
Article:
Li, Zhehan, Grace, David orcid.org/0000-0003-4493-7498 and Mitchell, Paul Daniel orcid.org/0000-0003-0714-2581 (2016) Traffic-aware cell management for green ultra-dense small cell networks. IEEE Transactions on Vehicular Technology. 7484757. pp. 2600-2614. ISSN 0018-9545

https://doi.org/10.1109/TVT.2016.2576404

Traffic-Aware Cell Management for Green Ultra-dense Small Cell Networks

Zhehan Li, David Grace, *Senior Member, IEEE,* and Paul Mitchell, *Senior Member, IEEE*

***Abstract*—To reduce the power consumption of 5G ultra-dense small cell networks, base stations can be switched to low power sleep modes when local traffic levels are low. In this paper, a novel sleep mode control algorithm is proposed to control such sleep modes. The algorithm innovates a concept called Traffic-Aware Cell Management (TACM). It involves cell division, cell death and cell migration to represent adaptations of networks, where state transitions of base stations are controlled. Direction of arrival is adopted for distributed decision making. The TACM algorithm aims at reducing the network power consumption while alleviating the impacts of applying sleep modes, such as mitigating system overheads and reducing user transmission power. The TACM algorithm is compared with a recent consolidated baseline scheme by simulation on networks with unbalanced traffic distributions and with base stations at random locations. In contrast, the TACM algorithm shows a significant improvement in mitigating system overheads due to no load information exchange overhead and up to 72 times less switching frequency. Up to 81% network power consumption can be reduced compared with the baseline scheme if considering high energy consumption of switching transient states. In addition, at a low traffic level, average uplink transmission power is reduced by 79% comparatively. Furthermore, the impact of important performance governing parameters of the TACM algorithm is analysed. The insensitivity to the estimation accuracy of direction of arrival is also demonstrated. The results show that the proposed TACM algorithm has a comprehensive advantage of power reduction and overhead mitigation over the baseline scheme.**

***Index Terms*—Energy saving, random topology, sleep modes, ultra-dense small cell networks**

I. Introduction

With the proliferation of the mobile device market, the dramatically increasing demand of mobile data services is imposing great pressure on existing wireless networks. As predicted in [1], global mobile data traffic is expected to grow at a compound annual growth rate of 61 percent from 2013 to 2018. One promising solution to cope with this trend is deployment of ultra-dense Small Cell Networks (SCNs) to increase system capacity in highly populated locations [2]. In SCNs, cell sizes are reduced to enhance spatial frequency reuse, meaning that there are higher densities of base stations which need to be lower power and cost less. Therefore, users can handover more frequently to obtain services from local base stations and use lower transmit power.

The authors are with the Department of Electronics, University of York, York, YO10 5DD, U.K.
E-mail: {zl837, david.grace, paul.mitchell}@york.ac.uk

Although small cell base stations are relatively low-power, the overall network power consumption is still considerable, given the rising quantity of base stations required. To reduce the operational expenditure (OPEX) of electricity consumption, operators are very interested in lowering the overall network power consumption while boosting the system capacity [3]. As a result, SCNs are not only required to be environmentally friendly, they can also be economically lucrative. To address the power consumption challenge, it is essential to consider various techniques such as energy efficient base station deployments, energy efficient protocols and architecture designs, opportunistic network access, smart grids [4], [5], etc. Another promising solution is the application of low-power sleep modes, which deactivate hardware modules within base stations for energy saving. As real traffic distributions are temporally and spatially variable in reality, it is possible to switch some Evolved Node Bs (eNBs) in a SCN to sleep modes where and when the local traffic levels are low. The former coverage areas of the sleeping eNBs are then compensated for by the other active eNBs.

A. Prior Work

Regarding the network components performing the computation, sleep mode control schemes can be categorised into centralised schemes and distributed schemes. When considering distributed solutions, there are also various approaches and different requirements about the infrastructure and the functionality of the sleep modes [6]. Activation and deactivation decisions of eNBs can either be made by eNBs themselves, their neighbour eNBs or their local User Equipment (UEs). Among these solutions, the design of sleep modes and the number of active hardware modules during sleep modes depend on whether sleeping eNBs are required to perform computation. We notice that some research has been carried out to suit the preferred distributed manner due to the inefficient management of centralised schemes. For example in [7], authors propose a distributed scheme that requires both active and sleeping base stations to receive load information from neighbour eNBs and control state transitions based on capacity thresholds. In [8], a similar algorithm is presented to control sleep modes in conventional cellular networks also needs load information exchange. In addition to the exchanged load information, coverage information is also required for decision making in a sleep mode control algorithm in [9]. A more consolidated scheme is proposed and analysed in [10], where an eNB considers both the load information of its neighbour eNBs and the handover information of the associated UEs to make switch-off decisions. Activation decisions are made by neighbour eNBs based on their load

conditions. However, all these schemes mentioned need some form of information exchange among neighbour eNBs, such as load information, which yields additional system overheads. The system overheads are not accordingly evaluated and not intended to be reduced.

In addition to the extra system overheads, the other cost is on the UE side, which is the increase of the average uplink transmission power. With the application of sleep modes, there are fewer active eNBs in a network so that the average distance between UEs and their corresponding associated eNBs is longer. This drives UEs to increase the transmission power to compensate for the path loss, which reduces their battery lives. Although the SCNs with much smaller cell sizes have reduced the transmission power of UEs dramatically, and the increase of the UE transmission power is negligible compared with the significant power reduction at the eNB side, it is essential to minimise this effect by making the active eNBs closer to UEs overall when designing a sleep mode control algorithm. However, this is rarely considered in prior work.

The aforementioned schemes have mostly been investigated on networks with a grid topology or other specific layouts. These network models are too idealised, especially for future self-organised SCNs, where more random placement should be introduced to verify the effectiveness of a specific sleep mode control scheme. On the other hand, based on the random placement of eNBs, crucial system configurations for sleep mode control such as Neighbour Cell List (NCL) should be initially self-configured by the eNBs autonomously when there is a change in the system, e.g. deployment of new eNBs. This is especially important for sleep mode control to be realised in a distributed way. As part of self-organisation, self-configuration is used to avoid manual configuration of a large number of eNBs in a SCN, which becomes more and more time-consuming and costly. However, the importance of NCL set-up for sleep mode control in SCNs of random topologies has only drawn limited attention in recent research.

B. Main Contributions

When designing a sleep mode control algorithm, the following issues have to be addressed:

- Determination of the best time to switch on-off
- Reduction of associated system overheads including information exchange overhead and switching overhead
- Placing active eNBs where the increase of the UE uplink transmission power can be minimised

The main contributions of the paper are summarised as:

- *DOA Utilisation:* In the novel sleep mode control algorithm proposed in this paper, Direction of Arrival (DOA) information of the UEs is used as one of the decision making criteria. DOA estimation has been studied over the past few decades and is widely used for smart antenna beamforming, which is expected to have more performance benefit in future networks. With the spread of multiple antenna systems, the availability of DOA information can be guaranteed and be used for sleep mode control in the SCNs investigated. Conventionally, DOA can be estimated by uniform sensor arrays as well as either isotropic or directional antenna arrays using various algorithms, e.g. MUSIC [11], which is compared and analysed with other algorithms in [12]. The experimental work done is as rich as the theoretical algorithms. For instance in [13], MUSIC is evaluated with an antenna array of 6 parasitic elements, which achieves an error of less than one degree under the conditions of a 20 dB signal-to-noise ratio (SNR), 1000 snapshots and a 500 kHz sampling rate. In spite of this, attention still has to be paid to the resolution property of DOA estimation which depends upon the SNR, the number of elements in the array, the number of snapshots, the array geometry and so on [12]. Although, DOA estimation is more effective for line-of-sight (LOS) paths, research is also ongoing to deal with the influence of non-line-of-sight (NLOS) paths [14], [15]. In the SCN scenario investigated in this paper, the transmission path between a UE and an eNB is usually short and is more likely to be LOS so that DOA estimation can potentially have high accuracy. The DOA estimation error is also quantified to demonstrate its effects on the algorithm performance are limited.
- *Algorithm Design:* To exempt the load information exchange between neighbour eNBs, DOA is used to make eNBs aware of the directions of areas with high traffic levels. In the algorithm, DOAs are classified, and distributed weight memories are created at all the eNBs. The memory of an eNB is cumulatively updated through its past experience. An activation decision of a sleeping eNB is made by an active eNB only based on its own historical DOA perception and its own average load estimation, which does not require the load information of neighbour eNBs and therefore mitigates the information exchange overhead. Deactivation decisions are also made by eNBs themselves without the necessity of considering the loads of neighbour eNBs. A hysteresis time is applied to ensure that cursory decisions are not made based on temporarily fluctuating traffic levels, which also reduces the switch on-off frequency. The algorithm enables actions to be taken according to the locations of hotspot areas where eNBs are more likely to be activated, reducing the average UE connecting distance and the average uplink transmission power as a result. With the joint efforts of all eNBs in a SCN, the SCN can therefore manage its cell configurations according to the varying environment. A comparison of the proposed algorithm is made to a recent consolidated scheme introduced in [10]. The virtues of our design over the comparison scheme, which are reducing system overheads and uplink transmission power, are examined. The potential reduction of the network power consumption contributed by the switching transient states is also evaluated.
- *Practical Aspects:* The algorithm is based on our early work [16] and is significantly generalised to suit the future SCNs with random placement of eNBs, based on which the NCLs are carefully defined. A mathematical model capturing this feature is adopted to model the network architecture. The logical diagram of the network structure is shown as in Fig. 1, where the ultra-dense

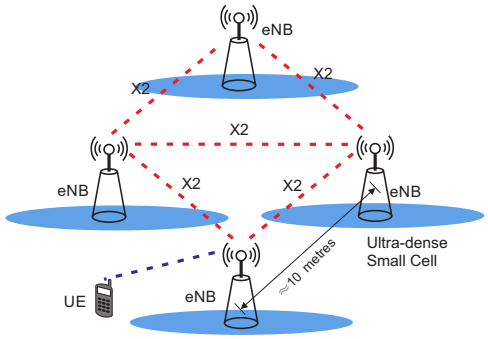


Fig. 1. The investigated ultra-dense SCN structure.

