SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

March 2022, Vol. 65 132304:1–132304:10
https://doi.org/10.1007/s11432-021-3306-4

On the local delay and energy efficiency under decoupled uplink and downlink in HetNets

Tianjie HUANG, Fu-Chun ZHENG* & Lifeng LAI

School of Electronic and Information Engineering, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, *China*

Received 21 April 2021/Revised 25 May 2021/Accepted 27 July 2021/Published online 14 February 2022

Abstract With the proliferation of the Internet of Things, uplink transmission performance has become more and more important. Moreover, with the traditional coupled uplink (UL)/downlink (DL) access (CUDA) mode, it is becoming hard to improve the system performance of UL in ultra dense heterogeneous networks (UDN, or HetNets). In this paper, we have conducted a UL performance comparison study for the decoupled UL/DL access (DUDA) mode over the CUDA mode in UDN. The key performance indicators of local delay and energy efficiency (EE) were investigated with respect to SIR threshold, ratio of small cell base station (SBS) and macro cell base station (MBS) densities, and fractional power control (FPC) factor. Numerical and simulation results confirm that the local delay is lower and the EE is higher under DUDA than under CUDA, and as such the DUDA mode is superior to the CUDA mode for the UL performance of UDN.

Keywords coupled and decoupled UL/DL access, HetNets, stochastic geometry, local delay, energy efficiency, UDN

Citation Huang T J, Zheng F-C, Lai L F. On the local delay and energy efficiency under decoupled uplink and downlink in HetNets. Sci China Inf Sci, 2022, 65(3): 132304, https://doi.org/10.1007/s11432-021-3306-4

1 Introduction

With the rapid development of wireless Internet, the architecture of ultra dense heterogeneous networks (UDN, or simply HetNets) has been adopted. HetNets consist of multiple base stations (BSs) with different transmit power levels and path loss characteristics, such as macro cell base stations (MBS) and small cell base stations (SBS). Due to the overlay of these BS tiers in HetNets, the user equipment (UE) should have the freedom to choose different cells for downlink (DL) and uplink (UL). In the traditional mode of coupled UL and DL access (CUDA), however, the UL and DL of the UE are always associated with the same BS, from which the UE receives the strongest signal in DL. Although the CUDA mode and HetNets can temporarily reduce network congestion to some extent, the UL transmission pressure of UE can still become challenging in the future [1, 2], in particular for the uplink centric broadband communications (UCBC) scenarios. The decoupled UL and DL access (DUDA) mode has therefore been explored recently for ultra dense small cell deployment.

In the DUDA mode, the DL connection can be the same as in the CUDA mode, however, the UL of UEs is likely to be with a geometrically closer SBS (e.g., its DL is still associated with an MBS). In fact, the higher the transmit power is, the larger the service radius is, thus the number of UE connections for each tier of cells in the CUDA mode is unequal, which leads to a serious load imbalance. In contrast, since the transmission power of the UEs is the same in DUDA, the service range of each BS in UL is equal, and therefore the load is much more balanced. Specifically, assuming that a UE is located at the edge of a macro cell in the CUDA mode, its UL path loss will be too large to transmit reliably. In a DUDA system, however, the UE will be connected to a closer SBS, suffering from much lower path loss, hence the recent increase in the wireless industry's interest in DUDA.

On the other hand, the deployment of ultra dense BSs in UDN makes energy efficiency (EE) an important performance indicator for the network performance, which reflects the requirements of green

* Corresponding author (email: fzheng@ieee.org)

communications as well as the networks' energy cost [3]. In addition, the local delay of uplink transmission is directly related not only to the reliability of communications but also to the corresponding overall end-to-end latency [4]. In this paper, we will therefore investigate both the local delay and EE in the CUDA and DUDA modes.

1.1 Related work

Currently, most existing research studies on local delay are focused on DL transmission in HetNets. Ref. [5] analyzed the impact of different handover hysteresis parameters on the local delay in dense networks. In [6], the local delay and EE based on the Poisson cluster process (PCP) were analyzed, while in [7] the closed-form expressions of EE and local delay were derived under the random discontinuous transmission (DTX) scheme. The effects of user mobility on local delay and EE were investigated based on the Poisson point process (PPP) and PCP in [8]. Ref. [9] analyzed the local delay and spectral efficiency for a full-duplex system. All these studies are for DL and have not considered the BS association issue. They have, however, inspired us to examine the local delay and EE for UL in this paper, especially in the context of CUDA and DUDA.

Some results on the DUDA mode have already been published in the literature. For example, a joint time division duplexing (TDD) and DUDA statistical model was applied in [10] to derive analytical expressions for the signal to interference ratio (SIR) and capacity, which shows that the DUDA mode offers a higher gain for both DL and UL. In [11], the coverage performance and spectral efficiency improve significantly by using Matern cluster process in DUDA under aerial-terrestrial heterogeneous cellular networks. Ref. [12] used a stochastic geometry based model to assess the improvement brought by the DUDA mode to TDD HetNets. Moreover, Ref. [13] modeled a location-dependent per-mobile power state in UL transmission by applying fractional power control (FPC), and the final results revealed that the DUDA mode significantly outperforms the CUDA mode in terms of spectrum efficiency (SE). Finally, the coverage and average throughput of HetNets under DUDA/CUDA mode was analyzed in [14]. However, none of the above studies on DUDA has considered the local delay, which, as mentioned earlier, is a key performance indicator in 5G and 6G networks. This forms the motivation of this paper.

