This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2672977, IEEE Transactions on Wireless Communications

Energy and Spectral Efficiency of Cellular Networks with Discontinuous Transmission

Peiliang Chang and Guowang Miao
KTH Royal Institute of Technology, Stockholm, Sweden
Email:{peiliang, guowang}@kth.se

***Abstract*—Cell discontinuous transmission (DTX) has been proposed as a solution to reduce energy consumption of cellular networks. This paper investigates the impact of network traffic load on spectral and energy efficiency of cellular networks with DTX. The SINR distribution as a function of traffic load is derived firstly. Then sufficient condition for ignoring thermal noise and simplifying the SINR distribution is investigated. Based on the simplified SINR distribution, the network spectral and energy efficiency as functions of network traffic load are derived. It is shown that the network spectral efficiency increases monotonically in traffic load, while the optimal network energy efficiency depends on the ratio of the sleep-mode power consumption to the active-mode power consumption of base stations. If the ratio is larger than a certain threshold, the network energy efficiency increases monotonically with network traffic load and is maximized when the network is fully loaded. Otherwise, the network energy efficiency firstly increases and then decreases in network traffic load. The optimal load can be identified with a binary search algorithm. The power ratio threshold depends solely on the path loss exponent α, e.g. 56% for $\alpha = 4$. All these analytic results are further validated by the numerical simulations.**

***Index Terms*—Green communication, cell discontinuous transmission, network traffic load, energy efficiency, network spectral efficiency.**

I. Introduction

Driven by the increasing usage of smart devices and mobile applications, the traffic of cellular networks has grown dramatically and this trend would continue in the future. It is forecasted that the global mobile traffic would increase by nearly tenfold from 2014 to 2019 [1]. Therefore network densification has been proposed to increase the network capacity by increasing the reuse of radio resources [2]. However, deploying more base stations (BSs) would lead to soaring energy consumption, which not only incurs severe environmental problems but also increases operation cost. It is therefore critical to increase the energy efficiency of cellular networks.

As indicated in [3], the energy consumption of BSs accounts for almost 60% of all the energy consumed by cellular networks. Different approaches have been proposed to reduce the energy consumption of BSs. One is to develop low-energy-consuming hardware and the other is to operate BSs to traffic demand. The latter is motivated by the fact that the existing BSs are deployed and operated to cater for the maximum traffic demand while the network traffic may vary in time [4]. BSs can be switched into lower energy consumption sleep mode when there is lower traffic demand to save energy.

According to time scale, there exist two types of traffic variation. One is the long-term traffic variation, for which the time scale is at level of hours and the average traffic intensity varies from hour to hour [5], [6]. The other is the short-term traffic varation, for which the time scale is at milliseconde level. Due to burst nature of traffic demand, a BS may have no traffic request for a short period [7]. Accordingly, BSs can have two levels of sleep mode: long-term deep sleep mode and short-term micro sleep mode. It takes long time, i.e. several minutes, for BSs to switched into deep sleep mode or wake up from deep sleep mode. Deep-sleep is used to adapt to the long-term traffic varation. BSs are switched into deep-sleep mode or waked up according to the variation intensity. On the other hand, the Orthogonal Frequency-Division Multiplexing (OFDM) technology provides the possibility for BSs to apply discontinuous transmission (cell DTX) techinique to adapt to the short-term traffic variation [7]–[9]. BSs can servce incoming traffic request in certain slots and then switched into micro sleep mode during idle slots.

A. Related Works and Motivation

Many research efforts have been devoted to studying BS sleeping operations. In [5], the authors studied the performance of a real network and proved the energy saving potential of dynamic BS on/off operation. The impact of BS on/off operation frequency on energy savings is investigated in [10] and it is shown that the daily traffic pattern plays a central role in the design of dynamic BS operation strategy. In [11], a theoretical framework for BS energy saving that encompasses dynamic BS operation and user association is proposed and the optimal user association and BS sleeping operation is investigated considering both energy saving and flow-level delay. In [12], the authors studied the design of energy efficient cellular networks through the employment of BS sleep mode strategies as well as small cells, and investigated the tradeoff issues associated with these techniques. A distributed switching-on/off based energy saving algorithm is proposed in [4]. These works mainly focus on BS long-term sleeping operations that leverage large-scale traffic variation.

On the other side, Cell DTX also receives much attention. Cell DTX has been long applied in the GSM network [13] by muting the transmission during the silence periods and it has been proposed and discussed in the standardization body of the LTE networks as a candidate to reduce interference and improve energy efficiency [7]. In [8], the feasibility of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2672977, IEEE Transactions on Wireless Communications

cell DTX in LTE networks was examined and it is shown that significant energy saving can be achieved with cell DTX. In [14], the authors studied the design aspects of cell DTX to mitigate inter-cell interference, such as the time scale of DTX operations and whether the DTX operations should be restricted to small cells only. The energy saving by jointly considering network deployment and cell DTX was studied in [15]. These works mainly focused on the power saving with cell DTX rather than the network energy efficiency. Furthermore, most of these works obtained study results through numerical simulations for certain specific scenarios and there is no general analytic expressions for the obtained results.

Multiple models can be used to model and analyze the performance of cellular networks, such as the Wyner model [16] and the hexagonal grid model [17]. Although the Wyner model is tractable, it fails to capture the principal characteristics of the realistic cellular networks [18]. Regarding the grid model, it becomes intractable when the network size increases. Furthermore, the grid model cannot describe the irregularity of the practical cellular networks, which is more and more common due to the increasing deployment of small cells, such as pico cells and femto cells. In this paper, the cellular network is modelled as a homogeneous Poisson point process (PPP) and the tools of statistical geometry [19] are utilized to analyze the performance of cellular networks. This model could provide both accurate and tractable results on the network performance [20]. This approach has been extensively applied to analyze the performance of cellular networks [20]–[23]. Using stochastic geometry, the authors developed general models for multi-cell signal-to-interference-plus-noise ratio (SINR) and compared the outage predictions to those of the grid model and of the actual network deployment in [20]. It was shown that modelling cellular networks with PPP results in a pessimistic performance evaluation of real cellular networks. In [21], this work is extended to the heterogeneous networks and the coverage probability in K-tier heterogeneous cellular networks is derived. With the help of stochastic geometry, the optimal BS densities in heterogeneous cellular networks to maximize the energy efficiency are derived in [22]. In [23], the authors investigated the impact of BS density and antenna numbers on the throughput and energy efficiency of small networks with multi-antenna BSs. All these works mainly focused on the network deployment aspects and studied the impact of BS densities. Different from these, we consider the operation of cellular networks that have been deployed and investigate the impact of traffic load on the network performance.

B. Contributions

In this paper, we investigate the impact of traffic load on network performance and endeavor to discover the explicit relationship between traffic load and spectral and energy efficiency of cellular networks using cell DTX. The main contributions of this paper can be summarized as follows:

1) Derive the network SINR distribution while considering network traffic load. Then we further derive network spectral and energy efficiency.

2) Present a sufficient condition for a cellular network to be interference-limited.

3) Analyze the impact of network traffic load on network spectral and energy efficiency.

4) Run numerical simulations to further confirm the analytic results.

C. Paper Organization

The rest of the paper is organized as follows. The system model is described in Section II. In Section III, the extended traffic-aware SINR distribution model is derived, and the sufficient condition for networks to be interference-limited is also investigated. Section IV studies the impact of traffic load on network spectral and energy efficiency. Then the numerical results are presented in Section V. Finally Section VI concludes this paper.

II. System Model and Performance Metrics

In this section, we first describe the system model and the necessary assumptions for the performance analysis. Then the network traffic load and power consumption model are explained. In the end, the performance metrics are described.

A. Network Model

We consider the downlink transmission in a network where both BSs and users are randomly distributed. The network is assumed to be homogeneous in terms of both traffic demand and BS distribution. The distribution of BSs is modelled with an ergodic PPP Φ_B with density λ_B. Note that we consider homogeneous networks and the case of heterogeneous networks is beyond the scope of this paper. Each user is associated to its closest BS. Thus the coverage area of each BS can be modelled using the Poisson Voronoi Tessellation (PVT) method. Fig. 1 illustrates an example of such a network. All the BSs are assumed to support DTX. The BS stays in active mode and transmits when there is any traffic request. Otherwise, it switches into sleep mode and does not transmit. The universal frequency reuse is applied and the system bandwidth is W. The users within each cell equally share the resources in an orthogonal manner. Only path loss and fast fading are considered. The link between a BS and a user is modelled as follows:

$$P_r = P_t C K r^{-\alpha} h = P_e r^{-\alpha} h, \tag{1}$$

where P_r, P_t, C, K, r, and α denote the receive power, the transmit power, the antenna gain, the path loss constant at unit distance, the distance between the BS and the user and the path loss exponent respectively. In order to simplify notations, the product $P_t C K$ is noted as P_e. The random variable h models Rayleigh fading, i.g. $h \sim exp(1)$. Here we assume that the signals from both the serving BS and the interfering BSs experience Rayleigh fading. The power control is out of the scope of this paper and all the active BSs are assumed to transmit with the same power P_t.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2672977, IEEE Transactions on Wireless Communications

