System-Level Modeling and Optimization of the Energy Efficiency in Cellular Networks - A Stochastic Geometry Framework

Marco Di Renzo, Alessio Zappone, Lam Thanh Tu, Merouane Debbah

System-Level Modeling and Optimization of the Energy Efficiency in Cellular Networks -- A Stochastic Geometry Framework

4 authors:

Marco Di Renzo
French National Centre for Scientific Research
315 PUBLICATIONS **8,018** CITATIONS

SEE PROFILE

Alessio Zappone
CentraleSupélec
138 PUBLICATIONS **3,099** CITATIONS

SEE PROFILE

Thanh Lam Tu
École Supérieure d'Electricité
29 PUBLICATIONS **237** CITATIONS

SEE PROFILE

mérouane Debbah
École Supérieure d'Electricité
810 PUBLICATIONS **25,961** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project 5Gto10G View project

Project Marie Curie ITN-H2020 5Gwireless project View project

System-Level Modeling and Optimization of the Energy Efficiency in Cellular Networks – A Stochastic Geometry Framework

Marco Di Renzo, *Senior Member, IEEE*, Alessio Zappone, *Senior Member, IEEE*,
Thanh Tu Lam, *Student Member, IEEE*, and Mérouane Debbah, *Fellow, IEEE*

Abstract—In this paper, we analyze and optimize the energy efficiency of downlink cellular networks. With the aid of tools from stochastic geometry, we introduce a new closed-form analytical expression of the potential spectral efficiency (bit/sec/m^2). In the interference-limited regime for data transmission, unlike currently available mathematical frameworks, the proposed analytical formulation depends on the transmit power and deployment density of the base stations. This is obtained by generalizing the definition of coverage probability and by accounting for the sensitivity of the receiver not only during the decoding of information data, but during the cell association phase as well. Based on the new formulation of the potential spectral efficiency, the energy efficiency (bit/Joule) is given in a tractable closed-form formula. An optimization problem is formulated and is comprehensively studied. It is mathematically proved, in particular, that the energy efficiency is a unimodal and strictly pseudo-concave function in the transmit power, given the density of the base stations, and in the density of the base stations, given the transmit power. Under these assumptions, therefore, a unique transmit power and density of the base stations exist, which maximize the energy efficiency. Numerical results are illustrated in order to confirm the obtained findings and to prove the usefulness of the proposed framework for optimizing the network planning and deployment of cellular networks from the energy efficiency standpoint.

Index Terms—Cellular Networks, Energy Efficiency, Poisson Point Processes, Stochastic Geometry, Optimization.

I. Introduction

The Energy Efficiency (EE) is regarded as a key performance metric towards the optimization of operational cellular networks, and the network planning and deployment

of emerging communication systems [1]. The EE is defined as a benefit-cost ratio where the benefit is given by the amount of information data per unit time and area that can be reliably transmitted in the network, i.e., the network spectral efficiency, and the cost is represented by the amount of power per unit area that is consumed to operate the network, i.e., the network power consumption. Analyzing and designing a communication network from the EE standpoint necessitate appropriate mathematical tools, which are usually different from those used for optimizing the network spectral efficiency and the network power consumption individually [2]. The optimization problem, in addition, needs to be formulated in a sufficiently simple but realistic manner, so that all relevant system parameters appear explicitly and the utility function is physically meaningful.

Optimizing the EE of a cellular network can be tackled in different ways, which include [1]: the design of medium access and scheduling protocols for optimally using the available resources, e.g., the transmit power; the use of renewable energy sources; the development of innovative hardware for data transmission and reception; and the optimal planning and deployment of network infrastructure. In the present paper, we focus our attention on optimizing the average number of Base Stations (BSs) to be deployed (or to be kept operational) per unit area and their transmit power. Henceforth, this is referred to as "system-level EE" optimization, i.e., the EE across the entire (or a large portion of the) cellular network is the utility function of interest.

System-level analysis and optimization are useful when the network operators are interested in optimizing the average performance across the entire cellular network. Hence, they are relevant for optimally operating current networks, and for deploying and planning future networks. In the first case, given an average number of BSs per unit area already deployed, they may provide information on the average number of BSs that can be switched off based on the average load of the network, and on their optimal transmit power to avoid coverage holes. In the second case, they may guide the initial deployment of cellular infrastructure that employs new types of BSs (e.g., powered by renewable energy sources), new transmission technologies (e.g., large-scale antennas), or that operate in new frequency bands (e.g., the millimeter-wave spectrum).

In the last few years, the system-level modeling and analysis of cellular networks have been facilitated by capitalizing on the mathematical tool of stochastic geometry and, more precisely,

on the theory of spatial point processes [3]-[5]. It has been empirically validated that, from the system-level standpoint, the locations of the BSs can be abstracted as points of a homogeneous Poisson Point Process (PPP) whose intensity coincides with the average number of BSs per unit area [6]. A comprehensive survey of recent results in this field of research is available in [7].

A relevant performance metric for the design of cellular networks is the Potential Spectral Efficiency (PSE), which is the network information rate per unit area (measured in bit/sec/m^2) that corresponds to the minimum signal quality for reliable transmission. Under the PPP modeling assumption, the PSE can be obtained in two steps: i) first by computing the PSE of a randomly chosen Mobile Terminal (MT) and by assuming a given spatial realization for the locations of the BSs and ii) then by averaging the obtained conditional PSE with respect to all possible realizations for the locations of the BSs and MTs. In the interference-limited regime, this approach allows one to obtain a closed-form expression of the PSE under the (henceforth called) *standard modeling assumptions*, i.e., single-antenna transmission, singular path-loss model, Rayleigh fading, fully-loaded BSs, cell association based on the highest average received power [3]. Motivated by these results, the PPP modeling approach for the locations of the BSs has been widely used to analyze the trade-off between the network spectral efficiency and the network power consumption, e.g., [8], as well as to minimize the network power consumption given some constraints on the network spectral efficiency or to maximize the network spectral efficiency given some constraints on the network power consumption [9]. The PPP modeling approach has been applied to optimize the EE of cellular networks as well. Notable examples for this field of research are [10]-[26]. A general study of the energy and spectral efficiencies of multi-tier cellular networks can be found in [27]. In the authors' opinion, however, currently available approaches for modeling and optimizing the system-level EE of cellular networks are insufficient and/or unsuitable for mathematical analysis. This is further elaborated in the next section.

A. Fundamental Limitations of Current Approaches for System-Level EE Optimization

We begin with an example that shows the limitations of the available analytical frameworks. In the interference-limited regime, under the standard modeling assumptions, the PSE is:

$$\text{PSE} = \lambda_{\text{BS}} B_W \log_2 \left(1 + \gamma_D\right) P_{\text{cov}}\left(\gamma_D\right)$$
$$\overset{(a)}{=} \frac{\lambda_{\text{BS}} B_W \log_2 \left(1 + \gamma_D\right)}{{}_2F_1\left(1, -2/\beta, 1 - 2/\beta, -\gamma_D\right)} \quad (1)$$

where λ_{BS} is the density of BSs, B_W is the transmission bandwidth, γ_D is the threshold for reliable decoding, $\beta > 2$ is the path-loss exponent, ${}_2F_1\left(\cdot, \cdot, \cdot, \cdot\right)$ is the Gauss hypergeometric function, $P_{\text{cov}}\left(\cdot\right)$ is the coverage probability defined in [3, Eq. (1)], and (a) follows from [3, Eq. (8)].

The main strength of (1) is its simple closed-form formulation. This is, however, its main limitation as well, especially as far as formulating meaningful system-level EE optimization

problems is concerned. Under the standard modeling assumptions, in fact, the network power consumption (Watt/m^2) is[1] $P_{\text{grid}} = \lambda_{\text{BS}}\left(P_{\text{tx}} + P_{\text{circ}}\right)$, where P_{tx} is the transmit power of the BSs and P_{circ} is the static power consumption of the BSs, which accounts for the power consumed in all hardware blocks, e.g., analog-to-digital and digital-to-analog converters, analog filters, cooling components, and digital signal processing [1]. The system-level EE (bit/Joule) is defined as the ratio between (1) and the network power consumption, i.e., $\text{EE} = \text{PSE}/P_{\text{grid}}$. Since the PSE in (1) is independent of the transmit power of the BSs, P_{tx}, and the network power consumption, P_{grid}, linearly increases with P_{tx}, we conclude that any EE optimization problems formulated based on (1) would result in the trivial optimal solution consisting of turning all the BSs off (the optimal transmit power is zero). In the context of multi-tier cellular networks, a similar conclusion has been obtained in some early papers on system-level EE optimization, e.g., [8], where it is shown that the EE is maximized if all macro BSs operate in sleeping mode. A system-level EE optimization problem formulated based on (1) would result, in addition, in a physically meaningless utility function, which provides a non-zero benefit-cost ratio, i.e., a strictly positive EE while transmitting zero power ($\text{EE}\left(P_{\text{tx}} = 0\right) = \text{PSE}/\left(\lambda_{\text{BS}} P_{\text{circ}}\right) > 0$). In addition, the EE computed from (1) is independent of the density of BSs. We briefly mention here, but will detail it in Section III, that the load model, i.e., the fully-loaded assumption, determines the conclusion that the EE does not depend on λ_{BS}. This assumption, however, does not affect the conclusion that the optimal P_{tx} is zero. This statement is made more formal in the sequel (see *Proposition 1* and *Corollary 1*). It is worth nothing that the conclusion that the PSE is independent of P_{tx} is valid regardless of the specific path-loss model being used[2]. It depends, on the other hand, on the assumptions of interference-limited operating regime and of having BSs that emit the same P_{tx}.

Based on these observations, we conclude that a new analytical formulation of the PSE that explicitly depends on the transmit power and density of the BSs, and that is tractable enough for system-level EE optimization is needed. From an optimization point of view, in particular, it is desirable that the PSE is formulated in a closed-form expression and that the resulting EE function is unimodal and strictly pseudo-concave in the transmit power (given the density) and in the density (given the transmit power) of the BSs. This would imply, e.g., that the first-order derivative of the EE with respect to the transmit power of the BSs (assuming the density given) would have a unique zero, which would be the unique optimal transmit power that maximizes the EE [2]. Similar conclusions would apply to the optimal density of the BSs for a given transmit power. Further details are provided in Section IV. In this regard, a straightforward approach to overcome the limitations of (1) would be to abandon the interference-limited assumption and to take the receiver noise into account. In this

---

[1]In the present paper, this holds true for Load Model 1 that is introduced in Section II-D.

[2]The reader may verify this statement by direct inspection of (4), where P_{tx} cancels out for any path-loss models.

case, the PSE would be formulated in terms of a single-integral that, in general, cannot be expressed in closed-form [3], [17, Eq. (9)]. This integral formulation, in particular, results in a system-level EE optimization problem that is not easy to tackle. This approach, in addition, has the inconvenience of formulating the optimization problem for an operating regime where cellular networks are unlikely to operate in practice.

B. State-of-the-Art on System-Level EE Optimization

We briefly summarize the most relevant research contributions on energy-aware design and optimization of cellular networks. Due to space limitations, we discuss only the contributions that are closely related to ours. A state-of-the-art survey on EE optimization is available in [2].

In [8], the authors study the impact of switching some macro BSs off in order to minimize the power consumption under some constraints on the coverage probability. Since the authors rely on the mathematical framework in (1), they conclude that all macro BSs need to be switched off to maximize the EE. In [9], the author exploits geometric programming to minimize the power consumption of cellular networks given some constraints on the network coverage and capacity. The EE is not studied. A similar optimization problem is studied in [11] and [17] for two-tier cellular networks but the EE is not studied either. As far as multi-tier cellular networks are concerned, an important remark is necessary. In the interference-limited regime, optimal transmit powers and densities for the different tiers of BSs may exist if the tiers have different thresholds for reliably decoding the data. The PSE, otherwise, is the same as that of single-tier networks, i.e., it is independent of the transmit power and density of the BSs. In [14], the authors study the EE of small cell networks with multi-antenna BSs. For some parameter setups, it is shown that an optimal density of the BSs exists. The EE, however, still decreases monotonically with the transmit power of the BSs, which implies that the EE optimization problem is not well formulated from the transmit power standpoint. More general scenarios are considered in [10], [12], [13], [15], [16], [18]-[25], but similar limitations hold. In some cases, e.g., [20], the existence and uniqueness of an optimal transmit power and density of the BSs are not mathematically proved or, e.g., in [24], the problem formulation has a prohibitive numerical complexity as it necessitates the computation of multiple integrals and infinite series. It is apparent, therefore, that a tractable approach for system-level EE optimization is missing in the open technical literature. In the present paper, we introduce a new definition of PSE that overcomes these limitations.

C. Research Contribution and Novelty

In the depicted context, the specific novel contributions made by this paper are as follows:

- We introduce a new closed-form analytical formulation of the PSE for interference-limited cellular networks (during data transmission), which depends on the transmit power and density of the BSs. The new expression of the PSE is obtained by taking into account the power sensitivity

of the receiver not only for data transmission but for cell association as well.
- Based on the new expression of the PSE, a new system-level EE optimization problem is formulated and comprehensively studied. It is mathematically proved that the EE is a unimodal and strictly pseudo-concave function in the transmit power given the BSs' density and in the BSs' density given the transmit power. The dependency of the optimal power as a function of the density and of the optimal density as a function of the power is discussed.
- A first-order optimal pair of transmit power and density of the BSs is obtained by using a simple alternating optimization algorithm whose details are discussed in the sequel. Numerical evidence of the global optimality of this approach is provided as well.
- Two load models for the BSs are analyzed and compared against each other. It is shown that they provide the same PSE but have different network power consumptions. Hence, the optimal transmit power and density of the BSs that maximize their EEs are, in general, different. Their optimal EEs and PSEs are studied and compared against each other.

The paper is organized as follows. In Section II, the system model is presented. In Section III, the new definition of PSE is introduced. In Section IV, the EE optimization problem is formulated and studied. In Section V, numerical results are shown. Finally, Section VI concludes the paper.

Notation: The main symbols and functions used in the present paper are reported in Table I.

II. SYSTEM MODEL

In this section, the network model is introduced. With the exception of the load model, we focus our attention on a system where the *standard modeling assumptions* hold. One of the main aims of the present paper is, in fact, to highlight the differences between currently available analytical frameworks and the new definition of PSE that is introduced. The proposed approach can be readily generalized to more advanced system models, such as that recently adopted in [5].

A. Cellular Network Modeling

A downlink cellular network is considered. The BSs are modeled as points of a homogeneous PPP, denoted by Ψ_{BS}, of density λ_{BS}. The MTs are modeled as another homogeneous PPP, denoted by Ψ_{MT}, of density λ_{MT}. Ψ_{BS} and Ψ_{MT} are independent of each other. The BSs and MTs are equipped with a single omnidirectional antenna. Each BS transmits with a constant power denoted by $\mathrm{P_{tx}}$. The analytical frameworks are developed for the typical MT, denoted by $\mathrm{MT_0}$, that is located at the origin (Slivnyak theorem [28, Th. 1.4.5]). The BS serving $\mathrm{MT_0}$ is denoted by $\mathrm{BS_0}$. The cell association criterion is introduced in Section II-C. The subscripts 0, i and n identify the intended link, a generic interfering link, and a generic BS-to-MT link. The set of interfering BSs is denoted by $\Psi_{\mathrm{BS}}^{(\mathrm{I})}$. As for data transmission, the network operates in the interference-limited regime, i.e., the noise is negligible compared with the inter-cell interference.

TABLE I
SUMMARY OF MAIN SYMBOLS AND FUNCTIONS USED THROUGHOUT THE PAPER.

