2x}{R_{\text{ave}}^2} \exp(-\pi x^2 \lambda_b \rho) dx \\ &= 1 - \frac{\exp(-\pi x^2 \lambda_b \rho)}{(-\pi \lambda_b \rho) R_{\text{ave}}^2} \Big|_0^{R_{\text{ave}}} \\ &= 1 + \frac{\lambda_u}{\lambda_b \rho N_i \exp\left(\frac{N_i \lambda_b \rho}{\lambda_u}\right)} - \frac{\lambda_u}{\lambda_b \rho N_i}. \end{aligned} \quad (58)$$

This completes the proof. \square

REFERENCES

- [1] C. Han, T. Harrold, S. Armour, I. Krikidis, S. Videv, P. M. Grant, H. Haas, J. S. Thompson, I. Ku, C.-X. Wang, T. A. Le, M. R. Nakhai, J. Zhang, and L. Hanzo, "Green radio: Radio techniques to enable energy-efficient wireless networks," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 46-54, Jun. 2011.
- [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] M.-S. Alouini and A. J. Goldsmith, "Area spectral efficiency of cellular mobile radio systems," *IEEE Trans. Vehicular Technology*, vol. 48, no. 4, pp. 1047-1066, Jul. 1999.
- [4] A. Fehske, G. Fettweis, J. Malmodin, and G. Biczok, "The global footprint of mobile communications: The ecological and economic perspective," *IEEE Commun. Mag.*, vol. 49, no. 8, pp. 55-62, Aug. 2011.
- [5] V. Mancuso and S. Alouf, "Reducing costs and pollution in cellular networks," *IEEE Commun. Mag.*, vol. 49, no. 8, pp. 63-71, Aug. 2011.
- [6] R. Q. Hu and Y. Qian, "An energy efficient and spectrum efficient wireless heterogeneous network framework for 5G systems," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 94-101, May 2014.
- [7] 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.
- [8] K. Zheng, X. Zhang, Q. Zheng, W. Xiang and L. Hanzo, "Quality-of-experience assessment and its application to video services in lte networks," *IEEE Wireless Commun.*, vol. 22, no. 1, pp. 70-78, Feb. 2015.
- [9] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, "Green cellular networks: A survey, some research issues and challenges," *IEEE Commun. Surveys Tutorials*, vol. 13, no. 4, pp. 524-540, 2011.
- [10] G. Auer *et al.*, "How much energy is needed to run a wireless network?," *IEEE Wireless Commun.*, vol. 18, no. 5, pp. 40-49, Oct. 2011.
- [11] S. Boiardi, A. Capone, and B. Sanso, "Planning for energy-aware wireless networks," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 156-162, Feb. 2014.
- [12] Z. Yang and Z. Niu, "Energy saving in cellular networks by dynamic RS-BS association and BS switching," *IEEE Trans. Vehicular Technology*, vol. 62, no. 9, pp. 4602-4614, Nov. 2013.
- [13] H. Holtkamp, G. Auer, S. Bazzi, and H. Haas, "Minimizing base station power consumption," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 2, pp. 297-306, Feb. 2014.
- [14] 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.
- [15] L. Zhao, G. Zhao, and T. O'Farrell, "Efficiency metrics for wireless communications," in *Proc. 2013 PIMRC* (London, UK), Sep. 8-11, 2013, pp. 2825-2829.
- [16] J. Akhtman and L. Hanzo, "Power versus bandwidth-efficiency in wireless communications: The economic perspective," in *Proc. VTC 2009-Fall* (Alaska, AK), Sep. 20-23, 2009, pp. 1-5.
- [17] 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.
- [18] 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. Vehicular Technology*, vol. 64, no. 9, pp. 4245-4256, Sep. 2015.
- [19] S. P. Weber, X. Yang, J. G. Andrews, and G. de Veciana, "Transmission capacity of wireless ad hoc networks with outage constraints," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4091-4102, Dec. 2005.
- [20] C. C. Wang, T. Q. S. Quek, and M. Kountouris, "Throughput optimization, spectrum allocation, and access control in two-tier femtocell networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 561-574, Apr. 2012.
- [21] 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.
- [22] T. D. Novlan, R. K. Ganti, A. Ghosh, and J. G. Andrews, "Analytical evaluation of fractional frequency reuse for heterogeneous cellular networks," *IEEE Trans. Commun.*, vol. 60, no. 7, pp. 2029-2039, Jul. 2012.
- [23] H. S. Dhillon, M. Kountouris, and J. G. Andrews, "Downlink MIMO HetNets: Modeling, ordering results and performance analysis," *IEEE Trans. Wireless Commun.*, vol. 12, no. 10, pp. 5208-5222, Oct. 2013.
- [24] H. 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.
- [25] 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.
- [26] J. Peng, P. Hong, K. Xue, "Energy-aware cellular deployment strategy under coverage performance constraints," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 69-80, Jan. 2015.
- [27] 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.
- [28] N. Deng, S. Zhang, W. Zhou, and J. Zhu, "A stochastic geometry approach to energy efficiency in relay-assisted cellular networks," in *Proc. GLOBECOM 2012* (Anaheim, CA), Dec. 3-7, 2012, pp. 3484-3489.
- [29] Y. Huang, X. Zhang, J. Zhang, J. Tang, Z. Su, and W. Wang, "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.
- [30] G. Grimmett and D. Stirzaker, *Probability and Random Processes* (Third Edition). Cambridge University Press: Oxford, 2004.
- [31] S. Lee and K. Huang, "Coverage and economy of cellular networks with many base stations," *IEEE Commun. Letters*, vol. 16, no. 7, pp. 1038-1040, Jul. 2012.
- [32] S. Srinivasa, *Modeling Interference in Uniformly Random Wireless Networks: Theory and Applications*. MSc Thesis, University of Notre Dame, USA, 2007.
- [33] R. L. Streit, *Poisson Point Processes: Imaging, Tracking and Sensing*. Springer, 2010.
- [34] J. Ferenc and Z. Néda, "On the size distribution of Poisson Voronoi cells," *Physica A: Statistical Mechanics and its Applications*, vol. 385, no. 2, pp. 518-526, 2007.
- [35] I.S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (7th edition). Academic Press, 2007.