SCNs are of the distributed structure and X2 air interface among eNBs is defined for sending messages. Furthermore, a nonuniform UE distribution model is chosen to reflect the existence of hotspots areas. Beyond the analysis in [16], the system overheads of implementing a sleep mode algorithm and the increase of UE transmission power are extensively evaluated. Owing to fewer active eNBs in a SCN, the impact on Quality of Service (QoS) of various traffic levels is considered. As reducing network power consumption and maintaining global QoS delivered by a SCN become two conflicting goals, the trade-offs are also analysed.

The remainder of the paper is organised as follows. In Section II, the details of the proposed Traffic-Aware Cell Management algorithm including DOA classification, weight update, cell management are described. The system models and assumptions are presented and illustrated in Section III. Section IV presents the simulation set-up and results of the proposed algorithm and the comparative schemes followed by explanation and analysis. Finally, conclusions are drawn in Section V.

II. Traffic-Aware Cell Management

The key point of designing a sleep mode control algorithm to achieve energy saving is to ensure that a SCN always provides just enough resources in appropriate areas in pace with the temporal and spatial traffic variations. Therefore, it is vital to only locate active eNBs delivering services at places with high service demands, which also prevents UEs from connecting far away causing severe interference and wasting transmission power. While meeting this requirement, information exchange should be reduced and the algorithm should be insensitive to temporary short-time traffic variations.

The Traffic-Aware Cell Management (TACM) algorithm proposed here is dedicated to achieving the above objectives. In the TACM algorithm, an eNB observes and updates its memory, based on which actions are taken from the action set consisting of cell division, cell migration and cell death. The details of the TACM algorithm are introduced in this section.

A. Observation

This observation process requires only active eNBs to estimate the DOAs of their admitted UEs. When UEs are preparing for file transmissions, DOA estimations are performed based on control signals containing different UE identifications. As noted earlier, DOA estimation is assumed to be ideal. The example in Fig. 2 (a) assumes that the eNB estimates the DOAs of all the UEs, each of which has one file arrival. The counted number of DOAs with 1 degree resolution as an example is plotted in Fig. 2 (b). It is obvious that the estimation of DOAs yields an angular distribution of the traffic perceived by the eNB and indicates the directions of higher traffic levels.

B. Memory Update

The memory update process is executed every time a new DOA estimation is performed. The DOA is further classified and the corresponding metrics are updated. In this process, an eNB can determine the directions of the areas with high traffic levels and take them as part of the decision making criteria.

The eNBs in a SCN network are denoted by a set $\mathcal{B} = \{B_1, B_2, \ldots, B_i\}$ ($i \in \mathbb{N}^*$) and the neighbour eNBs of a eNB B_i are represented by a set $\mathcal{N}_i = \{N_{i,1}, N_{i,2}, \ldots, N_{i,j}\}$, where $j \in \mathbb{N}^*$. To quantify the captured DOAs for each B_i, a weight memory ($\mathcal{W}_i = \{W_{i,1}, W_{i,2}, \ldots, W_{i,j}\}$) is needed and each memory unit in the weight memory is mapped to one of the neighbour eNBs $N_{i,j} \in \mathcal{N}_i$ of B_i:

$$\forall W_{i,j} \in \mathcal{W}_i : W_{i,j} \mapsto N_{i,j}. \tag{1}$$

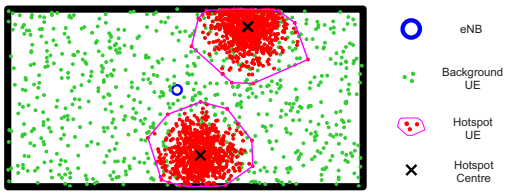
When B_i creates its own NCL, the angle of each $N_{i,j}$ relative to B_i, $D_{i,j} \in [0, 360)$ is acquired and recorded in degrees. Then for DOA classification, an interval $\mathcal{I}_{i,j}$ for each $W_{i,j}$ representing a sector with its angular bisector threading $N_{i,j}$ is generated and its centrally symmetric sector represented by an interval $\mathcal{I}^*_{i,j}$ is created as well:

$$\mathcal{I}_{i,j} = (D_{i,j} - \frac{\Delta\theta}{2}, D_{i,j} + \frac{\Delta\theta}{2}), \tag{2}$$

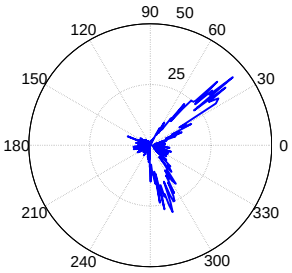
$$\mathcal{I}^*_{i,j} = (D_{i,j} - \frac{\Delta\theta}{2} + 180, D_{i,j} + \frac{\Delta\theta}{2} + 180), \tag{3}$$

where $\Delta\theta \in (0, 180]$ is a predefined parameter in degree representing the interval length (the effects of varying this parameter are further analysed in Section IV D3). In this way, B_i updates a weight $W_{i,j}$ for every neighbour eNB $N_{i,j}$, reflecting the traffic level in its representing direction relative to B_i. Like the example given in Fig. 3, B_3 denotes its first neighbour eNB B_2 as $N_{3,1}$ and creates an interval $\mathcal{I}_{3,1}$ for it. The angle of B_2 relative to B_3 is 354°, then $\mathcal{I}_{3,1}$ becomes (324°, 384°) according to Equation 2 with $\mathcal{I}^*_{3,1}$ being (504°, 564°) according to Equation 3 if $\frac{\Delta\theta}{2}$ is 30°.

When a UE is admitted by B_i for data transmission, the memory size ψ_i, defined as the total entries of UEs, is increased by 1 regardless of their DOAs. Meanwhile, the estimated DOA of the UE is represented by a relative angle $\alpha \in [0, 360)$. Afterwards, the entry of this UE is classified



(a)



(b)

Fig. 2. DOA estimation. (a) The eNB estimates the DOAs of all the UEs; (b) The angular distribution of the perceived traffic. The traffic level is the number of DOAs in the corresponding directions with 1 degree resolution.

into the intervals which it belongs to and the corresponding weights are updated respectively:

$$\psi_i^{(t+1)} = \psi_i^{(t)} + 1, \tag{4}$$

$$\exists k \in \{0, \pm 1\} : (\alpha + 360k) \in \mathcal{I}_{i,j} \Rightarrow W_{i,j}^{(t+1)} = W_{i,j}^{(t)} + 1, \tag{5a}$$

$$\exists k \in \{0, \pm 1\} : (\alpha + 360k) \in \mathcal{I}_{i,j}^{*} \Rightarrow W_{i,j}^{(t+1)} = W_{i,j}^{(t)} - 1, \tag{5b}$$

$$\forall k \in \{0, \pm 1\} : (\alpha + 360k) \notin (\mathcal{I}_{i,j} \cup \mathcal{I}_{i,j}^{*}) \Rightarrow W_{i,j}^{(t+1)} = W_{i,j}^{(t)}, \tag{5c}$$

where the superscript (t) is the prior time step of $(t+1)$. The $360k$ terms in Equation 5 exist in case that the ranges of $\mathcal{I}_{i,j}$ or $\mathcal{I}_{i,j}^{*}$ exceed 360° or fall below 0° ($\alpha + 360k$ is equal to α). As a result, the corresponding weight is increased by 1 if a DOA is classified in $\mathcal{I}_{i,j}$ or decreased by 1 if a DOA is classified in $\mathcal{I}_{i,j}^{*}$. Multiple weights may be updated for one entry if it is classified into multiple overlapping intervals.

In one case, the weight $W_{i,j}$ is relatively high if more DOAs are classified into the interval $\mathcal{I}_{i,j}$ indicating that the traffic level in the corresponding direction range $\mathcal{I}_{i,j}$ is high. In the other case, fewer DOAs classified into the interval $\mathcal{I}_{i,j}^{*}$ indicate less traffic offered in the opposite direction range of the one pointing to $N_{i,j}$, which also yields a higher weight $W_{i,j}$. Implicitly, as active eNBs should stay close to where there are high traffic levels, or stay far away from areas with low traffic levels, a relatively higher weight $W_{i,j}$ always implies that the corresponding direction range $\mathcal{I}_{i,j}$ points to an area where active eNBs should be located.