Symbol/Function	Definition
$\mathbb{E}\{\cdot\}$, $\Pr\{\cdot\}$	Expectation operator, probability measure
λ_{BS}, λ_{MT}	Density of base stations, mobile terminals
Ψ_{BS}, Ψ_{MT}, $\Psi_{\mathrm{BS}}^{(\mathrm{I})}$	PPP of base stations, mobile terminals, interfering base stations
BS_0, BS_i, BS_n	Serving, interfering, generic base station
P_{tx}, $\mathrm{P}_{\mathrm{circ}}$, $\mathrm{P}_{\mathrm{idle}}$	Transmit, circuits, idle power consumption of base stations
r_n, g_n	Distance, fading power gain of a generic link
$l(\cdot)$, L_n, L_0	Path-loss, shorthand of path-loss, path-loss of intended link
κ, $\beta > 0$	Path-loss constant, slope (exponent)
B_{W}, N_0	Transmission bandwidth, noise power spectral density
$\sigma_{\mathrm{N}}^2 = \mathrm{B}_{\mathrm{W}}\mathrm{N}_0$, $I_{\mathrm{agg}}(\cdot)$	Noise variance, aggregate other-cell interference
γ_{D}, γ_{A}	Reliability threshold for decoding, cell association
$\mathcal{L}(x) = 1 - (1 + x/\alpha)^{-\alpha}$, $\alpha = 3.5$	Probability that a base station is in transmission mode
$f_X(\cdot)$, $F_X(\cdot)$	Probability density/mass, cumulative distribution/mass function of X
$\mathbb{1}(\cdot)$, $_2F_1(\cdot,\cdot,\cdot,\cdot)$, $\Gamma(\cdot)$	Indicator function, Gauss hypergeometric function, gamma function
$\max\{x,y\}$, $\min\{x,y\}$	Maximum, minimum between x and y
$\Upsilon = {}_2F_1(-2/\beta, 1, 1 - 2/\beta, -\gamma_{\mathrm{D}}) - 1 \geq 0$	Shorthand
$\mathcal{Q}(x,y,z) = 1 - \exp\left(-\pi x(y/\eta)^{2/\beta}(1 + \Upsilon\mathcal{L}(z))\right)$	Shorthand with $\eta = \kappa\sigma_{\mathrm{N}}^2\gamma_{\mathrm{A}}$
SIR, $\overline{\mathrm{SNR}}$	Signal-to-interference-ratio, average signal-to-noise-ratio
$\mathrm{P}_{\mathrm{cov}}$, PSE, $\mathrm{P}_{\mathrm{grid}}$	Coverage, potential spectral efficiency, network power consumption
$\dot{z}_x(x,y)$, $\ddot{z}_x(x,y)$	First-order, second-order derivative with respect to x

B. Channel Modeling

For each BS-to-MT link, path-loss and fast-fading are considered. Shadowing is not explicitly taken into account because its net effect lies in modifying the density of the BSs [5]. All BS-to-MT links are assumed to be mutually independent and identically distributed (i.i.d.).

a) Path-Loss: Consider a generic BS-to-MT link of length r_n. The path-loss is $l(r_n) = \kappa r_n^{\beta}$, where κ and β are the path-loss constant and the path-loss slope (exponent). For simplicity, only the unbounded path-loss model is studied in the present paper. The analysis of more general path-loss models is an interesting but challenging generalization that is left to future research [29].

b) Fast-Fading: Consider a generic BS-to-MT link. The power gain due to small-scale fading is assumed to follow an exponential distribution with mean Ω. Without loss of generality, $\Omega = 1$ is assumed. The power gain of a generic BS-to-MT link is denoted by g_n.

C. Cell Association Criterion

A cell association criterion based on the highest average received power is assumed. Let $\mathrm{BS}_n \in \Psi_{\mathrm{BS}}$ denote a generic BS of the network. The serving BS, BS_0, is obtained as follows:

$$\begin{aligned}
\mathrm{BS}_0 &= \arg\max_{\mathrm{BS}_n \in \Psi_{\mathrm{BS}}}\{1/l(r_n)\} \\
&= \arg\max_{\mathrm{BS}_n \in \Psi_{\mathrm{BS}}}\{1/L_n\}
\end{aligned} \quad (2)$$

where the shorthand $L_n = l(r_n)$ is used. As for the intended link, $L_0 = \min_{r_n \in \Psi_{\mathrm{BS}}}\{L_n\}$ holds.

D. Load Modeling

Based on (2), several or no MTs can be associated to a generic BS. In the latter case, the BS transmits zero power, i.e., $\mathrm{P}_{\mathrm{tx}} = 0$, and, thus, it does not generate inter-cell interference. In the former case, on the other hand, two load models are studied and compared against each other. The main objective is to analyze the impact of the load model on the power consumption and EE of cellular networks. Further details are provided in the sequel. Let N_{MT} denote the number of MTs associated to a generic BS and B_{W} denote the transmission bandwidth available to each BS. If $\mathrm{N}_{\mathrm{MT}} = 1$, for both load models, the single MT associated to the BS is scheduled for transmission and the entire bandwidth, B_{W}, and transmit power, P_{tx}, are assigned to it.

a) Load Model 1: Exclusive Allocation of Bandwidth and Power to a Randomly Selected MT: If $\mathrm{N}_{\mathrm{MT}} > 1$, the BS randomly selects, at each transmission instance, a single MT among the N_{MT} associated to it. Also, the BS allocates the entire transmission bandwidth, B_{W}, and the total transmit power, P_{tx}, to it. The random scheduling of the MTs at each transmission instance ensures that, in the long term, all the MTs associated to a BS are scheduled for transmission.

b) Load Model 2: Equal Allocation of Bandwidth and Power Among All the MTs: If $\mathrm{N}_{\mathrm{MT}} > 1$, the BS selects, at each transmission instance, all the N_{MT} MTs associated to it. The BS equally splits the available transmission bandwidth, B_{W}, and evenly spreads the available transmit power, P_{tx}, among the N_{MT} MTs. Thus, the bandwidth and power are viewed as continuous resources by the BS's scheduler: each MT is assigned a bandwidth equal to $\mathrm{B}_{\mathrm{W}}/\mathrm{N}_{\mathrm{MT}}$ and the power

spectral density at the detector's (i.e., the typical MT, MT_0) input is equal to P_{tx}/B_W.

In the sequel, we show that the main difference between the two load models lies in the power consumption of the BSs. In simple terms, the more MTs are scheduled for transmission the higher the static power consumption of the BSs is. The analysis of general load models, e.g., based on a discrete number of resource blocks [5], is left to future research due to space limits.

E. Power Consumption Modeling

In the considered system model, the BSs can operate in two different modes: i) they are in idle mode if no MTs are associated to them and ii) they are in transmission mode if at least one MT is associated to them. The widespread linear power consumption model for the BSs is adopted [1], [30], which accounts for the power consumption due to the transmit power, P_{tx}, the static (circuit) power, P_{circ}, and the idle power, P_{idle}. If the BS is in idle mode, its power consumption is equal to P_{idle}. If the BS is in transmission mode, its power consumption is a function of P_{tx}, P_{circ}, and depends on the load model. Further details are provided in the sequel. In the present paper, based on physical considerations, the inequalities $0 \leq P_{idle} \leq P_{circ}$ are assumed.

III. A NEW ANALYTICAL FORMULATION OF THE PSE

In this section, we introduce and motivate a new definition of coverage probability, P_{cov}, and PSE, which overcomes the limitations of currently available analytical frameworks and is suitable for system-level optimization (see Section I-A). All symbols are defined in Table I.

Definition 1: Let γ_D and γ_A be the reliability thresholds for the successful decoding of information data and for the successful detection of the serving BS, BS_0, respectively. The coverage probability, P_{cov}, of the typical MT, MT_0, is defined as follows:

$$P_{cov}(\gamma_D, \gamma_A)$$
$$= \begin{cases} \Pr\left\{ SIR \geq \gamma_D, \overline{SNR} \geq \gamma_A \right\} & \text{if} \quad MT_0 \text{ is selected} \\ 0 & \text{if} \quad MT_0 \text{ is not selected} \end{cases}$$
(3)

where the Signal-to-Interference-Ratio (SIR) and the average Signal-to-Noise-Ratio ($\overline{SNR}$) can be formulated, for the network model under analysis, as follows:

$$SIR = \frac{P_{tx} g_0 / L_0}{\sum_{BS_i \in \Psi_{BS}^{(I)}} P_{tx} g_i / L_i \mathbb{1}(L_i > L_0)}$$
$$\overline{SNR} = \frac{P_{tx}/L_0}{\sigma_N^2}.$$
(4)

Remark 1: The definition of P_{cov} in (3) reduces to the conventional one if $\gamma_A = 0$ [3]. □

Remark 2: The average SNR, $\overline{SNR}$, in (4) is averaged with respect to the fast fading. The SIR depends, on the other hand, on fast fading. This choice is discussed in the sequel. □

Remark 3: The new definition of coverage probability, P_{cov}, in (3) is in agreement with the cell selection criterion specified
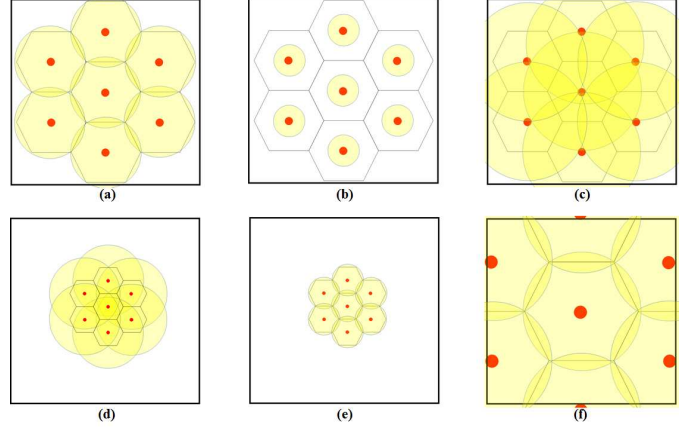

Fig. 1. Illustration of the interplay between P_{tx} and λ_{BS}. For simplicity, only a cluster of seven BSs is represented by keeping the size of the region of interest (square box) the same. The inter-site distance of the BSs (represented as red dots), i.e., the size of the hexagonal cells, is determined by λ_{BS}. The shape of the cells depends on the cell association in (2). The circular shaded disk (in light yellow) represents the actual coverage region of the BSs that is determined by P_{tx}: i) a MT inside the disk receives a sufficiently good signal to detect the BS and to get associated with it, ii) a MT outside the disk cannot detect the BS and is not in coverage. The sub-figures (a)-(c) are obtained by assuming the same λ_{BS} but a different P_{tx}. The sub-figures (d) and (e) are obtained by considering a λ_{BS} greater than that of sub-figures (a)-(c) but keeping the same P_{tx} as sub-figures (a) and (b), respectively. The sub-figure (f) is obtained by considering a λ_{BS} smaller than that of sub-figure (c) but keeping the same P_{tx} as it. We observe that, for a given λ_{BS}, the transmit power P_{tx} is appropriately chosen in sub-figures (a), (e) and (f). P_{tx} is, on the other hand, under-provisioned in sub-figure (b) and over-provisioned in sub-figures (c) and (d). In the first case, the MTs are not capable of detecting the BS throughout the entire cell, i.e., a high outage probability is expected. In the second case, the BSs emit more power than what is actually needed, which results in a high power consumption.

by the 3rd Generation Partnership Project (3GPP) [31, Sec. 5.2.3.2]. □

a) Motivation for the New Definition of P_{cov}: The motivation for the new definition of coverage probability originates from the inherent limitations of the conventional definition (obtained by setting $\gamma_A = 0$ in (3)), which prevents one from taking into account the strong interplay between the transmit power and the density of the BSs for optimal cellular networks planning. In fact, the authors of [3] have shown that, in the interference-limited regime, P_{cov} is independent of the transmit power of the BSs. If, in addition, a fully-loaded model is assumed, i.e., $\lambda_{MT}/\lambda_{BS} \gg 1$, then P_{cov} is independent of the density of BSs as well. This is known as the invariance property of P_{cov} as a function of P_{tx} and λ_{BS} [5]. The tight interplay between P_{tx} and λ_{BS} is, on the other hand, illustrated in Fig. 1, where, for ease of representation, an hexagonal cellular layout is considered. Similar conclusions apply to the PPP-based cellular layout studied in the present paper. In Fig. 1, it is shown that, for a given λ_{BS}, P_{tx} needs to be appropriately chosen in order to guarantee that, for any possible location of MT_0 in the cell, two conditions are fulfilled: i) the MT receives a sufficiently good signal quality, i.e., the average SNR is above a given threshold, γ_A, that ensures a successful cell association, i.e., to detect the presence (pilot signal) of the serving BS and ii) the BSs do not over-provision P_{tx}, which results in an unnecessary increase of the

power consumption. It is expected, therefore, that an optimal value of P_{tx} given λ_{BS} and an optimal value of λ_{BS} given P_{tx} that optimize EE exist [32].

b) Advantages of the New Definition of P_{cov}: The new definition of P_{cov} allows one to overcome the limitations of the conventional definition and brings about two main advantages. The first advantage originates from direct inspection of (4). In the conventional definition of P_{cov}, only the SIR is considered and the transmit power of the BSs, P_{tx}, cancels out between numerator and denominator. This is the reason why P_{cov} is independent of P_{tx}. In the proposed new definition, on the other hand, P_{tx} explicitly appears in the second constraint and does not cancel out. The density of the BSs, λ_{BS}, appears implicitly in the distribution of the path-loss of the intended link, L_0. The mathematical details are provided in the sequel. The second inequality, as a result, allows one to explicitly account for the interplay between P_{tx} and λ_{BS} (shown in Fig. 1). If λ_{BS} increases (decreases), in particular, L_0 decreases (increases) in statistical terms. This implies that P_{tx} can be decreased (increased) while still ensuring that the average SNR is above γ_A. The second advantage is that the new definition of P_{cov} is still mathematically tractable and the PSE is formulated in a closed-form expression. This is detailed in *Proposition 1*.

Remark 4: The new definition of P_{cov} in (3) is based on the actual value of L_0 because a necessary condition for the typical MT to be in coverage is that it can detect the pilot signal of at least one BS during the cell association. If the BS that provides the highest average received power cannot be detected, then any other BSs cannot be detected either. The second constraint on the definition of P_{cov}, in addition, is based on the average SNR, i.e., the SNR averaged with respect to the fast fading, because the cell association is performed based on long-term statistics, i.e., based on the path-loss in the present paper, in order to prevent too frequent handovers. □

Remark 5: Compared with the conventional definition of coverage based on the Signal-to-Interference+Noise-Ratio (SINR) [3], the new definition in (3) is conceptually different. Equation (3) accounts for the signal quality during both the cell association and data transmission phases. The definition of coverage based on the SINR, on the other hand, accounts for the signal quality only during the data transmission phase. In spite of this fundamental difference, P_{cov} in (3) may be interpreted as an approximation for the coverage probability based on the SINR, and, more precisely, as an alternative method to incorporate the thermal noise into the problem formulation. Compared with the coverage based on the SINR, however, the new definition in (3) accounts for the impact of thermal noise when it is the dominant factor, i.e., during the cell association phase when the inter-cell interference can be ignored as orthogonal pilot signals are used. □

Remark 6: Figure 1 highlights that the new definition of coverage in (3) is not only compliant with [31] but it has a more profound motivation and wider applicability. In PPP-based cellular networks, in contrast to regular grid-based network layouts, the size and shape of the cells are random. This implies that it is not possible to identify a relation, based on pure geometric arguments, between the cell size and the transmit power of the BSs that makes the constraint on $\overline{SNR}$

in (3) ineffective in practice. In equivalent terms, in this case, the threshold γ_A may turn out to be sufficiently small to render the constraint on $\overline{SNR}$ ineffective. This is, e.g., the approach employed in [32, Eq. (1)], where the relation between the transmit power and density of BSs is imposed a priori based on the path-loss. In practice, however, cellular networks are irregularly deployed, which makes the optimal relation between the transmit power and density of BSs difficult to identify because of the coexistence of cells of small and large sizes. The constraint on $\overline{SNR}$ in (3) allows one to take into account the interplay between the transmit power and density of BSs in irregular (realistic) cellular network deployments. □

A. Analytical Formulation of the PSE

In this section, we provide the mathematical definitions of the PSE for the two load models introduced in Section II-D. They are summarized in the following two lemmas, which constitute the departing point to obtain the closed-form analytical frameworks derived in Section III-B.