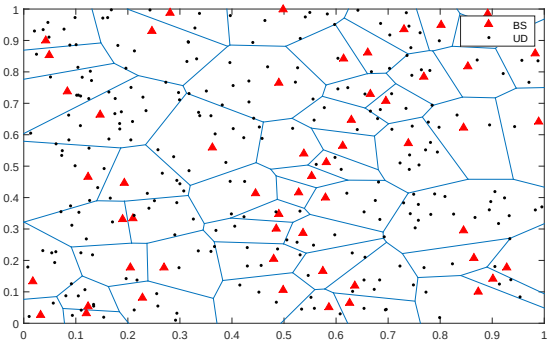

Figure 1. Distribution of base stations (BS) and user devices (UD) in a cellular network

With the above assumptions, the SINR of a downlink transmission from a user located at a distance of r from its associated BS b_0 can be expressed as

$$SINR = \frac{P_e r^{-\alpha} h_0}{\sum_{i \in \tilde{\Phi}_a \setminus b_0} P_e r_i^{-\alpha} h_i + \sigma_0^2} = \frac{P_e r^{-\alpha} h_0}{I_r + \sigma_0^2}, \quad (2)$$

where

$$I_r = \sum_{i \in \tilde{\Phi}_a \setminus b_0} P_e r_i^{-\alpha} h_i \quad (3)$$

represents the interference from the sets of active BSs $\tilde{\Phi}_a$ in the network, σ_0^2 is the thermal noise of the user device and r_i , h_i represent the distance from the user to the interfering BS i and the fast fading of the signal from BS i respectively.

B. Traffic Load Model

We consider packet based traffic request. The arrival of traffic request is modeled as a homogeneous temperal-spatial Poisson arrival process with intensity $\lambda_u(t)$ packets per second per square meter. Note that traffic density $\lambda_u(t)$ is constant during a short period, like one hour, but it may vary during a longer period, like from day time to deep in the night, [5], [6]. In this paper, we consider network performance during the period that $\lambda_u(t)$ keeps constant and network performance is stationary.

For a given BS, its load is defined as the percentage of utilized resources to satisfy its traffic requests. In order to maximize the available time for the sleep mode of cell DTX, it is assumed that the BS schedules all the bandwidth to serve its traffic requests to minimize its time in data transmission mode. Thus the BS load is modeled as the percentage of time that the BS is active. It is equivalent to the likelihood that the BS is active at any given instant. It should be noted that the more users that a BS serves, the more time it takes for the BS. Consequently, the BS is more likely to be active at any given instant. In general, due to the load-coupling among BSs caused by mutual interference, it is rather complex to identify an explicit relationship between the number of users served by a BS and its active probabilities [24].

In this paper, the coverage area sizes of all BSs follow the same distribution and they are on average the same [25].Furthermore, the traffic intensity is homogeneously distributed across different areas. Therefore, we assume that all BSs experience i.i.d traffic demands and the active probabilities of all BSs at an instant are the same. This active probability is used to model network traffic load ρ and also serves as an input to evaluate network performance. It is equivalent to the percentage of active BSs at a given instant, which can be easily measured by the network. As each BS independently decides its operation mode, the distribution of BSs after BS sleeping can be modelled as a thinned PPP $\tilde{\Phi}_a$ with a new BS density λ_a.

The relationship between the density of active BSs λ_a, the density of deployed BSs λ_B and the network load ρ can be expressed as

$$\lambda_a = \rho \lambda_B. \quad (4)$$

Remark. *For a given network topology, which is a realization of the PPP, the coverage area sizes of different BSs could be different, which leads to asymmetric traffic demands in each BS if users select their closest BSs as serving BS. Here, the active probability of each given BS is approximated with its expectation. As we focus on the macroscopic performance of network rather than each BS, this approximation is valid. This approach has been used and validated with real traffic data in [26]. Furthermore, in our numerical simulations, each user selects its closest BS as its serving BS and different BSs may have asymmetric traffic demands. The simulation results are very close to the results based on the proposed approximation of active probability.*

C. BS Dynamic DTX Operation and Power Consumption Model

BSs operate in two different modes: active mode and sleep mode. The dynamic DTX operations of BSs are illustrated in Fig. (2). When there is a traffic request, the BSs switch into the active mode and schedule all bandwidth to serve the traffic to shorten their active transmission time. Once the traffic request has been served, the BSs switch into DTX micro-sleep mode. The higher the load, i.e., resource utilization, of a BS is, the more time it stays in active mode. The mode switch time of cell DTX is expected to be very short, e.g., $35\mu s$ in [7], [8], which is much shorter than the length of transmission blocks, e.g., $1ms$, in LTE system. Therefore, we assume that the mode switch time is negligible and the BSs take no time to switch between active mode and micro-sleep mode.

The BS power consumption model proposed in [6] is adopted in this work. In the active mode, the BS power consumption is modelled as below:

$$P_a = \xi P_t + P_c, \quad (5)$$

where P_t and P_c are the transmit power and the circuit power respectively, and ξ is the inverse of the power amplifier efficiency. In the sleep mode, the BS turns off some components and consumes a smaller amount of power P_s. Define θ as

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2672977, IEEE Transactions on Wireless Communications

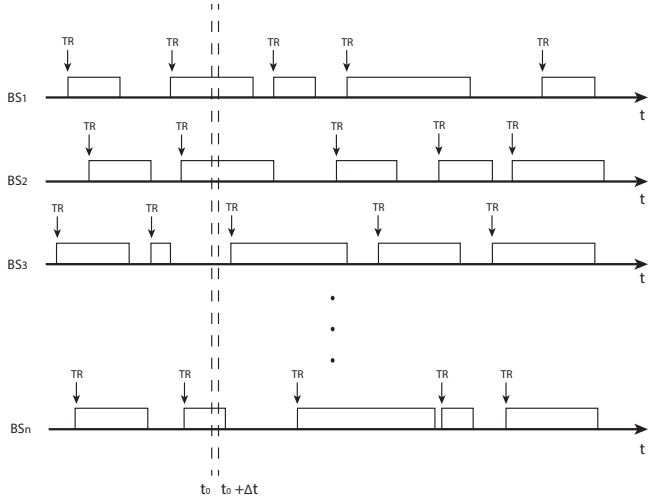

Figure 2. BS dynamic DTX operations

the ratio between the sleep-mode power consumption and the active-mode power consumption and we have

$$P_s = \theta P_a. \tag{6}$$

For a traffic load ρ, the average power consumption in unit area is

$$\mathbb{E}[P_u] = \lambda_a P_a + (\lambda_B - \lambda_a) P_s = \lambda_B P_a(\rho + (1-\rho)\theta). \tag{7}$$

Remark. *If longer mode switch time is considered, like several to hundreds of milliseconds, the BSs may not switch to micro-sleep mode if the interval between adjacent packets is shorter than the mode switch time to avoid delay in mode switching. In this case, the BSs keep their components on during idle time and consumes more energy than in the micro-sleep mode. The derived results on network energy efficiency need to be modified accordingly. The detailed analysis pf the impact of mode switching time and the corresponding traffic dynamics on network energy efficiency is beyond the scope of this paper and will be our future research.*

D. Performance Metrics

The focus of this paper is to analyze the impact of traffic load on network performance. The ergodic average network spectral and energy efficiency are used as metrics to evaluate network performance. Here the ergodic average is the spatial average in an infinite plan and it accounts for both the random Rayleigh fading and the random distribution of BSs and users. According to Shannon-Hartley theorem, the achievable spectral efficiency η_{SE} of a given link with SINR γ is defined as

$$\eta_{SE} = \log_2(1+\gamma). \tag{8}$$

Therefore the ergodic average link spectral efficiency is given as

$$\mathbb{E}[\eta_{SE}] = \mathbb{E}[\log_2(1+\gamma)], \tag{9}$$

where $\mathbb{E}[x]$ is the expectation of a random variable x. For a given traffic ρ, assume there exist N active BSs in a unit area. The ergodic average network spectral efficiency $\mathbb{E}[\eta_{ASE}]$ can be derived as

$$\begin{aligned}\mathbb{E}[\eta_{ASE}] &= \mathbb{E}\Big[\sum_{i=1}^{N}\log_2(1+\gamma_i)\Big] \\ &\overset{(a)}{=} \mathbb{E}[N]\,\mathbb{E}[\log_2(1+\gamma)] \\ &\overset{(b)}{=} \lambda_a\,\mathbb{E}[\eta_{SE}] \overset{(c)}{=} \rho\lambda_B\,\mathbb{E}[\eta_{SE}],\end{aligned} \tag{10}$$

where (a) follows from the approximation that the average area throughput is the product of average number of links per area and the average link throughput and this approximation will be validated by the simulation results in Section V;(b) follows from the fact that the average number of active BSs in unit area is λ_a and (c) is derived with (4). The unit of the network spectral efficiency is $bps/Hz/m^2$.

