Article:
Nie, W., Zhong, Y., Zheng, F-C. et al. (2016) HetNets with Random DTX Scheme: Local Delay and Energy Efficiency. IEEE Transactions on Vehicular Technology, 65 (8). pp. 6601-6613. ISSN: 0018-9545

https://doi.org/10.1109/TVT.2015.2477374

HetNets with Random DTX Scheme: Local Delay and Energy Efficiency

Weili Nie*, Yi Zhong†, Fu-Chun Zheng‡, Wenyi Zhang†, Timothy O'Farrell§

Abstract

Heterogeneous cellular networks (HetNets) are to be deployed for future wireless communication to meet the ever-increasing mobile traffic demand. However, the dense and random deployment of small cells and their uncoordinated operation raise important concerns about energy efficiency. On the other hand, discontinuous transmission (DTX) mode at the base station (BS) serves as an effective technology to improve energy efficiency of overall system. In this paper, we investigate the energy efficiency under finite local delay constraint in the downlink HetNets with the random DTX scheme. Using a stochastic geometry based model, we derive the local delay and energy efficiency in a general case and obtain closed-form expressions in some special cases. These results give insights into the effect of key system parameters, such as path loss exponents, BS densities, SIR threshold and mute probability on the system performance. We also provide the low-rate and high-rate asymptotic behavior of the maximum energy efficiency. It is analytically shown that it is less energy-efficient to apply random DTX scheme in the low-rate regime. In the high-rate regime, however, random DTX scheme is essential to achieve the finite local delay and higher energy efficiency. Finally, we extend the analysis to the load-aware DTX scheme where the mute probability depends on the user activity level.

Index Terms

Energy efficiency, HetNets, local delay, random DTX scheme, stochastic geometry.

*W. Nie is with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210093, China (e-mail: nieweili@seu.edu.cn).

†Y. Zhong and W. Zhang are with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230027, China (e-mail: geners@mail.ustc.edu.cn, wenyizha@ustc.edu.cn).

‡F.-C. Zheng is with the School of Systems Engineering, University of Reading, Reading, UK. (e-mail: f.zheng@reading.ac.uk).

§T. O'Farrell is with the Department of Electronic and Electrical Engineering, University of Sheffield, Sheffiled, UK. (e-mail: T.OFarrell@sheffield.ac.uk).

The research has been supported by the National Basic Research Program of China (973 Program) under grant 2012CB316004.

I. INTRODUCTION

Due to the explosive growth of mobile data traffic, an increasing portion of the mobile data and voice traffic is expected to be offloaded from the macrocell network onto other low power and low cost small cell networks [1], resulting in heterogeneous cellular networks (HetNets) [2]. HetNets comprise a conventional cellular network overlaid with a diverse set of lower-power base stations (BSs) or access points (APs), such as picocells, femtocells, WiFi APs and perhaps relays. Heterogeneity is expected to be a key feature of future cellular networks, and an essential means for providing higher end-user throughput as well as expanding its indoor and cell edge coverage. Nevertheless, the deployment of a large number of small cells overlaying the macrocells is not without new technical challenges [3].

First, designing green cellular networks has recently received great attention amongst network operators, regulatory bodies such as 3GPP and ITU and green communications research projects such as EARTH and GreenTouch [4]-[8]. The dense and random deployment of small cells and their uncoordinated operation raise important concerns about energy efficiency in the HetNets [9]. One of the major challenges is the incursion of inter-tier interference due to the aggressive spatial reuse. Discontinuous transmission (DTX) mode at BSs serves as a potential technology of managing interference and improving energy efficiency [5], [10], [11]. The main principle of DTX is to shut down some BS components in time periods without signal transmission to reduce interference and energy wastage of the system.

Delay is another critical performance indicator that reflects the quality-of-service (QoS) provided by a network and is directly related to the system reliability [12]. Generally, there are two kinds of delay in the wireless networks: the transmission delay, which refers to the time spent in transmitting data successfully, and the queueing delay, which mainly refers to the waiting time in one or more service queues. The local delay considered in this paper is a basic form of the transmission delay, thus our analysis of the local delay provides a lower bound for the overall system delay. Particularly, the local delay may be infinite for certain network parameters, which is called network contention phase transition [13]. If it is infinite, the network cannot provide any useful service to its users. Hence, it is crucial to take the local delay into consideration when evaluating the system performance.

A. Related Work and Motivations

Recently, a new general model for wireless node distribution based on stochastic geometry has been proposed [14]-[16] and the authors have shown a tractable and reasonably accurate solution for analyzing important metrics such as signal-to-interference-plus-noise ratio (SINR) coverage, average rate and rate coverage. In addition, the authors in [15] and [16] studied the stochastic geometry modeling and analysis in HetNets, but the delay which serves as an important system metric to evaluate the QoS performance has not been discussed yet.

The framework for analyzing the local delay was provided in [13], [17] where some concrete expressions and insights for the local delay in different scenarios were obtained. The work in [18] extended the results to the case of finite mobility. [19] proposed a new model to evaluate the local delay by using joint interference statistics, which characterized different degrees of temporal dependence. In [20], the optimal power control policies for different fading statistics were proposed to minimize the local delay. In [21], two MAC protocols FHMA and ALOHA are proposed to reduce the interference correlation and the local delay. All the above works are based on the homogeneous Poisson networks. In this work, we extend the framework to the heterogeneous case. Different from the homogeneous case in [17]-[21] where each receiver is at a fixed distance to the corresponding transmitter, we incorporate the cell association criterion, which is one of the most important characteristics in HetNets, into the analytical framework. Furthermore, another challenge is to analyze the impact of different system parameters from different tiers and to provide useful insights into network deployment and design.

Many works have evaluated the performance of DTX scheme in cellular networks. For instance, [22] and [23] showed that cell DTX is of key importance to achieve significant energy reduction in an LTE network. [24] analyzed the maximum achievable energy saving of the cell DTX by incorporating the cell DTX with a clean-slate network deployment. Nevertheless, no quantitative analysis of DTX scheme by using tools of stochastic geometry is performed in these papers. Similar work has analyzed random sleeping scheme in HetNets based on stochastic geometric model [25]. Different from the random sleeping scheme in [25], random DTX scheme enables sleep mode operations at BS side with a short enough time resolution, e.g. on millisecond level [10]. Besides, note that the impact of local delay on energy efficiency has not been discussed in previous work.

Since delay-energy tradeoff is one of four fundamental tradeoffs [26], it has gained extensive research in the literature. Nevertheless, it should be noted that most existing results are based on the queueing model and only consider a single-cell case, such as [27], [28]. In this paper, we discuss the energy efficiency under the finite local delay constraint and deal with the inter-cell interference in the HetNets by modeling the locations of BSs, both macrocells and small cells, as independent Poisson point processes (PPPs) with different intensities. In addition, using the tools of stochastic geometry, we investigate the question of whether the random DTX scheme applied in the HetNets is always beneficial in terms of local delay and energy efficiency.

B. Contributions and Organizations

Characterization of local delay and energy efficiency: We analytically derive the local delay of a typical user in a K-tier HetNet with the random DTX scheme. Based on the stochastic geometry framework, the expression is reasonably tractable and enjoys a high degree of generality. In addition, energy efficiency can be obtained from dividing the network throughput by the average area power consumption. These general results can be significantly simplified and even have closed forms in some special cases. Furthermore, we also give the expressions of local delay and energy efficiency in the K-tier HetNet with the load-aware DTX scheme.

Effect of key system parameters: Building upon the analytical results, we can obtain some insights into the effect of key system parameters on local delay and energy efficiency. For instance, local delay increases with SIR threshold while there exists an optimal SIR threshold to maximize energy efficiency. As the path loss exponent increases, local delay decreases and energy efficiency increases. The effect of BS densities heavily depends on power parameters and path loss exponents of different tiers. Moreover, the optimal mute probabilities in random DTX scheme that maximize local delay and energy efficiency, respectively, are generally different.

Asymptotic behavior of maximum energy efficiency: We consider two asymptotic regimes: the low-rate regime and the high-rate regime, and then derive the asymptotic expressions of the maximum energy efficiency and the optimal mute probability in random DTX scheme. Numerical results show that these asymptotic results are reasonably tight in their respective regimes. More importantly, the asymptotic behavior reveals that it is less energy-efficient to apply random DTX scheme in the low-rate regime while random DTX scheme is essential to achieve the finite local delay and higher energy efficiency in the high-rate regime.

In the remainder of this paper, Section II presents the system model. The local delay expressions in a general case and some special cases are derived in Section III. In Section IV, we consider energy efficiency and derive energy efficiency asymptotics in low-rate and high-rate regime. In Section V, the analysis is extended to the load-aware DTX scheme. Numerical results and discussions are provided in Section VI. Finally, Section VII concludes the paper.

II. System Model

A. Heterogeneous Network Model

We consider a HetNet composed by K independent network tiers. For notational ease, we denote $\mathcal{K} = \{1, 2, \cdots, K\}$. BSs across tiers differ in terms of deploying density λ_k, transmit power p_k, and path loss exponent α_k ($\alpha_k > 2$)[1]. The BS locations of each tier are modeled by an independent homogeneous PPP Φ_k. Note that the network is assumed to be static, i.e., all the BSs and users are generated at first and remain in the same position all the time. Denote by $\Phi = \bigcup_{k\in\mathcal{K}} \Phi_k$ the set of the overall BSs in the K-tier HetNet. BSs and users are assumed to be equipped with a single antenna.

Without loss of generality, we focus on the downlink analysis at a typical user in the origin $o \in \mathbb{R}^2$. For the propagation model, We consider a general power-law path loss model in which the signal transmitted by a BS located at x in the k^{th} tier undergoes a distance dependent path loss $l(x) = \|x\|^{-\alpha_k}$, where $\|x\|$ denotes the distance between x and the origin o. We assume that the power fading coefficients are spatially and temporally independent with exponential distribution of unit mean (i.e., Rayleigh fading assumption), and let $h_{t,x}$ be the power fading coefficient between BS located at x and the typical user in time slot t. Furthermore, since the BS density is typically high in the HetNets, we ignore the noise and focus on the interference-limited regime.