Remark 7: The PSE is defined from the perspective of the typical MT, MT_0 rather than from the perspective of the typical cell (or BS). This implies that the proposed approach allows one to characterize the PSE of the so-called Crofton cell, which is the cell that contains MT_0. This approach is commonly used in the literature and is motivated by the lack of results on the explicit distribution of the main geometrical characteristics of the typical cell of a Voronoi tessellation. Further details on the Crofton and typical cells are available in [34] and [35]. □

Let $\bar{N}_{MT}$ be the number of MTs that lie in the cell of the typical MT, MT_0, with the exception of MT_0. $\bar{N}_{MT}$ is a discrete random variable whose probability mass function in the considered system model can be formulated, in an approximated closed-form expression, as [33, Eq. (3)]:

$$f_{\bar{N}_{MT}}(u) = \Pr\{\bar{N}_{MT} = u\}$$
$$\approx \frac{3.5^{4.5}\Gamma(u + 4.5)(\lambda_{MT}/\lambda_{BS})^u}{\Gamma(4.5)\Gamma(u + 1)(3.5 + \lambda_{MT}/\lambda_{BS})^{u+4.5}}. \quad (5)$$

Remark 8: The probability mass function in (5) is an approximation because it is based on the widely used empirical expression of the probability density function of the area of the Voronoi cells in [36, Eq. (1)]. A precise formula for the latter probability density function is available in [37]. It is, however, not used in the present paper due to its mathematical intractability, as recently remarked in [26]. Throughout the rest of the paper, for simplicity, we employ the sign of equality ("=") in all the analytical formulas that rely *solely* on the approximation in (5). This is to make explicit that our analytical frameworks are not based on any other hidden approximations. □

Based on (5), a formal mathematical formulation for the PSE is given as follows.

Lemma 1: Let *Load Model 1* be assumed. The PSE (bit/sec/m²) can be formulated as shown in (6) at the top of this page.

Proof: It follows from the definition of PSE [5], where (a) originates from the fact that MT_0 is scheduled for transmission

$$\text{PSE}\left(\gamma_{\text{D}},\gamma_{\text{A}}\right) = \mathbb{E}_{\bar{\text{N}}_{\text{MT}}}\left\{\text{PSE}\left(\gamma_{\text{D}},\gamma_{\text{A}}|\bar{\text{N}}_{\text{MT}}\right)\right\}$$
$$\stackrel{(a)}{=}\lambda_{\text{MT}}\text{B}_{\text{W}}\log_2\left(1+\gamma_{\text{D}}\right)\Pr\left\{\text{SIR}\geq\gamma_{\text{D}},\overline{\text{SNR}}\geq\gamma_{\text{A}}\right\}\Pr\left\{\bar{\text{N}}_{\text{MT}}=0\right\}$$
$$+\sum_{u=1}^{+\infty}\lambda_{\text{MT}}\text{B}_{\text{W}}\log_2\left(1+\gamma_{\text{D}}\right)\frac{1}{u+1}\Pr\left\{\text{SIR}\geq\gamma_{\text{D}},\overline{\text{SNR}}\geq\gamma_{\text{A}}\right\}\Pr\left\{\bar{\text{N}}_{\text{MT}}=u\right\} \quad (6)$$
$$=\lambda_{\text{MT}}\text{B}_{\text{W}}\log_2\left(1+\gamma_{\text{D}}\right)\Pr\left\{\text{SIR}\geq\gamma_{\text{D}},\overline{\text{SNR}}\geq\gamma_{\text{A}}\right\}\sum_{u=0}^{+\infty}\frac{\Pr\left\{\bar{\text{N}}_{\text{MT}}=u\right\}}{u+1}.$$

$$\text{PSE}\left(\gamma_{\text{D}},\gamma_{\text{A}}\right) = \mathbb{E}_{\bar{\text{N}}_{\text{MT}}}\left\{\text{PSE}\left(\gamma_{\text{D}},\gamma_{\text{A}}|\bar{\text{N}}_{\text{MT}}\right)\right\}$$
$$\stackrel{(b)}{=}\sum_{u=0}^{+\infty}\lambda_{\text{MT}}\frac{\text{B}_{\text{W}}}{u+1}\log_2\left(1+\gamma_{\text{D}}\right)\Pr\left\{\text{SIR}\geq\gamma_{\text{D}},\overline{\text{SNR}}\geq\gamma_{\text{A}}\right\}\Pr\left\{\bar{\text{N}}_{\text{MT}}=u\right\} \quad (7)$$
$$=\lambda_{\text{MT}}\text{B}_{\text{W}}\log_2\left(1+\gamma_{\text{D}}\right)\Pr\left\{\text{SIR}\geq\gamma_{\text{D}},\overline{\text{SNR}}\geq\gamma_{\text{A}}\right\}\sum_{u=0}^{+\infty}\frac{\Pr\left\{\bar{\text{N}}_{\text{MT}}=u\right\}}{u+1}.$$

with unit probability if it is the only MT in the cell, while it is scheduled for transmission with probability $1/(u+1)$ if there are other u MTs in the cell. $\qquad\square$

Lemma 2: Let *Load Model 2* be assumed. The PSE (bit/sec/m^2) can be formulated as shown in (7) at the top of this page.

Proof: It follows from the definition of PSE [5], where (b) originates from the fact that MT_0 is scheduled for transmission with unit probability but the bandwidth is equally allocated among the MTs in the cell, i.e., each of the $u+1$ MTs is given a bandwidth equal to $\text{B}_{\text{W}}/(u+1)$. $\qquad\square$

Remark 9: By comparing (6) and (7), we note that the same PSE is obtained for both load models. This originates from the fact that P_{cov} in (3) is independent of the number of MTs in the cell. This property follows by direct inspection of (4) and has been used in the proof of *Lemma 1* and *Lemma 2*. As far as the first load model is concerned, this property originates from the fact that a single MT is scheduled at every transmission instance. It is, however, less intuitive for the second load model. In this latter case, as mentioned in Section II-D, P_{tx} and B_{W} are viewed as continuous resources by the BS's scheduler. The transmit power per unit bandwidth of both intended and interfering links is equal to $\text{P}_{\text{tx}}/\text{B}_{\text{W}}$. Regardless of the number of MTs available in the interfering cells, MT_0 "integrates" this transmit power per unit bandwidth over the bandwidth allocated to it, which depends on the total number of MTs in its own cell. Let the number of these MTs be $u+1$. Thus, the receiver bandwidth of MT_0 is $\text{B}_{\text{W}}/(u+1)$. This implies that the received power (neglecting path-loss and fast-fading) of both intended and interfering links is $\text{P}_{\text{rx}}=(\text{P}_{\text{tx}}/\text{B}_{\text{W}})\left(\text{B}_{\text{W}}/(u+1)\right)=\text{P}_{\text{tx}}/(u+1)$. As a result, the number of MTs, $u+1$, cancels out in the SIR of (4). Likewise, the received average SNR (neglecting the path-loss) is equal to $\text{P}_{\text{rx}}/(\text{N}_0\text{B}_{\text{W}}/(u+1))=(\text{P}_{\text{tx}}/(u+1))/(\text{N}_0\text{B}_{\text{W}}/(u+1))=\text{P}_{\text{tx}}/\sigma_{\text{N}}^2$, which is independent of the number of MTs, $u+1$, and agrees with the definition of average SNR in (4). In the next section, we show that the load models are not equivalent

in terms of network power consumption. $\qquad\square$

B. Closed-Form Expressions of PSE and P_{grid}

In this section, we introduce new closed-form analytical frameworks for computing the PSE. We provide, in addition, closed-form expressions of the network power consumption for the two load models under analysis. These results are summarized in the following three propositions.

Let N_{MT} be the number of MTs that lie in an arbitrary cell. The probability that the BS is in idle mode, $\mathbb{P}_{\text{BS}}^{(\text{idle})}$, and in transmission mode, $\mathbb{P}_{\text{BS}}^{(\text{tx})}$, can be formulated as follows [33, Prop. 1]:

$$\mathbb{P}_{\text{BS}}^{(\text{idle})} = \Pr\left\{\text{N}_{\text{MT}}=0\right\} = 1-\mathcal{L}\left(\lambda_{\text{MT}}/\lambda_{\text{BS}}\right)$$
$$\mathbb{P}_{\text{BS}}^{(\text{tx})} = \Pr\left\{\text{N}_{\text{MT}}\geq 1\right\} = 1-\mathbb{P}_{\text{BS}}^{(\text{idle})} = \mathcal{L}\left(\lambda_{\text{MT}}/\lambda_{\text{BS}}\right) \quad (8)$$

where $\mathcal{L}\left(\cdot\right)$ is defined in Table II. Using (8), PSE and P_{grid} are given in the following propositions.

Proposition 1: Consider either *Load Model 1* or *Load Model 2*. Assume notation and functions given in Tables I and II. The PSE (bit/sec/m^2) can be formulated, in closed-form, as follows:

$$\text{PSE}\left(\gamma_{\text{D}},\gamma_{\text{A}}\right) = \text{B}_{\text{W}}\log_2\left(1+\gamma_{\text{D}}\right)\frac{\lambda_{\text{BS}}\mathcal{L}\left(\lambda_{\text{MT}}/\lambda_{\text{BS}}\right)}{1+\Upsilon\mathcal{L}\left(\lambda_{\text{MT}}/\lambda_{\text{BS}}\right)} \quad (9)$$
$$\times\mathcal{Q}\left(\lambda_{\text{BS}},\text{P}_{\text{tx}},\lambda_{\text{MT}}/\lambda_{\text{BS}}\right).$$

Proof: See Appendix A. $\qquad\square$

Corollary 1: If $\gamma_{\text{A}}=0$, i.e., the conventional definition of P_{cov} is used, the PSE in (9) simplifies as follows:

$$\text{PSE}\left(\gamma_{\text{D}},\gamma_{\text{A}}=0\right) = \text{B}_{\text{W}}\log_2\left(1+\gamma_{\text{D}}\right)\frac{\lambda_{\text{BS}}\mathcal{L}\left(\lambda_{\text{MT}}/\lambda_{\text{BS}}\right)}{1+\Upsilon\mathcal{L}\left(\lambda_{\text{MT}}/\lambda_{\text{BS}}\right)}. \quad (10)$$

If, in addition, $\lambda_{\text{MT}}/\lambda_{\text{BS}}\gg 1$, the PSE in (9) reduces to (1).

Proof: It follows because $\mathcal{Q}\left(\cdot,\cdot,\cdot\right)=1$ if $\gamma_{\text{A}}=0$ and $\mathcal{L}\left(\lambda_{\text{MT}}/\lambda_{\text{BS}}\gg 1\right)\to 1$. $\qquad\square$

Remark 10: Corollary 1 substantiates the comments made above in this section about the need of a new definition of PSE,

$$\mathrm{EE}\left(\mathrm{P_{tx}}, \lambda_{\mathrm{BS}}\right) = \frac{\mathrm{PSE}}{\mathrm{P_{grid}}} = \frac{\mathrm{B_W} \log_2\left(1 + \gamma_{\mathrm{D}}\right) \mathcal{L}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) \mathcal{Q}\left(\lambda_{\mathrm{BS}}, \mathrm{P_{tx}}, \lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)}{\left[1 + \Upsilon\mathcal{L}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)\right]\left[\mathcal{L}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)\left(\mathrm{P_{tx}} + \mathrm{P_{circ}} - \mathrm{P_{idle}}\right) + \mathrm{P_{idle}} + \mathcal{M}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)\mathrm{P_{circ}}\right]}. \quad (14)$$

as well as the advantages of the proposed analytical formulation. In particular, (10) confirms that the PSE is independent of $\mathrm{P_{tx}}$ if $\gamma_{\mathrm{A}} = 0$ and that the PSE is independent of $\mathrm{P_{tx}}$ and λ_{BS} if fully-loaded conditions hold, i.e., $\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}} \gg 1$. $\square$

Proposition 2: Let *Load Model 1* be assumed. $\mathrm{P_{grid}}$ (Watt/m^2) can be formulated as follows:

$$\begin{aligned} \mathrm{P_{grid}^{(1)}} = &\lambda_{\mathrm{BS}}\left(\mathrm{P_{tx}} + \mathrm{P_{circ}}\right)\mathcal{L}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) \\ &+ \lambda_{\mathrm{BS}}\mathrm{P_{idle}}\left(1 - \mathcal{L}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)\right). \end{aligned} \quad (11)$$

Proof: The network power consumption is obtained by multiplying the average number of BSs per unit area, i.e., λ_{BS}, and the average power consumption of a generic BS, which is $\mathrm{P_{tx}} + \mathrm{P_{circ}}$ if the BS operates in transmission mode, i.e., with probability $\mathcal{L}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)$, and $\mathrm{P_{idle}}$ if the BS operates in idle mode, i.e., with probability $1 - \mathcal{L}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)$. $\square$

Proposition 3: Let *Load Model 2* be assumed. $\mathrm{P_{grid}}$ (Watt/m^2) can be formulated as follows:

$$\begin{aligned} \mathrm{P_{grid}^{(2)}} = &\lambda_{\mathrm{BS}}\mathrm{P_{tx}}\mathcal{L}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) \\ &+ \lambda_{\mathrm{MT}}\mathrm{P_{circ}} + \lambda_{\mathrm{BS}}\mathrm{P_{idle}}\left(1 - \mathcal{L}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)\right). \end{aligned} \quad (12)$$

Proof: It is similar to the proof of *Proposition 2*. The difference is that the power dissipation of a generic BS that operates in transmission mode is, in this case, equal to $\mathrm{P_{tx}} + \mathrm{P_{circ}} \sum_{u=1}^{+\infty} u \Pr\left\{\mathrm{N_{MT}} = u\right\} = \mathrm{P_{tx}} + \mathrm{P_{circ}}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)$, where $\mathrm{N_{MT}}$ is the number of MTs in the cell and the last equality follows from [33, Lemma 1]. $\square$

Remark 11: The power consumption models obtained in (11) and (12), which account for the transmit, circuits, and idle power consumption of the BSs, have been used, under some simplifying assumptions, in previous research works focused on the analysis of the EE of cellular networks. Among the many research works, an early paper that has adopted this approach under the assumption of fully-loaded BSs and of having a single active MT per cell is [8]. $\square$

Remark 12: Since $\mathcal{L}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) \leq \lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}$ for every $\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}} \geq 0$, we conclude that $\mathrm{P_{grid}^{(2)}} \geq \mathrm{P_{grid}^{(1)}}$ by assuming the same $\mathrm{P_{tx}}$ and λ_{BS} for both load models. This originates from the fact that, in the present paper, we assume that the circuits power consumption increases with the number of MTs that are served by the BSs. It is unclear, however, the best load model to be used from the EE standpoint, especially if $\mathrm{P_{tx}}$ and λ_{BS} are optimized to maximize their respective EEs. In other words, the optimal $\mathrm{P_{tx}}$ and λ_{BS} that maximize the EE of each load model may be different, which may lead to different optimal EEs. The trade-off between the optimal PSE and the optimal EE is analyzed numerically in Section V for both load models. $\square$

IV. System-Level EE Optimization: Formulation and Solution

In this section, we formulate a system-level EE optimization problem and comprehensively analyze its properties. For convenience of analysis, we introduce the following auxiliary function (LM = Load Model):

$$\begin{aligned} &\mathcal{M}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) \\ &= \begin{cases} 0 & \text{if } \mathrm{LM}-1 \text{ is assumed} \\ \lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}} - \mathcal{L}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) & \text{if } \mathrm{LM}-2 \text{ is assumed.} \end{cases} \end{aligned} \quad (13)$$

A unified formulation of the EE (bit/Joule) for the cellular network under analysis is provided in (14) shown at the top of this page, where the parameters of interest from the optimization standpoint, i.e., $\mathrm{P_{tx}}$ and λ_{BS}, are explicitly highlighted. In the rest of the present paper, all the other parameters are assumed to be given.

A. Preliminaries

For ease of presentation, we report some lemmas that summarize structural properties of the main functions that constitute (14). Some lemmas are stated without proof because they are obtained by simply studying the sign of the first-order and second-order derivatives of the function with respect to the variable of interest and by keeping all the other variables fixed. Functions of interest for this section are given in Table II. Also, we define $\Delta\mathrm{P} = \mathrm{P_{circ}} - \mathrm{P_{idle}} \geq 0$.