The ergodic average energy efficiency $\mathbb{E}[\eta_{EE}]$ is defined as the average number of bits that can be successfully transmitted with unit energy [27]–[29]. For a given traffic load ρ, the average network throughput in unit area is $\mathbb{E}[T_u] = W\,\mathbb{E}[\eta_{ASE}]$. Considering the power consumption model described in II-C, the average energy efficiency $\mathbb{E}[\eta_{EE}]$ is given as

$$\mathbb{E}[\eta_{EE}] = \frac{\mathbb{E}[T_u]}{\mathbb{E}[P_u]} = \frac{\rho W\,\mathbb{E}[\eta_{SE}]}{P_a(\rho + (1-\rho)\theta)}. \tag{11}$$

The unit of the average energy efficiency is $bits/Joule$.

III. SINR Distribution

In this section, we fisrt extend the existing model of SINR distribution by taking traffic load into consideration. Then the simplification of this extended model is discussed and a sufficient condition to simplify this model with adequate accuracy is obtained. Finally the impact of the traffic load on the SINR distribution is investigated.

A. Closed-form expression of SINR distribution

In [20], the SINR distribution of cellular networks has been studied by modelling the locations of BSs with homogeneous PPP. In the work of [20], the BSs are assumed to be always transmitting regardless of network traffic load. For the networks where the BSs have the capability of DTX, the BSs could switch into sleep mode when there is no traffic request. This would undoubtedly reduce the interference suffered by the users. Following the same approach of [20], the distribution of SINR as a function of traffic load can be derived as follows.

Proposition 1. *For a homogeneous cellular network with cell DTX, the cumulative density function (CDF) of the downlink transmission SINR under the traffic load ρ is*

$$P[SINR \le \gamma] = 1 - \pi\lambda_B \int_0^{\infty} e^{-\pi\lambda_B x(1+\rho\beta(\gamma,\alpha)) - \frac{1}{P_e}\gamma\sigma_0^2 x^{\alpha/2}} dx \tag{12}$$

where

$$\beta(\gamma,\alpha) = \gamma^{2/\alpha}\int_{\gamma^{-2/\alpha}}^{\infty}\frac{1}{1+y^{\alpha/2}}dy \tag{13}$$

Proof. See Appendix A. □

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2672977, IEEE Transactions on Wireless Communications

B. Sufficient condition for being interference-limited

According to (12), the double-integral needs to be calculated to obtain the CDF of the downlink transmission SINR. Furthermore, in order to calculate the spectral efficiency, another more integral is necessary (see IV-A). This leads to challenging computational complexity. By assuming the network to be interference-limited and neglecting the impact of thermal noise, a simplified form of the SINR distribution can be obtained as follows,

$$P[SINR \leq \gamma] = 1 - \frac{1}{1 + \rho\beta(\gamma, \alpha)}. \quad (14)$$

Neglecting the contribution of thermal noise would unavoidably result in loss of accuracy. In fact, the impact of thermal noise on SINR distribution relies on several factors, such as BS transmit power, BS density and path loss exponent. The loss of accuracy of this simplification can be described with following theorem.

Theorem 1. *Define the probabilities obtained with (12) and (14) as $P(\gamma)$ and $P_s(\gamma)$ respectively. For any small positive number ϵ, $0 \leq P(\gamma) - P_s(\gamma) \leq \epsilon$, if the following inequality is satisfied*

$$\lambda_B \geq \frac{-ln(\epsilon/2)}{\pi(1 + \rho\beta(\gamma, \alpha))} \Big(\frac{\gamma\sigma_0^2}{-P_e ln(1 - \epsilon/2)} \Big)^{(2/\alpha)}. \quad (15)$$

Proof. See Appendix B. □

Theorem 1 provides a sufficient condition for a network to be interference-limited and quantifies the required BS density for neglecting the impact of thermal noise with acceptable accuracy. Let's denote the right side of (15) as λ_0. When the BS density is larger than λ_0, the loss of simplified CDF $P_s(\gamma)$ is below ϵ. It can be easily proved that λ_0 is increasing in γ, α and ϵ, and decreasing in ρ.

The average inter site distance (ISD) of cellular networks is proportional to the inverse of the square root of BS density as $ISD \approx \frac{1}{\sqrt{\lambda_B}}$. According to Theorem 1, the ISD should be small enough to neglect the impact of thermal noise and the maximum ISD necessary for a network to be interference limited is decreasing in γ, α and ϵ, and decreasing in ρ. Using the system parameters in [30], the distribution of SIR and SINR with different ISDs are simulated and presented in Fig. 3. According to (14), the SIR distribution is independent with the BS densities. Thus only the CDF of SIR of one BS density is plotted. From this figure we can see that the CDF of SIR and the CDFs of SINR with small ISDs (ISD = 100 m, ISD= 316 m and ISD = 1000 m) are the same. The thermal noise has little impact on the SINR distribution. Even for the large ISD, 3160 m, the impact is still not so much and there is only a slight difference between SIR and SINR.

Using the same parameters and applying Theorem 1, Table I gives the maximum ISD necessary for the networks to be interference-limited for servral traffic loads. The ISDs in Table I are still comparable to those of the realistic networks [30]. As λ_0 is increasing in SINR γ, for SINRs smaller than 20 dB, the required ISD can be even larger to assure the accuracy of simplification.

Table I
MAXIMUM OF REQUIRED ISD FOR THE TOLERABLE LOSS BELOW 1%, $\gamma = 20dB, P_t = 43dBmW, \sigma_0^2 = -96dBmW, C = 17dB, K = -35dB$

Max ISD	$\alpha = 2.5$	$\alpha = 3$	$\alpha = 3.5$	$\alpha = 4$
$\rho = 0.1$	4298 m	758 m	242 m	108 m
$\rho = 0.5$	9392 m	1580 m	475 m	198 m
$\rho = 1$	13245 m	2213 m	660 m	272 m

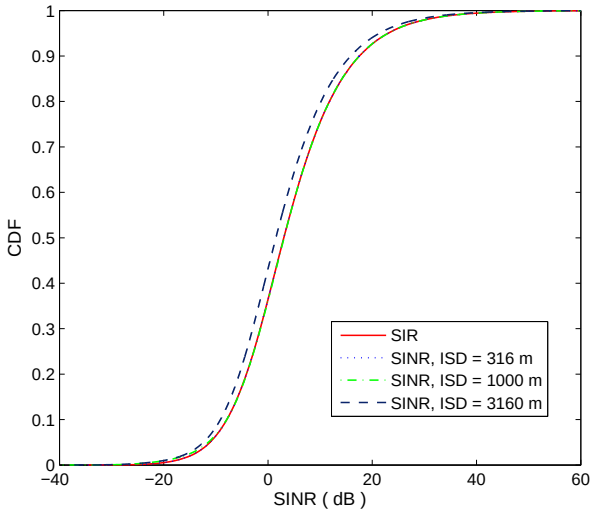

Figure 3. CDF of SINR with different ISDs, $P_t = 43dBmW, \sigma_0^2 = -96dBmW, C = 17dB, K = -35dB, \alpha = 3.5$ and $\rho = 0.6$

Based on these observations, we can say that the thermal noise has impact on SINR distribution only for extremely large ISDs. As more and more BS are deployed to cater for the increasing traffic demand the BS densities would become larger in the future. Therefore, we assume the network to be interference-limited.

C. Impact of traffic load on SINR distribution

Using Proposition 1 and Theorem 1, we can obtain the following proposition.

Proposition 2. *For interference-limited cellular networks with cell DTX, the CDF of the downlink SINR is monotonically increasing with network load ρ.*

Proof. According to Proposition 1, the CDF of the downlink SINR of interference-limited cellular networks with cell DTX is expressed as (14). Taking the derivative of CDF over ρ gives

$$\frac{\partial P[SINR \leq \gamma]}{\partial \rho} = \frac{\beta(\gamma, \alpha)}{(1 + \rho\beta(\gamma, \alpha))^2}. \quad (16)$$

As $\beta(\gamma, \alpha) > 0$, $\frac{\partial P[SINR \leq \gamma]}{\partial \rho}$ is positive. The proposition is proved. □

This proposition indicates that the higher the traffic load is the worse the link quality of an individual user is.

IV. NETWORK SPECTRAL AND ENERGY EFFICIENCY

In this section, we investigate the impact of traffic load on the network spectral and energy efficiency.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2672977, IEEE Transactions on Wireless Communications

A. Network Spectral Efficiency

In this subsection we investigate the impact of the network traffic load on the ergodic average network spectral efficiency.

In order to study the ergodic average spectral efficiency of cellular networks, the following lemma is needed.

Lemma 1. *For a positive random variable x with CDF $F_x(x)$ and probability density function (PDF) $f_x(x)$, its expectation can be calculated as follows,*

$$\mathbb{E}[x] = \int_0^\infty (1 - F_x(x))dx. \tag{17}$$

Proof. See Appendix C. □

With Lemma 1, the ergodic average spectral efficiency for a single downlink transmission and the ergodic average network spectral efficiency for the network are given in the following proposition.