An example of DOA estimation and weight update is given in Fig. 3. The newly arriving UE_1, UE_3 and UE_4 are classified to $\mathcal{I}_{1,1}$, $\mathcal{I}_{3,1}$ and $\mathcal{I}_{3,1}^{*}$ according to the relative angles of the DOAs and the reference zero angles. This is processed based on Equation 5. After that at time step $(t+1)$, the weight memory unit $W_{1,1}^{(t)}$ is increased by 1. $W_{3,1}^{(t)}$ remains the same because of one step of increase contributed by UE_3 and one step of decrease contributed by UE_4. However, UE_2 is not classified into any intervals and therefore no corresponding weights are changed for this entry. On the other hand, $\psi_1^{(t)}$ and $\psi_3^{(t)}$ are both increased by 2 at time step $(t+1)$ according to Equation 4.

For a certain eNB, its weights should be normalised to the overall traffic supported by it if it is required to know the relative traffic levels (irrespective of the absolute traffic level) in different direction ranges pointing to its neighbour eNBs. For this purpose, weight bias $\beta_{i,j}$ is defined as

$$\beta_{i,j}^{(t)} = \frac{W_{i,j}^{(t)}}{\psi_i^{(t)}}. \tag{6}$$

It is easy to find out that $\beta_{i,j}$ is also mapped to $N_{i,j}$ because of the association with $W_{i,j}$. $\beta_{i,j}$ also stands as a direction indicator of hotspot areas like $W_{i,j}$, but with normalisation reflecting the relative traffic level to a direction.

From the previous explanation, it is known that the direction range $\mathcal{I}_{i,j}$ with a corresponding higher weight $W_{i,j}$ points to the area where more traffic arrives. By normalising $W_{i,j}$ to the memory size ψ_i, the relative trends of service demands in areas in different directions are quantified by $\beta_{i,j}$. A higher value of bias signifies the direction where the cell supporter should be.

As the state transitions of eNBs change the coverage areas of the active eNBs, this may give rise to inconsistency between the historical memory and the existing traffic loaded by an eNB. To deal with this, the memory of an eNB should be updated after the action has been taken, making the memory more reliable for the next action to be taken. This is described in the following subsections.

C. Cell Division

Cell division helps an overloaded eNB activate one of its sleeping neighbour eNBs in the desired direction. It is triggered based on the local traffic levels, after which the updated memories help eNBs determine which neighbour eNB to be activated. Whenever the load L_i of B_i reaches a division threshold $DivL_{\text{th}}$ for a continuous hysteresis duration H_{div} timed by the hysteresis timer, cell division is triggered. The hysteresis duration H_{div} is to prevent B_i from being perturbed by temporary short-term traffic variations which may easily make the load of B_i exceed the division threshold, $DivL_{\text{th}}$. This reduces the ping-pong effect and therefore reduces the overall switching on-off frequency. As mentioned before, through cumulative DOA estimation and weight update from the past experience of B_i, a higher value of $W_{i,j}$ indicates that there is a high traffic level in the direction to $N_{i,j}$. After cell division is triggered, B_i selects and activates only one $N_{i,j} \in \mathcal{N}_i^{\text{off}}$ mapped to the largest weight, where $\mathcal{N}_i^{\text{off}}$ is the neighbour sleeping eNB set of B_i. As the NCL is already initialised after deployment, B_i sends an activation message via the air interface to its selected neighbouring eNB. A sleeping neighbour eNB has a minimum message receiving capability for control purposes and is activated when receiving

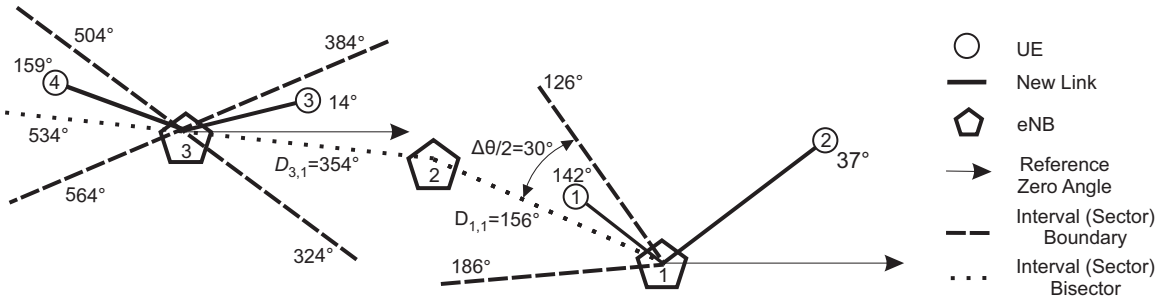


Fig. 3. An example of interval creation, DOA classification and weight update. B_2 is denoted as $N_{1,1}$ and $N_{3,1}$, indicating that it is the first neighbour eNB of B_1 and B_3, respectively. The intervals for neighbour eNBs are built. For instance, the angle of B_2 relative to B_3 is 354°, then $\mathcal{I}_{3,1}$ is (324°, 384°) according to Equation 2 with $\mathcal{I}^*_{3,1}$ being (504°, 564°) according to Equation 3 assuming $\frac{\Delta\theta}{2}$ to be 30°. UE_1 and UE_2 are admitted by B_1 while UE_3 and UE_4 are admitted by B_2. UE_1 is classified to $\mathcal{I}_{1,1}$ of B_1 while UE_3 and UE_4 are classified to $\mathcal{I}_{3,1}$ and $\mathcal{I}^*_{3,1}$ according to the relative angles of the DOAs and the reference zero angles according to Equation 5. Equation 5 is used for classification. Such as for UE_3 with 14° DOA, it is classified to $\mathcal{I}_{3,1}$, (324°, 384°), as 14 plus 360 is 374 with $k = 1$ from Equation 5b.

such messages. It is worth mentioning that such messages are only transmitted when an activation occurs and do not require the eNBs to frequently communicate with each other. Then, the activated eNB recovers normal operation and starts to broadcast reference signals to serve UEs again.

To provide sufficient resources by having more cells in a SCN, cell division occurs by activating one of the neighbour sleeping eNBs of an overloaded eNB. Then, B_i resets every element in $\mathcal{W}_i$ and ψ_i to zero, and empties the hysteresis timer. As shown in the example in Fig. 4 (a), B_1 is overloaded by its local traffic load and perceives a high traffic level from its right. It activates its right neighbour eNB B_3 to support the high traffic level. Consequently, B_1 and B_3 adjust their coverage areas following a certain UE association policy and more resources are delivered to the previous overloaded local area.

D. Cell Migration

Due to the lack of logical interfaces between non-neighbour eNBs in a distributed SCN, only neighbour eNBs can be activated in cell division to satisfy the neighbour areas in demand. Therefore, eNBs can be activated in the required directions but not at the expected distances. Moreover, SCNs only have spatial traffic variations sometimes so that only the locations of active eNBs have to be adjusted and the overall number of active eNBs may remain the same. By guaranteeing the correct locations of active eNBs, a combination of both desirable directions and distances, the average distance between UEs and eNBs and the resulting average UE uplink transmission power can be minimised.

Motivated by the above requirements, cell migration is designed. In cell migration, if an eNB perceives a relatively high traffic level in the direction interval mapped to a neighbour eNB, it will activate this neighbour eNB and switch off itself. In this way, the newly activated eNB is located nearer to the area with a higher traffic level. The migration process can be triggered taking $\beta_{i,j}$ into consideration when the local traffic level in a cell is not high, but of a biased angular distribution. More specifically, migration of a cell supported by B_i happens when there is a $\beta_{i,j}$ exceeding a bias threshold β_{th} and $N_{i,j} \in \mathcal{N}_i^{\text{off}}$. Meanwhile, the memory size ψ_i should be over a threshold ψ_{th} ensuring that a sufficiently large number of DOAs and a reliable past experience for reasoning are captured. After satisfying the conditions, the corresponding mapped neighbour eNB $N_{i,j}$ is activated by B_i. If there are multiple $\beta_{i,j}$ exceeding the bias threshold β_{th} with their mapped eNBs currently in sleep modes, B_i chooses the neighbour eNB with the largest $\beta_{i,j}$. Then B_i resets every element in $\mathcal{W}_i$ and ψ_i, and the hysteresis timer to zero. B_i switches to sleep mode afterwards.

After a migration step, the newly activated eNB supporting the cell may perceive less biased traffic because of the changes in DOAs. In another case, the migration process may have several steps until a closest eNB relative to the hotspot area is activated. However, finally, due to less biased traffic perceived by the newly activated eNB, the cell migration chain will stop and the network becomes stable again. In the example given in Fig. 4 (b), a weight bias of B_3, which is mapped to its neighbour eNB B_1, is over the threshold, indicating the hotspot area is to its left. B_3 activates B_1 and deactivates itself, making the active eNBs located nearer to the area with a higher traffic level. The cell migration and the changed DOAs of UE_2 and UE_4 reduce the bias of the traffic distribution relative to the new cell supporter, B_1. This forms a feedback loop from the network adaptations and the gradual migration process is terminated until angularly uniform traffic is perceived by the new cell supporter.