1.2 Contributions

The key intermediate metric of successful transmission probability (STP) for UL in the DUDA and CUDA modes has been derived by using the Laplace transform of aggregate interference and stochastic geometry theory. In addition, we have also examined the cell association probability and shown that the load becomes more balanced under DUDA than under CUDA.

The local delay and EE for UL are then analyzed with respect to the key system parameters such as SIR threshold, ratio of SBS and MBS densities, and FPC factor, showing that under CUDA the FPC strategy leads to the decrease of both local delay (desired) and EE (sacrificed), while under DUDA, the local delay can remain low without FPC and sacrificing the EE.

Numerical results show that, for the same parameters, the local delay is lower and the EE is higher under DUDA than under CUDA. The DUDA mode significantly outperforms the CUDA mode in terms of UL performance.

2 System model and analysis

Consider a K-tier HetNet, where each tier's BS locations follow an independent homogeneous PPP Φ_i $(i = 1, 2, \ldots, K)$ and P_i, λ_i, and α_i are the transmit power, deployment density, and path loss exponent of each tier, respectively. The total BSs density is therefore $\lambda_{\text{tot}} = \sum_{i=1}^{K} \lambda_i$. We assume the standard path loss model $l_i(x) = hx^{-\alpha_i}$ $(\alpha_i > 2)$, and denote $\mathcal{K} = \{1, 2, \ldots, K\}$, where h is the fading channel's power coefficient and follows an exponential distribution of unit mean (i.e., Rayleigh fading). This paper focuses on the uplink transmission, where UEs also follow an independent PPP Φ_U with density λ_U. In addition, full frequency reuse is applied at every BS.

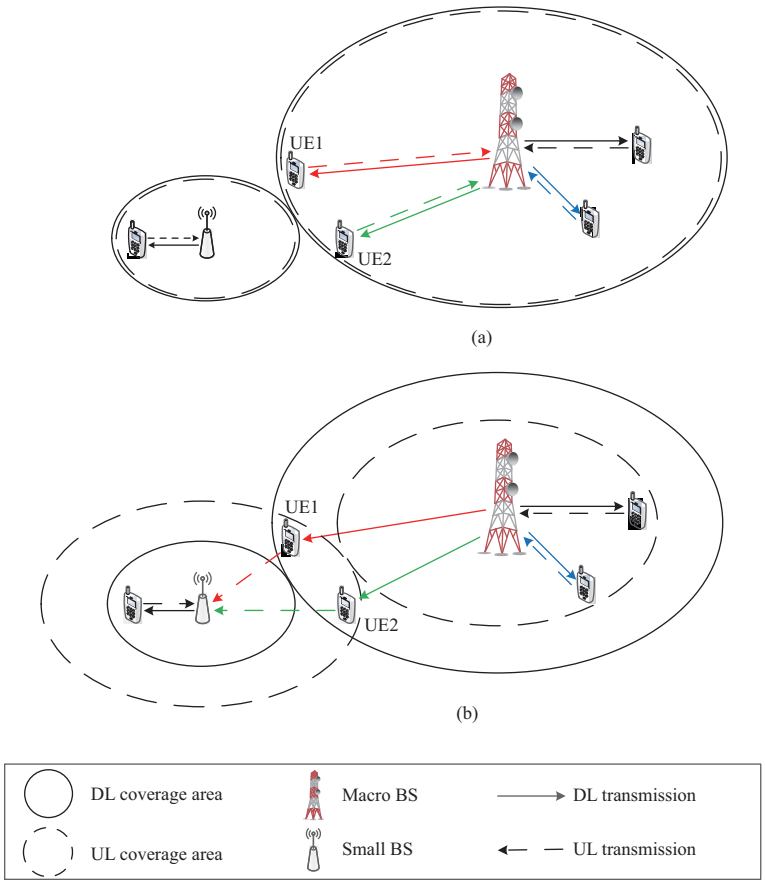

Figure 1 (Color online) Illustration of UL/DL transmission under (a) CUDA and (b) DUDA.

2.1 DUDA and CUDA modes

In the traditional association mode CUDA, a UE is associated with the same BS with the strongest DL received power for both UL and DL. In the DUDA mode under consideration in this paper[1)], however, a UE is associated with the geometrically nearest BS for UL transmission, while applying the same rule for DL as in the CUDA mode. Figure 1 illustrates the difference between the CUDA mode and the DUDA mode. In the CUDA mode, because of the higher transmission power of the macro BSs, its coverage area is much larger than that of the small BSs. Although UE1 and UE2 at the cell edge are closer to the small BS, they may still be associated with the macro BS. In contrast, in the DUDA mode, UE1 and UE2 may well be associated with the nearest small BS for UL. Obviously, compared with the CUDA mode, the DUDA mode can balance the load of the base stations and reduce the uplink distance of cell edge UEs.