We consider a cell association where each user is associated with the strongest BS in terms of average received signal strength (RSS) at the user. As a result, the tier-index of the serving BS is given by

$$k = \underset{j\in\mathcal{K}}{\operatorname{argmax}}\ p_j \|x_{j,0}\|^{-\alpha_j}, \tag{1}$$

[1]Each BS in the k^{th} tier uses the same transmit power $\{p_k\}_{k\in\mathcal{K}}$.

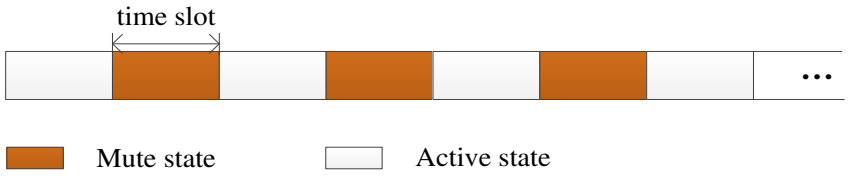

Fig. 1. A realization of the random DTX scheme at the BS. Each BS only transmits signal in the active state and shuts down some components when there is no signal transmission in the mute state.

where the fading coefficient is averaged out and $x_{j,0}$ denotes the location of the nearest BS in the j^{th} tier. Note that the following analytical results are also applicable to the biased association case [16] with a more complicated representation.

B. Random DTX Scheme

We assume that time is divided into discrete slots with equal duration and each transmission attempt occupies one time slot. Fig. 1 illustrates a realization of the random DTX scheme at the BS, where each BS has two transmission states in each time slot, i.e., the mute state and the active state.

In this work, we model the DTX mode in each tier and each time slot as an independent Bernoulli trial with a tunable parameter ζ_k $(0 \leq \zeta_k < 1)$ called mute probability[2]. Namely, a BS from the k^{th} tier in one time slot stay mute with probability ζ_k and stay active with probability $1-\zeta_k$, independently of the BS location and time slot (i.e., memoryless both spatially and temporally). Let $\Phi_{k,t}$ be the set of active BSs in the k^{th} tier in the time slot t. The interference at the typical user in the time slot t is

$$I_t = \sum_{j\in\mathcal{K}} \sum_{x_{j,i}\in\Phi_j\setminus\left\{x_{k,0}\right\}} p_j h_{t,x_{j,i}} \|x_{j,i}\|^{-\alpha_j} \mathcal{I}\left(x_{j,i}\in\Phi_{j,t}\right), \tag{2}$$

where $\mathcal{I}(\cdot)$ is the indicator function, and $I_{t,x_{j,i}} \triangleq p_j h_{t,x_{j,i}} \|x_{j,i}\|^{-\alpha_j} \mathcal{I}\left(x_{j,i}\in\Phi_{j,t}\right)$ denotes the interference from BS located at $x_{j,i}$ in time slot t. The signal-to-interference ratio (SIR) of the

[2]The mute probability in the DTX does not only depend on the traffic pattern, but also the transmission pattern, especially in LTE type of systems. For example, in an LTE system, it would be possible to use (or trade in) more subcarriers to obtain more idle time slots (during which DTX can be employed, increasing the mute probability). When the traffic is certain, by pseudo-randomizing the transmission of resource blocks, we could adjust the value of mute probability to some extent, which provides a practical possibility for the random DTX scheme.

typical user from its associated BS in the k^{th} tier and time slot t is

$$\text{SIR}_{k,t} = \frac{p_k h_{t,x_{k,0}} r_k^{-\alpha_k}}{I_t}, \tag{3}$$

where r_k denotes the distance between the associated BS in the k^{th} tier and the typical user.

C. Performance Metrics

Throughout the paper, we will focus on two performance metrics: local delay and energy efficiency.

1) Local Delay: We assume that if a transmission fails in a certain time slot, a retransmission will be conducted. The local delay is defined as the mean time (in number of time slots) until a packet is successfully received over a link between a transmitter and its receiver[3] [13].

Denote $\mathcal{C}_{k|\Phi}$ as the success event conditioned on the distance r_k and the point process Φ. Success events in different time slots are independent, so there is no need to add a time index to this event. By setting the SIR threshold to be θ, we denote the probability of a successful transmission conditioned on r_k and Φ as $\Pr\left(C_{k|\Phi}\right) = (1-\zeta_k)\Pr\left(\text{SIR}_{k,t} > \theta \,|r_k, \Phi\right)$. where the coefficient $1-\zeta_k$ exists because the associated BS stays active in any given time slot with probability $1-\zeta_k$. That is, the successful transmission occurs only when the associated BS stays active and the instantaneous SIR is greater than the SIR threshold. Accordingly, the number of time slots needed until success given r_k and Φ, denote by $\Delta_{k|\Phi}$, is a geometrically distributed random variable with mean $\Pr\left(\mathcal{C}_{k|\Phi}\right)^{-1}$. The expectation with respect to r_k and Φ yields the local delay when a typical user is associated with the k^{th} tier:

$$D_k = \mathbb{E}_{r_k,\Phi}\left(\mathbb{E}\left(\Delta_{k|\Phi}\right)\right) = \mathbb{E}_{r_k,\Phi}\left[\frac{1}{\Pr\left(\mathcal{C}_{k|\Phi}\right)}\right]. \tag{4}$$

Since the typical user is associated with at most one tier, from the law of total probability, the local delay is

$$D = \sum\nolimits_{k\in\mathcal{K}} \mathcal{A}_k D_k, \tag{5}$$

[3]In this paper, we do not specify the duration of one time slot on the condition that each transmission attempt only occupies one time slot.

where $\mathcal{A}_k$ is the per-tier association probability. Note that we assume BSs are fully-loaded so that whenever a BS is active it always has data to transmit. The local delay is thus the transmission delay, but not the queuing delay.

2) Energy Efficiency: We define the network energy efficiency as a ratio of network throughput to average area power consumption [25].

First, We assume that a fixed-rate transmission scheme is applied, e.g. in each time slot, as long as the SIR is above a threshold θ, each data packet can be successfully transmitted at the (normalized) rate $\log(1+\theta)$ nats per sec per Hz[4]. As a result, the network throughput is defined as the average number of successfully transmitted nats per sec per Hz per unit area, which is

$$\tau = D^{-1}\log(1+\theta)\sum_{k=1}^{K}(1-\zeta_k)\lambda_k, \tag{6}$$

where $D^{-1}\log(1+\theta)$ denotes the average number of successfully transmitted nats per sec per Hz over a single link and $\sum_{k=1}^{K}(1-\zeta_k)\lambda_k$ is the density of active BSs in the K-tier HetNet. Note that the network throughput can also be regarded as a measure of the area spectral efficiency.

Second, we adopt a linear BS power consumption model [29], namely, the power consumption per BS in the k^{th} tier is given by

$$P_{k,in} = \begin{cases} P_{k0} + \Delta_k p_k, & \text{if } p_k > 0, \\ P_{k,S}, & \text{if } p_k = 0. \end{cases} \tag{7}$$

where P_{k0} and $P_{k,S}$ denote the static power expenditure in the active mode and mute mode in the k^{th} tier respectively, and Δ_k denotes the slope of power consumption in the k^{th} tier. Accordingly, the average area power consumption is given by $P_a = \sum_{k=1}^{K}\lambda_k\left[(1-\zeta_k)(P_{k0}+\Delta_k p_k)+\zeta_k P_{k,S}\right]$. Therefore, the network energy efficiency is calculated as

$$\eta_{EE} = \frac{\tau}{P_a} = \frac{D^{-1}\log(1+\theta)\sum_{k=1}^{K}(1-\zeta_k)\lambda_k}{\sum_{k=1}^{K}\lambda_k\left[(1-\zeta_k)(P_{k0}+\Delta_k p_k)+\zeta_k P_{k,S}\right]}, \tag{8}$$

where the unit is nats/Joule/Hz.

[4]The physical meaning of the fixed-rate transmission is that the transmitter may not know the instantaneous SIR information and thus adopts a fixed modulation/coding strategy, such as QPSK and M-QAM, which leads to a fixed transmission rate in each time slot (or each resource block).

III. Local Delay

In this section, we first give a general result of the local delay in the HetNets and further consider some special cases with simpler expressions and important insights.

A. General Case and Main Result

We now provide the general result of the local delay, consisting of tier-specific BS density λ_k, transmit power p_k, path loss exponent α_k and mute probability ζ_k.

Theorem 1. *In a K-tier HetNet with Random DTX scheme and cell association based on the long-term RSS, the local delay is given by*

$$D=\sum_{k=1}^{K}\frac{2\pi\lambda_k}{1-\zeta_k}\int_0^{\infty} r\exp\left[-\sum_{j=1}^{K}\pi\lambda_j\left(\frac{p_j}{p_k}\right)^{\frac{2}{\alpha_j}} r^{\frac{2\alpha_k}{\alpha_j}}\left(1-(1-\zeta_j)\,\mathcal{Z}\left(\zeta_j,\alpha_j,\theta\right)\right)\right]dr, \tag{9}$$

where

$$\mathcal{Z}\left(\zeta_j,\alpha_j,\theta\right)=\int_1^{\infty}\frac{\theta}{u^{\frac{\alpha_j}{2}}+\theta\zeta_j}du. \tag{10}$$