Lemma 3: The function $\mathcal{L}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)$ fulfills the following properties with respect to λ_{BS} (assuming λ_{MT} fixed): i) $\mathcal{L}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) \geq 0$ for $\lambda_{\mathrm{BS}} \geq 0$; ii) $\mathcal{L}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) = 1$ if $\lambda_{\mathrm{BS}} \to 0$; iii) $\mathcal{L}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) = 0$ if $\lambda_{\mathrm{BS}} \to \infty$; iv) $\dot{\mathcal{L}}_{\lambda_{\mathrm{BS}}}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) \leq 0$ for $\lambda_{\mathrm{BS}} \geq 0$; v) $\ddot{\mathcal{L}}_{\lambda_{\mathrm{BS}}}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) \leq 0$ for $\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}} \geq 2\alpha/(\alpha-1) = 2.8$; and vi) $\ddot{\mathcal{L}}_{\lambda_{\mathrm{BS}}}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) \geq 0$ for $\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}} \leq 2\alpha/(\alpha-1) = 2.8$.

Lemma 4: As far as Load Model 2 is concerned, the function $\mathcal{M}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)$ fulfills the following properties with respect to λ_{BS} (assuming λ_{MT} fixed): i) $\mathcal{M}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) \geq 0$ for $\lambda_{\mathrm{BS}} \geq 0$; ii) $\mathcal{M}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) \to \infty$ if $\lambda_{\mathrm{BS}} \to 0$; iii) $\mathcal{M}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) = 0$ if $\lambda_{\mathrm{BS}} \to \infty$; iv) $\dot{\mathcal{M}}_{\lambda_{\mathrm{BS}}}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) \leq 0$ for $\lambda_{\mathrm{BS}} \geq 0$; and v) $\ddot{\mathcal{M}}_{\lambda_{\mathrm{BS}}}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) \geq 0$ for $\lambda_{\mathrm{BS}} \geq 0$.

Lemma 5: The function $\mathcal{Q}\left(\lambda_{\mathrm{BS}}, \mathrm{P_{tx}}, \lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)$ fulfills the following properties with respect to $\mathrm{P_{tx}}$: i) $\mathcal{Q}\left(\lambda_{\mathrm{BS}}, \mathrm{P_{tx}}, \lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) \geq 0$ for $\mathrm{P_{tx}} \geq 0$; ii) $\mathcal{Q}\left(\lambda_{\mathrm{BS}}, \mathrm{P_{tx}}, \lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) = 0$ if $\mathrm{P_{tx}} \to 0$; iii) $\mathcal{Q}\left(\lambda_{\mathrm{BS}}, \mathrm{P_{tx}}, \lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) = 1$ if $\mathrm{P_{tx}} \to \infty$; iv) $\dot{\mathcal{Q}}_{\mathrm{P_{tx}}}\left(\lambda_{\mathrm{BS}}, \mathrm{P_{tx}}, \lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) \geq 0$ for $\mathrm{P_{tx}} \geq 0$; and v) $\ddot{\mathcal{Q}}_{\mathrm{P_{tx}}}\left(\lambda_{\mathrm{BS}}, \mathrm{P_{tx}}, \lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) \leq 0$ for $\mathrm{P_{tx}} \geq 0$.

Proof: The result in v) follows from $\ddot{\mathcal{Q}}_{\mathrm{P_{tx}}}\left(\cdot, \cdot, \cdot\right)$ in Table II, because iv) and $\beta > 2$ hold. $\square$

Lemma 6: The function $\mathcal{Q}\left(\lambda_{\mathrm{BS}}, \mathrm{P_{tx}}, \lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)$ fulfills the following properties with respect to λ_{BS} (assuming λ_{MT} fixed): i) $\mathcal{Q}\left(\lambda_{\mathrm{BS}}, \mathrm{P_{tx}}, \lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) \geq 0$ for $\lambda_{\mathrm{BS}} \geq 0$; ii) $\mathcal{Q}\left(\lambda_{\mathrm{BS}}, \mathrm{P_{tx}}, \lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) = 0$ if $\lambda_{\mathrm{BS}} \to 0$; iii) $\mathcal{Q}\left(\lambda_{\mathrm{BS}}, \mathrm{P_{tx}}, \lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) = 1$ if $\lambda_{\mathrm{BS}} \to \infty$; iv)

TABLE II
SUMMARY OF MAIN AUXILIARY FUNCTIONS USED THROUGHOUT THE PAPER.

Function Definition

$$\mathcal{L}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) = 1 - \left(1 + (1/\alpha)\,\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)^{-\alpha}$$

$$\mathcal{M}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) = \lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}} - \mathcal{L}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)$$

$$\mathcal{Q}\left(\lambda_{\mathrm{BS}}, \mathrm{P}_{\mathrm{tx}}, \lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) = 1 - \exp\left(-\pi\lambda_{\mathrm{BS}}(\mathrm{P}_{\mathrm{tx}}/\eta)^{2/\beta}\left(1 + \Upsilon\mathcal{L}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)\right)\right)$$

$$\dot{\mathcal{Q}}_{\mathrm{P}_{\mathrm{tx}}}\left(\lambda_{\mathrm{BS}}, \mathrm{P}_{\mathrm{tx}}, \lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) = \pi\lambda_{\mathrm{BS}}(1/\eta)^{2/\beta}\left(2/\beta\right)\left(1 + \Upsilon\mathcal{L}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)\right)\mathrm{P}_{\mathrm{tx}}^{2/\beta-1}$$
$$\times \exp\left(-\pi\lambda_{\mathrm{BS}}(\mathrm{P}_{\mathrm{tx}}/\eta)^{2/\beta}\left(1 + \Upsilon\mathcal{L}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)\right)\right)$$

$$\ddot{\mathcal{Q}}_{\mathrm{P}_{\mathrm{tx}}}\left(\lambda_{\mathrm{BS}}, \mathrm{P}_{\mathrm{tx}}, \lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) = \pi\lambda_{\mathrm{BS}}(1/\eta)^{2/\beta}\left(2/\beta\right)\left(1 + \Upsilon\mathcal{L}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)\right)\mathrm{P}_{\mathrm{tx}}^{2/\beta-1}$$
$$\times\left[-\dot{\mathcal{Q}}_{\mathrm{P}_{\mathrm{tx}}}\left(\lambda_{\mathrm{BS}}, \mathrm{P}_{\mathrm{tx}}, \lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)\right]$$
$$+\pi\lambda_{\mathrm{BS}}(1/\eta)^{2/\beta}\left(2/\beta\right)\left(2/\beta - 1\right)\left(1 + \Upsilon\mathcal{L}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)\right)\mathrm{P}_{\mathrm{tx}}^{2/\beta-2}$$
$$\times \exp\left(-\pi\lambda_{\mathrm{BS}}(\mathrm{P}_{\mathrm{tx}}/\eta)^{2/\beta}\left(1 + \Upsilon\mathcal{L}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)\right)\right)$$

$$\dot{\mathcal{L}}_{\lambda_{\mathrm{BS}}}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) = -\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}^2\right)\left(1 + (1/\alpha)\,\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)^{-(\alpha+1)}$$

$$\dot{\mathcal{M}}_{\lambda_{\mathrm{BS}}}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) = -\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}^2\right)\left[1 - \left(1 + (1/\alpha)\,\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)^{-(\alpha+1)}\right]$$

$$\dot{\mathcal{Q}}_{\lambda_{\mathrm{BS}}}\left(\lambda_{\mathrm{BS}}, \mathrm{P}_{\mathrm{tx}}, \lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) = \pi(\mathrm{P}_{\mathrm{tx}}/\eta)^{2/\beta}\left[1 + \Upsilon\mathcal{L}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) + \Upsilon\lambda_{\mathrm{BS}}\dot{\mathcal{L}}_{\lambda_{\mathrm{BS}}}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)\right]$$
$$\times \exp\left(-\pi\lambda_{\mathrm{BS}}(\mathrm{P}_{\mathrm{tx}}/\eta)^{2/\beta}\left(1 + \Upsilon\mathcal{L}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)\right)\right)$$

$$\ddot{\mathcal{L}}_{\lambda_{\mathrm{BS}}}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) = \left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}^3\right)\left(1 + (1/\alpha)\,\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)^{-(\alpha+1)}$$
$$\times\left[2 - (1+\alpha)\,(1/\alpha)\,\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\left(1 + (1/\alpha)\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)^{-1}\right]$$

$$\ddot{\mathcal{M}}_{\lambda_{\mathrm{BS}}}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) = 2\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}^3\right)\left[1 - \left(1 + (1/\alpha)\,\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)^{-(\alpha+1)}\right]$$
$$+ (1+\alpha)\,(1/\alpha)\left(\lambda_{\mathrm{MT}}^2/\lambda_{\mathrm{BS}}^4\right)\left(1 + (1/\alpha)\,\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)^{-(\alpha+2)}$$

$$\mathcal{S}_{\mathcal{P}}\left(\mathrm{P}_{\mathrm{tx}}\right) = \mathcal{L}\left(\frac{\lambda_{\mathrm{MT}}}{\lambda_{\mathrm{BS}}}\right)\left[\frac{\mathcal{Q}(\lambda_{\mathrm{BS}},\mathrm{P}_{\mathrm{tx}},\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}})}{\dot{\mathcal{Q}}_{\mathrm{P}_{\mathrm{tx}}}(\lambda_{\mathrm{BS}},\mathrm{P}_{\mathrm{tx}},\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}})} - (\mathrm{P}_{\mathrm{tx}} + \Delta\mathrm{P})\right] - \mathrm{P}_{\mathrm{circ}}\mathcal{M}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)$$

$$\mathcal{S}_{\mathcal{D}}\left(\lambda_{\mathrm{BS}}\right) = \frac{\mathrm{P}_{\mathrm{circ}}}{\dot{\mathcal{L}}_{\lambda_{\mathrm{BS}}}(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}})}\left(\mathcal{L}\left(\frac{\lambda_{\mathrm{MT}}}{\lambda_{\mathrm{BS}}}\right)\dot{\mathcal{M}}_{\lambda_{\mathrm{BS}}}\left(\frac{\lambda_{\mathrm{MT}}}{\lambda_{\mathrm{BS}}}\right) - \dot{\mathcal{L}}_{\lambda_{\mathrm{BS}}}\left(\frac{\lambda_{\mathrm{MT}}}{\lambda_{\mathrm{BS}}}\right)\mathcal{M}\left(\frac{\lambda_{\mathrm{MT}}}{\lambda_{\mathrm{BS}}}\right)\right)$$
$$+ \Upsilon\mathcal{L}^2\left(\frac{\lambda_{\mathrm{MT}}}{\lambda_{\mathrm{BS}}}\right)(\mathrm{P}_{\mathrm{tx}} + \Delta\mathrm{P}) + \Upsilon\mathrm{P}_{\mathrm{circ}}\mathcal{L}^2\left(\frac{\lambda_{\mathrm{MT}}}{\lambda_{\mathrm{BS}}}\right)\frac{\dot{\mathcal{M}}_{\lambda_{\mathrm{BS}}}(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}})}{\dot{\mathcal{L}}_{\lambda_{\mathrm{BS}}}(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}})}$$
$$- \frac{\mathcal{L}(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}})\dot{\mathcal{Q}}_{\lambda_{\mathrm{BS}}}(\lambda_{\mathrm{BS}},\mathrm{P}_{\mathrm{tx}},\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}})}{\dot{\mathcal{L}}_{\lambda_{\mathrm{BS}}}(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}})\mathcal{Q}(\lambda_{\mathrm{BS}},\mathrm{P}_{\mathrm{tx}},\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}})}\left(1 + \Upsilon\mathcal{L}\left(\frac{\lambda_{\mathrm{MT}}}{\lambda_{\mathrm{BS}}}\right)\right)$$
$$\times\left[\mathcal{L}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)(\mathrm{P}_{\mathrm{tx}} + \Delta\mathrm{P}) + \mathrm{P}_{\mathrm{idle}} + \mathrm{P}_{\mathrm{circ}}\mathcal{M}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right)\right]$$

$\dot{\mathcal{Q}}_{\lambda_{\mathrm{BS}}}\left(\lambda_{\mathrm{BS}}, \mathrm{P}_{\mathrm{tx}}, \lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) \geq 0$ for $\lambda_{\mathrm{BS}} \geq 0$; and v) $\ddot{\mathcal{Q}}_{\lambda_{\mathrm{BS}}}\left(\lambda_{\mathrm{BS}}, \mathrm{P}_{\mathrm{tx}}, \lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) \leq 0$ for $\lambda_{\mathrm{BS}} \geq 0$.

Proof: The result in iv) follows from $\dot{\mathcal{Q}}_{\lambda_{\mathrm{BS}}}\left(\cdot,\cdot,\cdot\right)$ in Table II because, for $\lambda_{\mathrm{BS}} \geq 0$, $\mathcal{L}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) + \lambda_{\mathrm{BS}}\dot{\mathcal{L}}_{\lambda_{\mathrm{BS}}}\left(\lambda_{\mathrm{MT}}/\lambda_{\mathrm{BS}}\right) \geq 0$. This latter inequality holds true because $1 + x\left(1 + 1/\alpha\right) \leq \left(1 + x/\alpha\right)^{(\alpha+1)}$ for $x \geq 0$. The result in v) follows without explicitly computing $\ddot{\mathcal{Q}}_{\lambda_{\mathrm{BS}}}\left(\cdot,\cdot,\cdot\right)$ because $\dot{\mathcal{Q}}_{\lambda_{\mathrm{BS}}}\left(\cdot,\cdot,\cdot\right)$ in Table II is the composition of two increasing and concave functions in λ_{BS}, i.e., the function in the square brackets in the first row and the exponential function in the second row. $\square$

Lemma 7: The EE in (14) fulfills the following properties with respect to P_{tx} and λ_{BS}: i) $\mathrm{EE}\left(\mathrm{P}_{\mathrm{tx}}, \lambda_{\mathrm{BS}}\right) = 0$ if $\mathrm{P}_{\mathrm{tx}} \to 0$ or $\lambda_{\mathrm{BS}} \to 0$; and ii) $\mathrm{EE}\left(\mathrm{P}_{\mathrm{tx}}, \lambda_{\mathrm{BS}}\right) = 0$ if $\mathrm{P}_{\mathrm{tx}} \to \infty$ or $\lambda_{\mathrm{BS}} =\to \infty$.

Proof: This immediately follows from *Lemmas 3-6*. $\square$

B. Optimal Transmit Power Given the Density of the BSs

In this section, we analyze whether there exists an optimal and unique transmit power, $\mathrm{P}_{\mathrm{tx}}^{(\mathrm{opt})}$, that maximizes the EE formulated in (14), while all the other parameters, including λ_{BS},

are fixed and given. In mathematical terms, the optimization problem can be formulated as follows:

$$\max_{\mathrm{P}_{\mathrm{tx}}} \quad \mathrm{EE}\left(\mathrm{P}_{\mathrm{tx}}, \lambda_{\mathrm{BS}}\right)$$
$$\text{subject to} \quad \mathrm{P}_{\mathrm{tx}} \in \left[\mathrm{P}_{\mathrm{tx}}^{(\mathrm{min})}, \mathrm{P}_{\mathrm{tx}}^{(\mathrm{max})}\right]. \tag{15}$$

where $\mathrm{P}_{\mathrm{tx}}^{(\mathrm{min})} \geq 0$ and $\mathrm{P}_{\mathrm{tx}}^{(\mathrm{max})} \geq 0$ are the minimum and maximum power budget of the BSs, respectively. One may assume, without loss of generality, $\mathrm{P}_{\mathrm{tx}}^{(\mathrm{min})} \to 0$ and $\mathrm{P}_{\mathrm{tx}}^{(\mathrm{max})} \to \infty$.

The following theorem completely characterizes the solution of (15).