Proposition 3. *For a homogeneous cellular network with cell DTX, the ergodic average spectral efficiency for a single link is*

$$\mathbb{E}[\eta_{SE}] = \int_0^\infty \frac{1}{1 + \rho\beta(2^\tau - 1, \alpha)} d\tau, \tag{18}$$

and the ergodic average network spectral efficiency for the network is

$$\mathbb{E}[\eta_{ASE}] = \lambda_B \rho \int_0^\infty \frac{1}{1 + \rho\beta(2^\tau - 1, \alpha)} d\tau. \tag{19}$$

Proof. According to definition (8), the link spectral efficiency is positive. Using Lemma 1, the expectation of link spectral efficiency η_{SE} is calculated as

$$\begin{aligned} \mathbb{E}[\eta_{SE}] &= \int_0^\infty (1 - F_{\eta_{SE}}(x))dx \\ &= \int_0^\infty P[\eta_{SE} > x]dx. \end{aligned} \tag{20}$$

Plugging (8) into (20) gives

$$\begin{aligned} \mathbb{E}[\eta_{SE}] &= \int_0^\infty P[\log_2(1 + SINR) > \tau]d\tau \\ &= \int_0^\infty P[SINR > 2^\tau - 1]d\tau \\ &\overset{(a)}{=} \int_0^\infty \frac{1}{1 + \rho\beta(2^\tau - 1, \alpha)} d\tau \end{aligned} \tag{21}$$

where (a) follows from (14).

With the definition in (10), the ergodic average network spectral efficiency is

$$\mathbb{E}[\eta_{ASE}] = \rho\lambda_B \int_0^\infty \frac{1}{1 + \rho\beta(2^\tau - 1, \alpha)} d\tau.$$

This finishes the proof of Proposition 3. □

With Proposition 3, the impact of traffic load on average link and network spectral efficiency can be described with following proposition.

Proposition 4. *The average link spectral efficiency is decreasing in traffic load while the average network spectral efficiency is increasing in traffic load. The average network spectral efficiency is maximized when the network is fully loaded.*

Proof. The proposition can be proved by taking the derivatives of average link and network spectral efficiency over traffic load. The derivative of average link spectral efficiency over traffic load is

$$\frac{\partial \mathbb{E}[\eta_{SE}]}{\partial \rho} = \int_0^\infty \frac{-\beta(2^\tau - 1, \alpha)}{(1 + \rho\beta(2^\tau - 1, \alpha))^2} d\tau < 0. \tag{22}$$

Thus the average link spectral efficiency $\mathbb{E}[\eta_{SE}]$ decreases as the network load ρ increases. The derivative of average network spectral efficiency over traffic load is

$$\frac{\partial \mathbb{E}[\eta_{ASE}]}{\partial \rho} = \int_0^\infty \frac{\lambda_B}{(1 + \rho\beta(2^\tau - 1, \alpha))^2} d\tau > 0. \tag{23}$$

Therefore the network spectral efficiency $\mathbb{E}[\eta_{ASE}]$ is monotonically increasing with the network load ρ. This finishes the proof. □

The impacts of network load on spectral efficiency and network spectral efficiency are validated by the simulation results (Fig. 7 and Fig. 8) in section V.

B. Energy Efficiency

In this section we study the impact of traffic load on the average energy efficiency of cellular networks with cell DTX.

By plugging (18) into (11), the ergodic average energy efficiency can be easily derived as follows,

$$\mathbb{E}[\eta_{EE}] = \int_0^\infty \frac{W\rho}{P_a(1 + \rho\beta(2^\tau - 1, \alpha))(\rho + (1 - \rho)\theta)} d\tau. \tag{24}$$

The impact of traffic load on the average energy efficiency can be described with following theorem.

Theorem 2. *For cellular networks with cell DTX, the average energy efficiency is a strictly quasi-concave function of the network traffic load. If*

$$\theta = \frac{P_s}{P_a} \geq \frac{\Omega(\alpha)}{1 + \Omega(\alpha)},$$

the average energy efficiency is monotonically increasing with network load ρ. Otherwise there exists a unique traffic load ρ^* that is less than 1 and maximizes the average energy efficiency. The average energy efficiency increases with traffic load on $[0, \rho^*)$ and decreases with traffic load on $(\rho^*, 1]$. ρ^* is the solution of the following equation

$$g(\rho^*) = \int_0^\infty \frac{\theta - \rho^{*2}\beta(1 - \theta)}{(1 + \rho^*\beta)^2} d\tau = 0. \tag{25}$$

And $\Omega(\alpha)$ is defined as

$$\Omega(\alpha) = \frac{\int_0^\infty \frac{\beta(2^\tau - 1, \alpha)}{(1 + \beta(2^\tau - 1, \alpha))^2} d\tau}{\int_0^\infty \frac{1}{(1 + \beta(2^\tau - 1, \alpha))^2} d\tau}. \tag{26}$$

Proof. See Appendix D. □

The relationship between the traffic load and the normalized average energy efficiency for different power ratio θ is illustrated in Fig. 4. It is clear that for large power ratio θ, the average energy efficiency is increasing in traffic load. While for small θ, the average energy efficiency first increases and then decreases in traffic load. This can be explained by

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2672977, IEEE Transactions on Wireless Communications

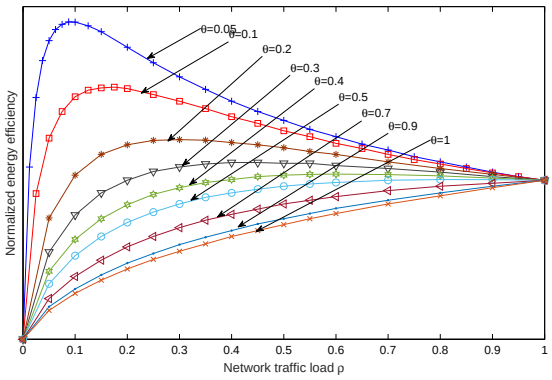

Figure 4. Distribution of normalized average energy efficiency with traffic load (Numerical result), $\alpha = 4$

analysing the relationship between the increase of network spectral efficiency and the increase of network power consumption as the traffic load increases. According to (7), the average area network power consumption increases linearly in traffic load and the increasing rate is $\lambda_B P_a(1-\theta)$. The smaller θ is, the larger the increasing rate is. According to (10), the average area throughput also increases in traffic load. However, according to (22), the increasing rate $\lambda_B \mathbb{E}[\eta_{SE}]$ is decreasing in traffic load as the average link quality deteriorates when the traffic load increases. The minimum increasing rate of the average area throughput is attained when the network is fully loaded, i.e., $\rho = 1$. At network load ρ_0, the average area spectral efficiency is $\mathbb{E}[\eta_{ASE}](\rho_0)$, the average area power consumption is $\mathbb{E}[P_u](\rho_0)$ and the average network energy efficiency is $\mathbb{E}[\eta_{EE}](\rho_0) = \frac{W\,\mathbb{E}[\eta_{ASE}](\rho_0)}{\mathbb{E}[P_u](\rho_0)}$. If network load increases by $\Delta\rho$, the average area spectral efficiency increases by $\Delta\rho\lambda_B\,\mathbb{E}[\eta_{SE}](\rho_0)$ and the average area power consumption increases by $\Delta\rho\lambda_B P_a(1-\theta)$. The corresponding average network energy efficiency is $\mathbb{E}[\eta_{EE}](\rho_0+\Delta\rho) = \frac{W(\mathbb{E}[\eta_{ASE}](\rho_0)+\Delta\rho\lambda_B\,\mathbb{E}[\eta_{SE}](\rho_0))}{\mathbb{E}[P_u](\rho_0)+\Delta\rho\lambda_B P_a(1-\theta)}$. If

$$\begin{aligned}\frac{W\,\mathbb{E}[\eta_{SE}](\rho_0)}{P_a(1-\theta)} &= \frac{\Delta\rho\lambda_B W\,\mathbb{E}[\eta_{SE}](\rho_0)}{\Delta\rho\lambda_B P_a(1-\theta)} \\ &\geq \frac{W\,\mathbb{E}[\eta_{ASE}](\rho_0)}{\mathbb{E}[P_u](\rho_0)} = \mathbb{E}[\eta_{EE}](\rho_0),\end{aligned} \tag{27}$$

then $\mathbb{E}[\eta_{EE}](\rho_0+\Delta\rho) \geq \mathbb{E}[\eta_{EE}](\rho_0)$, i.e., the average energy efficiency is increasing at ρ_0. For a given θ, $\frac{W\,\mathbb{E}[\eta_{SE}](\rho_0)}{P_a(1-\theta)}$ is minimized at $\rho = 1$. If $\frac{W\,\mathbb{E}[\eta_{SE}](1)}{P_a(1-\theta)} \geq \mathbb{E}[\eta_{EE}](1)$, then the average energy efficiency is always increasing on $[0,1]$ and it is maximized at $\rho = 1$. If $\frac{W\,\mathbb{E}[\eta_{SE}](1)}{P_a(1-\theta)} < \mathbb{E}[\eta_{EE}](1)$, there must exist $\rho^* \in (0,1)$ such that $\frac{W\,\mathbb{E}[\eta_{SE}](\rho^*)}{P_a(1-\theta)} = \mathbb{E}[\eta_{EE}](\rho^*)$ as both $\frac{W\,\mathbb{E}[\eta_{SE}](\rho)}{P_a(1-\theta)}$ and $\mathbb{E}[\eta_{EE}](\rho)$ are continuous functions of ρ and $\mathbb{E}[\eta_{EE}](0) = 0$ while $\frac{W\,\mathbb{E}[\eta_{SE}](0)}{P_a(1-\theta)} > 0$. In this case, the average energy efficiency increases in load on $[0,\rho^*)$ and decreases in load on $(\rho^*,1]$. It is maximized at $\rho = \rho^*$. This will be further validated by the simulation results in Section V.