2.1.1 *Probability of cell association*

The loads of the BSs can be represented by the probability of cell association, which are different in the CUDA and DUDA modes. In the DUDA mode, as the uplink the UEs are associated with the geometrically nearest BSs, the uplink association probability is related to BS density rather than DL transmit power and path loss of BSs. In the rest of this paper, we use superscript 'D' and 'C' to denote the DUDA and CUDA mode, respectively. As BS uplink association in DUDA is based on UE-BS distance and the BSs of each tier follow a PPP, the probability that a UE is associated with an ith tier BS in the DUDA mode is proportional to the corresponding BS density and is therefore given by

$$A_i^{\mathrm{D}} = \frac{\lambda_i}{\lambda_{\mathrm{tot}}}, \tag{1}$$

1) We consider the scenario of UCBC here (e.g., machine vision in IoT systems). The DUDA mode can also mean that, in other scenarios (e.g., video streaming), the UE is associated with the nearest BS for the DL, which is of course important, but not the focus of this paper.

where, as mentioned above, λ_i is the BS density for Tier i and $\lambda_{\text{tot}} = \sum_{j=1}^{K} \lambda_j$ is the total BS density of all tiers.

In the CUDA mode, on the other hand, the probability that a UE is associated with an ith tier BS is based on maximum received power (with small scale fading averaged out, hence the shortest distance), as given by [15]

$$A_i^{\text{C}} = 2\pi\lambda_i \int_0^{\infty} r\mathrm{e}^{-\pi \sum_{j=1, j\neq i}^{\mathcal{K}} \lambda_j (\frac{P_j}{P_i})^{\frac{2}{\alpha_j}} r^{\frac{2\alpha_i}{\alpha_j}} - \lambda_i \pi r^2} \mathrm{d}r. \tag{2}$$

2.1.2 *Probability density function (PDF) of association distance*

Let R_i be the distance between a UE and its serving BS in the ith tier. In the DUDA mode, since a UE selects the serving BS based on the distance between itself and the BSs, the PDF of R_i^{D} is only related to the densities of BSs and is given by [13]

$$f_{R_i^{\text{D}}}(r) = 2\pi\lambda_{\text{tot}} r\mathrm{e}^{-\pi\lambda_{\text{tot}} r^2}, \tag{3}$$

while the PDF of R_i^{C} in the CUDA mode is [15]

$$f_{R_i^{\text{C}}}(r) = \frac{2\pi\lambda_i}{A_i^{\text{C}}} r\mathrm{e}^{-\pi \sum_{j\in\mathcal{K}} \lambda_j (P_j/P_i)^{2/\alpha_j} r^{2\alpha_i/\alpha_j}}. \tag{4}$$

2.2 Power model

In the traditional CUDA mode, load imbalance will cause the distance from the edge UEs to the MBSs to be greater than those of the other UEs, which seriously affects the communications quality and degrades the transmission performance of UL. FPC strategies are normally applied to solve such a problem [13]. In this paper, the transmit power of a UE associated with an ith tier BS by using FPC is defined as $P_{T\_u} = P_0 r^{\alpha_i \beta}$, where P_0 is the baseline transmit power (a constant), r the distance between a UE and its serving BS, α_i the path loss factor of the ith tier, and $\beta \in [0, 1]$ the FPC compensation factor to represent the degree of path loss compensation. When a UE is far away from the serving BS, the UE can increase its transmit power (i.e., the β value) to improve the STP, subject to the maximum transmit power [16].

As a result, the total power consumption of a UE can be modeled as $P_{\text{tot}} = P_c + P_{T\_u}/\rho$, where P_c is the circuit power consumption and ρ is the power amplifier efficiency.

2.3 Average STP

This paper focuses on the uplink transmission and assumes that a typical BS in the ith tier is located at the origin, and the distance between a UE and the typical BS is r. According to the FPC strategy and the standard path loss model, the received power at the typical BS from the UE is $P_{i\_R\_u}(r) = P_0 h r^{\alpha_i(\beta-1)}$. Given the very high density of BS and UE in UDN, the interferences received from other UEs are much larger than the noise, and the noise is neglected from now on. The received SIR at the typical BS in the ith tier is therefore given by

$$\text{SIR}_i = \frac{P_0 h r^{\alpha_i(\beta-1)}}{\sum_{j\in\mathcal{K}, j\neq i} I_j}, \tag{5}$$

where I_j is the interference from other UEs in the jth tier. We assume a mobile environment in this paper, and as a result both the received signal power and the interference power in (5) vary from one transmission slot to the next due to user mobility (i.e., time-variant channel fading coefficients) as well as the transmission status change of other interfering users.

Assuming full frequency reuse, for an uplink between a UE and the typical BS there can be only one interfering UE at most from each of the other BSs at a time. The density of interfering UEs is hence equal to the density of BS λ_j in their respective associated tiers, and the expression of I_j is given by

$$I_j = \sum_{v\in\Phi_j} P_0 h {R_v}^{\alpha_j \beta} {y_v}^{-\alpha_i}, \tag{6}$$

where R_v is the distance from the interfering UE v to its serving BS and y_v is the distance between the UE v and the typical BS at the origin.

When the received SIR at a BS is larger than its pre-set threshold θ, it constitutes a successful transmission. The probability of such an event is termed STP in the literature. Therefore, the average STP in the ith tier can be defined as $\psi_i = \Pr[\mathrm{SIR}_i > \theta]$, and can be determined as follows.