Theorem 1: Let $\mathcal{S}_{\mathcal{P}}\left(\cdot\right)$ be the function defined in Table II. The EE in (14) is a unimodal and strictly pseudo-concave function in P_{tx}. The optimization problem in (15) has a unique solution given by $\mathrm{P}_{\mathrm{tx}}^{(\mathrm{opt})} = \max\left\{\mathrm{P}_{\mathrm{tx}}^{(\mathrm{min})}, \min\left\{\mathrm{P}_{\mathrm{tx}}^*, \mathrm{P}_{\mathrm{tx}}^{(\mathrm{max})}\right\}\right\}$, where $\mathrm{P}_{\mathrm{tx}}^*$ is the only stationary point of the EE in (14) that is obtained as the unique solution of the following equation:

$$\dot{\mathrm{EE}}_{\mathrm{P}_{\mathrm{tx}}}\left(\mathrm{P}_{\mathrm{tx}}^*, \lambda_{\mathrm{BS}}\right) = \mathrm{P}_{\mathrm{idle}} - \mathcal{S}_{\mathcal{P}}\left(\mathrm{P}_{\mathrm{tx}}^*\right) = 0$$
$$\Leftrightarrow \mathcal{S}_{\mathcal{P}}\left(\mathrm{P}_{\mathrm{tx}}^*\right) = \mathrm{P}_{\mathrm{idle}}. \tag{16}$$

Proof: See Appendix B. □

C. Optimal Density Given the Transmit Power of the BSs

In this section, we analyze whether there exists an optimal and unique density of BSs, $\lambda_{\mathrm{BS}}^{(\mathrm{opt})}$, that maximizes the EE formulated in (14), while all the other parameters, including P_{tx}, are fixed and given. In mathematical terms, the optimization problem can be formulated as follows:

$$
\begin{aligned}
\max_{\lambda_{\mathrm{BS}}} \quad & \mathrm{EE}\left(\mathrm{P}_{\mathrm{tx}}, \lambda_{\mathrm{BS}}\right) \\
\text{subject to} \quad & \lambda_{\mathrm{BS}} \in \left[\lambda_{\mathrm{BS}}^{(\mathrm{min})}, \lambda_{\mathrm{BS}}^{(\mathrm{max})}\right]
\end{aligned}
\tag{17}
$$

where $\lambda_{\mathrm{BS}}^{(\mathrm{min})} \geq 0$ and $\lambda_{\mathrm{BS}}^{(\mathrm{max})} \geq 0$ are the minimum and maximum allowed density of the BSs, respectively. One may assume, without loss of generality, $\lambda_{\mathrm{BS}}^{(\mathrm{min})} \to 0$ and $\lambda_{\mathrm{BS}}^{(\mathrm{min})} \to \infty$.

The following theorem completely characterizes the solution of (17).

Theorem 2: Let $\mathcal{S}_{\mathcal{D}}\left(\cdot\right)$ be the function defined in Table II. The EE in (14) is a unimodal and strictly pseudo-concave function in λ_{BS}. The optimization problem in (17) has a unique solution given by $\lambda_{\mathrm{BS}}^{(\mathrm{opt})} = \max\left\{\lambda_{\mathrm{BS}}^{(\mathrm{min})}, \min\left\{\lambda_{\mathrm{BS}}^{*}, \lambda_{\mathrm{BS}}^{(\mathrm{max})}\right\}\right\}$, where $\lambda_{\mathrm{BS}}^{*}$ is the only stationary point of the EE in (14) that is obtained as the unique solution of the following equation:

$$
\begin{aligned}
\dot{\mathrm{EE}}_{\lambda_{\mathrm{BS}}}\left(\mathrm{P}_{\mathrm{tx}}, \lambda_{\mathrm{BS}}^{*}\right) &= \mathcal{S}_{\mathcal{D}}\left(\lambda_{\mathrm{BS}}^{*}\right) - \mathrm{P}_{\mathrm{idle}} = 0 \\
&\Leftrightarrow \mathcal{S}_{\mathcal{D}}\left(\lambda_{\mathrm{BS}}^{*}\right) = \mathrm{P}_{\mathrm{idle}}.
\end{aligned}
\tag{18}
$$

Proof: See Appendix C. □

D. On the Dependency of Optimal Transmit Power and Density of the BSs

The optimal transmit power and BSs' density that maximize the EE are obtained from the unique solutions of (16) and (18), respectively. These equations, however, cannot be further simplified and, therefore, explicit analytical expressions for $\mathrm{P}_{\mathrm{tx}}^{(\mathrm{opt})}$ and $\lambda_{\mathrm{BS}}^{(\mathrm{opt})}$ cannot, in general, be obtained. This is an inevitable situation when dealing with EE optimization problems, and, indeed, a closed-form expression of the optimal transmit power for simpler EE optimization problems does not exist either [1]. In some special cases, the transmit power can be implicitly expressed in terms of the Lambert-W function, which, however, is the solution of a transcendental equation [2]. Notable examples of these case studies include even basic point-to-point communication systems without interference [40]. Based on these considerations, it seems hopeless to attempt finding explicit analytical expressions from (16) and (18), respectively. However, thanks to the properties of the EE function, i.e., unimodality and strict pseudo-concavity, proved in *Theorem 1* and *Theorem 2*, $\mathrm{P}_{\mathrm{tx}}^{(\mathrm{opt})}$ and $\lambda_{\mathrm{BS}}^{(\mathrm{opt})}$ can be efficiently computed with the aid of numerical methods that are routinely employed to obtain the roots of non-linear scalar equations, e.g., the Newton's method [42]. For example, the unique solutions of (16) and (18) may be obtained by using the functions `FSolve` in Matlab and `NSolve` in Mathematica. *Theorem 1* and *Theorem 2* are, however, of paramount importance, since they state that an optimum maximizer exists and is unique.

Even though explicit analytical formulas for $\mathrm{P}_{\mathrm{tx}}^{(\mathrm{opt})}$ and $\lambda_{\mathrm{BS}}^{(\mathrm{opt})}$ cannot be obtained, it is important to understand how these optimal values change if any other system parameter changes. For instance, two worthwhile questions to answer are: "*How does $\mathrm{P}_{\mathrm{tx}}^{(\mathrm{opt})}$ change as a function of λ_{BS}?*" and "*How does $\lambda_{\mathrm{BS}}^{(\mathrm{opt})}$ change as a function of P_{tx}?*". These questions are relevant to optimize the deployment of cellular networks from the EE standpoint, since they unveil the inherent interplay between transmit power and density of BSs discussed in Section III and illustrated in Fig. 1. A general answer to these two questions is provided in the following two propositions.

Proposition 4: Let $\overline{\mathrm{P}}_{\mathrm{tx}}^{*}$ be the unique solution of (16) if $\lambda_{\mathrm{BS}} = \overline{\lambda}_{\mathrm{BS}}$. Let the optimal P_{tx} according to *Theorem 1* be $\overline{\mathrm{P}}_{\mathrm{tx}}^{(\mathrm{opt})} = \max\left\{\mathrm{P}_{\mathrm{tx}}^{(\mathrm{min})}, \min\left\{\overline{\mathrm{P}}_{\mathrm{tx}}^{*}, \mathrm{P}_{\mathrm{tx}}^{(\mathrm{max})}\right\}\right\}$. Let $\overline{\overline{\lambda}}_{\mathrm{BS}} \lesseqgtr \overline{\lambda}_{\mathrm{BS}}$ be another BSs' density. Let $\dot{\mathrm{EE}}_{\mathrm{P}_{\mathrm{tx}}}\left(\cdot, \cdot\right)$ be the first-order derivative in (16). The following holds:

$$
\overline{\overline{\mathrm{P}}}_{\mathrm{tx}}^{(\mathrm{opt})} \lesseqgtr \overline{\mathrm{P}}_{\mathrm{tx}}^{(\mathrm{opt})} \Leftrightarrow \dot{\mathrm{EE}}_{\mathrm{P}_{\mathrm{tx}}}\left(\overline{\mathrm{P}}_{\mathrm{tx}}^{(\mathrm{opt})}, \overline{\overline{\lambda}}_{\mathrm{BS}}\right) \lesseqgtr 0.
\tag{19}
$$

Proof: *Theorem 1* states that the EE function has a single stationary point that is its unique global maximizer. In mathematical terms, this implies $\dot{\mathrm{EE}}_{\mathrm{P}_{\mathrm{tx}}}\left(\mathrm{P}_{\mathrm{tx}}, \lambda_{\mathrm{BS}}\right) > 0$ if $\mathrm{P}_{\mathrm{tx}} < \mathrm{P}_{\mathrm{tx}}^{*}$ and $\dot{\mathrm{EE}}_{\mathrm{P}_{\mathrm{tx}}}\left(\mathrm{P}_{\mathrm{tx}}, \lambda_{\mathrm{BS}}\right) < 0$ if $\mathrm{P}_{\mathrm{tx}} > \mathrm{P}_{\mathrm{tx}}^{*}$ for every $\lambda_{\mathrm{BS}} \geq 0$. Therefore, the optimal transmit power needs to be increased (decreased) if the first-order derivative of the EE is positive (negative). Based on this, (19) follows because $\min\left\{\cdot, \cdot\right\}$ and $\max\left\{\cdot, \cdot\right\}$ are increasing functions. □

Proposition 5: Let $\overline{\lambda}_{\mathrm{BS}}^{*}$ be the unique solution of (18) if $\mathrm{P}_{\mathrm{tx}} = \overline{\mathrm{P}}_{\mathrm{tx}}$. Let the optimal λ_{BS} according to *Theorem 2* be $\overline{\lambda}_{\mathrm{BS}}^{(\mathrm{opt})} = \max\left\{\lambda_{\mathrm{BS}}^{(\mathrm{min})}, \min\left\{\overline{\lambda}_{\mathrm{BS}}^{*}, \lambda_{\mathrm{BS}}^{(\mathrm{max})}\right\}\right\}$. Let $\overline{\overline{\mathrm{P}}}_{\mathrm{tx}} \lesseqgtr \overline{\mathrm{P}}_{\mathrm{tx}}$ be another transmit power. Let $\dot{\mathrm{EE}}_{\lambda_{\mathrm{BS}}}\left(\cdot, \cdot\right)$ be the first-order derivative in (18). The following holds:

$$
\overline{\overline{\lambda}}_{\mathrm{BS}}^{(\mathrm{opt})} \lesseqgtr \overline{\lambda}_{\mathrm{BS}}^{(\mathrm{opt})} \Leftrightarrow \dot{\mathrm{EE}}_{\lambda_{\mathrm{BS}}}\left(\overline{\overline{\mathrm{P}}}_{\mathrm{tx}}, \overline{\lambda}_{\mathrm{BS}}^{(\mathrm{opt})}\right) \lesseqgtr 0.
\tag{20}
$$

Proof: It follows from *Theorem 2*, similar to the proof of *Proposition 4*. □

Remark 13: It is worth mentioning that the approach utilized to prove *Proposition 4* and *Proposition 5* is applicable to study the dependency of $\mathrm{P}_{\mathrm{tx}}^{(\mathrm{opt})}$ and $\lambda_{\mathrm{BS}}^{(\mathrm{opt})}$, respectively, with respect to any other system parameters. The findings in *Proposition 4* and *Proposition 5* are especially relevant for cellular network planning. Let us consider, e.g., (19). By simply studying the sign of the first-order derivative $\dot{\mathrm{EE}}_{\mathrm{P}_{\mathrm{tx}}}\left(\cdot, \cdot\right)$, one can identify, with respect to an optimally deployed cellular network, the set of BSs' densities that would require to increase or decrease the transmit power while still operating at the optimum. In Section V, numerical examples are shown to highlight that $\mathrm{P}_{\mathrm{tx}}^{(\mathrm{opt})}$ may either decrease or increase as λ_{BS} increases or decreases. □

E. Joint Optimization of Transmit Power and Density of the BSs

In Sections IV-B and IV-C, either λ_{BS} or P_{tx} are assumed to be given, respectively. In practical applications, however, it is important to identify the optimal pair $\left(\mathrm{P}_{\mathrm{tx}}^{(\mathrm{opt})}, \lambda_{\mathrm{BS}}^{(\mathrm{opt})}\right)$

TABLE III
ALTERNATING OPTIMIZATION OF THE EE.

Algorithm
`Let` $P_{tx} \in \left[P_{tx}^{(min)}, P_{tx}^{(max)} \right]$; $\lambda_{BS} \in \left[\lambda_{BS}^{(min)}, \lambda_{BS}^{(max)} \right]$;
`Set` $\lambda_{BS} = \overline{\lambda}_{BS}^{(opt)} \in \left[\lambda_{BS}^{(min)}, \lambda_{BS}^{(max)} \right]$ (initial guess); $V = 0$; $\epsilon > 0$;
Repeat
$\quad V_0 = V$;
$\quad \overline{P}_{tx}^* \leftarrow \dot{EE}_{P_{tx}} \left(P_{tx}, \overline{\lambda}_{BS}^{(opt)} \right) = 0; \quad \overline{P}_{tx}^{(opt)} = \max \left\{ P_{tx}^{(min)}, \min \left\{ \overline{P}_{tx}^*, P_{tx}^{(max)} \right\} \right\};$ (16)
$\quad \overline{\lambda}_{BS}^* \leftarrow \dot{EE}_{\lambda_{BS}} \left(\overline{P}_{tx}^{(opt)}, \lambda_{BS} \right) = 0; \quad \overline{\lambda}_{BS}^{(opt)} = \max \left\{ \lambda_{BS}^{(min)}, \min \left\{ \overline{\lambda}_{BS}^*, \lambda_{BS}^{(max)} \right\} \right\};$ (18)
$\quad V = EE \left(\overline{P}_{tx}^{(opt)}, \overline{\lambda}_{BS}^{(opt)} \right);$ (14)
Until $\lvert V - V_0 \rvert / V \leq \epsilon$;
`Return` $P_{tx}^{(opt)} = \overline{P}_{tx}^{(opt)}$; $\lambda_{BS}^{(opt)} = \overline{\lambda}_{BS}^{(opt)}$.

that *jointly* maximizes the EE in (14). This joint optimization problem can be formulated as follows:

$$\max_{P_{tx}, \lambda_{BS}} \quad EE \left(P_{tx}, \lambda_{BS} \right)$$
$$\text{subject to} \quad P_{tx} \in \left[P_{tx}^{(min)}, P_{tx}^{(max)} \right], \ \lambda_{BS} \in \left[\lambda_{BS}^{(min)}, \lambda_{BS}^{(max)} \right] \tag{21}$$

where a notation similar to that used in (15) and (17) is adopted.

In *Theorem 1* and *Theorem 2*, we have solved the optimization problem formulated in (21) with respect to P_{tx} for a given λ_{BS} and with respect to λ_{BS} for a given P_{tx}, respectively. By leveraging these results, a convenient approach for tackling (21) with respect to P_{tx} and λ_{BS} is to utilize the alternating optimization method, which iteratively optimizes P_{tx} for a given λ_{BS} and λ_{BS} for a given P_{tx} until convergence of the EE in (14) within a desired level of accuracy [41, Proposition 2.7.1]. The algorithm that solves (21) based on the alternating optimization method is reported in Table III. Its convergence and optimality properties are summarized as follows.

Proposition 6: Let $\overline{P}_{tx}^{(opt)}(m)$, $\overline{\lambda}_{BS}^{(opt)}(m)$, and $EE(m)$ be P_{tx}, λ_{BS} and EE obtained from the algorithm in Table III at the mth iteration, respectively. The sequence $EE(m)$ is monotonically increasing and converges. In addition, every limit point of the sequence $\left(\overline{P}_{tx}^{(opt)}(m), \overline{\lambda}_{BS}^{(opt)}(m) \right)$ fulfills the Karush-Kuhn-Tucker (KKT) first-order optimality conditions of the problem in (21).

Proof: At the end of each iteration of the algorithm in Table III, the value of EE does not decrease. The sequence $EE(m)$, hence, converges, because the EE in (14) is a continuous function over the compact feasible set of the problem in (21) and, thus, it admits a finite maximum by virtue of the Weierstrass extreme value theorem [41]. From [41, Proposition 2.7.1], the alternating optimization method fulfills the KKT optimality conditions, provided that i) the objective and constraint functions are differentiable, ii) each constraint function depends on a single variable, and iii) each subproblem has a unique solution. The first and second requirements follow by direct inspection of (21). The third requirement is ensured by *Theorem 1* and *Theorem 2*. $\square$

TABLE IV
SETUP OF PARAMETERS (UNLESS OTHERWISE STATED). IT IS WORTH NOTHING THAT THE SETUP $\gamma_D = \gamma_A$ CONSTITUTES JUST A CASE STUDY AND THAT THE MAIN FINDINGS OF THE PRESENT PAPER HOLD TRUE FOR EVERY $\gamma_A > 0$.