Now let us study the impact of θ on network energy efficiency. $\frac{W\,\mathbb{E}[\eta_{SE}](\rho)}{P_a(1-\theta)}$ is increasing in power ration θ while $\mathbb{E}[\eta_{EE}](\rho)$ is decreasing in power ratio θ. Consider a power ratio θ_0 such that

$$\frac{W\,\mathbb{E}[\eta_{SE}](1)}{P_a(1-\theta_0)} = \mathbb{E}[\eta_{EE}](1) = \int_0^\infty \frac{W}{P_a(1+\beta(2^\tau-1,\alpha))}d\tau. \tag{28}$$

If $\theta > \theta_0$, then $\frac{W\,\mathbb{E}[\eta_{SE}](1)}{P_a(1-\theta)} > \mathbb{E}[\eta_{EE}](1)$, and $\frac{W\,\mathbb{E}[\eta_{SE}](\rho)}{P_a(1-\theta)} > \mathbb{E}[\eta_{EE}](\rho)$ holds for all $\rho \in [0,1]$. The increase of power consumption is always relatively slower than that of network throughput and the average energy efficiency always increases in traffic load. If $\theta < \theta_0$, then $\frac{W\,\mathbb{E}[\eta_{SE}](1)}{P_a(1-\theta)} < \mathbb{E}[\eta_{EE}](1)$, and there must exist $\rho^* \in (0,1)$ such that $\frac{W\,\mathbb{E}[\eta_{SE}](\rho^*)}{P_a(1-\theta)} = \mathbb{E}[\eta_{EE}](\rho^*)$. The increase of network throughput is relatively faster than that of power consumption in low load zone $[0,\rho^*)$ but slower in high load zone $(\rho^*,1]$. Consequently, the average energy efficiency first increases and then decreases in traffic load. According to (18), $\mathbb{E}[\eta_{SE}](1)$ is only determined by the path loss exponent α. Therefore, θ_0 that solves (28) is only determined by α. This explains why the threshold of power ratio θ is only determined by the path loss exponent α in Theorem 2. As larger α leads to higher increasing rate of network spectral efficiency, the threshold of θ for larger α is smaller than that of smaller α. By numerically calculating (26), the threshold is found to be 56.2% for $\alpha = 4$ and 59.3% for $\alpha = 3$.

Optimal traffic load for maximizing energy efficiency: According to Theorem 2, the average energy efficiency is always maximized when the network is fully loaded if $\theta > \frac{\Omega(\alpha)}{1+\Omega(\alpha)}$. Otherwise the optimal traffic load ρ^* is less than 1 and it is affected by the power ratio θ. The relationship between θ and the optimal traffic load ρ^* can be described with the following proposition.

Proposition 5. *If $\theta < \frac{\Omega(\alpha)}{1+\Omega(\alpha)}$, the optimal traffic load ρ^* for maximizing the average energy efficiency is monotonically increasing in the power ratio θ.*

Proof. See Appendix E. □

This proposition indicates that the lower θ is, the lower the optimal traffic load for maximizing the energy efficiency is.

If $\theta < \frac{\Omega(\alpha)}{1+\Omega\alpha}$, the optimal traffic load ρ^* is determined by equation (25). However, there is no explicit expression of ρ^*. As indicated in Appendix D, the function $g(\rho)$ is monotonically decreasing in ρ. As a result, the binary search method in [31] can be used to find the optimal traffic load for maximizing the energy efficiency. With this algorithm, the optimal traffic load ρ^* for $\alpha = 4$ is numerically calculated and illustrated in Fig. 5. It is clear that the optimal traffic load ρ^* increases in power ratio θ and it becomes 1 when $\theta \geq 56\%$.

V. Numerical Results

In this section, the above analysis on the performance of cellular networks with DTX are demonstrated with Monte Carlo simulations. For the simulations, the BSs and the users are randomly placed as two independent PPPs. The size of the simulated area is $100Km^2$ and the number of BSs is 1000. The number of users ranges from 100 to 4000 to simulate different traffic loads. The system bandwidth is $10MHz$. Regarding

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2672977, IEEE Transactions on Wireless Communications

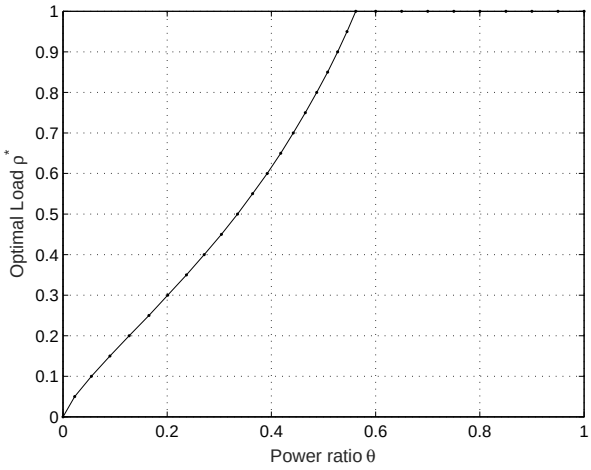

Figure 5. Distribution of optimal traffic load with power ratio, $\alpha = 4$

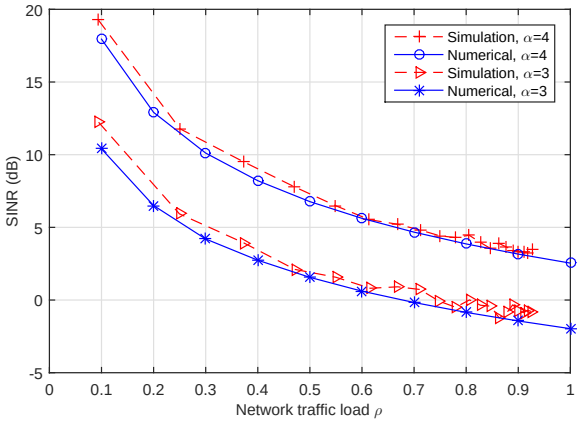

Figure 6. Distribution of average SINR under different network load

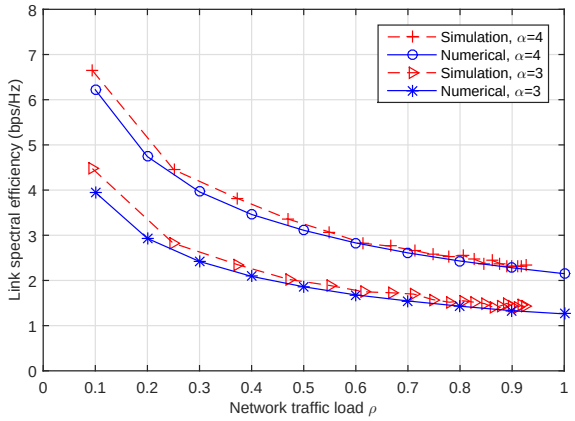

Figure 7. Distribution of average link spectral efficiency under different network load

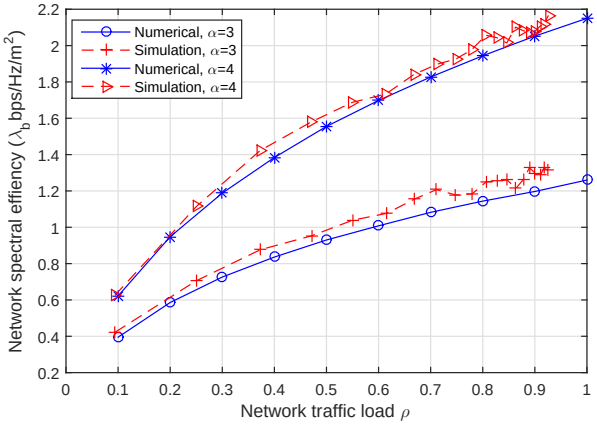

Figure 8. Distribution of average network spectral efficiency under different network load

the BS power consumption, there exists various simulation settings in different works [15], [23], [32]. In our simulations, the power consumption in the active mode is normalized to be 1 and that in the sleep mode is normalized by the active-mode power, ranging from 0.05 to 1. In order to demonstrate the impact of the radio propagation environment, two path loss exponents ($\alpha = 3$ and 4) are simulated. Besides the simulation results, the performances metrics are also calculated according to their expressions in Section IV.