Proof: According to the definition of local delay, the probability of successful transmission conditioned on r_k and Φ can be evaluated as

$$\begin{aligned}
&\Pr\left(\mathcal{C}_{k|\Phi}\right)\overset{(a)}{=}(1-\zeta_k)\Pr\left(\mathrm{SIR}_{k,t}>\theta\,|r_k,\Phi\right)\\
&\overset{(b)}{=}(1-\zeta_k)\Pr\left(p_k h_{t,x_{k,0}} r_k^{-\alpha_k}>\theta I_t\,|r_k,\Phi\right)\\
&\overset{(c)}{=}(1-\zeta_k)\,\mathbb{E}_{I_t}\left[\exp\left(-\frac{\theta r_k^{\alpha_k} I_t}{p_k}\right)\middle| r_k,\Phi\right]\\
&\overset{(d)}{=}(1-\zeta_k)\,\mathbb{E}_{\left\{I_{t,x_{j,i}}\right\}}\left[\prod_{j\in\mathcal{K}}\prod_{x_{j,i}\in\Phi_j\setminus\left\{x_{k,0}\right\}}\exp\left(-\frac{\theta r_k^{\alpha_k} I_{t,x_{j,i}}}{p_k}\right)\middle| r_k,\Phi\right]\\
&\overset{(e)}{=}(1-\zeta_k)\prod_{j\in\mathcal{K}}\prod_{x_{j,i}\in\Phi_j\setminus\left\{x_{k,0}\right\}}\mathcal{L}_{I_{t,x_{j,i}}}\left(-\frac{\theta r_k^{\alpha_k}}{p_k}\middle| r_k,\Phi_j\right),
\end{aligned} \tag{11}$$

where (a) is from the definition of $\Pr\left(\mathcal{C}_{k|\Phi}\right)$, (b) follows from the SIR expression in (3), (c) follows from the fact that the power fading coefficient $h_{t,x_{k,0}}$ is exponentially distributed with unit mean, (d) follows from the interference expression in (2), that is, $I_t=\sum_{j\in\mathcal{K}}\sum_{x_{j,i}\in\Phi_j\setminus\left\{x_{k,0}\right\}} I_{t,x_{j,i}}$, (e) follows from the independence property of different interfering links, and $\mathcal{L}_{I_{t,x_{j,i}}}\left(s|\,\Phi_j\right)$

denotes the Laplace transformation of $I_{t,x_{j,i}}$ conditioned on Φ_j. From (2), using the definition of Laplace transformation yields,

$$\begin{aligned}
&\mathcal{L}_{I_{t,x_{j,i}}}\left(s|\,\Phi_j\right)=\mathbb{E}_{I_{t,x_{j,i}}}\left[\exp\left(-sI_{t,x_{j,i}}\right)\big|\,\Phi_j\right]\\
&=\mathbb{E}_{h_{t,x_{j,i}}}\left[\exp\left(-sp_jh_{t,x_{j,i}}\|x_{j,i}\|^{-\alpha_j}\mathcal{I}\left(x_{j,i}\in\Phi_{j,t}\right)\right)\big|\,\Phi_j\right]\\
&=\mathbb{E}_{h_{t,x_{j,i}}}\left[(1-\zeta_j)\exp\left(-sp_jh_{t,x_{j,i}}\|x_{j,i}\|^{-\alpha_j}\right)+\zeta_j\right]\\
&\overset{(a)}{=}\frac{1-\zeta_j}{1+sp_j\|x_{j,i}\|^{-\alpha_j}}+\zeta_j,
\end{aligned}\tag{12}$$

where (a) follows from the fact that $h_{t,x_{j,i}}$ is exponentially distributed with unit mean. From (4), the conditional local delay given r_k can be evaluated as

$$\begin{aligned}
&D_k\left(r_k\right)=\mathbb{E}_{\Phi}\left[\frac{1}{\Pr\left(\mathcal{C}_{k|\Phi}\right)}\right]\\
&\overset{(a)}{=}\frac{1}{1-\zeta_k}\prod_{j\in\mathcal{K}}\mathbb{E}_{\Phi_j}\left[\prod_{x_{j,i}\in\Phi_j\setminus\left\{x_{k,0}\right\}}\frac{1}{\mathcal{L}_{I_{x_{j,i}}}\left(-\frac{\theta r_k^{\alpha_k}}{p_k}\middle|\,r_k,\Phi_j\right)}\right],
\end{aligned}\tag{13}$$

where (a) follows from the independence property of BS point processes among tiers. By substituting (12) into (13), and applying the probability generating functional (PGFL) of the PPP [30], we obtain

$$D_k\left(r_k\right)=\frac{1}{1-\zeta_k}\prod_{j\in\mathcal{K}}\exp\left[2\pi\lambda_j\left(1-\zeta_j\right)\int_{b_j}^{\infty}\frac{r}{\frac{p_k}{\theta r_k^{\alpha_k}p_j}r^{\alpha_j}+\zeta_j}dr\right],\tag{14}$$

where the integration limits are from b_j to ∞ since the closest interference from the j^{th} tier is at least at a distance $b_j=\left(p_jp_k^{-1}\right)^{1/\alpha_j}r_k^{\alpha_k/\alpha_j}$. Applying a change of variable $u=\left(r_k^{\alpha_k}p_jp_k^{-1}\right)^{-\frac{2}{\alpha_j}}r^2$ yields,

$$D_k\left(r_k\right)=\frac{1}{1-\zeta_k}\exp\left[\sum_{j=1}^{K}\pi\lambda_j\left(1-\zeta_j\right)\left(\frac{p_j}{p_k}\right)^{\frac{2}{\alpha_j}}r_k^{\frac{2\alpha_k}{\alpha_j}}\mathcal{Z}\left(\zeta_j,\alpha_j,\theta\right)\right].\tag{15}$$

According to [16, Lemma 3], the pdf of the distance r_k is given by

$$f_{r_k}\left(r\right)=\frac{2\pi\lambda_k}{\mathcal{A}_k}r\exp\left[-\pi\sum_{j=1}^{K}\lambda_j\left(\frac{p_j}{p_k}\right)^{\frac{2}{\alpha_j}}r^{\frac{2\alpha_k}{\alpha_j}}\right].\tag{16}$$

Thus, deconditioning on r_k yields,

$$D_k = \frac{2\pi\lambda_k}{(1-\zeta_k)\,\mathcal{A}_k}\int_0^\infty r\exp\left[-\sum_{j=1}^{K}\pi\lambda_j\left(\frac{p_j}{p_k}\right)^{\frac{2}{\alpha_j}} r^{\frac{2\alpha_k}{\alpha_j}}\left(1-(1-\zeta_j)\,\mathcal{Z}\left(\zeta_j,\alpha_j,\theta\right)\right)\right]dr, \quad (17)$$

Finally, substituting (17) into (5) yields the desired result in (9). ■

Although Theorem 1 does not give a closed-form expression, the integral is fairly easy to evaluate. For some special cases where significant simplification is possible, key insights from these simple local delay expressions can be obtained.

B. Special Case and Insights

First, we assume $\{\alpha_k\} = \alpha$, then Theorem 1 can be simplified in the following corollary.

Corollary 1. *When $\{\alpha_k\} = \alpha$, i.e., the path loss exponents of all the tiers are identical, the local delay is given by*

$$D = \frac{\sum_{k=1}^{K}\frac{\lambda_k p_k^{\frac{2}{\alpha}}}{1-\zeta_k}}{\sum_{j=1}^{K}\lambda_j p_j^{\frac{2}{\alpha}}\left[1-(1-\zeta_j)\,\mathcal{Z}\left(\zeta_j,\alpha,\theta\right)\right]}. \quad (18)$$

Proof: By letting $\{\alpha_k\} = \alpha$ in (9), we obtain

$$D = \sum_{k=1}^{K}\frac{2\pi\lambda_k}{1-\zeta_k}\int_0^\infty r\exp\left[-r^2\sum_{j=1}^{K}\pi\lambda_j\left(\frac{p_j}{p_k}\right)^{\frac{2}{\alpha}}\left(1-(1-\zeta_j)\,\mathcal{Z}\left(\zeta_j,\alpha,\theta\right)\right)\cdot\right]dr. \quad (19)$$

Since $\int_0^\infty 2re^{-Ar^2}dr = \frac{1}{A}$, we can easily get the desired result in (18). ■

Since from (10) we observe that $\mathcal{Z}\left(\zeta_j,\alpha,\theta\right)$ is a monotonically increasing function of θ, this expression reveals that the local delay D increases as the SIR threshold θ increases. This coincides with our intuition that when SIR threshold increases, the probability that a transmission fails in a certain time slot increases, hence the number of retransmissions needed increases. Besides, since $\mathcal{Z}\left(\zeta_j,\alpha,\theta\right) \to \infty$ as $\theta \to \infty$, it is observed that when θ exceeds some threshold value, the local delay will be infinite. As a result, we could define the critical SIR threshold as

$$\theta_c \triangleq \sup\left\{\theta : D\left(\theta\right) < \infty\right\}, \quad (20)$$

and also define a phase transition in the sense that the finite local delay cannot be achieved for a given $\{\zeta_k\}$ provided $\theta \geqslant \theta_c$.

Furthermore, when $\{\zeta_k\} = \zeta$, i.e., equal mute probability, the local delay expression can be further simplified.

Corollary 2. *When $\{\alpha_k\} = \alpha$ and $\{\zeta_k\} = \zeta$, i.e., the path loss exponents and mute probabilities of all the tiers are both identical, the local delay is given by*

$$D = \frac{1}{1-\zeta} \cdot \frac{1}{1-(1-\zeta)\,\mathcal{Z}\left(\zeta,\alpha,\theta\right)}. \tag{21}$$

Specially, when $\zeta = 0$, we have

$$D = \frac{1}{1-\frac{2}{\alpha-2}\theta}. \tag{22}$$

On the other hand, when $0 < \zeta < 1$, we have

$$D = \frac{1}{1-\zeta} \cdot \frac{1}{1-(1-\zeta)\left[C\left(\alpha\right)\theta^{\frac{2}{\alpha}}\zeta^{\frac{2}{\alpha}-1} - \frac{1}{\zeta}H_{\alpha}\left(-\frac{1}{\theta\zeta}\right)\right]}, \tag{23}$$

where $C\left(\alpha\right) = 1/\mathrm{sinc}\left(\frac{2}{\alpha}\right)$, $H_{\alpha}\left(x\right) \triangleq {}_2F_1\left(1, \frac{2}{\alpha}; 1+\frac{2}{\alpha}; x\right)$ and ${}_2F_1\left[\cdot\right]$ denotes the Gauss hypergeometric function.

Proof: By letting $\{\zeta_k\} = \zeta$ in (18), we can get (21) directly. Subsequently, we consider two different subcases to evaluate $\mathcal{Z}\left(\zeta,\alpha,\theta\right)$ and obtain a closed-form expression of D.