Lemma 1. The average STP for a typical BS in the ith tier for UL under DUDA and CUDA is, respectively, given by

$$\psi_i^{\mathrm{D}} = \int_0^{\infty} 2\pi\lambda_{\mathrm{tot}} r \mathrm{e}^{-\pi\lambda_{\mathrm{tot}} r^2} \prod_{j\in\mathcal{K}} L_{I_j}^{\mathrm{D}}(z)\mathrm{d}r, \tag{7}$$

and

$$\psi_i^{\mathrm{C}} = \int_0^{\infty} \frac{2\pi\lambda_i}{A_i^{\mathrm{C}}} r \mathrm{e}^{-\pi\sum_{j\in\mathcal{K}}\lambda_j(\frac{P_j}{P_i})^{\frac{2}{\alpha_j}} r^{\frac{2\alpha_i}{\alpha_j}}} \prod_{j\in\mathcal{K}} L_{I_j}^{\mathrm{C}}(z)\mathrm{d}r, \tag{8}$$

where $z = \theta P_0^{-1} r^{\alpha_i(1-\beta)}$, and the Laplace function of interference can be expressed by

$$L_{I_j}^{\mathrm{D}}(z) = \int_0^{\infty} \exp\left(\frac{2\lambda_j\pi^2}{-\alpha_i}\csc\left(\frac{2\pi}{\alpha_i}\right)\left(\frac{\theta R_v^{\frac{2\beta\alpha_j}{\alpha_i}}}{r^{2(\beta-1)}}\right)\right) \times 2\pi\lambda_{\mathrm{tot}} R_v \mathrm{e}^{-\pi\lambda_{\mathrm{tot}} {R_v}^2}\mathrm{d}R_v, \tag{9}$$

and

$$\begin{aligned} L_{I_j}^{\mathrm{C}}(z) = &\int_0^{\infty} \exp\left(\frac{2\lambda_j\pi^2}{-\alpha_i}\csc\left(\frac{2\pi}{\alpha_i}\right)\left(\frac{\theta R_v^{\frac{2\beta\alpha_j}{\alpha_i}}}{r^{2(\beta-1)}}\right)\right) \\ &\times \frac{2\pi\lambda_i R_v}{A_i^{\mathrm{C}}} \exp\left(-\pi\sum_{j\in\mathcal{K}}\lambda_j\left(\frac{P_j}{P_i}\right)^{\frac{2}{\alpha_j}} r^{\frac{2\alpha_i}{\alpha_j}}\right)\mathrm{d}R_v. \end{aligned} \tag{10}$$

Proof. See Appendix A.

2.4 Local delay and EE

Local delay. As in the literature, the local delay of UEs under both CUDA and DUDA is defined as the average number of time slots required to successfully transmit a packet across a wireless link. If a data packet cannot be successfully decoded at the BSs, i.e., the SIR is lower than the threshold, then the UEs will retransmit the data packet until the BSs can successfully decode. So the number of time slots for a successful transmission is a geometrically distributed random variable, and the average local delay (for CUDA or DUDA, as appropriate) in the ith tier can easily be calculated as

$$D_i = \frac{1}{\psi_i}. \tag{11}$$

According to the law of total probability, the average local delay for the UDN can be written by

$$D = \sum_{i\in\mathcal{K}} \frac{\mathcal{A}_i}{\psi_i}. \tag{12}$$

EE. The EE of HetNets in UL is defined as the ratio of a UE's average network throughput to a UE's average power consumption. The average throughput [7] in ith tier, i.e., the number of successfully transmitted nats per unit time per unit bandwidth, is given by

$$\tau_i = {D_i}^{-1}\ln(1+\theta). \tag{13}$$

The average power consumption of a UE in the ith tier under CUDA and DUDA can be derived as

$$P_{\mathrm{tot},i}^{\mathrm{C}} = P_c + \rho^{-1}\int_0^{\infty} P_0 r^{\alpha_i\beta} f_{R_i^{\mathrm{C}}}(r)\mathrm{d}r$$

Table 1 System parameters

Parameter	value
Macro BS density (m^{-2}), λ_1	$1/(\pi 500^2)$
Small BS density (m^{-2}), λ_2	$4/(\pi 500^2)$
BS path loss factor, α_i	4
Macro BS transmit power (dBm), P_1	43
Small BS transmit power (dBm), P_2	21
Received SIR threshold (dB), θ	5
PA efficiency, ρ	0.5
UE baseline transmit power (dBm), P_0	10
UE circuit power consumption (dBm), P_c	13

$$= P_c + \rho^{-1} \int_0^\infty \frac{2\pi\lambda_i P_0 r^{\alpha_i\beta}}{A_i^{\mathrm{C}}} r \mathrm{e}^{-\pi \sum_{j\in\mathcal{K}} \lambda_j (\frac{P_j}{P_i})^{\frac{2}{\alpha_j}} r^{\frac{2\alpha_i}{\alpha_j}}} \mathrm{d}r, \tag{14}$$

and

$$\begin{aligned} P_{\mathrm{tot},i}^{\mathrm{D}} &= P_c + \rho^{-1} \int_0^\infty P_0 r^{\alpha_i\beta} f_{R_i^{\mathrm{D}}}(r) \mathrm{d}r \\ &= P_c + \rho^{-1} \int_0^\infty 2\pi\lambda_{\mathrm{tot}} r P_0 r^{\alpha_i\beta} \mathrm{e}^{-\pi\lambda_{\mathrm{tot}} r^2} \mathrm{d}r. \end{aligned} \tag{15}$$