Fig. 6 shows how the average SINR changes as the traffic load increases. Firstly, it is also shown that the simulation results and the numerical calculation results follow the same trend, although there is a minor gap between them. The results tell that the higher the traffic load is, the lower the average SINR is. This is due to the fact as the traffic load increases, there will be more active BSs in the network, which brings in stronger interference. Therefore the average SINR deteriorates. Comparing the results with different path loss exponents, we can find that the higher the exponent is, the better the SINR distribution is. This is due the fact that the relative fading between the interference signals and the serving signals is more significant in an environment with larger path loss exponent. As the link spectral efficiency is a monotonically increasing function of SINR, the similar results can be found for the link spectral efficiency (see Fig. 7).

Fig. 8 shows the impact of traffic load on network spectral efficiency. The vertical axis is normalized with the BS density λ_B. Unlike SINR and link spectral efficiency, the network spectral efficiency increases as the network traffic load increases. The maximum network spectral efficiency can be achieved when the network is fully loaded. This is resulted from the fact that as the traffic load increases, in spite of the deterioration of single link quality, the frequency reuse factor increases. The later linearly contributes to the increase of network spectral efficiency and overcomes the loss resulted from the link quality deterioration.

Fig. 9 illustrates the change of average energy efficiency as the traffic load increases. The vertical axis is normalized with the system bandwidth W and the power consumption P_a of active BSs. The relationship between the average energy efficiency and the network traffic load is highly influenced by the ratio θ of the sleep-mode power to the active-mode power. For small ratios, the energy efficiency would first increase and then decrease as the traffic load increases. There exists

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2672977, IEEE Transactions on Wireless Communications

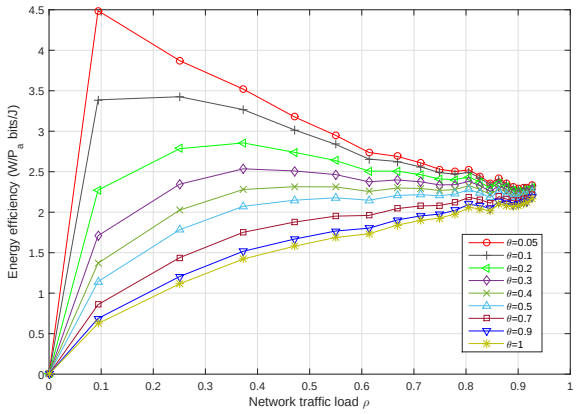

Figure 9. Distribution of average energy efficiency with traffic load (Simulation result), $\alpha = 4$

a load ρ^* that can maximize the energy efficiency. The higher the ratio is, the higher the optimal load ρ^* is. While for large ratios the energy efficiency would increases as the traffic load increases. The energy efficiency is maximized when the network is fully loaded. This is consistent with our analytic result in Section IV-B.

Comparing Fig. 8 and Fig. 9 we can find that when the power ratio θ is large, both the average network spectral and energy efficiency increase in traffic load and they are maximized when the network is fully loaded. However, when θ is small, the trends of the average network spectral and energy efficiency are different. The network spectral efficiency is always increasing in traffic load and is maximized with a full load while the energy efficiency first increases and then decreases in traffic load is maximized with a load smaller than 1. The smaller θ is, the larger the discrepancy between the maximization of the two metrics is.

VI. Conclusion and Future Work

In this paper we have investigated the relationship between network performance and network traffic load for networks with cell DTX. The network is analyzed with theories of stochastic geometry. The SINR distribution as a function of traffic load is derived firstly and a sufficient condition for the networks to be interference-limited is presented. Based on the simplified SINR distribution, analytical expressions are obtained to describe the impact of the network load on the performances, including link spectral efficiency, network spectral and energy efficiency. It is shown that as the network load increases, the average link spectral efficiency decreases while the network spectral efficiency increases. The network energy efficiency is strictly quasi-concave on the network load and the relative power consumption in the sleep mode plays a key role. For small sleep-mode power consumption, the energy efficiency would first increase and then decrease as the network load increases. If the sleep-mode power consumption is larger than a threshold, the energy efficiency would monotonically increase as the network load increases, and the maximum energy efficiency is achieved when the network is fully loaded.

This paper considers only homogeneous networks. The problem would become more complicated in heterogeneous networks where user association, cell-range expansion and cross-tier interference management are considered. The performance analysis of cell DTX in heterogeneous networks will be performed in our future work.

VII. Acknowledgment

The authors would like to thank the China Scholarship Council for its sponsor of this work. The authors would also like to thank the anonymous reviewers for their valuable and constructive comments that contribute to improve the quality of this paper.

Appendix A
Proof of Proposition 1

The proof of this proposition is developed with the same approach as [20]. The difference is to take the impact of network load into consideration. Consider a homogeneous cellular network with load ρ and the density of deployed BSs is λ_B. With cell DTX, the sleeping BSs can immediately switch into active mode when there is a traffic request. Therefore the density of serving BSs for the users is λ_B, regardless the network load ρ. However, the density of the active BSs λ_a is tightly related to the network load by by (4). For a given downlink transmission, the density of interfering BSs is that of active BSs. As a result, the serving BSs is a PPP Φ_B with density λ_B while the interfering BSs is a thinned PPP $\tilde{\Phi}_a$ with density λ_a. The probability density function (PDF) of the distance r between a user and its serving BS is

$$f_r(r) = 2\lambda_B \pi r e^{-\lambda_B \pi r^2}. \tag{29}$$

The CDF of the downlink transmission SINR under network load ρ is $F_\gamma(\gamma, \rho) = P[SINR \le \gamma]$. Conditioning on the distance from the user to the BS r, the CDF $F_\gamma(\gamma, \rho)$ can be expressed as

$$\begin{aligned} F_\gamma(\gamma,\rho) &= \mathbb{E}_r[P[SINR \le \gamma | r]] \\ &= \int_0^\infty P[SINR \le \gamma | r] f_r(r) dr \\ &\overset{(a)}{=} \int_0^\infty P\Big[\frac{P_e r^{-\alpha} h_0}{I_r + \sigma_0^2} \le \gamma | r\Big] 2\lambda_B \pi r e^{-\lambda_B \pi r^2} dr \\ &= \int_0^\infty P\Big[h_0 \le \frac{1}{P_e} \gamma r^\alpha (I_r + \sigma_0^2) | r\Big] 2\lambda_B \pi r e^{-\lambda_B \pi r^2} dr, \end{aligned} \tag{30}$$

where (a) follows by replacing the definition of SINR with (2) and the distribution of r with (29).

Following the assumption that $h \sim exp(1)$ in sub-section II-A, the conditioned CDF of SINR on the distance r for the user to its serving BS can be expressed as

$$\begin{aligned} P[h_0 \le \frac{1}{P_e}\gamma r^\alpha (I_r + \sigma_0^2) | r] &= \mathbb{E}_{I_r}\Big[P\Big[h_0 \le \frac{1}{P_e}\gamma r^\alpha (I_r + \sigma_0^2) | r, I_r\Big]\Big] \\ &= \mathbb{E}_{I_r}\Big[1 - e^{-\frac{1}{P_e}\gamma r^\alpha (I_r + \sigma_0^2)} | r, I_r\Big] \\ &= 1 - e^{-\frac{\sigma_0^2}{P_e}\gamma r^\alpha} \mathcal{L}_{I_r}(\frac{1}{P_e}\gamma r^\alpha), \end{aligned} \tag{31}$$

where $\mathcal{L}_{I_r}(s)$ is the Laplace transform of random variable I_r evaluated as s conditioned on the distance r for the user to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2672977, IEEE Transactions on Wireless Communications

the serving BS. With the definition of Laplace transform, we have

$$\begin{aligned}\mathcal{L}_{I_r}(s) &= E_{I_r}[e^{-sI_r}] \\ &\overset{(a)}{=} \mathbb{E}_{\tilde{\Phi}_a,h_i}\left[exp(-sP_e\sum_{i\in\tilde{\Phi}_a\setminus b_0} h_i r_i^{-\alpha})\right] \\ &= \mathbb{E}_{\tilde{\Phi}_a,\{h_i\}}\left[\prod_{i\in\tilde{\Phi}_a\setminus b_0} exp(-sP_e h_i r_i^{-\alpha})\right] \\ &\overset{(b)}{=} \mathbb{E}_{\tilde{\Phi}_a}\left[\prod_{i\in\tilde{\Phi}_a\setminus b_0} \mathbb{E}_{h_i}[exp(-sP_e h_i r_i^{-\alpha})]\right] \\ &\overset{(c)}{=} \mathbb{E}_{\tilde{\Phi}_a}\left[\prod_{i\in\tilde{\Phi}_a\setminus b_0} \frac{1}{1+sP_e r_i^{-\alpha}}\right] \\ &\overset{(d)}{=} exp(-2\pi\lambda_a\int_r^\infty (1-\frac{1}{1+sP_e x^{-\alpha}})xdx),\end{aligned} \tag{32}$$

where (a) follows from the definition of I_r (3), (b) and (c) are obtained from the assumption that all the interference signals experience i.i.d Rayleigh fading with mean 1 and (d) follows from the probability generating function (PGFL) [19] of the PPP $\tilde{\Phi}_a$ of the interfering BSs with density λ_a. This is unlike the situation that all the BSs are transmitting in [20]. Plugging $s=\frac{1}{P_e}\gamma r^\alpha$ in (32) gives

$$\begin{aligned}\mathcal{L}_{I_r}(\frac{1}{P_e}\gamma r^\alpha) &= exp(-2\pi\lambda_a\int_r^\infty (1-\frac{1}{1+\gamma r^\alpha x^{-\alpha}})xdx) \\ &= exp(-2\pi\lambda_a\int_r^\infty \frac{\gamma}{\gamma+(\frac{x}{r})^\alpha}xdx) \\ &\overset{(a)}{=} exp(-\pi\lambda_a r^2\beta(\gamma,\alpha)),\end{aligned} \tag{33}$$

where

$$\beta(\gamma,\alpha)=\gamma^{2/\alpha}\int_{\gamma^{-2/\alpha}}^\infty \frac{1}{1+y^{\alpha/2}}dy,$$