Parameter	Value
β	3.5
$\kappa = \left(4\pi f_c / 3 \cdot 10^8 \right)^2$	f_c = 2.1 GHz
N_0	-174 dBm/Hz
B_W	20 MHz
P_{circ}	51.14 dBm [8]
P_{idle}	48.75 dBm [8]
P_{tx}	43 dBm [8]
$\lambda_{BS} = 1/ \left(\pi R_{cell}^2 \right)$ BSs/m^2	R_{cell} = 250 m
$\lambda_{MT} = 1/ \left(\pi R_{MT}^2 \right) = 121$ MTs/km^2	R_{MT} = 51.29
$\gamma_D = \gamma_A$	5 dB

Remark 14: The optimization problems in *Theorem 1* and *Theorem 2* can be efficiently solved by using the Newton's method, which allows one to find the root of real-valued objective functions via multiple iterations of increasing accuracy and at a super-linear (i.e., quadratic if the initial guess is sufficiently close to the actual root) convergence rate [42]. The properties of convergence of the alternating maximization algorithm in Table III to a stationary point of the objective function in (21) are discussed in [41, Proposition 2.7.1]. Under mild assumptions that hold for the specific problem at hand, the algorithm in Table III is locally q-linearly convergent to a local maximizer of the objective function provided that the initial guess is sufficiently close to the actual root [43, Section 2]. Further details can be found in [43]. $\square$

In Section V, numerical evidence of the global optimality of the algorithm in Table III is given as well. In addition, numerical results on the average (with respect to the initial guess) number of iterations as a function of the tolerance of convergence, $\epsilon > 0$, are illustrated.

V. NUMERICAL RESULTS

In this section, we show numerical results to validate the proposed analytical framework for computing the PSE and
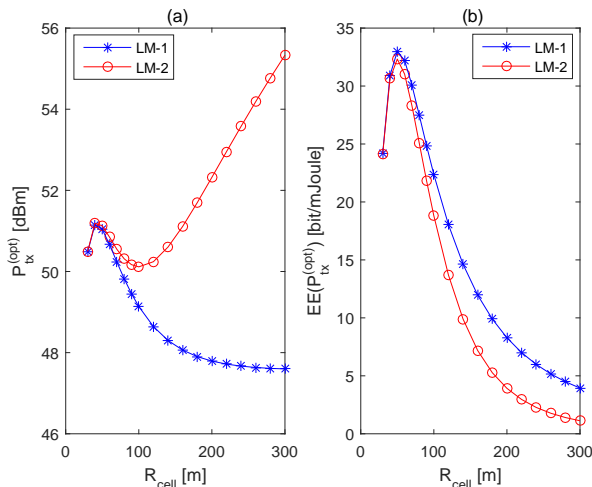
Fig. 2. Optimal transmit power (a) and energy efficiency (b) versus R_{cell}. Solid lines: Optimum from *Theorem 1*. Markers: Optimum from a brute-force search of (15). Special case with $\beta = 6.5$ and $\lambda_{MT} = 21$ MTs/km^2.
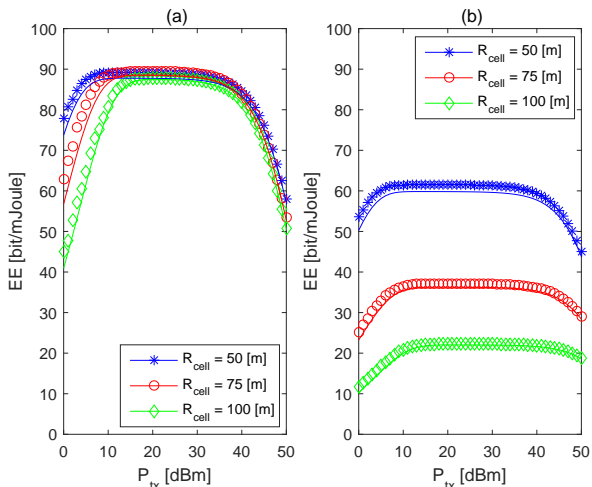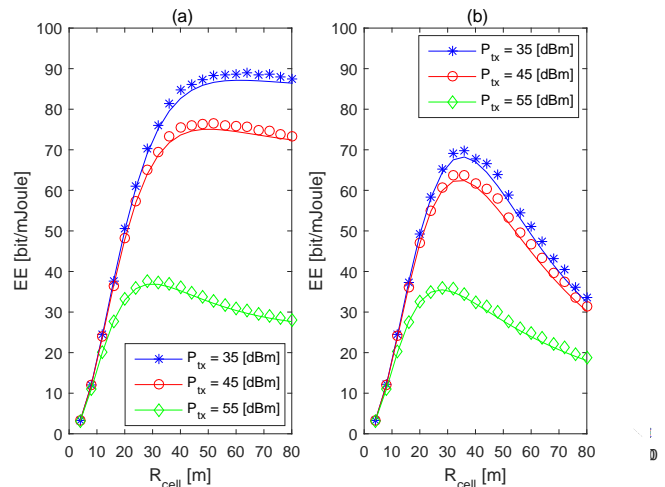

Fig. 4. Energy efficiency versus R_{cell} for Load Model 1 (a) and Load Model 2 (b). Solid lines: Mathematical Framework from (14). Markers: Monte Carlo simulations.
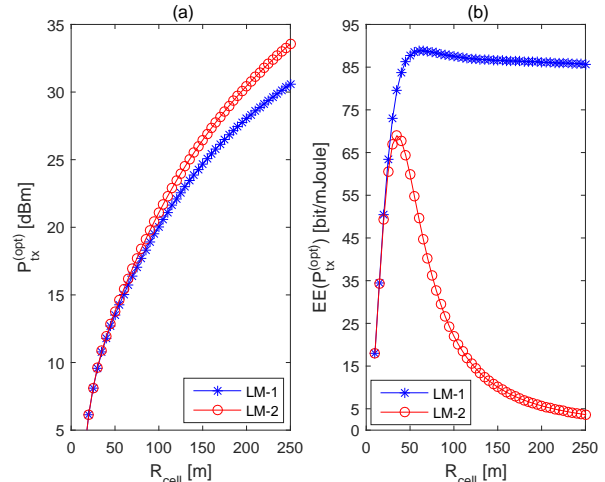

Fig. 3. Energy efficiency versus the transmit power for Load Model 1 (a) and Load Model 2 (b). Solid lines: Framework from (14). Markers: Monte Carlo simulations.

Fig. 5. Optimal transmit power (a) and energy efficiency (b) versus R_{cell}. Solid lines: Optimum from *Theorem 1*. Markers: Optimum from a brute-force search of (15). LM-1: Load Model 1 and LM-2: Load Model 2.

EE, as well as to substantiate the findings originating from the analysis of the system-level EE optimization problems as a function of the transmit power and density of the BSs. Unless otherwise stated, the simulation setup is summarized in Table IV. For ease of understanding, the BSs' density is represented via the inter-site distance (R_{cell}) defined in Table IV. A similar comment applies to the density of the MTs that is expressed in terms of their average distance (R_{MT}). As far as the choice of the setup of parameters is concerned, it is worth mentioning that the power consumption model is in agreement with [8] and [30]. The density of the MTs coincides with the average density of inhabitants in France.

a) Validation Against Monte Carlo Simulations: In Figs. 3 and 4, we validate the correctness of (14) against Monte Carlo simulations. Monte Carlo results are obtained by simulating several realizations, according to the PPP model, of the cellular network and by empirically computing the PSE

according to its definition in (6) and (7), as well as the power consumption based on the operating principle described in the proofs of *Proposition 2* and *Proposition 3*. It is worth mentioning that, to estimate the PSE, only the definitions in the first line of (6) and (7) are used. The results depicted in Figs. 3 and 4 confirm the good accuracy of the proposed mathematical approach. They highlight, in addition, the unimodal and pseudo-concave shape of the EE as a function of the transmit power, given the BSs' density, and of the BSs' density, given the transmit power. If the same transmit power and BSs' density are assumed for both load models, we observe, as expected, that the first load model provides a better EE than the second load model.

b) Validation of Theorem 1 and Theorem 2: In Figs. 5 and 6, we compare the optimal transmit power and BSs' density obtained from *Theorem 1* and *Theorem 2*, i.e., by computing the unique zero of (16) and (18), respectively,
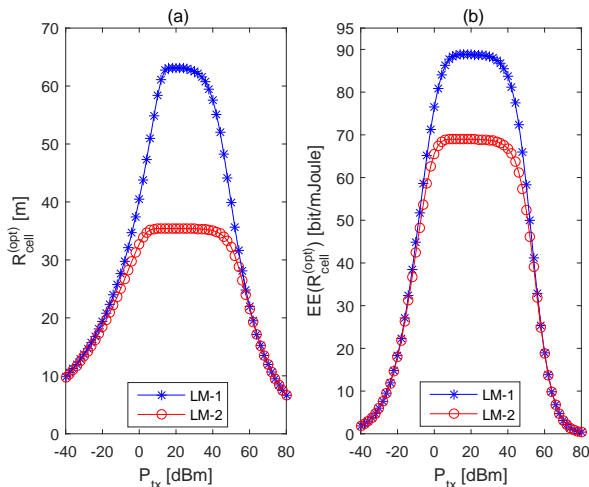
Fig. 6. Optimal density of BSs (R_{cell}) (a) and energy efficiency (b) versus the transmit power. Solid lines: Optimum from *Theorem 2*. Markers: Optimum from a brute-force search of (17). LM-1: Load Model 1, LM-2: Load Model 2.
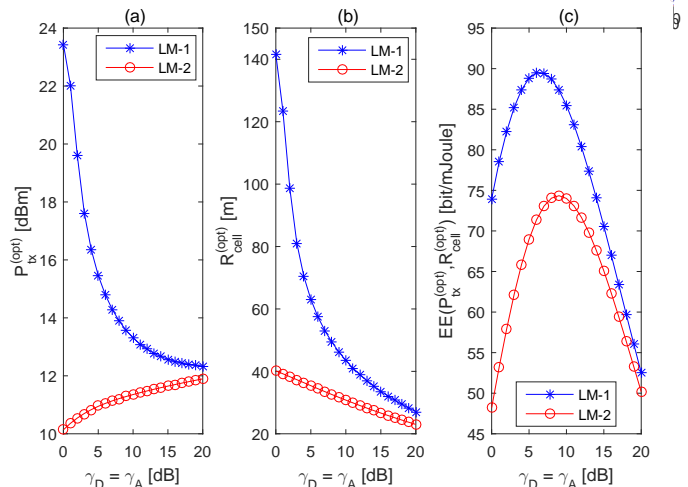

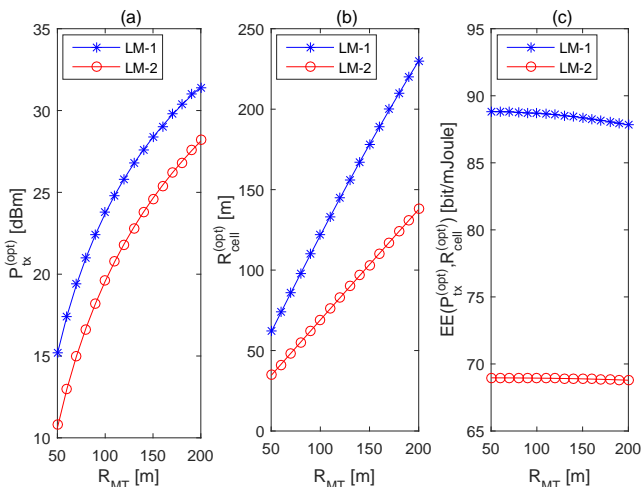Fig. 8. Optimal transmit power (a), density of BSs (R_{cell}) (b), and energy efficiency (c) versus the reliability thresholds ($\gamma_D = \gamma_A$). Solid lines: Optimum from the algorithm in Table III. Markers: Optimum from a brute-force search of (21). LM-1: Load Model 1 and LM-2: Load Model 2.

Fig. 7. Optimal transmit power (a), density of BSs (R_{cell}) (b), and energy efficiency (c) versus the density of MTs (R_{MT}). Solid lines: Optimum from the algorithm in Table III. Markers: Optimum from a brute-force search of (21). LM-1: Load Model 1 and LM-2: Load Model 2.
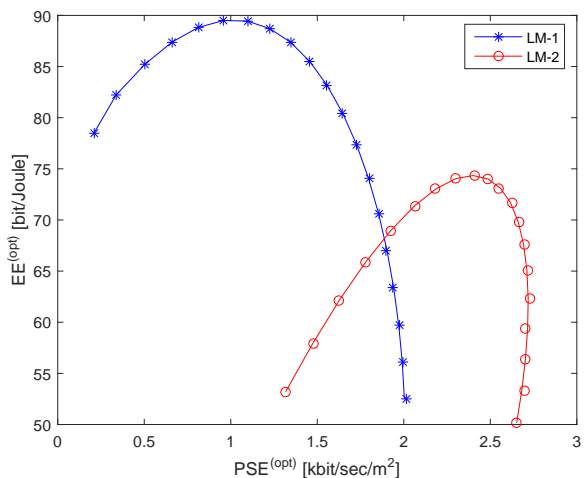
Fig. 9. Analysis of the EE vs. PSE trade-off. Solid lines: Optimum from the algorithm in Table III. Markers: Optimum from a brute-force search of (21). LM-1: Load Model 1 and LM-2: Load Model 2.

against a brute-force search of the optimum of (15) and (17), respectively. We observe the correctness of *Theorem 1* and *Theorem 2* for the load models analyzed in the present paper. Figures 5 and 6, in addition, confirm two important remarks that we have made throughout this paper. The first is that a joint pair of transmit power and BSs' density exists. This is highlighted by the fact that the EE evaluated at the optimal transmit power, given the BSs' density, and at the optimal BSs' density, given the transmit power, is still a unimodal and pseudo-concave function. This motivates one to use the alternating optimization algorithm proposed in Section IV-E. The second is related to the difficulty of obtaining an explicit closed-form expression of the optimal transmit power as a function of the BSs' density and of the BSs' density as a function of the transmit power. Figure 6(a), for example,

clearly shows that the behavior of the optimal transmit power is not monotonic as a function of the BSs' density. This is in contrast with heuristic optimization criteria based on the coverage probability metric [32]. Figure 5(a), on the other hand, provides more intuitive trends according to which the optimal transmit power increases as the density of the BSs decreases. This is, however, just a special case that is parameter-dependent. A counter-example is, in fact, illustrated in Fig. 2, where, for a different set of parameters, it is shown that the optimal transmit power may increase, decrease and then increase again as a function of the average inter-site distance of the BSs (R_{cell}). In this case, the density of the MTs coincides with the average density of inhabitants in Sweden and a large path-loss exponent is assumed to highlight the peculiar performance trend. These numerical examples clearly substantiate the importance of *Theorem 1* and *Theorem 2*, and
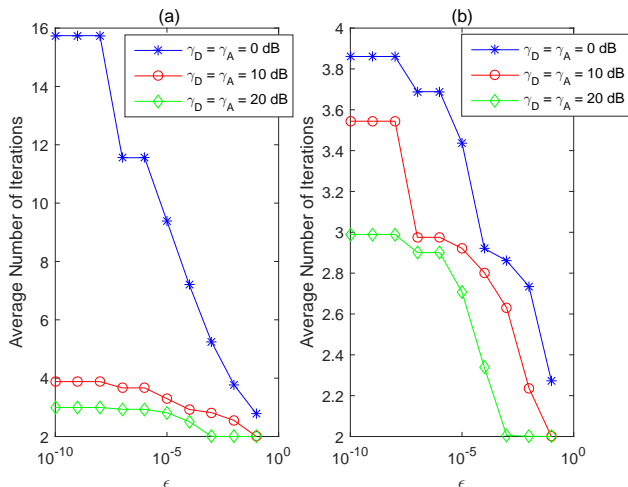
Fig. 10. Number of iterations of the algorithm in Table III as a function of $\epsilon > 0$. The number of iterations is averaged (15000 trials) over the initial guess $\lambda_{\mathrm{BS}} = \overline{\lambda}_{\mathrm{BS}}^{(\mathrm{opt})} \in \left[\lambda_{\mathrm{BS}}^{(\mathrm{min})}, \lambda_{\mathrm{BS}}^{(\mathrm{max})}\right]$. (a) Load Model 1 and (b) Load Model 2. Setup: $\mathrm{R}_{\mathrm{cell}}^{(\mathrm{min})} = 10$ m, $\mathrm{R}_{\mathrm{cell}}^{(\mathrm{max})} = 2000$ m, $\mathrm{P}_{\mathrm{tx}}^{(\mathrm{min})} = -20$ dBm, $\mathrm{P}_{\mathrm{tx}}^{(\mathrm{max})} = 60$ dBm.