Hence, the EE (unit: nats/Joule/Hz) is given by

$$\eta_{\mathrm{EE}} = \sum_{i\in\mathcal{K}} \mathcal{A}_i \frac{\tau_i}{P_{\mathrm{tot},i}}. \tag{16}$$

3 Numerical and simulation results

In order to facilitate clear comparison, this paper only considers a two-tier HetNet, including macro cells (Tier m) and small cells (Tier s). Furthermore, the FPC factors of macro and small BSs have the same value. Table 1 lists all the main parameter values (similar to [17,18] where appropriate). Both analytical (numerical, "ana" in the figures) and simulation ("sim") results are presented and they match each other very well, verifying the theoretical derivations.

3.1 Load and distance balance

The BS load is reflected by the probability of it being associated by the UEs, which can be calculated by (1) and (2). On the other hand, the distance distribution of UEs reflects the association difference under CUDA and DUDA. The average association distance from UEs to BSs in the ith tier under CUDA and DUDA is, respectively, given by $\bar{d}_i^{\mathrm{D}} = \int_0^\infty r f_{R_i^{\mathrm{D}}}(r)\mathrm{d}r$ and $\bar{d}_i^{\mathrm{C}} = \int_0^\infty r f_{R_i^{\mathrm{C}}}(r)\mathrm{d}r$. Based on Table 1, these have been calculated numerically and shown in Table 2 below. Since the transmission power of the MBS is hundreds of times that of the SBS, the coverage area of the macro cell is far greater than that of the small cell under CUDA, causing severe load imbalance. Although the density of SBSs is larger than that of MBSs, under CUDA the association probability of the macro cells is still higher than that of the small cells. However, in the DUDA mode, the UEs are associated with the nearest base station for UL, so the association probability is only related to the base station density.

Table 2 indicates that the coverage area of the macro cells is much larger, and the distance between the UE located at the edge of a macro cell and the serving BSs is correspondingly larger too in CUDA, so is the average association distance. On the other hand, the DUDA mode has overcome these issues and can therefore balance the load much better.

3.2 Effect of SIR threshold

Figure 2 presents the theoretical and simulation results for the average local delay as a function of SIR threshold under CUDA and DUDA when $\beta = 0$ and $\alpha_i = \alpha$ ($i \in \mathcal{K}$). In this case, from (12), the local

Table 2 Comparison of load and association distance under two modes

Comparison item	Association probability		Average distance (m)	
	CUDA	DUDA	CUDA	DUDA
SBS	0.24	0.8	109	198
MBS	0.76	0.2	386	198
Average	–	–	319	198

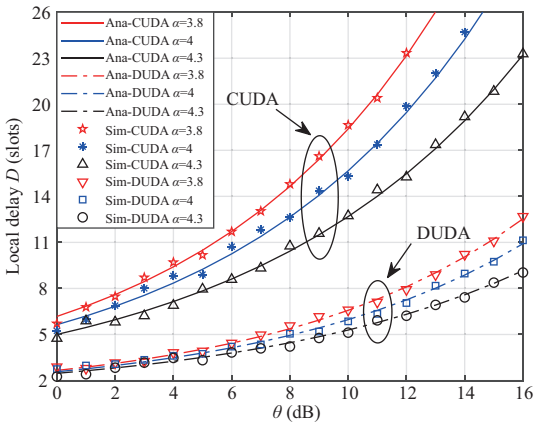

Figure 2 (Color online) Local delay D as a function of SIR threshold θ: theoretical and simulation results, where $\beta = 0$, and $\alpha_i = \alpha$ $(i \in \mathcal{K})$.

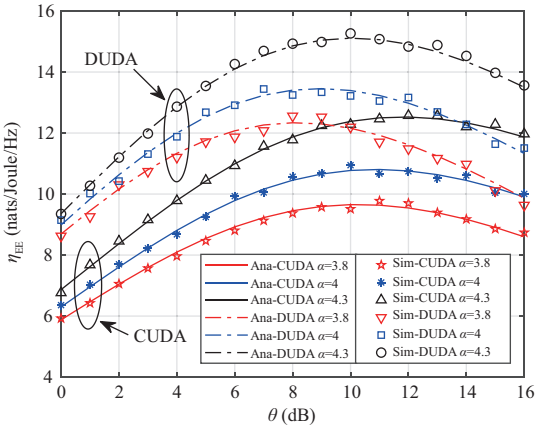

Figure 3 (Color online) Energy efficiency η_{EE} as a function of the SIR threshold θ under theoretical analysis and simulation, where $\beta = 0$, and $\alpha_i = \alpha$ $(i \in \mathcal{K})$.