(a) is derived with a change of variable $x=ry^{1/2}\gamma^{1/\alpha}$. Plugging (33) into (31) gives

$$P[h_0 \le \frac{\gamma r^\alpha(I_r+\sigma_0^2)}{P_e}|r] = 1-e^{-\frac{\sigma_0^2}{P_e}\gamma r^\alpha-\pi\lambda_a r^2\beta(\gamma,\alpha)}. \tag{34}$$

Plugging (34) into(30) gives

$$\begin{aligned}F_\gamma(\gamma,\rho) &= \int_0^\infty 2\pi r\lambda_B e^{-\lambda_B\pi r^2}(1-e^{-\frac{\sigma_0^2}{P_e}\gamma r^\alpha-\pi\lambda_a r^2\beta(\gamma,\alpha))}dr \\ &\overset{(a)}{=} 1-\pi\lambda_B\int_0^\infty e^{-\lambda_B\pi x(1+\rho\beta(\gamma,\alpha))-\frac{\sigma_0^2}{P_e}\gamma x^{\alpha/2}}dx\end{aligned} \tag{35}$$

where (a) is derived by changing r with $r=x^{1/2}$ and changing λ_a with $\lambda_a=\rho\lambda_B$. This completes the proof.

Appendix B
Proof of Theorem 1

Define two functions $f(x)$ and $g(x)$ as

$$f(x)=e^{-\pi\lambda_B x(1+\rho\beta(\gamma,\alpha))}, x\ge 0$$

and

$$g(x)=e^{-\frac{\gamma\sigma_0^2 x^{\alpha/2}}{P_e}}, x\ge 0.$$

Both $f(x)$ and $g(x)$ are continuous functions and they are monotonically decreasing as x increases. And we have $0<f(x)\le 1$ and $0<g(x)\le 1$. According to (12), $P(\gamma)=1-\lambda_B\pi\int_0^\infty f(x)g(x)dx$. By neglecting the contribution of receiver noise, we have $P_s(\gamma)=1-\lambda_B\pi\int_0^\infty f(x)dx=1-\frac{1}{1+\rho\beta(\gamma,\alpha)}$. Thus we have$P(\gamma)-P_s(\gamma)=\lambda_B\pi\int_0^\infty f(x)(1-g(x))dx>0$. For a given small positive number ϵ, there exists $x_0=\frac{-ln(\epsilon/2)}{\lambda_B\pi(1+\rho\beta(\gamma,\alpha))}$ that $f(x_0)=\epsilon/2$. As a result,

$$\begin{aligned}P(\gamma) &= 1-\lambda_B\pi\int_0^\infty f(x)g(x)dx \\ &= 1-\lambda_B\pi\left(\int_0^{x_0} f(x)g(x)dx+\int_{x_0}^\infty f(x)g(x)dx\right) \\ &\le 1-\lambda_B\pi\int_0^{x_0} f(x)g(x)dx \\ &< 1-g(x_0)\lambda_B\pi\int_0^{x_0} f(x)dx \\ &= 1-\frac{1}{1+\rho\beta(\gamma,\alpha)}g(x_0)(1-f(x_0))\end{aligned}$$

$$\begin{aligned}P(\gamma)-P_s(\gamma) &< \frac{1}{1+\rho\beta(\gamma,\alpha)}(1-g(x_0)(1-f(x_0))) \\ &< 1-g(x_0)(1-f(x_0)) \\ &= 1-g(x_0)(1-\epsilon/2)\end{aligned}$$

Since $1-(1-\epsilon/2)^2=1-1+\epsilon-1/4\epsilon^2<\epsilon$, a sufficient condition for $P(\gamma)-P_s(\gamma)<\epsilon$ to be true is

$$g(x_0)\ge 1-\epsilon/2 \tag{36}$$

By solving (36) we can get

$$x_0\le(\frac{-P_e ln(1-\epsilon/2)}{\gamma\sigma_0^2})^{(2/\alpha)} \tag{37}$$

Insert $x_0=\frac{-ln(\epsilon/2)}{\lambda_B\pi(1+\rho\beta(\gamma,\alpha))}$ into (37) and the following condition is derived

$$\lambda_B\ge\frac{-ln(\epsilon/2)}{\pi(1+\rho\beta(\gamma,\alpha))}(\frac{\gamma\sigma_0^2}{-P_e ln(1-\epsilon/2)})^{(2/\alpha)}$$

This completes the proof.

Appendix C
Proof of Lemma 1

For a positive random variable with PDF $f_x(x)$, its expectation is defined as

$$\mathbb{E}[x]=\int_0^\infty x f_x(x)dx.$$

Consider a function $g(y)=1$. We have $x=\int_0^x g(y)dy$. As a result,

$$\begin{aligned}\mathbb{E}[x] &= \int_0^\infty x f_x(x)dx \\ &= \int_0^\infty\int_0^x f_x(x)dydx \overset{(a)}{=} \int_0^\infty\int_y^\infty f_x(x)dxdy \\ &= \int_0^\infty (1-F_x(y))dy \overset{(b)}{=} \int_0^\infty (1-F_x(x))dx,\end{aligned}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2017.2672977, IEEE Transactions on Wireless Communications

where (a) follows by changing the order of integration and (b) is derived by changing y with x. The desired result is derived.

Appendix D
Proof of Theorem 2

In order to simplify the notations, we note $\beta(2^\tau - 1, \alpha)$ as β in this proof.

Taking the derivative of $\mathbb{E}[\eta_{EE}]$ over ρ, we can get

$$\frac{\partial \mathbb{E}[\eta_{EE}]}{\partial \rho} = \int_0^\infty \frac{W(\theta - \rho^2\beta(1-\theta))}{P_a(1+\rho\beta)^2(\theta+\rho(1-\theta))^2} d\tau.$$

This derivative is denoted as $f(\rho)$. Define a function $g(\rho)$ of ρ as

$$g(\rho) = \int_0^\infty \frac{\theta - \rho^2\beta(1-\theta)}{(1+\rho\beta)^2} d\tau.$$

We have

$$f(\rho) = g(\rho)\frac{W}{P_a(\theta+\rho(1-\theta))^2}.$$

As $\frac{W}{P_a(\theta+\rho(1-\theta))^2}$ is strictly positive, the function $f(\rho)$ and $g(\rho)$ have the same sign. The derivative of $g(\rho)$ over ρ is

$$\begin{aligned}\frac{\partial g}{\partial \rho} &= \int_0^\infty \frac{-2\rho\beta(1-\theta)(1+\rho\beta) - 2\beta(\theta - \rho^2\beta(1-\theta))}{(1+\rho\beta)^3} d\tau \\ &= \int_0^\infty \frac{-2(\rho\beta(1-\theta)+\beta\theta)}{(1+\rho\beta)^3} d\tau.\end{aligned}$$

As $\frac{-2(\rho\beta(1-\theta)+\beta\theta)}{(1+\rho\beta)^3}$ is strictly negative, $\frac{\partial g}{\partial \rho}$ is negative. Thus $g(\rho)$ is continuous and monotonically decreasing in ρ. Based on the fact that $\rho \in [0,1]$, we have

$$g_{max} = g(0) = \int_0^\infty \theta d\tau,$$

and

$$g_{min} = g(1) = \int_0^\infty \frac{\theta}{(1+\beta)^2} d\tau - \int_0^\infty \frac{\beta(1-\theta)}{(1+\beta)^2} d\tau.$$

It is clear that g_{max} is positive.

Now we study the sign of g_{min}. If g_{min} is positive, the following inequality must be true.

$$\theta > \frac{\Omega(\alpha)}{1+\Omega(\alpha)}, \tag{38}$$

where Ω is defined as (26). In this case $g(\rho)$ is positive on $[0,1]$. So is $f(\rho)$. Thus the ergodic average energy efficiency is increasing in traffic load ρ. If the inequality (38) is false, which means $\theta \leq 1 - \frac{1}{1+\Omega(\alpha)}$, we have $g_{min} \leq 0$. As $g(\rho)$ is continuous and decreasing in ρ, there exists one unique ρ^* that makes $g(\rho)$ to be zero. This ρ^* also makes $f(\rho)$ to be zero. For $\rho \in [0, \rho^*)$, $f(\rho) > 0$ and for $\rho \in (\rho^*, 1]$, $f(\rho) < 0$. As a result, the average energy efficiency is increasing on $[0, \rho^*)$ while decreasing on $[\rho^*, 1]$. According to the definition of quasi-concave function, it can be easily derived that the average energy efficiency is a quasi-concave function of the traffic load. This completes the proof.