E. Cell Death

To provide just enough resources in a SCN, fewer eNBs should be active when the overall traffic level is low. Cell death enables eNBs with low loads to be switched off in order to reduce the overall network power consumption. As an inverse process of cell division, when the load L_i of B_i is not greater than a death threshold $DieL_{\text{th}}$ for a continuous hysteresis duration H_{die} timed by the hysteresis timer, B_i switches itself to sleep mode, removing the cell supported by it. Similarly, H_{die} exists to prevent perturbation from temporary short-term traffic variations. With $DieL_{\text{th}}$, it can be chosen as a pair for cell death when considering $DivL_{\text{th}}$ and H_{div} as a pair of parameters for cell division. After cell death, B_i resets every element in $\mathcal{W}_i$ and ψ_i, and clears the hysteresis timer to zero

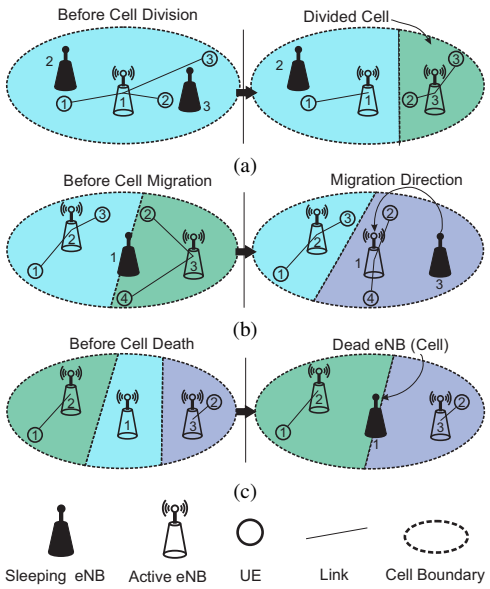


Fig. 4. The examples of the moments of the actions in the TACM algorithm. (a) Cell division: B_1 activates B_3, dividing the bigger cell into two small cells when B_1 is overloaded and perceives that the hotspot area is to the right; (b) Cell migration: B_3 perceives that the angular traffic distribution is biased to the left, then it activates B_1 and deactivates itself, moving the cell supporter closer to the hotspot area; (c) Cell death: B_1 switches to sleep mode when it has little load, killing the cell it served.

before switching to sleep mode. Before switching to sleep mode, B_i broadcasts a sleeping message. Any eNBs receiving such message will mark it as in sleep mode if it is in their NCLs. Fig. 4 (c) shows an example of cell death, in which B_1 is of a low load, and deactivates itself, reducing the overall network power consumption. It is worth mentioning that the death of a cell may also lead to cell migration owning to the biased traffic perceived by other eNBs.

F. Algorithm Summary

As a result of the mechanism introduced, active eNBs observe the SCN situation by estimating DOAs and cumulatively update their respective memories, which quantify the historical experiences of the eNBs. Network adaptation decisions are made in a distributed way and each eNB reasons the action to be taken independently without exchanging load information among neighbour eNBs, which mitigates the system overheads and becomes one of the virtues of the algorithm. Moreover, the existence of hysteresis timers reduces the ping-pong effect caused by temporary short-term traffic variations, resulting in a lower switching on-off frequency, which can potentially reduce the energy consumption of transient states. Another advantage of TACM is contributed by the directional division and migration, which move active eNBs closer to hotspot areas with higher service demands and therefore reduce the UE average transmission power.

The TACM algorithm is summarised in the pseudo code as given below:

for a $B_i \in \mathcal{B}^{\text{on}}$, B_i **do**
  Estimate DOA of the admitted UE with service request
  if $\exists k \in \{0, \pm 1\} : (\alpha + 360k) \in \mathcal{I}_{i,j}$ **then**
    $W_{i,j}^{(t+1)} \leftarrow W_{i,j}^{(t)} + 1$
  else if $\exists k \in \{0, \pm 1\} : (\alpha + 360k) \in \mathcal{I}_{i,j}^{*}$ **then**
    $W_{i,j}^{(t+1)} \leftarrow W_{i,j}^{(t)} - 1$
  else
    $W_{i,j}^{(t+1)} \leftarrow W_{i,j}^{(t)}$
  end if
  $\psi_i^{(t+1)} \leftarrow \psi_i^{(t)} + 1$
  $\beta_{i,j}^{(t+1)} \leftarrow \frac{W_{i,j}^{(t+1)}}{\psi_i^{(t+1)}}$

  Cell Division (for $\mathcal{B}^{\text{on}}$):
  if $L_i \geq DivL_{\text{th}}$ continuously for H_{div} **then**
    if $\mathcal{N}_i^{\text{off}} \neq \varnothing$ **then**
      Find $N_{i,aim} \leftarrow \arg\max_{N_{i,j} \in \mathcal{N}_i^{\text{off}}} W_{i,j}^{(t+1)}$
      Activate $N_{i,aim}$
      B_i resets every element in $\mathcal{W}_i$ and ψ_i, and the hysteresis timer to zero which initialises the memory
    else
      B_i resets the hysteresis timer to zero
    end if
  end if

  Cell Migration (for $\mathcal{B}^{\text{on}}$):
  if $(\psi_i^{(t+1)} \geq \psi_{\text{th}}) \wedge (\mathcal{N}_i^{\text{off}} \neq \varnothing)$ **then**
    Find $\beta_{i,\max}^{(t+1)} \leftarrow \max_{N_{i,j} \in \mathcal{N}_i^{\text{off}}} \beta_{i,j}^{(t+1)}$
    if $\beta_{i,\max}^{(t+1)} > \beta_{\text{th}}$ **then**
      Find $N_{i,\text{aim}} \leftarrow \arg\max_{N_{i,j} \in \mathcal{N}_i^{\text{off}}} \beta_{i,j}^{(t+1)}$
      B_i activates $N_{i,\text{aim}}$
      B_i resets every element in $\mathcal{W}_i$ and ψ_i, and the hysteresis timer to zero which initialises the memory
      B_i switches to sleep mode
    end if
  end if

  Cell Death:
  if $L_i \leq DieL_{\text{th}}$ continuously for H_{die} **then**
    B_i resets every element in $\mathcal{W}_i$ and ψ_i, and the hysteresis timer to zero which initialises the memory
    B_i switches to sleep mode
  end if
end for

Cell Division and Migration (for $\mathcal{B}^{\text{off}}$):
for a $B_i \in \mathcal{B}^{\text{off}}$, B_i **do**
  if B_i receives activation message from B_j where $B_i \in \mathcal{N}_j$ **then**
    B_i switches to active mode
  end if
end for

III. System Models and Assumptions

A. Network Architecture

The network investigated in this paper is in a bounded square area $\mathcal{A} \subset \mathbb{R}^2$. To introduce the randomness of actual base station deployment which better reflects practical networks, especially SCNs, a 2D Poisson Point Process (PPP) with the intensity λ_{p} can be considered to model the spatial

distribution of the eNBs in SCNs [17], [18]. However, as a significant weakness, the independence of PPP may allow eNBs to be located very close to each other, which is not practical in reality. To solve this problem, we apply a repulsive dependent thinning to PPP the same as the one in [19], rendering a type II Matérn Hard-core Process (MHP) with a hard-core distance σ_{b} as a result [19]–[21]. Therefore, $\overline{|\mathcal{B}|}$, the cardinality mean of $\mathcal{B}$ can be calculated as

$$\overline{|\mathcal{B}|} = \frac{|\mathcal{A}|\,[1-\exp(-\lambda_{\mathrm{p}}\pi\sigma_{\mathrm{b}}^2)]}{\pi\sigma_{\mathrm{b}}^2}, \tag{7}$$

where $|\mathcal{A}|$ is the area of $\mathcal{A}$.

To reflect the spatially heterogeneous distribution of the traffic in reality, the UE distribution is modelled by two tiers, the background tier and the hotspot tier, with a hotspot ratio γ. In the background tier, $1-\gamma$ of the total active UEs are scattered randomly and uniformly in $\mathcal{A}$ as the background UEs. In the hotspot tier, UEs are further equally divided into several hotspot groups and UEs in each group independently conform to a 2D normal distribution truncated into $\mathcal{A}$, of which the standard deviation is denoted as σ_{n}. Each group centre is randomly and uniformly scattered in $\mathcal{A}$ and keeps at least σ_{n} away from the network boundary and σ_{h} metres away from other group centres.

In Fig. 5, an example instance of the network architecture is shown, where the small circles representing eNBs are scattered in the square area $\mathcal{A}$. The background UEs are depicted as squares and the hotspot UEs are represented as dots with the centres of their own hotspot groups represented as crosses. The convex hulls of the point sets of the hotspot groups are depicted as polygons for the recognition of each hotspot group. The related parameters of the example and the simulation set-up are given in Section IV. The other aspects of the figure are illustrated in the following subsections.