delay can be simplified as

$$D^{\mathrm{C}} = \sum_{i \in K} \mathcal{A}_i^{\mathrm{C}} \left(\frac{(2\pi/\alpha)\csc(2\pi/\alpha)\,\theta^{2/\alpha}\,(\lambda_{\mathrm{tot}}/\lambda_i)}{\sum_{j \in K}(\lambda_j/\lambda_i)\,(P_j/P_i)^{2/\alpha}} + 1 \right), \tag{17}$$

and

$$D^{\mathrm{D}} = (2\pi/\alpha)\csc(2\pi/\alpha)\,\theta^{2/\alpha} + 1. \tag{18}$$

The above simplified formulas explain theoretically why the average local delay grows exponentially with the SIR threshold. Physically, the increase of threshold reduces the STP, which leads to a growth in local delay. From Figure 2, we can see that the local delay under DUDA is 50%–60% lower than that under CUDA, and the disparities between the two modes widen with the increasing θ. The reason is again that the DUDA mode can balance the load of BSs and improve the STP for UL.

Figure 3 illustrates the EE as a function of the SIR threshold under theoretical analysis and simulation when $\beta = 0$ and $\alpha_i = \alpha$ $(i \in \mathcal{K})$. Clearly, with the increase of the SIR threshold, the EE increases first and then decreases, i.e., there exist maximum values. From (17) and (18), the local delay is a monotonically increasing function with θ, and this makes sense physically: the higher the threshold is, the lower the STP becomes for a time slot, and the more retransmissions it takes to reach a successful transmission. Note, however, that the EE is determined by both local delay $D_i(\theta)$ and the transmission rate $\ln(1+\theta)$. With an increasing SIR threshold, the rate increases faster than the local delay, leading to an increasing EE, but up to a point, after which the increase in local delay overtakes the increase in rate (due to the logarithmic nature of rate), causing the EE to decrease. This means that a maximum EE can be achieved by selecting an optimal SIR threshold. Moreover, Figure 3 clearly shows that EE under DUDA is always superior to that under CUDA due to a lower local delay under DUDA.

Comparing Figures 2 and 3, we can observe a trade-off among rate, local delay, and EE. By suffering some more delay via increasing the SIR threshold, we can enjoy some higher transmission rate and EE — until a certain point (i.e., the optimal SIR threshold). After such a point, no benefit would ensue. The value of the optimal threshold depends upon network parameters such as path loss exponent, BS density, and association mode. Finally, over the whole range threshold, the DUDA mode greatly outperforms the CUDA mode.

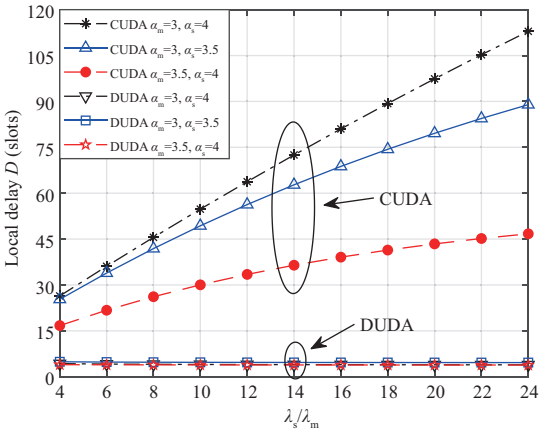

Figure 4 (Color online) Local delay D as a function of the ratio of SBS and MBS densities.

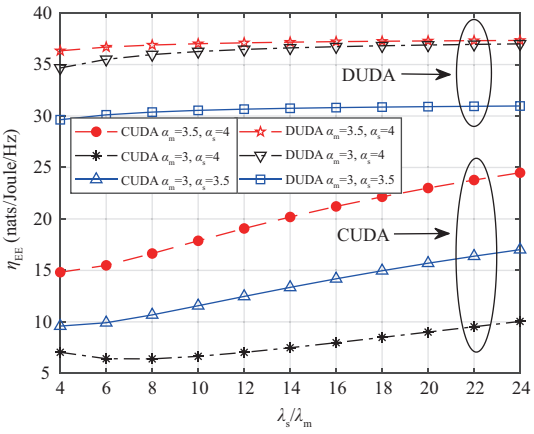

Figure 5 (Color online) Energy efficiency η_{EE} as a function of the ratio of SBS and MBS densities.

3.3 Effect of BS density

Figure 4 shows the local delay D as a function of the ratio of SBS and MBS densities, assuming that the density of MBSs remains static. For the same path loss exponent, the local delay under CUDA is higher and increases with the density of BSs in Tier s. The reason is that the interference under CUDA is larger than that under DUDA because of the load imbalance. For CUDA, the increase of local delay with the BS density in Tier s in Figure 4 is because, with the increase of SBS density, the increase of interference level at the MBSs is the dominating factor. For DUDA, on the other hand, the local delay has changed little with the increase of the BS density in Tier s, because the decrease of distance (i.e., the higher received power level at the SBSs) has canceled out the increase in interference level (leading to largely unchanged SIR).

In Figure 5, the EE under DUDA is higher than that under CUDA, due to the higher local delay under CUDA. Moreover, the EE in the DUDA mode remains high with the increase of the ratio of SBS and MBS densities. The results of Figures 4 and 5 reveal that the DUDA mode brings greater benefits compared with the CUDA mode in UDN (a key feature of 5G and B5G networks).