First, when $\zeta = 0$, we have $\mathcal{Z}\left(\zeta,\alpha,\theta\right) = \frac{2}{\alpha-2}\theta$, and substituting it into (21), we can readily get (22). Second, when $0 < \zeta < 1$, from (10), $\mathcal{Z}\left(\zeta,\alpha,\theta\right)$ can be rewritten as

$$\begin{aligned}\mathcal{Z}\left(\zeta,\alpha,\theta\right) &= \int_0^{\infty} \frac{\theta}{u^{\frac{\alpha}{2}}+\theta\zeta}du - \int_0^{1} \frac{\theta}{u^{\frac{\alpha}{2}}+\theta\zeta}du \\ &= C\left(\alpha\right)\theta^{\frac{2}{\alpha}}\zeta^{\frac{2}{\alpha}-1} - \frac{1}{\zeta}{}_2F_1\left(1, \frac{2}{\alpha}; 1+\frac{2}{\alpha}; -\frac{1}{\theta\zeta}\right).\end{aligned} \tag{24}$$

Substituting (24) into (21) gives the desired result in (23). ■

The local delay is now independent of the BS transmit power p_k and BS density λ_k. It is also independent of the number of tiers K. This means that when we randomly add new BSs or increase transmit power of any tiers, the local delay remains the same. Intuitively, when adding new BSs, for instance, the average distance between a user and its associated BS is shorter, which increases the RSS; nevertheless, the user will experience a larger inter-cell interference. In this case, the above two effects cancel out each other, thus the SIR statistics do not change

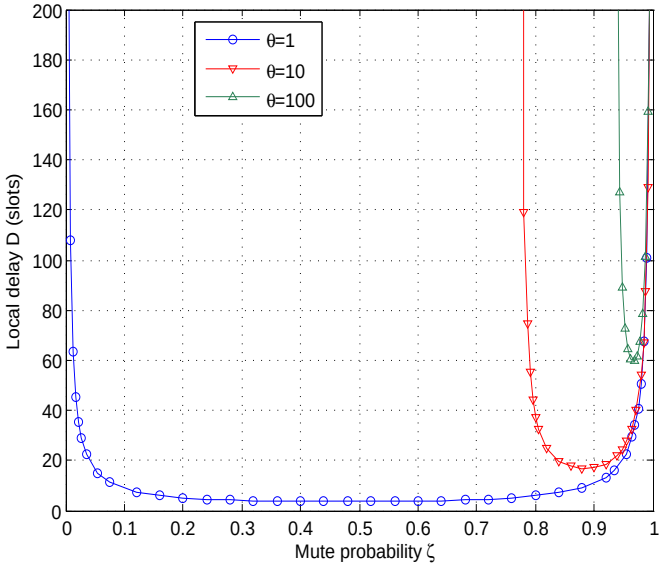

Fig. 2. The local delay D as a function of mute probability ζ for $\theta = 1, 10, 100$ and $\alpha = 4$. The delay is given in (21).

though more BSs are deployed. This invariance property coincides with the observations in [15], [16], where the performance metric under consideration is coverage probability. Besides, since $\mathcal{Z}(\zeta, \alpha, \theta)$ is a monotonically decreasing function of α, from (21) the local delay D decreases with path loss exponent α. This confirms our intuition that with a larger path loss exponent each BS is more likely to become isolated, hence higher path loss reduces the inter-cell interference and improves SIR.

By varying the values of θ and ζ, we get the curves in Fig. 2. Similar to the definition of θ_c, here we could define the critical mute probability as $\zeta_c \triangleq \inf\{\zeta : D(\zeta) < \infty\}$. From Fig. 2, we observe that the critical mute probability ζ_c increases when increasing θ, and particularly $\zeta_c \to 1$, as $\theta \to \infty$. This can be explained that with a larger SIR threshold, the outage probability in each time slot increases, which leads to a larger local delay. In order to keep the local delay finite, we should increase the mute probability (e.g., let more BSs stay mute in one time slot) to reduce the interference correlation [21].

The following proposition provides insights on the effect of random DTX scheme on local delay.

Proposition 1. *Without random DTX scheme applied in the HetNets, i.e., $\zeta = 0$, there exists a phase transition where $\theta_c = \frac{\alpha-2}{2}$. Instead, with random DTX scheme applied in the HetNets,*

i.e., $0<\zeta<1$*, there is no phase transition if* $\zeta_c<\zeta<1$ *where the critical mute probability* ζ_c *satisfies*

$$\zeta_c \sim 1-\frac{\theta^{-\frac{2}{\alpha}}}{C\left(\alpha\right)},\ \ \theta\to\infty. \tag{25}$$

Proof: When $\zeta=0$, from (22) and (20) we readily get the value of θ_c. When $0<\zeta<1$, in order to demonstrate that there is no phase transition if $\zeta_c<\zeta<1$, we just need to find a finite $\zeta_c\in(0,1)$ for the limit $\theta\to\infty$. From the definition of Gauss hypergeometric function, we have $\lim_{\theta\to\infty}H_\alpha\left(-\frac{1}{\theta\zeta}\right)=1$, and since $\zeta_c\to 1$, as $\theta\to\infty$, from (23) we find

$$D\sim\frac{1}{1-\zeta}\cdot\frac{1}{1-\left(1-\zeta\right)C\left(\alpha\right)\theta^{\frac{2}{\alpha}}},\ \theta\to\infty. \tag{26}$$

Thus, solving the equation $1-\left(1-\zeta_c\right)C\left(\alpha\right)\theta^{\frac{2}{\alpha}}=0$ gives (25). As $\theta\to\infty$, $1-\frac{\theta^{-\frac{2}{\alpha}}}{C(\alpha)}\in(0,1)$. Hence there exists a mute probability $\zeta\in(\zeta_c,1)$ that keeps a finite local delay for any θ. ■

From this proposition, we can draw a conclusion that for any SIR threshold θ, it is essential for random DTX scheme to ensure no phase transition. Particularly in the high-θ regime, we could effectively adjust the mute probability ζ to keep the local delay finite.

IV. Energy Efficiency

In this section, we will evaluate the network energy efficiency in the HetNets, which will be shown to depend critically on the BS power consumption model in (7), local delay and random DTX scheme.

By substituting (9) into (8), we can obtain a general expression of the energy efficiency η_{EE}. Besides, since (9) is not in closed form, numerical evaluation has to be employed to calculate η_{EE}, which will be shown in the next section. In order to obtain analytical results about the dependence of energy efficiency on some key parameters, such as mute probability and SIR threshold, in the following, we mainly focus on the special case of equal path loss exponent and equal mute probability, i.e., $\{\alpha_k\}=\alpha$ and $\{\zeta_k\}=\zeta$. Accordingly, the energy efficiency is given by

$$\eta_{EE}=\frac{\left(1-\zeta\right)^2\left[1-\left(1-\zeta\right)\mathcal{Z}\left(\zeta,\alpha,\theta\right)\right]\log\left(1+\theta\right)\sum_{k=1}^{K}\lambda_k}{\sum_{k=1}^{K}\lambda_k\left[\left(1-\zeta\right)\left(P_{k0}+\Delta_k p_k\right)+\zeta P_{k,S}\right]}. \tag{27}$$

where $\mathcal{Z}\left(\zeta,\alpha,\theta\right)$ is from (10).

Note that η_{EE} is a monotonically increasing function of path loss exponent α. This is consistent with the observation in Section III that the local delay D decreases when increasing α, which leads to higher energy efficiency.

A. The Effect of the BS Density

Intuitively, when we increase the BS density, the network throughput will increase. However, the area power consumption also increases with BS density. To this end, the following proposition gives the effect of the BS density on the energy efficiency.

Proposition 2. *The energy efficiency is a monotonically increasing function with λ_k if*

$$P_{k,ave} < \frac{\sum_{j=1,j\neq k}^{K} P_{j,ave}\lambda_j}{\sum_{j=1,j\neq k}^{K} \lambda_j}, \tag{28}$$

where $P_{k,ave} \triangleq (1-\zeta)(P_{k0}+\Delta_k p_k)+\zeta P_{k,S}$ for $k \in \mathcal{K}$. Otherwise, the energy efficiency is a monotonically decreasing function with λ_k.

Proof: From (27) we have $\eta_{EE}(\lambda_k) = \frac{\mathcal{F}\sum_{k=1}^{K}\lambda_k}{\sum_{k=1}^{K}\lambda_k P_{k,ave}}$, where $\mathcal{F}$ is positive and independent of λ_k. Hence the condition (28) is derived by investigating the derivative of $\eta_{EE}(\lambda_k)$ with respect to λ_k. ■

From this result, we observe that whether increasing the k^{th} tier's BS intensity λ_k will improve the energy efficiency depends on the BS power parameters in the k^{th} tier. Besides, since the local delay is independent of $\{\lambda_k\}$ in such a case, we can improve the energy efficiency by properly adjusting the BS density in each tier according to condition (28) without increasing the local delay.

B. The Effect of the SIR Threshold

In order to study the effect of SIR threshold θ on the energy efficiency, we give the following proposition.

Proposition 3. *Given a mute probability ζ, the optimal SIR threshold θ^* that maximizes the network energy efficiency satisfies*

$$\mathcal{Z}(\zeta,\alpha,\theta^*) + (1+\theta^*)\log(1+\theta^*)\,\mathcal{Z}'(\zeta,\alpha,\theta^*) = \frac{1}{1-\zeta}, \tag{29}$$

where $\mathcal{Z}(\zeta,\alpha,\theta)$ *is from (10) and*

$$\mathcal{Z}'(\zeta,\alpha,\theta)=\int_1^{\infty}\frac{u^{\frac{\alpha}{2}}}{\left(u^{\frac{\alpha}{2}}+\theta\zeta\right)^2}du. \tag{30}$$

In particular, when $\zeta=0$*, i.e., random DTX scheme is not applied in the HetNets, the bounds of the optimal SIR threshold* θ^* *is given by*

$$\theta^*\in\left(\sqrt{\frac{\alpha}{2}}-1,\frac{\alpha-2}{4}\right). \tag{31}$$

Proof: When ζ is fixed, from (27) we have $\eta_{EE}(\theta)=\mathcal{G}\left[1-(1-\zeta)\mathcal{Z}(\zeta,\alpha,\theta)\right]\log(1+\theta)$, where $\mathcal{G}$ is positive and independent of θ. In order to maximize the energy efficiency under the finite local delay constraint, we can formulate an optimization problem as

$$\begin{aligned}\max_{\theta}\quad & \eta_{EE}(\theta)\\ \text{s.t.}\ \ & 0<\theta<\theta_c\end{aligned} \tag{32}$$

Since $\eta_{EE}(0)=\eta_{EE}(\theta_c)=0$, then there exists an optimal SIR threshold $\theta^*\in(0,\theta_c)$, and the value of θ^* is obtained from the derivative of $\eta_{EE}(\theta)$. By letting $\frac{\partial\eta_{EE}(\theta)}{\partial\theta}=0$, we get (29). Specially, by letting $\zeta=0$ in (29), the optimal SIR threshold θ^* satisfies

$$(1+\theta^*)\log(1+\theta^*)+\theta^*=\frac{\alpha-2}{2}. \tag{33}$$

By applying the inequalities $\frac{x}{1+x}<\log(1+x)<x$ to (33), we get the following inequalities that

$$\theta^*<\frac{\alpha-2}{2}-\theta^*<(1+\theta^*)\theta^*. \tag{34}$$

From (34), the bounds in (31) are obtained, both of which satisfy the constraint in (32). ■

From this result, we observe that the optimal SIR threshold θ^* depends on the mute probability ζ and path loss exponent α. Since we assume fixed-rate transmission, i.e., $\mathcal{R}=\log(1+\theta)$, this result also gives a guideline that the transmit rate in each time slot could be adjusted to maximize the network energy efficiency, and the optimal transmit rate $\mathcal{R}^{opt}$ in terms of energy effciency is given by $\mathcal{R}^{opt}=\log(1+\theta^*)$.