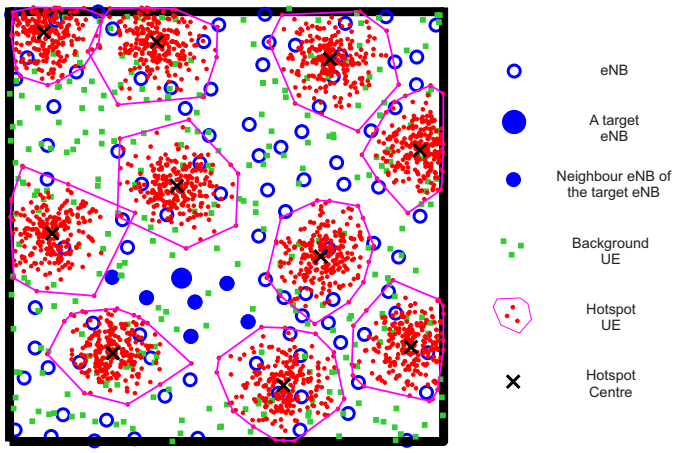


Fig. 5. A network example: The side length of the square area $\mathcal{A}$ is 100 metres. The small circles represent the eNBs. The small solid circles are the neighbour eNBs of the big solid circle. Background UEs are depicted as squares and 10 groups of hotspot UEs are depicted as dots. Hotspot UEs in each hotspot group are encircled by the corresponding convex hull depicted by the polygons. The crosses are the centres of the hotspot groups.

B. Neighbour Cell List

As mentioned before, the algorithm proposed relies on activating neighbour eNBs, which should be carefully determined especially in a SCN with a random topology. In recent practice, the Neighbour Cell List (NCL) is manually configured when deploying a new network, which has many drawbacks due to the sensitivity and the dynamics of the radio propagation environments [22]. Furthermore, to construct a self-organised network as one of the future requirements, a newly deployed eNB has to be capable of automatically self-configuring the NCL and continuously self-optimising it. Among the most common algorithms for this purpose, an eNB autonomously scans the broadcast pilot signals from adjacent cells and creates an initial NCL based on the signal-to-interference-plus-noise ratio (SINR) [23], [24], after which the eNB maintains the NCL based on the SINR and other information collected by UEs [24], [25]. Since the self-organisation of the NCL goes beyond the scope of the study in this paper, it is assumed that each eNB $B_i \in \mathcal{B}$ initialises the NCL $\mathcal{N}_i$ of B_i based on the pilot signals and the NCL is not updated during the lifetime of B_i. Specifically, B_i determines the six strongest pilot signals and selects the corresponding six eNBs as the neighbour eNBs, which is equivalent to seeking the six eNBs with the lowest path loss. These six neighbour eNBs constitute the NCL $\mathcal{N}_i = \{N_{i,1}, N_{i,2}, \ldots, N_{i,j}, \ldots, N_{i,6}\}$, which has the same size as the one in an ideal hexagonal grid architecture. In Fig. 5 for example, the eNBs marked by the solid small circles are added to the NCL of the eNB denoted by the big solid circle.

C. Antenna Model, Link Model and DOA Related Assumptions

To distinguish the DOAs coming from two opposite directions, eNBs can be equipped with circular antenna arrays, or sectorised with multiple antenna chains. Other antenna array solutions to enable discriminability for opposite directions are also available. The antenna arrays are assumed to contain a sufficient number of antenna elements for the DOA estimation capability of resolving multiple users simultaneously. However, as strategies of antenna selection of eNBs or smart antenna beamforming are beyond the scope of this paper, the circular antenna arrays are assumed to have omnidirectional and azimuthal radiation patterns. On the other hand, all the UEs are assumed to use isotropic antennas.

The SCNs investigated are assumed to be deployed in urban areas such as stations and airports, so the WINNER II model B3 in [26] is used for path loss calculation. As the average inter-site distance of eNBs in SCNs is ultra-small (10 metres in the investigated case), it can be known from the link model that the probability of having NLOS is very low. Hence, the effect of NLOS paths is neglected in DOA estimation in the proposed algorithm. The effective range of the path loss model is from 5 to 100 metres. Although there are cases that UEs are located at the distances lower than the five-metre bound, this path loss model is the best we can find for the ultra-dense small cell scenario investigated. Moreover, in terms of the main issue investigated (sleep mode operation), such a range does not have significant impacts on the results. Thus,

we assume that a transmitter has sufficient transmission power to send accurate messages to a receiver at such distances.

D. Power Consumption Model

For network power consumption evaluation, the base station power model in [27] is adopted. As a single transceiver chain is assumed, the eNB power consumption model can be simplified to

$$P_i = \begin{cases} P_0 + \Delta_\mathrm{p} P_\mathrm{out}, & \text{if active,} \\ P_\mathrm{s}, & \text{if asleep,} \end{cases} \tag{8}$$

where P_i is the instantaneous power of an eNB, B_i, which depends on its state and load. P_0 is the static power of the active mode and P_s is the sleep mode power. Δ_p is the slope of the load-dependent power consumption, P_out, which is assumed linear to the eNB load and reaches P_max at the maximum load. The power model here is an approximation of the power consumption of all hardware modules, e.g. power amplifiers, radio fronts, baseband processing, DC-DC and main supply. The data of femtocell base stations in [27] is used ($P_0 = 4.8\ W$, $P_\mathrm{s} = 2.9\ W$, $\Delta_\mathrm{p} = 8.0$, $P_\mathrm{max} = 0.05\ W$), for the SCNs investigated. Then, the instantaneous power P_i calculated with the assigned values is used for evaluating the instantaneous network power consumption, defined as

$$P_\mathrm{net} = \frac{\sum_{i=1}^{|\mathcal{B}|} P_i}{|\mathcal{A}|}. \tag{9}$$

As handover is not introduced, when an eNB is to be switched to a sleep mode, it waits until the end of the data transmission and stops to admit future UEs. During this state, P_0 should be used to model the instantaneous power. The energy consumed during transient states can also been considered. The transient states are short-period states before an eNB is completely switched on or completely switched to sleep modes. The switch-on transient state is modelled by a constant power consumption $P_\mathrm{tran}^\mathrm{on}$ and the total time needed to complete the switch-on transition $T_\mathrm{tran}^\mathrm{on}$. On the other hand, the transient state of turning an eNB to a sleep mode can be modelled by a power consumption $P_\mathrm{tran}^\mathrm{off}$ and the required time $T_\mathrm{tran}^\mathrm{off}$ [28]. Then the network power consumption without considering transient states (Equation 9) can be updated by taking the switching times into account:

$$\overline{P_\mathrm{net}} = \frac{\int_0^T P_\mathrm{net}(t)\,|\mathcal{A}|\,\mathrm{d}t + n_\mathrm{on} P_\mathrm{tran}^\mathrm{on} T_\mathrm{tran}^\mathrm{on} + n_\mathrm{off} P_\mathrm{tran}^\mathrm{off} T_\mathrm{tran}^\mathrm{off}}{T\,|\mathcal{A}|}, \tag{10}$$

where T is a period of time under consideration, n_on and n_off are the times of switching on and switching to sleep modes in the network during T, respectively.

The average transmission power of UEs $\overline{P_\mathrm{UE}^\mathrm{Tx}}$ over T is defined as

$$\overline{P_\mathrm{UE}^\mathrm{Tx}} = \frac{\sum_{m=1}^{M} \sum_{l=1}^{L_m} p_{m,l} t_{m,l}}{\sum_{m=1}^{M} \sum_{l=1}^{L_m} t_{m,l}}, \tag{11}$$

where $p_{m,l}$ is the transmission power of UE_m during the transmission of the file l and $t_{m,l}$ is its corresponding elapsed time during file transmission. L_m is the number of files transmitted by UE_m during T and M is the number of UEs in the network.

IV. Performance Evaluation

A. Simulation Set-up

The simulation of a specific sleep mode operation scheme mainly consists of two parts: the radio resource management and the switching mechanism. The metrics to evaluate the effectiveness of these two parts are the delivered QoS and the sleep mode related performance. As the radio resource management for QoS improvement falls out the scope of the paper and the sleep mode operation focused on does not show significant differences between the uplink and the downlink, only the uplink is simulated. The downlink transmission power is modelled assuming that the downlink load is the same as the uplink. As standardised in [29], the minimum resource allocation unit is a Physical Resource Block (PRB), which is mapped to a Virtual Resource Block (VRB). Four consecutive VRBs selected as a spectrum bundle are assigned to a UE during data transmission. The spectrum resource bundle with the highest SINR is allocated. A UE is blocked if there are no free resource bundles with SINR above a threshold. After successful spectrum allocation, the UE occupies the resource bundle assigned by the eNB until the completion of the file transmission.

Active eNBs are required to broadcast reference signals and UEs estimate the received power. Then they are associated with, and acquire resources from, the active eNBs sending the reference signals with the strongest received power. In this way, the cell sizes are automatically adjusted after an activation or a deactivation process. UE handover is assumed to take place when there is no file being transmitted and the basic strongest received power policy is adopted. Handover of UEs during data transmission is not explicitly considered in this paper, but in any case will not fundamentally change the impact of the sleep mode operation algorithm. Therefore, when an eNB determines that it needs to switch into the sleep mode, it stops broadcasting the reference signal. Thus, it ceases associating with the newly arriving UEs and waits until the data transmission of all currently associated UEs has finished.

To create temporal variations of the hotspot traffic, the simulation duration is divided into 10 periods, during each of which half of the hotspot groups are randomly chosen to be active with files arriving according to the aforementioned traffic model while the others remain inactive. The total number of active hotspot UEs in the network is fixed at any time point during the simulation (γ of the total active UEs). The mean file arrival rate in the traffic model is varied to get different traffic load levels. The parameters of the system models and the simulation are summarised in Table I.