3.4 Effect of power compensation factor

In Figure 6, the average local delay under DUDA is always lower than that under CUDA and the highest decrease of 60% is achieved when $\beta = 0$ (i.e., when there is no FPC). Also, the EE and local delay under CUDA both decrease with the increase of the FPC factor. The reasons are as follows. When the distance between UEs and BSs becomes greater, the compensation power becomes higher. Under CUDA, the distance distribution of UEs is unbalanced so that the local delay can only be lowered by increasing the FPC factor (i.e., by increasing the UE transmit power, leading to a lower EE). As a result, under CUDA, although the EE decreases (therefore sacrificed), the FPC strategy will have to be adopted to achieve a lower local delay in UL transmission.

In the DUDA mode, however, the local delay always remains below that of the CUDA mode with the increase of FPC factor, but the best case happens when $\beta = 0$ (i.e., no FPC). This means that DUDA can reduce the local delay of the UL transmission without having to increase the UE transmit power (i.e., having to sacrifice the EE). This can be explained as follows. Since under DUDA the interference caused by the power compensation becomes larger under the balanced load, the FPC strategy does not help, and it even increases the local delay. Figure 6 indicates that the FPC strategy, which decreases the EE, should be abandoned under DUDA, and overall the DUDA mode significantly outperforms the CUDA mode in terms of both the local delay and EE.

4 Conclusion

In this paper, we have conducted a UL performance comparison study for the decoupled and coupled UL/DL modes (i.e., DUDA and CUDA modes) in ultra dense HetNets based on stochastic geometry.

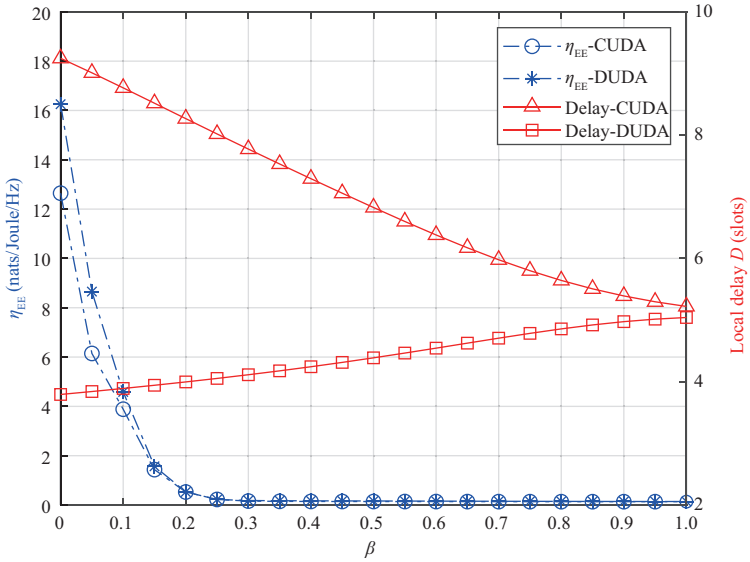

Figure 6 (Color online) Local delay D and energy efficiency η_{EE} as a function of power conpensation factor β.

A general K-tier HetNet model of UL communication is constructed for these two modes. We have investigated the local delay and EE with respect to the SIR threshold, ratio of SBS and MBS densities, and FPC factor, and demonstrated that the local delay is lower and the EE is always much higher in the DUDA mode than in the CUDA mode under the same system parameters. For future work, one option is to investigate the local delay and EE under clustered BS distributions.

Acknowledgements This work was supported in part by National Major Research and Development Program of China (Grant No. 2020YFB1805005) and Shenzhen Science and Technology Program (Grant Nos. KQTD20190929172545139, JCYJ201803061718-15699).