C. The Effect of the Mute Probability

In the following, we will investigate how the mute probability ζ in the random DTX scheme will affect the network energy efficiency, and whether applying the random DTX scheme in the HetNets will be always beneficial. Assume $0 < \zeta < 1$, i.e., random DTX scheme is applied in the HetNets, then by substituting (24) into (27), the energy efficiency can be rewritten as

$$\begin{aligned}\eta_{EE}(\zeta) = \frac{v_0(1-\zeta)^2}{v_1 - v_2\zeta} \log(1+\theta) \cdot \\ \left[1 - (1-\zeta)\left(C(\alpha)\,\theta^{\frac{2}{\alpha}}\zeta^{\frac{2}{\alpha}-1} - \frac{1}{\zeta}H_\alpha\left(-\frac{1}{\theta\zeta}\right)\right)\right],\end{aligned} \tag{35}$$

where $v_0 = \sum_{k=1}^{K} \lambda_k$, $v_1 = \sum_{k=1}^{K} \lambda_k (P_{k0} + \Delta_k p_k)$ and $v_2 = \sum_{k=1}^{K} \lambda_k (P_{k0} + \Delta_k p_k - P_{k,S})$.

Although the energy efficiency in (35) is closed-form and amenable to numerical evaluation, it is also desirable to find a simpler estimate that lends itself to a more direct interpretation of the benefits of random DTX scheme. As a result, we analyze the energy efficiency in two asymptotic regimes in the following.

1) Low-Rate Regime: Here, we investigate the asymptotic characteristics of the energy efficiency when $\theta \to 0$. Note that the limit $\theta \to 0$ refers to the low-rate regime since in this limit, the transmit rate $\mathcal{R}$ goes to zero.

The following proposition gives the effect of the mute probability on the energy efficiency in the low-rate regime.

Proposition 4. *In the low-rate regime, the optimal mute probability is given by* $\zeta_{\theta\to 0}^{opt} \to 0$*, and the maximum energy efficiency is*

$$\eta_{EE}^{\max} \sim \frac{v_0}{v_1}\log(1+\theta), \quad \theta \to 0. \tag{36}$$

Proof: From (24), for a general $\alpha > 2$, we have

$$C(\alpha)\,\theta^{\frac{2}{\alpha}}\zeta^{\frac{2}{\alpha}-1} - \frac{1}{\zeta}H_\alpha\left(-\frac{1}{\theta\zeta}\right) \sim \frac{2}{\alpha-2}\theta, \quad \theta \to 0. \tag{37}$$

Substituting (37) into (35) yields

$$\begin{aligned}\eta_{EE}(\zeta) &\sim \frac{v_0(1-\zeta)^2}{v_1 - v_2\zeta}\left[1 - \frac{2\,(1-\zeta)}{\alpha-2}\theta\right]\log\,(1+\theta) \\ &\sim \frac{v_0(1-\zeta)^2}{v_1 - v_2\zeta}\log\,(1+\theta)\,, \quad \theta \to 0.\end{aligned} \tag{38}$$

Besides, by substituting (37) into (23), we have

$$\begin{aligned}D &\sim \frac{1}{1-\zeta}\cdot\frac{1}{1-(1-\zeta)\,\frac{2}{\alpha-2}\theta} \\ &\sim \frac{1}{1-\zeta} + \frac{2}{\alpha-2}\theta, \quad \theta \to 0.\end{aligned} \tag{39}$$

This means that the local delay D is finite for all $\zeta \in (0,1)$ in this limit. As a result, in the low-rate regime, from (38) and (39), we can formulate the energy efficiency maximization problem as

$$\begin{aligned}\max_{\zeta} \quad & \frac{v_0(1-\zeta)^2}{v_1 - v_2\zeta}\log\,(1+\theta) \\ \text{s.t.} \quad & 0 < \zeta < 1.\end{aligned} \tag{40}$$

Let $g\,(\zeta) = \frac{(1-\zeta)^2}{v_1 - v_2\zeta}$ where $0 < \zeta < 1$ and $v_1 > v_2 > 0$, then by calculation, the derivative of $g\,(\zeta)$ over ζ is negative. Thus the maximum energy efficiency is obtained from $\zeta \to 0$, and inserting the asymptotic value $\zeta = 0$ into (38) gives the desired result in (36). ■

Asymptotically, from this proposition, we observe that in the low-rate regime, it will be less energy-efficient to apply random DTX scheme in the HetNets. In addition, we should note that the maximum energy efficiency in (36) does not depend on the path loss exponent α. This can be explained by noting that although increasing the path loss exponent α will improve the local delay, the improvement of the local delay in the low-rate regime is very marginal.

The low-rate asymptotic behavior of $\eta_{EE}^{\max}$ and its numerical results are shown in Fig. 3. It can be observed that the asymptotic value is quite tight as SIR threshold θ is small. Also, when $\rho = \frac{\lambda_2}{\lambda_1}$ increases, the asymptote becomes tighter. Furthermore, the maximum energy efficiency $\eta_{EE}^{\max}$ (with respect to ζ) increases logarithmically with SIR threshold θ. It is because that the dominant factor that affects the energy efficiency in this limit is transmit rate rather than the local delay, then increasing SIR threshold θ will increase transmit rate, which leads to higher energy efficiency.

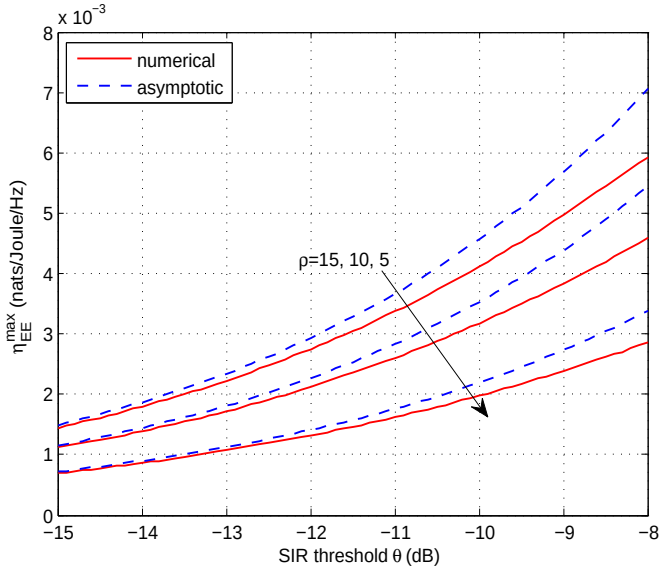

Fig. 3. The low-rate asymptotic behavior of the maximum energy efficiency $\eta_{EE}^{\max}$ (with respect to ζ) in a two-tier HetNets consisting of macrocells and picocells, where $\rho \triangleq \frac{\lambda_2}{\lambda_1} = 5, 10, 15$ and $\alpha = 4$.

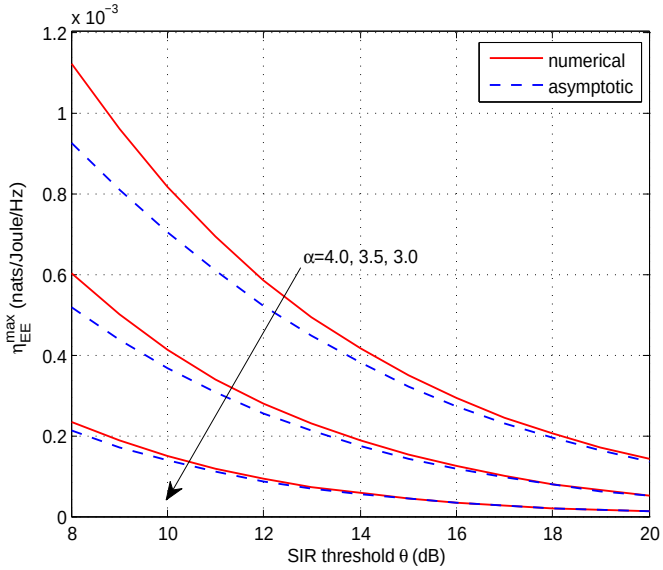

Fig. 4. The high-rate asymptotic behavior of the maximum energy efficiency $\eta_{EE}^{\max}$ (with respect to ζ) in a two-tier HetNets consisting of macrocells and picocells, where $\lambda_2 = 5\lambda_1$ and $\alpha = 3.0, 3.5, 4.0$.

2) High-Rate Regime: Here, we investigate the asymptotic characteristics of the energy efficiency when $\theta \to \infty$. In this regime, the outage probability in each time slot goes to 1 while the transmit rate goes to infinity. Thus, it is crucial to select proper mute probability so that we keep the local delay finite while maximizing the energy efficiency in this limit.

The following proposition gives the effect of the mute probability on the energy efficiency in the high-rate regime.