The proposed Traffic-Aware Cell Management (TACM) algorithm is simulated based on the models introduced in Section III and the above simulation set-up. The arguments of the TACM are selected to get a similar QoS in terms of blocking probability and delay as the system with all eNBs always active. The values of the parameters of the algorithm are listed in Table III. It is also compared with a consolidated baseline scheme of sleep mode control proposed in [10] to demonstrate the advantages of the TACM algorithm.

TABLE I
THE PARAMETERS OF THE SYSTEM MODELS AND THE SIMULATION SET-UP

$\mathcal{A}$ side length	100 metres
λ_p	20000/km^2
σ_b	5 metres
Average inter-site distance	10 metres
λ_a	150000/km^2
γ	0.8
σ_n	5 metres
σ_h	25 metres
Number of hotspot groups	10
eNB antenna gain	5 dBi
Link level evaluation	Truncated Shannon bound [30]
Carrier frequency	3.5 GHz
System bandwidth	20 MHz
Number of PRBs	100
PRB bandwidth	180 kHz
Uplink resource allocation	Type 0 [29]
Minimum SINR	5 dB
Power control	Open loop, 25 dB SNR
Noise figure	5 dB
Temperature	300 K
Noise floor	-96 dBm
Traffic Model	Poisson process, 4 Mbits file size [31]
Number of simulation run	100

The comparative baseline scheme is transplanted to suit the self-organised SCN with the random topology. To achieve best energy saving, the load information of neighbour eNBs and the handover information coming from associated UEs are needed so that a notion of network impact can be defined and calculated. It requires active eNBs to get load information from their neighbour eNBs, which needs frequent load information exchange and therefore yields additional system overhead. In the scheme, an eNB determines to switch off when the network impact is below a level computed by a load threshold minus a half of a hysteresis margin, where the load threshold is set to 12% and the hysteresis margin is 16% to mitigate the inefficient switching on-off [10]. An eNB decides to activate a neighbour sleeping eNB when the load of itself exceeds half of the hysteresis margin plus its corresponding recorded load. The values are chosen to make the provided QoS similar to the system with no sleep mode control and achieve reduction of inefficient switching on-off to some extent. More details about the baseline scheme can be found in [10].

B. System Overhead Comparison

Two aspects of system overheads generated by the TACM algorithm and the baseline algorithm are compared. As exchanging information among eNBs is always accompanied by the requirement of additional signalling, including information messages themselves and control signals, the frequency of the load information transmission is quantified in Section IV B1. Additionally in Section IV B2, the switching on-off frequency is also investigated since state transitions of an eNB take time and may consume extra energy. The state transitions of an eNB also compel other eNBs around to change their operating parameters, creating extra system overheads.

1) Load Information Transmission Overhead: As load information acquirement among neighbour eNBs is assumed in the baseline scheme, the frequency of load information transmission is quantified by assuming that an eNB sends the load message whenever its load is varied and sending a message from an eNB to a neighbour eNB is counted as one process of load information transmission. In Fig. 6 (a), the average frequency of load information transmission is plotted against the average network traffic density, where the curve of the baseline algorithm shows an increasing trend because there are more load variations at an eNB when the overall traffic level increases. Obviously, the TACM algorithm does not require any load information exchange among neighbour eNBs so that the frequency is always zero, the same as the system without sleep mode operation. The TACM algorithm surpasses the baseline algorithm prominently in this aspect.

2) On-off Switching Overhead: Fig. 6 (b) shows the average frequency of switching on-off where either switching on or switching off is counted. Although all sleep mode control schemes produce some system overheads due to the switching of eNBs, it can be reduced if an algorithm is comprehensively better. As can be seen from the simulation results, even with the application of the hysteresis margin in the baseline algorithm, the switching on-off frequency is very large. This is because that without averaging the load of an eNB over a period of time, the instant load can easily exceeds or falls below certain levels. With hysteresis durations predefined in the TACM algorithm, eNBs monitor their loads and make sure they are continuously above or below the corresponding thresholds in the hysteresis durations. In this way, TACM algorithm achieves up to 72 times less switching frequency compared with the baseline scheme, which thereby mitigates the system overhead dramatically.

C. Power Comparison

As one of the major metrics concerned in the conventional research into sleep mode control, network power consumption of both two algorithms is compared in Section IV C1. Moreover, as power control is involved in the system, the increase of UE transmission power on account of fewer active eNBs in the SCN along with sleep mode operation should be minimised, which is also investigated and illustrated in Section IV C2.

1) Network Power Consumption: Network power consumption investigated in Fig. 7 (a) and (b) is defined as the overall average power consumed by eNBs in the network per square kilometre, which is computed using the network power consumption model introduced in Section III D. As there is a trade-off between QoS and network power consumption, both schemes are first configured to ensure that the QoS is not degraded compared with no sleep mode application and then the power reduction is maximised (more details are given in Section IV D1). It is noticeable from Fig. 7 (a) that the network power consumption reduced by the TACM algorithm is more than 34% (equivalent to 86% less overall active time on average) when the traffic level is below 0.1 Gbps/MHz/km^2

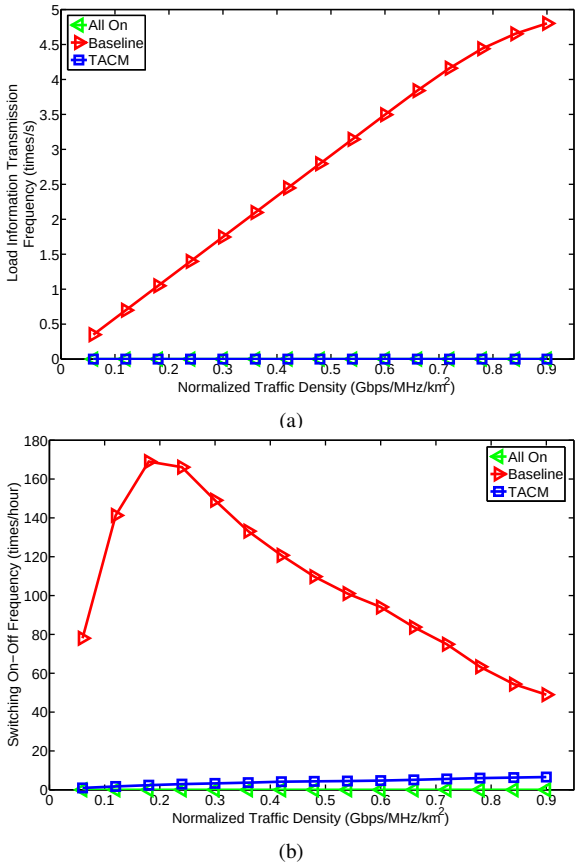


Fig. 6. Comparisons of the TACM algorithm and the baseline algorithm on system overheads. (a) The average frequency of load information transmission against network traffic density; (b) The average frequency switching on-off against network traffic density.

compared with no sleep mode control. Without the necessity of load information exchange, however, the TACM algorithm also reduces more network power consumption than the baseline algorithm, which is up to 81% more around 0.36 Gbps/MHz/km^2. This result is obtained by assuming a relatively high energy consumption during transient states (with parameters given in Table II). As can be seen from Fig. 7 (b), the average network power consumption of the baseline scheme is strongly affected by how much energy is consumed during transient states since its corresponding switching frequency is very high. Comparatively, the network power consumption generated by the TACM algorithm does not vary a lot versus the energy consumption of transient states (only the high level version is shown for clarity) since it mitigates the switching on-off overheads significantly while the high switching frequency of the baseline scheme may increase the overall network power consumption if the energy consumption of transient states are not negligible. Therefore, the proposed TACM algorithm also outperforms the baseline algorithm in terms of the switching frequency reduction and the resulting robustness to the potential cost of the energy consumption introduced by transient states.

It is worth mentioning that the algorithm can still save some power even when the overall network load is heavy. This is because the spatial traffic distribution in the SCN is extremely

TABLE II
ASSUMPTIONS OF THE TRANSIENT STATES

Level of Energy Consumption	$P_{\text{tran}}^{\text{on}}$	$T_{\text{tran}}^{\text{on}}$	$P_{\text{tran}}^{\text{off}}$	$T_{\text{tran}}^{\text{off}}$
High	$2P_0$	2 s	P_0	0.2 s
Medium	$2P_0$	1 s	P_0	0.1 s
None	$2P_0$	0 s	P_0	0 s

heterogeneous in the investigated scenario with hotspot ratio γ equal to 0.8 so that some of the eNBs can be switched to sleep modes without affecting the overall QoS if the traffic levels are not high in their local areas. The network power consumption with all eNBs deactivated is denoted by the dashed line at the bottom of Fig. 7 (a), which actually represents the lower bound of the network power consumption.