highlight the complexity of the optimization problem that is analyzed and successfully solved in the present paper.

c) Validation of the Alternating Optimization Algorithm in Table III: In Figs. 7 and 8, we provide numerical evidence of the convergence of the alternating optimization algorithm introduced in Section IV-E towards the global optimum of the optimization problem formulated in (21). The study is performed by computing the joint optimal transmit power and BSs' density as a function of the density of the MTs (Fig. 7) and of the reliability thresholds (Fig. 8). We observe a very good agreement between the algorithm in Table III and a brute-force search of the optimum of (21). Similar studies have been conducted as a function of other system parameters, but they are not reported in the present paper due to space limitations.

d) Comparison Between Load Model 1 and 2: With the exception of Figs. 3 and 4, all the figures reported in this section illustrate the achievable EE of the two load models analyzed in the present manuscript when they operate at their respective optima. Based on the obtained results, we conclude that, for the considered system setup, the first load model outperforms the second one in terms of EE. Figures 7 and 8 show, for example, that this may be obtained by transmitting a higher power but, at the same time, by reducing the deployment density of the BSs. It is worth mentioning that, even though both load models provide the same PSE and serve, in the long time-horizon, all the MTs of the network, they have one main difference: the MTs under the first load model experience a higher latency (i.e., the MTs experience a longer delay before being served, since they are randomly chosen among all the available MTs in the cell), since a single MT is served at any time instance. We evince, as a result, that the higher EE provided by the first load model is obtained at the price of increasing the MTs' latency. The analysis and optimization of energy-efficient cellular networks with latency constraints is, therefore, an important generalization of the

study conducted in the present paper.

e) Analysis of the EE vs. PSE Trade-Off: In Fig. 9, we illustrate the trade-off between EE and PSE, which is obtained by setting the transmit power and density of the BSs at the optimal values that are obtained by solving the optimization problem in (21) with the aid of the algorithm in Table III. Figure 9 provides a different view of the comparison between Load Model 1 and 2 introduced in Section II-D. The Load Model 1 is a suitable choice to obtain a high EE at low-medium PSEs, while the Load Model 2 is a more convenient option torequired to converge within obtain a good EE at medium-high PSEs. Based on these results, the optimization of the EE vs. PSE trade-off constitutes an interesting generalization of the study carried out in the present paper.

f) Convergence Analysis of the Maximization Algorithm in Table III: Motivated by *Remark 14*, Fig. 10 shows the average number of iterations of the alternating optimization algorithm in Table III as a function of the convergence accuracy ϵ. We observe that the algorithm necessitates more iterations for Load Model 1. In general, however, we observe that the number of iterations that are required to converge within the defined convergence accuracy is relatively small.

VI. CONCLUSION

In the present paper, we have introduced a new closed-form analytical expression of the potential spectral efficiency of cellular networks. Unlike currently available analytical frameworks, we have shown that the proposed approach allows us to account for the tight interplay between transmit power and density of the base stations in cellular networks. Therefore, the proposed approach is conveniently formulated for the optimization of the network planning of cellular networks, by taking into account important system parameters. We have applied the new approach to the analysis and optimization of the energy efficiency of cellular networks. We have mathematically proved that the proposed closed-form expression of the energy efficiency is a unimodal and strictly pseudo-concave function in the transmit power, given the density, and in the density, given the transmit power of the base stations. Under these assumptions, as a result, a unique transmit power and density of the base stations exist, which can be obtained by finding the unique zero of a simple non-linear function that is provided in a closed-form expression. All mathematical derivations and findings have been substantiated with the aid of numerical simulations. We argue that the applications of the proposed approach to the system-level modeling and optimization of cellular networks are countless and go beyond the formulation of energy efficiency problems.

Extensions and generalizations of the analytical and optimization frameworks proposed in the present paper, include, but are not limited to, the system-level analysis and optimization of i) the energy efficiency versus spectral efficiency trade-off, ii) uplink cellular networks, iii) three-dimensional network topologies with elevated base stations and spatial blockages, iv) cache-enabled cellular networks, v) cellular networks with network slicing, vi) cellular networks with

renewable energy sources and energy harvesting, and vii) multi-tier (heterogeneous) cellular networks.

APPENDIX A
PROOF OF PROPOSITION 1

Under the assumption that MT_0 is selected, from (3) and (4), we have:

$$P_{cov}(\gamma_D, \gamma_A) = \Pr\left\{ \frac{g_0/L_0 \mathbb{1}\left(L_0 \leq P_{tx}/(\gamma_A \sigma_N^2)\right)}{\sum_{BS_i \in \Psi_{BS}^{(I)}} g_i/L_i \mathbb{1}(L_i > L_0)} \geq \gamma_D \right\}$$

$$= \int_0^{P_{tx}/(\gamma_A \sigma_N^2)} \underbrace{\Pr\left\{ \frac{g_0/x}{\sum_{BS_i \in \Psi_{BS}^{(I)}} g_i/L_i \mathbb{1}(L_i > x)} \geq \gamma_D \right\}}_{\mathcal{G}(\gamma_D; x)} f_{L_0}(x)\, dx$$

(22)

where $f_{L_0}(x) = 2\pi\lambda_{BS}\left(\kappa^{2/\beta}\beta\right)^{-1} x^{2/\beta-1} e^{-\pi\lambda_{BS}(x/\kappa)^{2/\beta}}$ is the probability density function of L_0 that is obtained by applying the displacement theorem of PPPs [5, Eq. (21)]. It is worth mentioning that (22) is exact if the Crofton cell is considered, while it is an approximation if the typical cell is considered (see *Remark 7* for further details).

The probability term, $\mathcal{G}(\cdot; \cdot)$, in the integrand function of (22) can be computed as follows:

$$\mathcal{G}(\gamma_D; x) \overset{(a)}{=} \exp\left(-\int_x^{+\infty} \left(1 + \frac{y}{x\gamma_D}\right)^{-1} 2\pi\lambda_{BS}^{(tx)} \frac{y^{2/\beta-1}}{\kappa^{2/\beta}\beta} dy \right)$$

$$\overset{(b)}{=} \exp\left(-\pi\lambda_{BS}^{(tx)}(x/\kappa)^{2/\beta}\Upsilon \right)$$

(23)

where (a) follows from the probability generating functional theorem of PPPs [3] by taking into account that, based on (8), the interfering BSs constitute a PPP of intensity equal to $\lambda_{BS}^{(tx)} = \lambda_{BS}\mathbb{P}_{BS}^{(tx)} = \lambda_{BS}\mathcal{L}(\lambda_{MT}/\lambda_{BS})$, and (b) follows by solving the integral. The intensity of the interfering PPP, $\lambda_{BS}^{(tx)}$, is obtained by taking into account that only that BSs that are in transmission mode contribute to the inter-cell interference. The analytical expression of $\lambda_{BS}^{(tx)}$ is, in particular, obtained with the aid of the independent thinning theorem of PPPs, similar to [5] and [38]. The impact of the spatial correlation that exists among the BSs that operate in transmission mode [39], is, on the other hand, postponed to future research.

By inserting (23) in (22) and by applying some changes of variable, we obtain:

$$P_{cov}(\gamma_D, \gamma_A) = \pi\lambda_{BS}\kappa^{-2/\beta}$$

$$\times \int_0^{\left(P_{tx}/(\gamma_A\sigma_N^2)\right)^{2/\beta}} \exp\left(-\pi\lambda_{BS}\kappa^{-2/\beta}(1 + \Upsilon\mathcal{L}(\lambda_{MT}/\lambda_{BS}))z \right) dz.$$

(24)

The proof follows from (6) and (7) with the aid of some simplifications and by using the identity $\sum_{u=0}^{+\infty}(u+1)^{-1}\Pr\{\bar{N}_{MT} = u\} = (\lambda_{MT}/\lambda_{BS})^{-1}\mathcal{L}(\lambda_{MT}/\lambda_{BS})$ [33, Proposition 2].

APPENDIX B
PROOF OF THEOREM 1

In this section, we are interested in the functions that depend on P_{tx}. For ease of writing, we adopt the simplified notation: $P_{tx} \to P$, $\mathcal{L}(\cdot) \to \mathcal{L}$, $\mathcal{M}(\cdot) \to \mathcal{M}$, $\mathcal{Q}(\cdot, P_{tx}, \cdot) \to \mathcal{Q}(P)$, $\dot{\mathcal{Q}}_{P_{tx}}(\cdot, P_{tx}, \cdot) \to \dot{\mathcal{Q}}(P)$, $P_{circ} = P_c$, $P_{idle} = P_i$, $EE(P_{tx}, \cdot) \to EE(P)$, and $\dot{EE}_{P_{tx}}(P_{tx}, \cdot) \to \dot{EE}(P)$. A similar notation is adopted for higher-order derivatives with respect to P.

The stationary points of (14) are the zeros of the first-order derivative of $EE(\cdot)$ with respect to P. From (14), we obtain $\dot{EE}(P) = 0 \Leftrightarrow P_i - \mathcal{S}_{\mathcal{P}}(P) = 0$, which can be re-written as follows:

$$\underbrace{\mathcal{Q}(P)\Big/\dot{\mathcal{Q}}(P) - P}_{W_{left}(P)} = \underbrace{\Delta P + P_i/\mathcal{L} + P_c\mathcal{M}/\mathcal{L}}_{W_{right}}.$$

(25)

With the aid of some algebraic manipulations and by exploiting *Lemmas 3-6*, the following holds: i) $W_{right} \geq 0$ is a non-negative function that is independent of P, ii) $W_{left}(P) \geq 0$ is a non-negative and increasing function of P, i.e., $\dot{W}_{left}(P) \geq 0$, since $\mathcal{Q}(P) \geq 0$ and $\ddot{\mathcal{Q}}(P) \leq 0$ from *Lemma 5*, iii) $W_{left}(P \to 0) = 0$ and $W_{left}(P \to \infty) = \infty$. This implies that $W_{left}(\cdot)$ and W_{right} intersect each other in just one point. Therefore, a unique stationary point, P^*, exists. Also, $\dot{EE}(P) > 0$ for $P < P^*$ and $\dot{EE}(P) < 0$ for $P > P^*$. Finally, by taking into account the constraints on the transmit power, it follows that the unique optimal maximizer of the EE is $P^{(opt)} = \max\left\{P^{(min)}, \min\left\{P^*, P^{(max)}\right\}\right\}$, since $P \in \left[P^{(min)}, P^{(max)}\right]$. This concludes the proof.

APPENDIX C
PROOF OF THEOREM 2

In this section, we are interested in the functions that depend on λ_{BS}. For ease of writing, we adopt the simplified notation: $\lambda_{BS} \to \lambda$, $\mathcal{L}(\cdot/\lambda_{BS}) \to \mathcal{L}(\lambda)$, $\mathcal{M}(\cdot/\lambda_{BS}) \to \mathcal{M}(\lambda)$, $\mathcal{Q}(\lambda_{BS}, \cdot, \cdot/\lambda_{BS}) \to \mathcal{Q}(\lambda)$, $\dot{\mathcal{Q}}_{\lambda_{BS}}(\lambda_{BS}, \cdot, \cdot/\lambda_{BS}) \to \dot{\mathcal{Q}}(\lambda)$, $P_{circ} = P_c$, $P_{idle} = P_i$, $EE(\cdot, \lambda_{BS}) \to EE(\lambda)$, $\dot{EE}_{\lambda_{BS}}(\cdot, \lambda_{BS}) \to \dot{EE}(\lambda)$, $P_{tx} \to P$. Similar notation applies to higher-order derivatives.

The proof is split in two parts: i) $\lambda_{MT}/\lambda \geq 2.8$ and ii) $\lambda_{MT}/\lambda \leq 2.8$. This is necessary because, from *Lemma 3*, $\mathcal{L}(\cdot)$ is concave in λ if $\lambda_{MT}/\lambda \geq 2.8$ and convex in λ if $\lambda_{MT}/\lambda \leq 2.8$.

a) Case Study $\lambda_{MT}/\lambda \geq 2.8$: The stationary points of (14) are the zeros of the first-order derivative of $EE(\cdot)$ with respect to λ. From (14), we obtain $\dot{EE}(\lambda) = 0 \Leftrightarrow \mathcal{S}_{\mathcal{D}}(\lambda) - P_i = 0$. This stationary equation can be re-written as follows

$(W_{\text{right}}(\lambda) = \sum_{\ell=1}^{5} W_\ell(\lambda))$:

$$\underbrace{\mathrm{P_i}}_{W_{\text{left}}} = \underbrace{-\left(\mathcal{L}(\lambda)\big/\dot{\mathcal{L}}(\lambda)\right)\left(\dot{\mathcal{Q}}(\lambda)\big/\mathcal{Q}(\lambda)\right)}_{W_1(\lambda)}$$

$$\times \underbrace{[1 + \Upsilon\mathcal{L}(\lambda)]\left[\mathcal{L}(\lambda)(\mathrm{P}+\Delta\mathrm{P}) + \mathrm{P_i} + \mathrm{P_c}\mathcal{M}(\lambda)\right]}_{W_2(\lambda)}$$

$$+ \underbrace{\mathrm{P_{circ}}\left(\dot{\mathcal{M}}(\lambda)\,\mathcal{L}(\lambda)\big/\dot{\mathcal{L}}(\lambda) - \mathcal{M}(\lambda)\right)}_{W_3(\lambda)}$$

$$+ \underbrace{\Upsilon\mathcal{L}^2(\lambda)(\mathrm{P}+\Delta\mathrm{P})}_{W_4(\lambda)} + \underbrace{\Upsilon\mathrm{P_c}\mathcal{L}^2(\lambda)\dot{\mathcal{M}}(\lambda)\big/\dot{\mathcal{L}}(\lambda)}_{W_5(\lambda)}.$$

$$(26)$$

With the aid of some algebraic manipulations and by exploiting *Lemmas 3-6*, the following holds: i) $W_{\text{left}} \geq 0$ is a non-negative function that is independent of λ, ii) $W_{\text{right}}(\lambda) \geq 0$ is a non-negative function of λ, since $W_\ell(\lambda) \geq 0$ for $\ell = 1, \dots, 5$ if $\lambda_{\text{MT}}/\lambda \geq 2.8$. In particular, $W_3(\lambda) \geq 0$ if $\lambda_{\text{MT}}/\lambda \geq 1.4$ and $W_\ell(\lambda) \geq 0$ for $\lambda \geq 0$ if $\ell = 1, 2, 4, 5$, iii) $W_{\text{right}}(\lambda \to 0) = \infty$ and $W_{\text{right}}(\lambda \to \infty) = 0$. This implies that W_{left} and $W_{\text{right}}(\cdot)$ would intersect each other in just a single point if W_{right} is a decreasing function in λ, i.e., $\dot{W}_{\text{right}}(\lambda) \leq 0$ for $\lambda_{\text{MT}}/\lambda \geq 2.8$. A sufficient condition for this to hold is that $W_\ell(\cdot)$ for $\ell = 1, \dots, 5$ are decreasing functions in λ, i.e., $\dot{W}_\ell(\lambda) \leq 0$ for $\lambda_{\text{MT}}/\lambda \geq 2.8$. This holds to be true and can be proved as follows. $\dot{W}_2(\lambda) \leq 0$ for $\lambda \geq 0$ and $\dot{W}_4(\lambda) \leq 0$ for $\lambda \geq 0$ because $\mathcal{L}(\cdot)$ and $\mathcal{M}(\cdot)$ are decreasing functions in λ (see *Lemma 3* and *Lemma 4*). $\dot{W}_3(\lambda) \leq 0$ for $\lambda \geq 0$ and $\dot{W}_5(\lambda) \leq 0$ for $\lambda \geq 0$ immediately follow by inserting into them the first-order derivatives of $\mathcal{L}(\cdot)$ and $\mathcal{M}(\cdot)$ with respect to λ and with the aid of simple algebraic manipulations. Less evident is the behavior of $W_1(\cdot)$ as a function of λ. Using some algebra, the first-order derivative satisfies the following:

$$\dot{W}_1(\lambda)\left(\mathcal{Q}(\lambda)\dot{\mathcal{L}}(\lambda)\right)^2 = \underbrace{-\mathcal{L}(\lambda)\dot{\mathcal{L}}(\lambda)\mathcal{Q}(\lambda)\ddot{\mathcal{Q}}(\lambda)}_{A_1(\lambda)}$$

$$+ \underbrace{\left(-\dot{\mathcal{L}}^2(\lambda)\mathcal{Q}(\lambda)\dot{\mathcal{Q}}(\lambda)\right)}_{A_2(\lambda)}$$

$$+ \underbrace{\mathcal{L}(\lambda)\dot{\mathcal{L}}(\lambda)\dot{\mathcal{Q}}^2(\lambda)}_{A_3(\lambda)}$$

$$+ \underbrace{\mathcal{L}(\lambda)\ddot{\mathcal{L}}(\lambda)\mathcal{Q}(\lambda)\dot{\mathcal{Q}}(\lambda)}_{A_4(\lambda)}.$$

$$(27)$$

A sufficient condition for $W_1(\cdot)$ to be a decreasing function in λ is that $A_\ell(\lambda) \leq 0$ for $\ell = 1, \dots, 4$. From *Lemmas 3-6*, this can be readily proved. In particular, $A_\ell(\lambda) \leq 0$ for $\lambda \geq 0$ if $\ell = 1, 2, 3$ and $A_4(\lambda) \leq 0$ for $\lambda_{\text{MT}}/\lambda \geq 2.8$. Therefore, a unique stationary point, λ^*, exists. Also, $\dot{\text{EE}}(\lambda) > 0$ for $\lambda < \lambda^*$ and $\dot{\text{EE}}(\lambda) < 0$ for $\lambda > \lambda^*$. Finally, by taking into account the constraints on the density of BSs, it follows that the unique optimal maximizer of the EE is $\lambda^{(\text{opt})} = \max\left\{\lambda^{(\min)}, \min\left\{\lambda^*, \lambda^{(\max)}\right\}\right\}$, since $\lambda \in \left[\lambda^{(\min)}, \lambda^{(\max)}\right]$.

b) Case Study $\lambda_{\text{MT}}/\lambda \leq 2.8$: As for this case study, we leverage a notable result in fractional optimization [2]: the ratio between a i) non-negative, differentiable and concave function, and a ii) positive, differentiable and convex function is a pseudo-concave function. It is, in addition, a unimodal function with a finite maximizer if the ratio vanishes when the variable of interest (i.e., the BSs' density) tends to zero and to infinity. As for the case study under analysis, the EE in (14) can be re-written, by neglecting unnecessary constants that are independent of λ and do not affect the properties of the function, as follows:

$$\text{EE}(\lambda) =$$
$$\frac{\mathcal{Q}(\lambda)}{[1 + \Upsilon\mathcal{L}(\lambda)]\left[(\mathrm{P}+\Delta\mathrm{P}) + \mathrm{P_i}/\mathcal{L}(\lambda) + \mathrm{P_c}\mathcal{M}(\lambda)/\mathcal{L}(\lambda)\right]}.$$

$$(28)$$

From *Lemma 6*, the numerator of (28) is a non-negative, differentiable, increasing and concave function for $\lambda \geq 0$. From *Lemma 7*, the EE in (28) tends to zero if $\lambda \to 0$ and $\lambda \to \infty$. Therefore, a sufficient condition to prove the unimodality and pseudo-concavity of the EE is to show that the denominator of (28) is a positive, differentiable and convex function in λ for $\lambda_{\text{MT}}/\lambda \leq 2.8$. From *Lemma 3* and *Lemma 4*, the first two properties are immediately verified. To complete the proof, the convexity of the denominator of (28) needs to be analyzed.

Let $\text{Den}(\cdot)$ be the denominator of (28). Let us introduce the function $\mathcal{K}(\lambda) = 2\dot{\mathcal{L}}^2(\lambda)\big/\mathcal{L}(\lambda) - \ddot{\mathcal{L}}(\lambda)$. The second-order derivative of $\text{Den}(\cdot)$, as a function of λ, is as follows:

$$\ddot{\text{Den}}(\lambda) = \underbrace{\Upsilon(\mathrm{P}+\Delta\mathrm{P})\ddot{\mathcal{L}}(\lambda)}_{D_1(\lambda)} + \underbrace{\Upsilon\mathrm{P_c}\ddot{\mathcal{M}}(\lambda)}_{D_2(\lambda)}$$

$$+ \underbrace{\left(\mathrm{P_c}/\mathcal{L}^2(\lambda)\right)\left(2\lambda_{\text{MT}}/\lambda^3\right)}_{D_3(\lambda)}\underbrace{\left(\mathcal{L}(\lambda) + \lambda\dot{\mathcal{L}}(\lambda)\right)}_{D_4(\lambda)}$$

$$+ \underbrace{\mathrm{P_c}\left(\mathcal{M}(\lambda)/\mathcal{L}^2(\lambda)\right)\mathcal{K}(\lambda)}_{D_5(\lambda)} + \underbrace{\left(\mathrm{P_c}/\mathcal{L}(\lambda)\right)\mathcal{K}(\lambda)}_{D_6(\lambda)}$$

$$+ \underbrace{\left(\mathrm{P_i}/\mathcal{L}^2(\lambda)\right)\mathcal{K}(\lambda)}_{D_7(\lambda)}.$$

$$(29)$$

A sufficient condition for proving that $\text{Den}(\cdot)$ is a convex function in λ is to show that $D_\ell(\lambda) \geq 0$ for $\ell = 1, 2, \dots, 7$ and $\mathcal{K}(\lambda) \geq 0$ if $\lambda_{\text{MT}}/\lambda \leq 2.8$. This can be proved as follows. $D_1(\lambda) \geq 0$ for $\lambda_{\text{MT}}/\lambda \leq 2.8$ follows from *Lemma 3*. $D_\ell(\lambda) \geq 0$ for $\ell = 2, 5$ if $\lambda \geq 0$ follows from *Lemma 4*. $D_\ell(\lambda) \geq 0$ for $\ell = 3, 6, 7$ if $\lambda \geq 0$ follows from *Lemma 3*. $D_4(\cdot)$ and $\mathcal{K}(\cdot)$ require deeper analysis. Define $\xi = \lambda_{\text{MT}}/\lambda$. $D_4(\cdot)$ and $\mathcal{K}(\cdot)$ are positive functions in ξ if:

$$D_4(\xi) \geq 0 \Leftrightarrow$$
$$\underline{D}_4(\xi) = 1 - (1+\xi/\alpha)^{-\alpha} - x(1+\xi/\alpha)^{-(\alpha+1)} \geq 0 \quad (30)$$

$$\mathcal{K}(\xi) \geq 0 \Leftrightarrow$$
$$\underline{\mathcal{K}}(\xi) = (1+\xi/\alpha)^{-\alpha} \quad (31)$$
$$+ \left[2 + (1+1/\alpha)x\right]\left[2 - (1-1/\alpha)x\right]^{-1} \geq 1.$$

By direct inspection of (30) and (31), it is not difficult to prove the following: i) $\underline{D}_4(\xi \to 0) = 0$ and $\underline{\dot{D}}_4(\xi) \geq 0$ for $\xi \geq 0$, and ii) $\underline{\mathcal{K}}(\xi \to 0) = 1$ and $\underline{\dot{\mathcal{K}}}(\xi) \geq 0$ for $\xi \leq 2.8$. These two conditions imply $D_4(\lambda) \geq 0$ for $\lambda \geq 0$ and $\mathcal{K}(\lambda) \geq 0$ for $\lambda_{\mathrm{MT}}/\lambda \leq 2.8$. This concludes the proof.

REFERENCES

[1] S. Buzzi, Chih-Lin I, T. E. Klein, H. V. Poor, C. Yang, and A. Zappone, "A survey of energy-efficient techniques for 5G networks and challenges ahead", *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 697-709, Apr. 2016.

[2] A. Zappone and E. Jorswieck, "Energy efficiency in wireless networks via fractional programming theory", *Found. Trends Commun. Inf. Theory*, vol. 11, no. 3-4, pp. 185-396, Jun. 2015

[3] 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.

[4] 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.

[5] M. Di Renzo, W. Lu, and P. Guan, "The intensity matching approach: A tractable stochastic geometry approximation to system-level analysis of cellular networks", *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 5963-5983, Sep. 2016.

[6] W. Lu and M. Di Renzo, "Stochastic geometry modeling of cellular networks: Analysis, simulation and experimental validation", *ACM MSWiM*, pp. 179-188, Nov. 2015.

[7] H. ElSawy, A. K. Sultan-Salem, M.-S. Alouini, and M. Z. Win, "Modeling and analysis of cellular networks using stochastic geometry: A tutorial", *IEEE Proc.*, vol. 19, no. 1, pp. 167-203, Jan. 2017.

[8] 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.

[9] T. Kwon, "Spatial topology adjustment for minimizing multicell network power consumption", *arXiv:1403.4346*, Mar. 2014. [Online]. Available: https://arxiv.org/pdf/1403.4346.pdf.

[10] 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.

[11] D. Cao, S. Zhou, and Z. Niu, "Optimal combination of base station densities for energy-efcient two-tier heterogeneous cellular networks", *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4350-4362, Sep. 2013.

[12] S. Luo, R. Zhang, and T. J. Lim, "Optimal power and range adaptation for green broadcasting", *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4592-4603, Sep. 2013.

[13] M. Wildemeersch, T. Q. S. Quek, C. H. Slump, and A. Rabbachin, "Cognitive small cell networks: Energy efficiency and trade-offs", *IEEE Trans. Commun.*, vol. 61, no. 9, pp. 4016-4028, Sep. 2013.

[14] C. Li, J. Zhang, and K. B. Letaief, "Throughput and energy efficiency analysis of small cell networks with multi-antenna base stations", *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2505-2517, May 2014.

[15] 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.

[16] C.-Y. Chang, W. Liao, H.-Y. Hsieh, and D.-S. Shiu, "On optimal cell activation for coverage preservation in green cellular networks", *IEEE Trans. Mob. Comp.*, vol. 13, no. 11, pp. 2580-2591, Nov. 2014.

[17] 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.

[18] X. Ge, B. Yang, J. Ye, G. Mao, C.-X. Wang, and T. Han, "Spatial spectrum and energy efciency of random cellular networks", *IEEE Trans. Commun.*, vol. 63, no. 3, pp. 1019-1030, Mar. 2015.

[19] C. Liu, B. Natarajan, and H. Xia, "Small cell base station sleep strategies for energy efciency", *IEEE Trans. Veh. Techol.*, vol. 65, no. 3, pp. 1652-1661, Mar. 2016.

[20] T. Zhang, J. Zhao, L. An, and D. Liu, "Energy efficiency of base station deployment in ultra dense hetnets: A stochastic geometry analysis", *IEEE Wireless Commun. Lett.*, vol. 5, no. 2, pp. 184-187, Apr. 2016.

[21] J. B. Rao and A. O. Fapojuwo, "An analytical framework for evaluating spectrum/energy efficiency of heterogeneous cellular networks", *IEEE Trans. Veh. Techol.*, vol. 65, no. 5, pp. 3568-3584, May 2016.

[22] Z. Chen, L. Qiu, and X. Liang, "Area spectral efficiency analysis and energy consumption minimization in multiantenna Poisson distributed networks", *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4862-4874, Jul. 2016.

[23] L. Li, M. Peng, C. Yang, and Y. Wu, "Optimization of base-station density for high energy-efficient cellular networks with sleeping strategies", *IEEE Trans. Wireless Commun.*, vol. 65, no. 9, pp. 7501-7514, Sep. 2016.

[24] A. Shojaeifard, K.-K. Wong, K. A. Hamdi, E. Alsusa, D. K. C. So, and J. Tang, "Stochastic geometric analysis of energy-efficient dense cellular networks", *IEEE Access*, vol. 5, no. 3, pp. 455-469, Mar. 2017.

[25] P. Chang and G. Miao, "Energy and spectral efficiency of cellular networks with discontinuous transmission", *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2991-3002, May 2017.

[26] A. Alam, P. Mary, J.-Y. Baudais, and X. Lagrange, "Asymptotic analysis of area spectral efficiency and energy efficiency in PPP networks with SLNR precoder", *IEEE Trans. Commun.*, vol. 65, no. 7, pp. 3172–3185, July 2017.

[27] S. Mukherjee, *Analytical Modeling of Heterogeneous Cellular Networks: Geometry, Coverage, and Capacity*, Cambridge University Press, 1st ed., Feb. 2014.

[28] F. Baccelli and B. Blaszczyszyn, *Stochastic Geometry and Wireless Networks, Part I: Theory*, Now Publishers, Sep. 2009.

[29] J. Liu, M. Sheng, L. Liu, and J. Li, "Effect of densication on cellular network performance with bounded pathloss model', *IEEE Commun. Lett.*, vol. 21, no. 2, pp. 346-439, Feb. 2017.

[30] G. Auer et al., "How much energy is needed to run a wireless network?", *IEEE Wireless Commun. Mag.*, vol. 18, no. 5, pp. 40-49, Oct. 2011.

[31] 3GPP TS36.304, "3rd generation partnership project; technical specification group radio access network, evolved universal terrestrial radio access (E-UTRA); user equipment (UE) procedures in idle mode". [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/Specification Details.aspx?specificationId=2432.

[32] S. Bhaumik, G. Narlikar, S. Chattopadhyay, and S. Kanugovi, "Breathe to stay cool: Adjusting cell sizes to reduce energy consumption", *ACM SIGCOMM, Green Networking Workshop*, pp. 41-46, Aug. 2010.

[33] S. M. Yu and S.-L. Kim, "Downlink capacity and base station density in cellular networks", *IEEE Workshop on Spatial Stochastic Models for Wireless Networks*, pp. 1-7, May 2013. [Online]. Available: http://arxiv.org/pdf/1109.2992.pdf.

[34] B. Yu, S. Mukherjee, H. Ishii, and L. Yang, "Dynamic TDD support in the LTE-B enhanced local area architecture", *IEEE Global Commun. Conf. – Workshops*, pp. 585591, Dec. 2012.

[35] A. Rajanna and M. Haenggi, "Enhanced cellular coverage and through-put using rateless codes", *IEEE Trans. Commun.*, vol. 65, no. 5, pp. 1899-1912, May 2017.

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

[37] P. Calka, "Precise formulae for the distributions of the principal geometric characteristics of the typical cells of a two-dimensional Poisson voronoi tessellation and a Poisson line process", *Adv. Appl. Prob.*, vol. 35, pp. 551-562, Sep. 2003.

[38] H. S. Dhillon, R. K. Ganti, J. G. Andrews, "Load-aware modeling and analysis of heterogeneous cellular networks", *IEEE Trans. Wireless Commun.*, vol. 12, no. 4, pp. 1666-1677, Apr. 2013.

[39] A. Shojaeifard, K. A. Hamdi, E. Alsusa, D. K. C. So, and J. Tang, "A unified model for the design and analysis of spatially-correlated load-aware HetNets", *IEEE Commun.*, vol. 62, no. 1, pp. 1-16, Nov. 2014.

[40] C. Isheden, Z. Chong, E. Jorswieck, and G. Fettweis, "Framework for link-level energy efficiency optimization with informed transmitter", *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2946-2957, Aug. 2012.

[41] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Sep. 1999.

[42] J. P. Crouzeix and J. A. Ferland, "Algorithms for generalized fractional programming", *Springer Mathematical Programming*, vol. 52, no. 1-3, pp. 191-207, May 1991.

[43] J. C. Bezdek and R. J. Hathaway, "Convergence of alternating optimization", *J. Neural, Parallel & Scientific Computations*, vol. 11, no. 4, pp. 351-368, Dec. 2003.