Proposition 5. *In the high-rate regime, the optimal mute probability is given by* $\zeta_{\theta\to\infty}^{opt} = 1-\frac{2\theta^{-\frac{2}{\alpha}}}{3C(\alpha)}$, *and the maximum energy efficiency is*

$$\eta_{EE}^{\max} \sim \frac{4v_0\theta^{-\frac{4}{\alpha}}\log\left(1+\theta\right)}{27[C\left(\alpha\right)]^2\left(v_1-v_2\right)}, \quad \theta\to\infty. \tag{41}$$

Proof: In order to keep the local delay finite, from Proposition 1, we have $\zeta_c < \zeta < 1$, where ζ_c is given by (25). As a result, given a coefficient $\varepsilon \in (0,1)$, the mute probability that keeps a finite local delay has a form that

$$\zeta \sim 1 - \frac{\varepsilon\theta^{-\frac{2}{\alpha}}}{C\left(\alpha\right)}, \quad \theta\to\infty. \tag{42}$$

Since $\lim_{\theta\to\infty} H_\alpha\left(-\frac{1}{\theta\zeta}\right) = 1$, substituting (42) into (35) yields

$$\eta_{EE}(\varepsilon) \sim \varepsilon^2\left(1-\varepsilon\right)\frac{v_0\theta^{-\frac{4}{\alpha}}\log\left(1+\theta\right)}{[C\left(\alpha\right)]^2\left(v_1-v_2\right)}, \quad \theta\to\infty. \tag{43}$$

Similarly, the energy efficiency maximization problem in the high-rate regime can be formulated as

$$\begin{aligned} \max_{\varepsilon} \quad & \varepsilon^2\left(1-\varepsilon\right)\frac{v_0\theta^{-\frac{4}{\alpha}}\log\left(1+\theta\right)}{[C\left(\alpha\right)]^2\left(v_1-v_2\right)} \\ \text{s.t.} \quad & 0<\varepsilon<1. \end{aligned} \tag{44}$$

Obviously, the optimal solution of (44) is $\varepsilon^{opt} = \frac{2}{3}$. By substituting $\varepsilon^{opt} = \frac{2}{3}$ into (42) and (43), we can obtain the asymptotic expressions of the optimal mute probability and the maximum energy efficiency. ■

This result shows that in the high-rate regime, the mute probability could be optimized to maximize the energy efficiency. Since the optimal mute probability $\zeta_{\theta\to\infty}^{opt} \neq 0$, it will be more energy-efficient to apply random DTX scheme in the HetNets. Actually, the high-rate asymptotics

of non-random DTX scheme applied in any tier do not exist since there is a phase transition where $\theta_c = \frac{\alpha-2}{2}$ when $\zeta = 0$.

The high-rate asymptotic behavior of $\eta_{EE}^{\max}$ and its numerical result are shown in Fig. 4. It can be seen that the asymptotic value is quite tight when the SIR threshold θ is large. Also, when α decreases, the asymptote becomes tighter. Besides, the maximum energy efficiency decreases with the increment of the SIR threshold θ, since the dominant fact that affects the energy efficiency in this limit is the local delay, and increasing SIR threshold θ will increase local delay, which leads to lower energy efficiency. Also, since $C(\alpha)$ is a monotonically decreasing function of α when $\alpha > 2$, the maximum energy efficiency increases with the path loss exponent α, which demonstrates a significant difference from the low-rate case.

V. Extension to Load-aware DTX Schemes

In this section, we extend the analysis of random DTX scheme to incorporate the impact of user activity. That is, instead of randomly switching BSs off, the BSs stay mute when their activity levels are relatively low. Similarly to the strategic sleeping in [25], we model the load-aware DTX scheme in each tier and each time slot as a function $y_k : [0,1] \mapsto [0,1]$, which means that if the activity level of the BS in the k^{th} tier is a_k, then it stays mute with probability $y_k(a_k)$ and stays active with probability $1 - y_k(a_k)$, independently of the BS location and time slot. Compared with the random DTX scheme, the load-aware DTX scheme goes one step further and models a network that is adaptive to the fluctuating activity levels within the location [25]. In this sense, this DTX scheme strategy can be seen as a load-aware policy and can incorporate traffic profile in the performance analysis.

Using the load-aware DTX scheme, the local delay that captures the user activity is provided in the following theorem.

Theorem 2. *In a K-tier HetNet with load-aware DTX scheme and cell association based on the long-term RSS, the local delay of the active user is given by*

$$\tilde{D} = \sum_{k=1}^{K} \frac{\mathbb{E}[a_k]}{\mathbb{E}[a_k] - \mathbb{E}[a_k y_k]} \mathcal{T}_k\left(\mathbb{E}[y_1], \mathbb{E}[y_2], \cdots, \mathbb{E}[y_K]\right), \tag{45}$$

where $\mathbb{E}[y_k] = \int_0^1 y_k(x) f_{a_k}(x)\, dx$ *and*

$$\mathcal{T}_k(v_1, v_2, \cdots, v_K) \triangleq 2\pi\lambda_k \int_0^\infty r \exp\left[-\sum_{j=1}^{K} \pi\lambda_j \left(\frac{p_j}{p_k}\right)^{\frac{2}{\alpha_j}} r^{\frac{2\alpha_k}{\alpha_j}} \left(1-(1-v_j)\,\mathcal{Z}(v_j, \alpha_j, \theta)\right)\right] dr. \tag{46}$$

Proof: The first step is to condition on the activity of a typical cell a_k. Next, according to the definition of local delay, the probability of successful transmission conditioned on r_k, Φ and a_k is given by

$$\begin{aligned}\Pr\left(\tilde{\mathcal{C}}_{k|\Phi}\right) &\overset{(a)}{=} \frac{1}{\mathbb{E}[a_k]} \int_0^1 x\left(1-y_k(x)\right) \Pr\left(\mathrm{SIR}_{k,t} > \theta \,|r_k, \Phi, a_k\right) f_{a_k}(x)\, dx \\ &= \frac{\mathbb{E}[a_k] - \mathbb{E}[a_k y_k]}{\mathbb{E}[a_k]} \Pr\left(\mathrm{SIR}_{k,t} > \theta \,|r_k, \Phi\right),\end{aligned} \tag{47}$$

where (a) is from the probability of successful transmission weighted over the active user links and $f_{a_k}(x)$ denotes the pdf of random variable a_k. The rest of the proof follows the same steps as in the proof of Theorem 2 by substituting ζ_k with $\mathbb{E}[y_k]$. ■

The average area power consumption after employing the load-aware DTX scheme is given by

$$P_a = \sum\nolimits_{k=1}^{K} \lambda_k \left\{\left(1-\mathbb{E}[y_k]\right)\left(P_{k0} + \Delta_k p_k\right) + \mathbb{E}[y_k]\, P_{k,S}\right\}. \tag{48}$$

By substituting (45) and (48) into (8), we could obtain the general expression of energy efficiency. In contrast to Theorem 1, we observe that most of the above analytical results in the case of random DTX scheme also apply to the load-aware DTX scheme. For instance, the insights into the impact of path loss exponents, BS densities and SIR threshold on local delay and energy efficiency remain unchanged. The main difference is that the mute probability in the random DTX scheme is an independent and tunable parameter, while, in the load-aware DTX scheme, the mute probability depends on the user activity level and thus could not be adjusted independently.

In order to compare the load-aware DTX scheme with the random DTX scheme, we now let $\mathbb{E}[y_k] = \zeta_k$ for fairness. The improvement of the load-aware DTX scheme is provided in the following proposition.

Proposition 6. *The load-aware DTX scheme improves the local delay of the active user if it*

satisfies the following inequality

$$\mathbb{E}\left[a_k y_k\right] < \mathbb{E}\left[a_k\right]\mathbb{E}\left[y_k\right]. \tag{49}$$

Proof: From Theorem 1, the local delay with random DTX scheme can be rewritten as $D = \sum_{k=1}^{K} \frac{1}{1-\zeta_k}\mathcal{T}_k\left(\zeta_1, \zeta_2, \cdots, \zeta_K\right)$. Note that $\mathbb{E}\left[y_k\right] = \zeta_k$, the sufficient condition for $\tilde{D} < D$ is $\frac{\mathbb{E}[a_k]}{\mathbb{E}[a_k]-\mathbb{E}[a_k y_k]} < \frac{1}{1-\zeta_k}$, which directly leads to (49). ■

From this proposition, we observe that for fixed $\mathbb{E}\left[y_k\right]$, in order to minimize the value of $\mathbb{E}\left[a_k y_k\right]$, we need to match large values of y_k with low activity. Therefore, by assuming that $y_k\left(a_k\right)$ is a strictly monotonically decreasing function of a_k, this guarantees that the load-aware DTX scheme will result in better performance than the random DTX scheme. In addition, the decreasing property of $y_k\left(a_k\right)$ suggests the intuitive policy that a large fraction of BSs are switched off when the user activity is low.

We conclude that compared to the actual DTX deployment, random DTX scheme can serve as a suboptimal DTX strategy (or baseline DTX strategy) since it completely neglects the dependence of mute probability on the BS load (or users traffic pattern).

VI. Numerical Results

In this section, we present the simulation results to validate our analysis and evaluate the local delay and energy efficiency performance under some general cases to complement our theoretical results. For clarity, we restrict our presented results to an interference-limited two-tier HetNet consisting of macro and pico BSs. Unless otherwise stated, the power parameters are chosen according to [29], namely, $p_1 = 20$, $p_2 = 0.13$, $P_{10} = 130$, $P_{20} = 6.8$, $\Delta_1 = 4.7$ and $\Delta_2 = 4.0$, where the subscripts '1' and '2' corresponds to macrocell and picocell networks, respectively. As for the static power expenditure in the mute mode of a BS in the k^{th} tier, we assume that $P_{k,S} = 0.75P_{k0}$, where $k = 1, 2$.