2) UE Transmission Power: Since the application of a sleep mode results in fewer active eNBs in the network, UEs increase their transmission power on average as a result of connecting to more distant eNBs during file transmission due to power control. Therefore, the average transmission power of UEs in this case is greater than that when sleep modes are not applied. To explore whether it is considerable compared to the power consumed by the eNBs, the average transmission power of UEs is plotted in Fig. 7 (c) using the UE transmission power model mentioned in Section III D. As observed in the figure, for each sleep mode control algorithm, fewer active eNBs in the network correspond to higher average UE transmission power, implying a larger mean distance between eNBs and UEs. This signifies that there is trade-off between the reduction in the network power consumption and the UE transmission power for a specific algorithm. However, although the transmission power of UEs is higher on average than the scheme with no sleep mode control when the traffic level is low, the energy radiated by all the UEs in the network during a certain period of time never exceeds 0.2% of the energy consumed by all the eNBs for both algorithms, meaning that the overall UE transmission power consumption is negligible relative to the amount of power consumption reduced by the sleep mode control algorithms.

On the other hand, considering the battery lifetime of the mobile devices, the increase of UE transmission power should be minimised by reducing the average distance between eNBs and UEs. Thanks to the directional activation and migration of the TACM algorithm, it is more likely to locate active eNBs at the places with high data service demands than the baseline algorithm. For this reason, the average transmission power of UEs can be 79% (computed with power in mW) less than that of the baseline algorithm at a low traffic level, which significantly improves the user experience.

D. Impact of Performance Related Parameters

In this section, the impact of the parameter variations of the TACM algorithm are demonstrated. The tuning parameters of the TACM algorithm are varied to reveal the trade-off between QoS and network power reduction. The other critical parameters are also varied to observe how they influence the

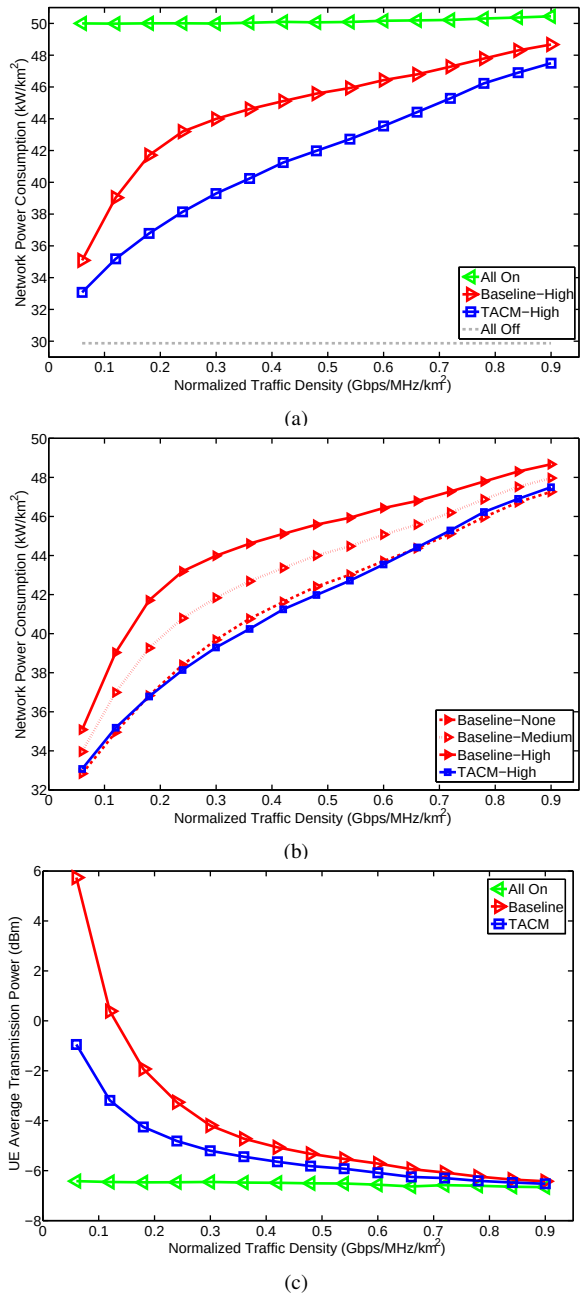


Fig. 7. Comparisons of the TACM algorithm and the baseline algorithm on power consumption. (a) Average network power consumption against network traffic density; (b) Average network power consumption assuming different energy consumptions during transient states against network traffic density; (c) Average UE transmission power against network traffic density.

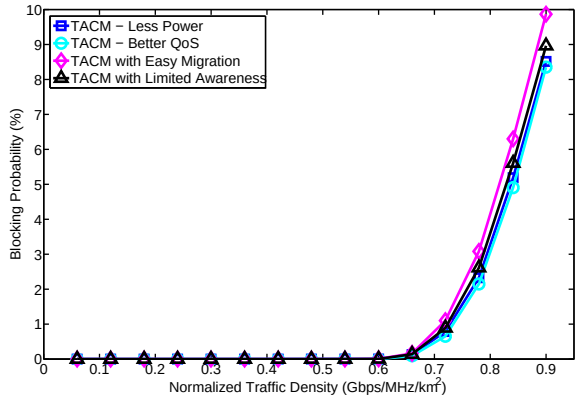
effectiveness of the TACM algorithm. The resulting impact in each case is classified into easy migration and limited awareness. Easy migration indicates that the cell migration process occurs too frequently due to the lower thresholds, ψ_{th} and β_{th}. Limited awareness means that the awareness of an eNB about the correct directions of areas with high traffic levels is limited because of the inappropriate angular interval, $\Delta\theta$. The effects of DOA estimation error on the awareness of eNBs are also analysed.

The QoS performance is shown in Fig. 8. Since the UE average transmission power shows an inverse trend with the network power consumption as demonstrated before, only the network power consumption is shown in Fig. 9. Due to the exemption of load information exchange delivered by the TACM algorithm, only the switching on-off frequency is taken into account when analysing system overheads in Fig. 10. The details of the figures are explained later in this section. All the parameters for each variation are summarised in Table III.

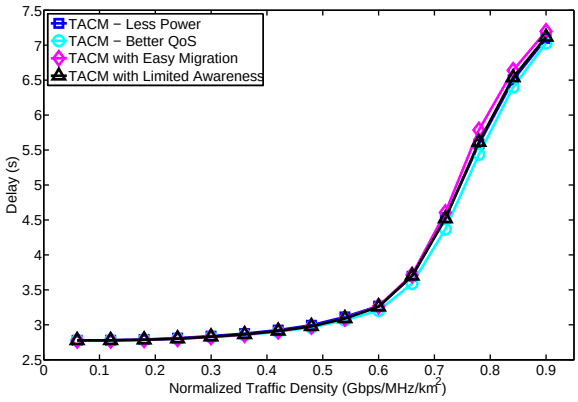
1) Impact of the Tuning Parameters: The tuning parameters are H_{div}, $DivL_{\text{th}}$, H_{die} and $DieL_{\text{th}}$. H_{div} and $DivL_{\text{th}}$ are designed as a pair to control the cell division while H_{die} and $DieL_{\text{th}}$ control the cell death as a pair. To summarise, bigger H_{div} and $DivL_{\text{th}}$ make cell division less likely to happen, leading to fewer active eNBs in a SCN. Bigger H_{die} and smaller $DieL_{\text{th}}$ make cell death harder to be triggered, leading to fewer sleeping eNBs in a SCN. The tuning parameters in the simulation are varied to first ensure the QoS (blocking probability and delay) is similar to the no sleep mode case. Then they are varied to reduce more power consumption until QoS just starts to degrade. In practical networks, this can be easily done through a software defined architecture with a feedback of the QoS measurements.

In the better QoS version of TACM with the arguments given in Table III, these pairs of parameters are varied to keep more eNBs active in the network. As shown in Fig. 8, the blocking probability and delay is improved a little at the cost of highly increased network power as shown in Fig. 9 compared with the less power version of the TACM algorithm. This is achieved by making cell division happen more easily and making cell death a little harder to be triggered. More remarkable is the fact that the QoS improvement gain brought by more active eNBs in the network is limited compared with the larger network power consumption.

2) Impact of Easy Migration: As illustrated before, ψ_{th} is defined to ensure that cell migration is only triggered based on a sufficiently large amount of DOA records. Like β_{th}, it also controls the frequency of cell migration. When ψ_{th} and β_{th} are set to low values such as the example of easy migration (as shown in Table III) based on the better QoS version, weight bias and memory size may exceed the thresholds more easily due to small random traffic variations and easy migration occurs then. As shown in Fig. 8, easy migration makes the original TACM degrade the QoS and switch more eNBs to sleep modes (as shown in Fig. 9). Since there is a trade-off between QoS and network power reduction, it is more convincing to compare the TACM with easy migration to the less power version of TACM, which is configured by varying the tuning parameters without unreliable migrations. Not surprisingly, the reduced power version of the TACM algorithm shows both better QoS and higher network power reduction than the TACM with easy migration. This indicates that the tuning parameters should be used to tune the relationship between QoS and network power reduction if required instead of ψ_{th} and β_{th}. Furthermore, as shown in Fig. 10, the redundant migration processes also increase the overall switching on-off times, which add to the system overheads. ψ_{th} should be set big to ensure a big database acquired, where 100 is enough and β_{th} should be relatively high, say 0.1, to reflect a high angular traffic bias. However, the performance



(a)



(b)

Fig. 8. Comparisons of the different versions of the TACM algorithm on QoS. (a) Average blocking probability against network traffic density; (b) Average delay against network traffic density.