Appendix E
Proof of Proposition 5

According to Theorem 2, the optimal traffic load ρ^* satisfies the equation (25) if $\theta < \frac{\Omega(\alpha)}{1+\Omega(\alpha)}$. Therefore we have

$$\theta = \frac{\int_0^\infty \frac{\rho^{*2}\beta}{(1+\rho^*\beta)^2} d\tau}{\int_0^\infty \frac{1+\rho^{*2}\beta}{(1+\rho^*\beta)^2} d\tau}.$$

Define $A = \int_0^\infty \frac{\rho^{*2}\beta}{(1+\rho^*\beta)^2} d\tau$, $B = \int_0^\infty \frac{1+\rho^{*2}\beta}{(1+\rho^*\beta)^2} d\tau$ and $C = \int_0^\infty \frac{1}{(1+\rho^*\beta)^2} d\tau$. Thus we have $\theta = \frac{A}{B}$. The derivatives of A and B over ρ^* are

$$A' = \frac{\partial A}{\partial \rho^*} = \int_0^\infty \frac{2\rho^*\beta(1+\rho^*\beta)^2 - 2\rho^{*2}\beta^2(1+\rho^*\beta)}{(1+\rho^*\beta)^4} d\tau,$$

$$B' = \frac{\partial B}{\partial \rho^*} = \int_0^\infty \frac{2\rho^*\beta(1+\rho^*\beta)^2 - 2(1+\rho^{*2}\beta)\beta(1+\rho^*\beta)}{(1+\rho^*\beta)^4} d\tau.$$

Consequently, the derivative of θ over ρ^* can be derived as

$$\begin{aligned}\frac{\partial \theta}{\partial \rho^*} &= \frac{BA' - AB'}{B^2} \\ &= \frac{CA' + AA' - AB'}{B^2} \\ &= \frac{CA' + A\int_0^\infty \frac{2\beta}{(1+\rho^*\beta)^3} d\tau}{B^2} > 0.\end{aligned}$$

Thus θ is a monotonically increasing function of ρ^*. According to [33], its inverse function $\rho^*(\theta)$ exist. And the derivative of $\rho^*(\theta)$ over θ is always positive. As a result, ρ^* is monotonically increasing in θ. This completes the proof.

References

[1] Cisco, "Global mobile data traffic forecast update, 2014-2019," *Cisco Visual Networking Index (VNI) Forecast*, 2015.

[2] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. Sukhavasi, C. Patel, and S. Geirhofer, "Network densification: the dominant theme for wireless evolution into 5g," *Communications Magazine, IEEE*, vol. 52, no. 2, pp. 82–89, February 2014.

[3] C. Han, T. Harrold, S. Armour, I. Krikidis, S. Videv, P. M. Grant, H. Haas, J. Thompson, I. Ku, C.-X. Wang, T. A. Le, M. Nakhai, J. Zhang, and L. Hanzo, "Green radio: radio techniques to enable energy-efficient wireless networks," *Communications Magazine, IEEE*, vol. 49, no. 6, pp. 46–54, June 2011.

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

[5] E. Oh, B. Krishnamachari, X. Liu, and Z. Niu, "Toward dynamic energy-efficient operation of cellular network infrastructure," *Communications Magazine, IEEE*, vol. 49, no. 6, pp. 56–61, 2011.

[6] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume *et al.*, "How much energy is needed to run a wireless network?" *Wireless Communications, IEEE*, vol. 18, no. 5, pp. 40–49, 2011.

[7] J.-F. Cheng, H. Koorapaty, P. Frenger, D. Larsson, and S. Falahati, "Energy efficiency performance of lte dynamic base station downlink dtx operation," in *Vehicular Technology Conference (VTC Spring), 2014 IEEE 79th*, May 2014, pp. 1–5.

[8] P. Frenger, P. Moberg, J. Malmodin, Y. Jading, and I. Godor, "Reducing energy consumption in lte with cell dtx," in *Vehicular Technology Conference (VTC Spring), 2011 IEEE 73rd*, May 2011, pp. 1–5.

[9] L. M. Correia, D. Zeller, O. Blume, D. Ferling, Y. Jading, I. Gódor, G. Auer, and L. Van Der Perre, "Challenges and enabling technologies for energy aware mobile radio networks," *Communications Magazine, IEEE*, vol. 48, no. 11, pp. 66–72, 2010.

[10] M. A. Marsan, L. Chiaraviglio, D. Ciullo, and M. Meo, "On the effectiveness of single and multiple base station sleep modes in cellular networks," *Computer Networks*, vol. 57, no. 17, pp. 3276–3290, 2013.

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

[12] Y. S. Soh, T. Q. Quek, M. Kountouris, and H. Shin, "Energy efficient heterogeneous cellular networks," *Selected Areas in Communications, IEEE Journal on*, vol. 31, no. 5, pp. 840–850, 2013.

[13] J. Scourias, *Overview of GSM: The global system for mobile communications*. University of Waterloo, Computer Science Department, 1996.

[14] S. Falahati, J.-F. Cheng, H. Koorapaty, and D. Larsson, "Base station downlink dtx designs for interference mitigation in high-performance lte networks," in *Vehicular Technology Conference (VTC Spring), 2014 IEEE 79th*, May 2014, pp. 1–5.

[15] S. Tombaz, S. wook Han, K. W. Sung, and J. Zander, "Energy efficient network deployment with cell dtx," *Communications Letters, IEEE*, vol. 18, no. 6, pp. 977–980, June 2014.

[16] A. Wyner, "Shannon-theoretic approach to a gaussian cellular multiple-access channel," *Information Theory, IEEE Transactions on*, vol. 40, no. 6, pp. 1713–1727, Nov 1994.

[17] G. Miao, J. Zander, K. W. Sung, and S. B. Slimane, *Fundamentals of Mobile Data Networks*. Cambridge University Press, 2016.

[18] J. Xu, J. Zhang, and J. Andrews, "On the accuracy of the wyner model in cellular networks," *Wireless Communications, IEEE Transactions on*, vol. 10, no. 9, pp. 3098–3109, September 2011.

[19] S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic geometry and its applications*. John Wiley & Sons, 2013.

[20] J. Andrews, F. Baccelli, and R. Ganti, "A tractable approach to coverage and rate in cellular networks," *Communications, IEEE Transactions on*, vol. 59, no. 11, pp. 3122–3134, November 2011.

[21] H. Dhillon, R. Ganti, F. Baccelli, and J. Andrews, "Modeling and analysis of k-tier downlink heterogeneous cellular networks," *Selected Areas in Communications, IEEE Journal on*, vol. 30, no. 3, pp. 550–560, April 2012.

[22] D. Cao, S. Zhou, and Z. Niu, "Optimal combination of base station densities for energy-efficient two-tier heterogeneous cellular networks," *Wireless Communications, IEEE Transactions on*, vol. 12, no. 9, pp. 4350–4362, 2013.

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

[24] I. Siomina and D. Yuan, "Analysis of cell load coupling for lte network planning and optimization," *Wireless Communications, IEEE Transactions on*, vol. 11, no. 6, pp. 2287–2297, 2012.

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

[26] B. Blaszczyszyn, M. Jovanovicy, and M. K. Karray, "How user throughput depends on the traffic demand in large cellular networks," in *Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), 2014 12th International Symposium on*. IEEE, 2014, pp. 611–619.

[27] G. Miao and G. Song, *Energy and Spectrum Efficient Wireless Network Design*. Cambridge University Press, 2014.

[28] G. Miao, N. Himayat, G. Li, and S. Talwar, "Distributed interference-aware energy-efficient power optimization," *Wireless Communications, IEEE Transactions on*, vol. 10, no. 4, pp. 1323–1333, April 2011.

[29] G. Miao, N. Himayat, G. Y. Li, and S. Talwar, "Low-complexity energy-efficient scheduling for uplink ofdma," *Communications, IEEE Transactions on*, vol. 60, no. 1, pp. 112–120, 2012.

[30] 3GPP, "Further advancements for e-utra physical layer aspects," *TR36.814 (V9.0.0)*, March, 2010.

[31] G. Miao, N. Himayat, and G. Y. Li, "Energy-efficient link adaptation in frequency-selective channels," *Communications, IEEE Transactions on*, vol. 58, no. 2, pp. 545–554, 2010.

[32] Y. S. Soh, T. Quek, M. Kountouris, and H. Shin, "Energy efficient heterogeneous cellular networks," *Selected Areas in Communications, IEEE Journal on*, vol. 31, no. 5, pp. 840–850, May 2013.

[33] R. A. Adams and C. Essex, *Calculus: a complete course*. Addison-Wesley, 1999, vol. 4.