In Fig. 5, we show local delay and energy efficiency as a function of picocell BS density λ_2 with different values of picocell path loss exponent α_2. First, since we assume that $\zeta_1 = \zeta_2 = 0.5$, it can be seen that when $\alpha_1 = \alpha_2 = 3.5$, the local delay remains constant with different values of picocell BS density, which verifies our theoretical analysis. Second, the local delay decreases as the picocell BS density increases when the picocell BSs experience higher path loss, i.e., $\alpha_2 > \alpha_1$, while it increases as more picocell BSs are added when $\alpha_2 < \alpha_1$. As we have discussed before,

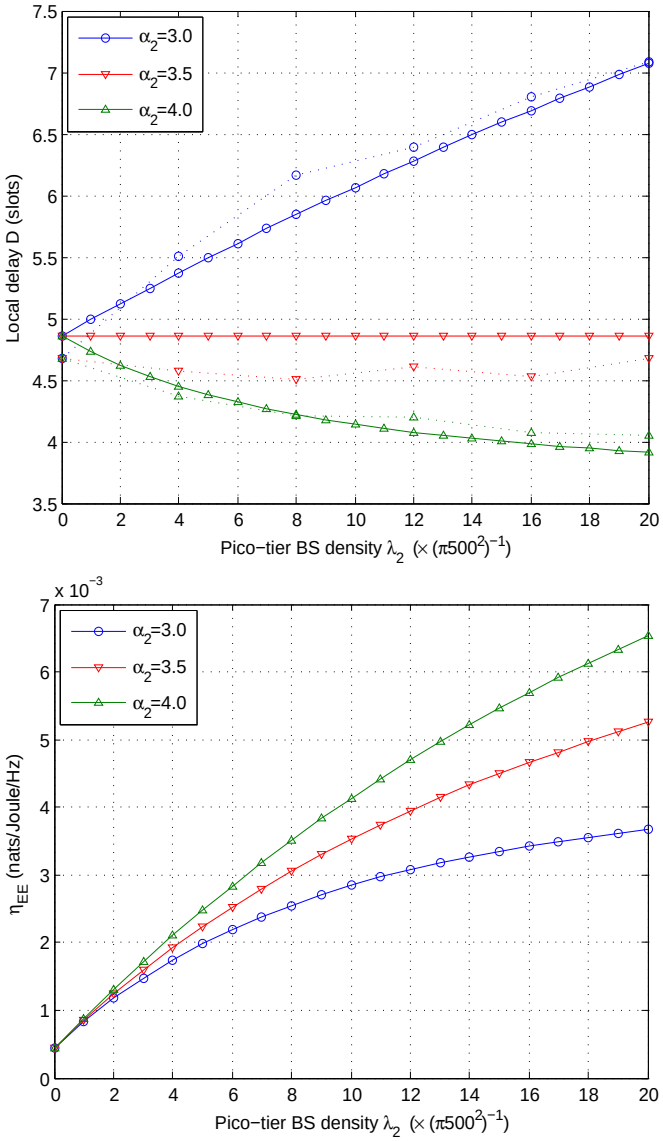

Fig. 5. Local delay and energy efficiency for varying BS density and path loss exponent of picocells in a two-tier HetNet, where $\alpha_1 = 3.5$, $\zeta_1 = \zeta_2 = 0.5$, $\theta = 0$dB and $\lambda_1 = \frac{1}{\pi 500^2}$.

adding more picocell BSs will cause a shorter average association distance and also a higher inter-tier interference, which cancels each other out if $\alpha_1 = \alpha_2$. Intuitively, if $\alpha_1 < \alpha_2$, the effect of a shorter average association distance becomes dominant and eventually improves the SIR; nevertheless, if $\alpha_1 > \alpha_2$, the effect of a higher inter-tier interference becomes dominant and eventually reduces the SIR.

In order to verify the theoretical analysis, we also show the Monte Carlo simulations of local

delay with dotted curves. In the Monte Carlo simulations, we chose a spatial window, which is a square of 20km $\times$ 20km, and generated two independent PPPs of BS locations with their respective densities. The final simulation results were obtained by averaging 50000 independent realizations. We observe that the simulation results match numerical results well and their gaps lie within the tolerable range. There are mainly two reasons for the existence of the gaps. First, the spatial window and the number of independent realizations are both relatively small due to the runtime constraint. Second, the local delay in [21] presents a heavy-tailed distribution and thus a large number of independent realizations are needed to calculate the mean value, which might be very time-consuming. This phenomenon indirectly illustrates the significance of theoretical analysis.

As for the network energy efficiency, we observe that it increases with the increment of the picocell BS density for all the three cases, but the increasing rates of these three curves are different. To explain this, we first take the case of $\alpha_1 = \alpha_2$ as a baseline. Since $P_{2,ave} < P_{1,ave}$ with the given power parameters of macrocell and picocell, according to Proposition 2, the energy efficiency increases with the increment of the picocell BS density λ_2. In addition, if $\alpha_1 < \alpha_2$, the local delay decreases when increasing λ_2, which further improves the energy efficiency, while if $\alpha_1 > \alpha_2$, the local delay increases when increasing λ_2, which instead deteriorates the energy efficiency to some extent. The above results indicate that if adding a picocell network tier overlaying the existing macrocell network tier, new picocell BSs are better deployed with a higher path loss exponent, from the perspective of both local delay and energy effiicency.

In Fig. 6, we show local delay and energy efficiency as a function of the picocell mute probability ζ_2 with different values of picocell BS density λ_2. First, for different picocell BS densities, there exists an optimal picocell mute probability $\zeta_2^* \approx \zeta_1$ that minimizes the local delay. This indicates that in terms of local delay, we had better keep the same mute probability in different tiers. Besides, it is observed that the optimal local delay under different picocell BS densities is also identical. It is because of the invariance property we have observed from the theoretical analysis when $\alpha_1 = \alpha_2$ and $\zeta_1 = \zeta_2$. However, when $\zeta_1 \neq \zeta_2$, we observe that local delay increases with picocell BS density λ_2 and the gap $|\zeta_2 - \zeta_1|$. This can be explained as follows. If $\zeta_2 < \zeta_2^*$, the interference correlation of different time slots becomes larger, while if $\zeta_2 > \zeta_2^*$, the probability that the associated BS is active becomes smaller, both of which will lead to the higher local delay. As picocell BS density λ_2 increases, one of the above two effects

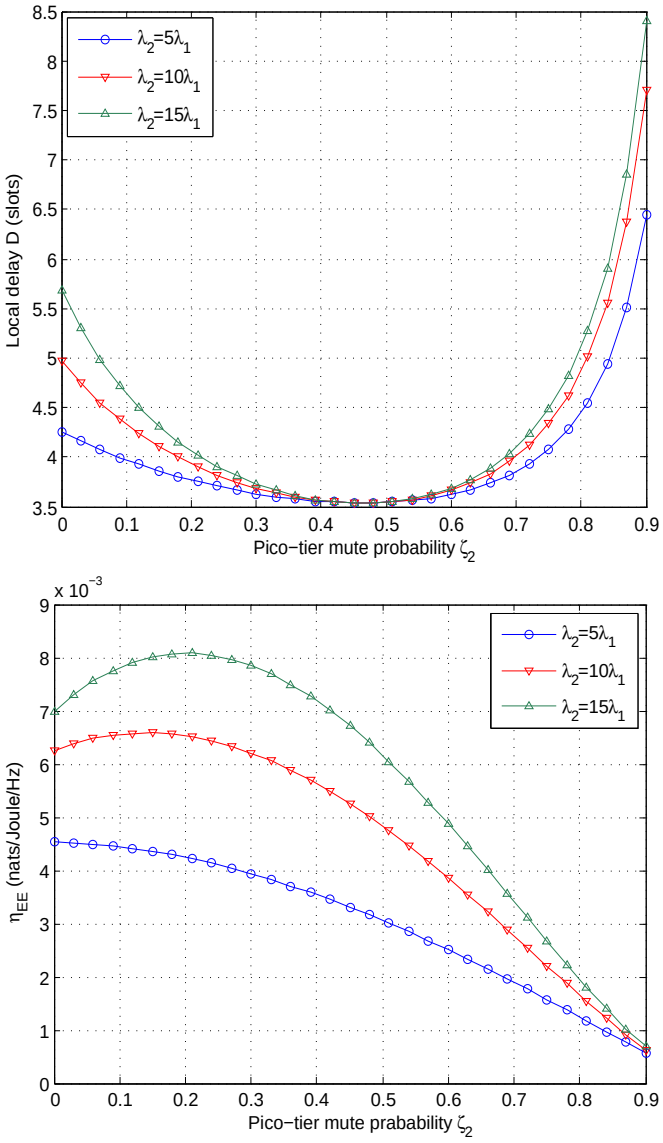

Fig. 6. Local delay and energy efficiency for varying mute probability and BS density of picocells in a two-tier HetNet, where $\zeta_1 = 0.5$, $\alpha_1 = \alpha_2 = 4.0$, $\theta = 0$dB and $\lambda_1 = \frac{1}{\pi 500^2}$.

becomes increasingly dominant, respectively, so as to further increase the local delay.

When observing the energy efficiency performance, for a given picocell BS density λ_2, there also exists an optimal mute probability $\zeta_2^*(\lambda_2)$ that maximizes the network energy efficiency. Furthermore, $\zeta_2^*(\lambda_2)$ is an increasing function of λ_2, which shows a large difference from that of the local delay. We can explain this that from (8) there are two parts in the expression that affect the network energy efficiency, namely, the local delay D and the unit time slot based energy