References

1 Guo W S, Liakata M, Mosquera G, et al. Big data methods for ultra dense-network deployment. In: Ultra-Dense Networks for 5G and Beyond: Modelling, Analysis, and Applications. Hoboken: Wiley, 2019. 203–230
2 Wong V W S, Schober R, Ng D W K. Key Technologies for 5G Wireless Systems. Cambridge: Cambridge University Press, 2017
3 Li P, Shen Y, Sahito F, et al. BS sleeping strategy for energy-delay tradeoff in wireless-backhauling UDN. Sci China Inf Sci, 2019, 62: 042303
4 Feng D Q, Lai L F, Luo J J, et al. Ultra-reliable and low-latency communications: applications, opportunities and challenges. Sci China Inf Sci, 2021, 64: 120301
5 Kose A, Han C, Foh C H, et al. Impact of mobility on communication latency and reliability in dense HetNets. In: Proceedings of the 89th Vehicular Technology Conference, 2019
6 Dong X J, Zheng F C, Zhu X, et al. On the local delay and energy efficiency of clustered HetNets. IEEE Trans Veh Technol, 2019, 68: 2987–2999
7 Nie W, Zhong Y, Zheng F C, et al. HetNets with random DTX scheme: local delay and energy efficiency. IEEE Trans Veh Technol, 2016, 65: 6601–6613
8 Dong X J, Zheng F C, Liu R X, et al. On the local delay and energy efficiency of HetNets with user mobility. In: Proceedings of IEEE International Conference on Communications, 2018
9 Marandi L, Naslcheraghi M, Ghorashi S A, et al. Delay analysis in full-duplex heterogeneous cellular networks. IEEE Trans Veh Technol, 2019, 68: 9713–9721
10 Lahad B, Ibrahim M, Lahoud S, et al. Joint modeling of TDD and decoupled uplink/downlink access in 5G HetNets with multiple small cells deployment. IEEE Trans Mobile Comput, 2021, 20: 2395–2411
11 Arif M, Wyne S, Navaie K, et al. Decoupled downlink and uplink access for aerial terrestrial heterogeneous cellular networks. IEEE Access, 2020, 8: 111172
12 Lahad B, Ibrahim M, Lahoud S, et al. Analytical evaluation of decoupled uplink and downlink access in TDD 5G HetNets. In: Proceedings of the 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications, 2018
13 Zhang L, Nie W L, Feng G, et al. Uplink performance improvement by decoupling uplink/downlink access in HetNets. IEEE Trans Veh Technol, 2017, 66: 6862–6876
14 Sial M N, Ahmed J. A realistic uplink-downlink coupled and decoupled user association technique for K-tier 5G HetNets. Arabian J Sci Eng, 2018, 44: 2185–2204
15 Jo H S, Sang Y J, Xia P, et al. Heterogeneous cellular networks with flexible cell association: a comprehensive downlink SINR analysis. IEEE Trans Wirel Commun, 2012, 11: 3484–3495
16 Zhang J, Xiang L, Ng D W K, et al. Energy efficiency evaluation of multi-tier cellular uplink transmission under maximum power constraint. IEEE Trans Wirel Commun, 2017, 16: 7092–7107
17 Novlan T D, Dhillon H S, Andrews J G. Analytical modeling of uplink cellular networks. IEEE Trans Wirel Commun, 2013, 12: 2669–2679

18 Nie W L, Zheng F C, Wang X M, et al. User-centric cross-tier base station clustering and cooperation in heterogeneous networks: rate improvement and energy saving. IEEE J Sel Areas Commun, 2016, 34: 1192–1206

Appendix A Proof of the Lemma 1

The average STP is related to the channel fading and UE distance. From the STP definition, the average STP of UL under DUDA is given by

$$\begin{aligned}\psi_i^{\mathrm{D}} &= \Pr(\mathrm{SIR}_i^{\mathrm{D}} > \theta)\\ &\overset{(a)}{=} \int_0^{\infty} \Pr\left(\frac{P_0 h r^{\alpha_i(\beta-1)}}{\sum_{j=1}^{K} I_j} > \theta\right) f_{R_i^{\mathrm{D}}}(r)\mathrm{d}r\\ &= \int_0^{\infty} \Pr\left(h > \frac{-\theta \sum_{j=1}^{K} I_j}{P_0 r^{\alpha_i(\beta-1)}}\right) f_{R_i^{\mathrm{D}}}(r)\mathrm{d}r\\ &\overset{(b)}{=} \int_0^{\infty} \prod_{j\in\mathcal{K}} L_{I_j}^{\mathrm{D}}(z) f_{R_i^{\mathrm{D}}}(r)\mathrm{d}r, \end{aligned} \tag{A1}$$

where (a) and (b) take expectation with respect to distance r and channel power fading coefficient h, respectively. Moreover, the PDF of h is $f_h(h) = \mathrm{e}^{-h}$. The Laplace function of interference in (19) under DUDA can then be derived by using the stochastic geometry theory:

$$\begin{aligned} L_{I_j}^{\mathrm{D}}(z) &= E_{I_j}\left[\mathrm{e}^{-zI_j}\right]\\ &= E_h\left[\int_0^{\infty} \exp\left(-z \sum_{v\in\Phi_j} P_0 h {R_v}^{\alpha_j\beta} {y_v}^{-\alpha_i}\right) f_{R_j^{\mathrm{D}}}(R_v)\mathrm{d}R_v\right]\\ &= \int_0^{\infty} \prod_{v\in\Phi_j} \frac{1}{1+zP_0{R_v}^{\alpha_j\beta} y_v^{-\alpha_i}} f_{R_j^{\mathrm{D}}}(R_v)\mathrm{d}R_v\\ &\overset{(c)}{=} \int_0^{\infty} \exp\left(-2\pi\lambda_j \int_0^{\infty}\left(1-\frac{1}{1+zP_0{R_v}^{\alpha_j\beta} y^{-\alpha_i}}\right) y\mathrm{d}y\right) \times f_{R_j^{\mathrm{D}}}(R_v)\mathrm{d}R_v\\ &= \int_0^{\infty} \exp\left(\frac{2\lambda_j\pi^2}{-\alpha_i}\csc\left(\frac{2\pi}{\alpha_i}\right)\left(\frac{\theta {R_v}^{\frac{2\beta\alpha_j}{\alpha_i}}}{r^{2(\beta-1)}}\right)\right) f_{R_j^{\mathrm{D}}}(R_v)\mathrm{d}R_v, \end{aligned}$$

where (c) follows from the probability generating functional (PGFL) of PPP[2)]. The average STP in the CUDA mode can be derived in a similar fashion and is omitted here for brevity.

2) Stoyan D, Kendall W, Mecke J. Stochastic Geometry and Its Applications. 2nd ed. Hoboken: Wiley, 1996.