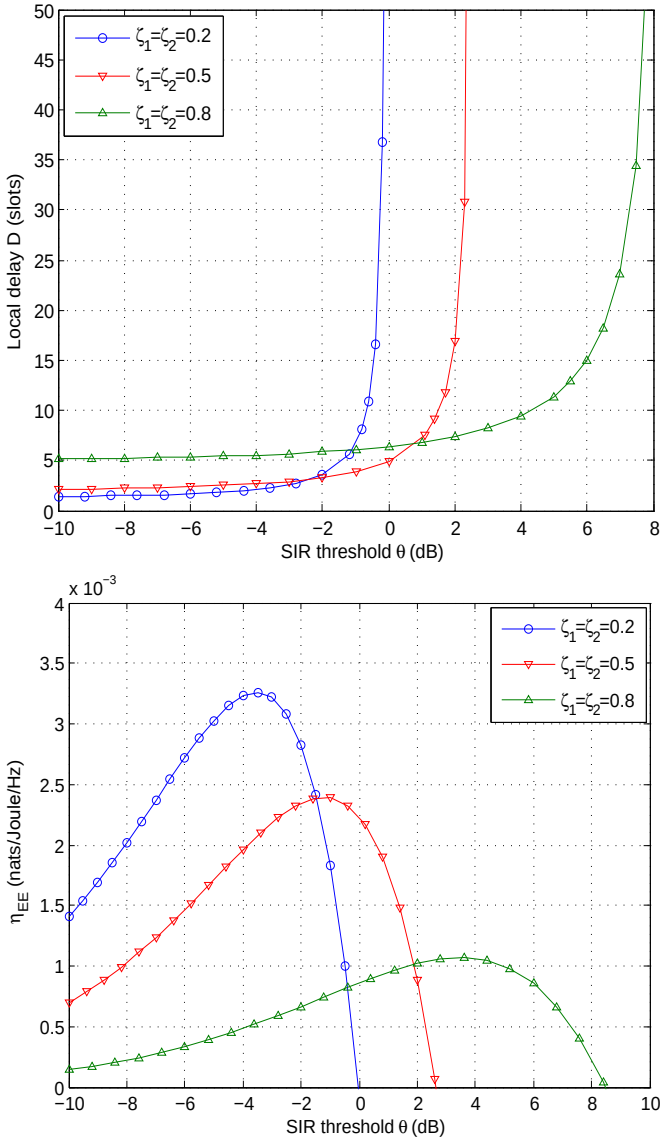

Fig. 7. Local delay and energy efficiency for varying SIR threshold and mute probability of picocells in a two-tier HetNet, where $\alpha_1 = \alpha_2 = 3.5$, $\lambda_2 = 5\lambda_1 = 5\left(\frac{1}{\pi 500^2}\right)$.

efficiency, denote by $\eta_{EE,t}$. To minimize D we should have $\zeta_2^* = \zeta_1$, and to maximize $\eta_{EE,t}$ we let $\zeta_2^* = 0$. This well gives the reason why the optimal mute probability $\zeta_2^*(\lambda_2)$ actually lies in the region $(0, \zeta_1)$. As λ_2 increases, the local delay D increases and its effect on the energy efficiency strengthens, thus we can slightly increase $\zeta_2^*(\lambda_2)$ to improve the network energy efficiency. Based on the above results, we should consider the tradeoff of local delay and energy efficiency when designing an appropriate picocell mute probability ζ_2 in practice.

In Fig. 7, we show local delay and energy efficiency as a function of SIR threshold θ with different values of mute probability ζ where $\zeta_1 = \zeta_2 = \zeta$. First, the local delay increases with the increment of the SIR threshold, which verifies our theoretical analysis. Also, it is observed that the critical SIR threshold $\theta_c(\zeta)$ is an increasing function of ζ. Intuitively, when ζ increases, the interference correlation reduces, thus the finite local delay D can be achieved by the larger SIR threshold θ. However, we observe that the local delay in the limit $\theta \to 0$, denote by $D_{\theta\to 0}$, indeed increases when increasing ζ. From (39), we have $D_{\theta\to 0} \approx \frac{1}{1-\zeta}$, which means that $D_{\theta\to 0}$ mainly depends on the probability that the associated BS is active rather than the interference correlation. Particularly, the points when $\theta = -10$dB matches the analytical result in (39) well.

Next, it is shown that for different values of mute probability ζ, there exists an optimal SIR threshold $\theta^*(\zeta)$ that maximizes the energy efficiency. As a result, $0 < \theta < \theta^*(\zeta)$ is called the rate-limited regime, and $\theta^*(\zeta) < \theta < \theta_c(\zeta)$ is called the delay-limited regime. According to Proposition 3, the optimal SIR threshold $\theta^*(\zeta)$ satisfies (29), which verifies our theoretical result. Besides, since $\eta_{EE,t}$ decreases with ζ, we can see from this figure that the maximum energy efficiency $\eta_{EE}^{\max}(\zeta)$ decreases with ζ. Also, as ζ increases, $\theta_c(\zeta)$ increases, and thus the rate-limited regime increases, this accurately explained the fact that the optimal SIR threshold $\theta^*(\zeta)$ increases with ζ.

VII. Conclusions

In this paper, we analytically derived the local delay and energy efficiency of a typical user in a K-tier HetNet with the random DTX scheme. Furthermore, the simplified or closed-form expressions in some special cases were obtained. The theoretical analysis, along with the numerical results gave some insights into the effect of key system parameters on local delay and energy efficiency, such as path loss exponents, BS densities, SIR threshold and mute probability. In addition, we investigated the low-rate and high-rate asymptotic behavior of the maximum energy efficiency. The asymptotics revealed that it is less energy-efficient to apply random DTX scheme in the low-rate regime while it could achieve the finite local delay and is more energy-efficient to apply random DTX scheme in the high-rate regime. Finally, we extended the analysis to the load-aware DTX scheme, which is more realistic.

REFERENCES

[1] T. Q. S. Quek, G. de la Roche, I. Guvenc, and M. Kountouris, *Small Cell Networks: Deployment, PHY Techniques, and Resource AllocaBon,* Cambridge University Press, 2013.

[2] Qualcomm, "LTE advanced: heterogeneous networks," white paper, Jan. 2011.

[3] D. Lopez-Perez, I. Guvenc, G. de la Roche, M. Kountouris, T. Q. S. Quek, and J. Zhang, "Enhanced intercell interference coordination challenges in heterogeneous networks," *IEEE Wireless Commun. Mag.*, vol. 18, no. 3, pp. 22-30, June 2011.

[4] 3GPP TR 32.826, Telecommunication management; Study on Energy Savings Management (ESM), (Release 10), Mar 2010. *Available:* http://www.3gpp.org/ftp/Specs/html-info/32826.htm.

[5] L. M. Correia, D. Zeller, O. Blume, D. Ferling, Y. Jading, I. Godor, G. Auer, and L. van der Perre, "Challenges and enabling technologies for energy aware mobile radio networks," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 66-72, Nov. 2010.

[6] J. He, P. Loskot, T. O'Farrell, V. Friderikos, S. Armour, and J. Thompson, "Energy efficient architectures and techniques for Green Radio access networks," in *Proc. 5th International ICST Conference on Communications and Networking in China (CHINACOM)*, Aug. 2010, Invited Paper.

[7] W. Guo and T. O'Farrell, "Dynamic Cell Expansion with Self-Organizing Cooperation", *IEEE J. Sel. Areas Commun.*, vol. 31, no. 5, pp. 851-860, May 2013.

[8] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, "Green cellular networks: a survey, some research issues and challenges," *IEEE Commun. Surveys & Tutorials*, vol. 13, no. 4, pp. 524-540, Fourth Quarter, 2011.

[9] J. Hoydis, M. Kobayashi, and M. Debbah, "Green small-cell networks," *IEEE Veh. Technol. Mag.*, vol. 6, no. 1, pp. 37-43, Mar. 2011.

[10] T. Bohn, *et al.*, "D4.1: Most promising tracks of green radio technologies," INFSO-ICT-247733 EARTH, Tech. Rep., Dec. 2010. [Online]. Available: https://www.ictearth.eu/publications/deliverables/deliverables.html.

[11] B. Rao and A. Fapojuwo, "A survey of energy efficient resource management techniques for multicell cellular networks," *IEEE Commun. Surveys & Tutorials*, vol. 16, no. 1, pp. 154-180, First Quarter 2014.

[12] J. G. Andrews, N. Jindal, M. Haenggi, R. Berry, S. Jafar, D. Guo, S. Shakkottai, R. Heath, M. Neely, S. Weber, and A. Yener, "Rethinking information theory for mobile ad hoc networks," *IEEE Commun. Mag.*, vol. 46, no. 12, pp. 94-101, Dec. 2008.

[13] F. Baccelli and B. Blaszczyszyn, "A new phase transition for local delays in MANETs," in *Proc. 2010 IEEE INFOCOM.*

[14] 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.

[15] 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.

[16] H.-S. Jo, Y. J. Sang, P. Xia, and J. G. Andrews, "Heterogeneous cellular networks with flexible cell association: a comprehensive downlink SINR analysis," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3484-3495, Oct. 2012.

[17] M. Haenggi, "The local delay in Poisson networks," *IEEE Trans. Inf. Theory*, vol. 59, pp. 1788-1802, Mar. 2013.

[18] Z. Gong and M. Haenggi, "The local delay in mobile Poisson networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4766-4777, Sept. 2013.

[19] K. Gulati, R. Ganti, J. G. Andrews, B. Evans, and S. Srikanteswara, "Characterizing decentralized wireless networks with temporal correlation in the low outage regime," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3112-3125, Sept. 2012.

[20] X. Zhang and M. Haenggi, "Delay-optimal power control policies," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3518-3527, Oct. 2012.

[21] Y. Zhong, W. Zhang, and M. Haenggi, "Managing interference correlation through random medium access," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 928-941, Feb. 2014.

[22] P. Frenger, P. Moberg, J. Malmodin, Y. Jading, and I. Godor, "Reducing energy consumption in LTE with cell DTX," in *Proc. 2011 IEEE VTC Spring*.

[23] K. Hiltunen, "Utilizing eNodeB sleep mode to improve the energy-efficiency of dense LTE networks," in *Proc. 2013 Personal Indoor and Mobile Radio Communications (PIMRC)*.

[24] S. Tombaz, S. Han, K. Sung, and J. Zander, "Energy efficient network deployment with cell DTX," *IEEE Commun. Letters*, vol. 18, no. 6, pp. 977-980, June 2014.

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

[26] C. Yan, S. Zhang, S. Xu, and G. Y. Li, "Fundamental trade-offs on green wireless networks," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 30-37, June 2011.

[27] M. J. Neely, "Optimal energy and delay tradeoffs for multiuser wireless downlinks," *IEEE Trans. Inf. Theory*, vol. 53, no. 9, pp. 3095-3113, Sept. 2007.

[28] J. Wu, S. Zhou, and Z. Niu, "Traffic-aware base station sleeping control and power matching for energy-delay tradeoffs in green cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 8, pp. 4196-4209, Aug. 2013.

[29] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?," *IEEE Wireless Commun. Mag.*, vol. 18, no. 5, pp. 40-49, Oct. 2011.

[30] D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and Its Applications,* 2nd edition. John Wiley and Sons, 1996.