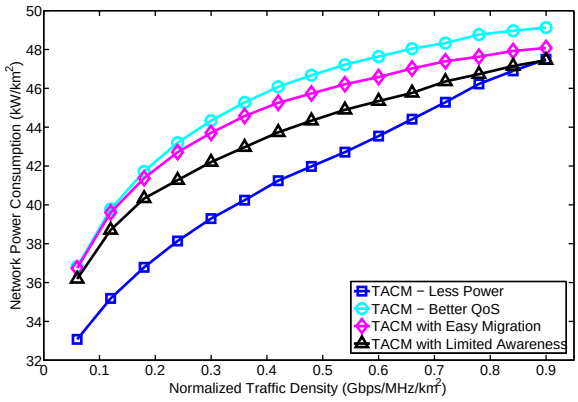


Fig. 9. Comparisons of the different versions of the TACM algorithm on network power consumption.

of the TACM algorithm is not that sensitive to the variations of these two parameters. Only in the extreme case in the given example, there are notable effects.

3) Impact of Limited Awareness: The other parameter, $\Delta\theta$, is critical to cell division and cell migration because it determines how to classify the DOAs and how to update the weights, which are decision-making materials of eNBs. A low value of $\Delta\theta$ leads to the situation that an eNB may be too concentrated on the specific small direction ranges without considering the traffic in the angularly adjacent directions.

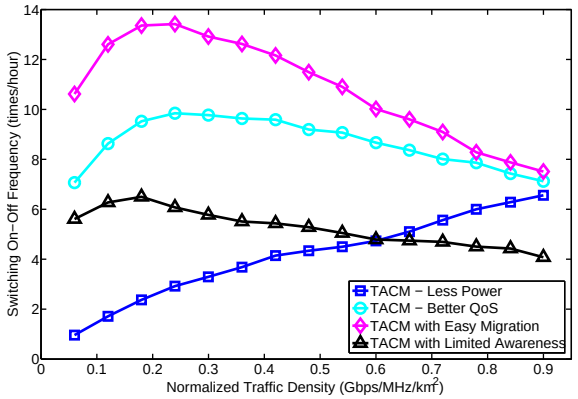


Fig. 10. Comparisons of the different versions of the TACM algorithm on the switching on-off frequency.

TABLE III
SUMMARY OF THE TACM ALGORITHM PARAMETERS

TACM Parameters	TACM - Less Power	TACM - Better QoS	TACM with Easy Migration	TACM with Limited Awareness
H_{div}	5s	4s	4s	4s
$DivL_{\text{th}}$	8%	4%	4%	4%
H_{die}	40s	50s	50s	50s
$DieL_{\text{th}}$	0%	0%	0%	0%
$\Delta\theta$	40°	40°	40°	4°
ψ_{th}	100	100	10	100
β_{th}	0.1	0.1	0.01	0.1

Moreover, an eNB is easier to be enticed by the random traffic spikes from the narrow direction ranges leading to useless cell migration. Comparatively, an over-large $\Delta\theta$ may result in too large a interval $\mathcal{I}_{i,j}$ for $N_{i,j}$, which may not effectively reveal the hotspot directions. This means that an eNB may consider a wide direction range when classifying DOAs and update the corresponding weight. This is likely to result in a situation where an eNB regards the direction of a neighbour eNB as the direction of the area with a high traffic level even if the two directions are far apart. These two situations both affect the accuracy of the orientational awareness about the areas of high traffic levels, which are defined as limited awareness. An example of the TACM with limited awareness is given by setting a small $\Delta\theta$ based on the better QoS version of TACM and its arguments are listed in Table III. Similarly, comparing it with the less power version, it shows slightly worse QoS but significantly higher network power consumption as shown in Fig. 8 and Fig. 9, respectively. This reveals the importance of choosing a medium and appropriate $\Delta\theta$. However, there is a relatively big range for choosing $\Delta\theta$ and the performance only begins to decline in extreme cases. Due to this feature, the TACM algorithm is actually very strong in defending inappropriate DOA classifications.

4) Impact of the DOA Estimation Error: In some applications in the past, the accuracy of DOA estimation has proved problematic [14], [15]. However, as mentioned before, the probability of having NLOS is low in the investigated

SCN scenario and the accuracy of DOA estimation can be very high. Moreover, the approach of using DOAs in the TACM algorithm is somewhat different as a relatively coarse estimation is required to identify the sectors which UEs are located in. From the aforementioned performance discussion, the TACM algorithm is shown to be fault-tolerant in terms of DOA classification. To further check this assumption, the impact of DOA estimation error is briefly investigated. A normal distribution with zero mean is adopted to model the DOA estimation error as an example. It is applied on the DOAs which are intended to be classified into sectors with an angular interval $\Delta\theta$ (explained in section 2.2). Precision and recall are utilised to evaluate the effect of the DOA estimation error. Precision is defined as the number of 'true positives' (i.e. the number of DOAs correctly classified into the interval) divided by the total number of DOAs classified into the interval. Recall is defined as the number of 'true positives' divided by the total number of DOAs that actually ought to be classified into the interval. In the case investigated, recall has the same tendency as precision so that only precision is shown in Fig. 11 (a) and (b).

Fig. 11 (a) gives an example of classification precision against the variations of the standard deviation of the normal distribution σ_e when setting the angular interval $\Delta\theta$ to 40°. From the figure, it is shown that the probability of having a wrong DOA classification is low when the DOA estimation error is low for a reasonable angular interval (40°), e.g. 80% precision corresponds to a standard deviation of 10°. In Fig. 11 (b), classification precision is plotted against different angular intervals revealing that precision of DOA classification is always high if the angular interval is chosen to be medium or high. With reasonable angular intervals chosen in the TACM algorithm, the effect of DOA estimation errors is not significant, meaning that an ideal DOA estimation assumption is sufficient. This can be confirmed by setting $\Delta\theta$ to an appropriate value to avoid the low precision. If σ_e is 10°, with $\Delta\theta$ set to 40°, almost 99.7% of DOAs have absolute errors smaller than 30 degrees.

With the error applied in the simulation, the resulting DOA estimation error does not affect the performance of the original TACM algorithm, but it starts to affect the overall performance of its less power version, where the system overhead produced by switching on-off is especially exacerbated relatively. However, generally speaking, as the effect of DOA estimation error in the performance of the TACM algorithm is limited even when σ_e is not small, the TACM proves robust against the DOA estimation inaccuracy if $\Delta\theta$ is properly chosen. Furthermore, DOA estimation is not used for positioning UEs in this case, but is used to reflect the relative arriving angles of signals, which better reveal the radio environments.

V. Conclusion

This paper has proposed a novel distributed sleep mode operation algorithm for 5G green ultra-dense Small Cell Networks with a random topology. The algorithm employs an innovative concept called Traffic-Aware Cell Management (TACM) to control cell activities, either cell division, cell death or cell migration, which accompany state transitions of eNBs. The TACM algorithm enables each eNB to be aware of the directions of hotspot areas with the application of Direction of Arrival, which completely removes the necessity of load information exchange for sleep mode control required by other traditional sleep mode control schemes. The simulation results based on Small Cell Networks with highly heterogeneous traffic distributions show that the TACM algorithm achieves more than 34% reduction in network power consumption when the traffic level is low, which is equivalent to 86% less overall active time compared with the system without sleep mode control. Compared with a consolidated baseline algorithm, the TACM algorithm significantly mitigates the system overheads and notably reduces the increase of UE transmission power. Assuming high energy consumption of transient states, the TACM algorithm can reduce the network power consumption by up to 81% compared with the baseline scheme. The TACM algorithm is also shown to be insensitive to the direction of arrival estimation inaccuracy.

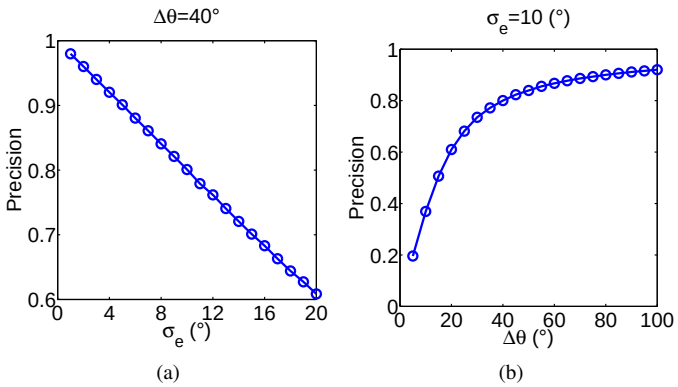


Fig. 11. An investigation of DOA estimation precision. (a) Precision of DOA classification affected by DOA estimation error with fixed classification angular interval against the standard deviation of the error distribution; (b) Precision of DOA classification affected by DOA estimation error with fixed error standard deviation against the variations of the angular interval.

References

[1] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2013-2018," Cisco, Report, Feb. 2014.
[2] J. Hoydis, M. Kobayashi, and M. Debbah, "Green small-cell networks," *IEEE Veh. Technol. Mag.*, vol. 6, no. 1, pp. 37–43, Mar. 2011.
[3] E. Oh, B. Krishnamachari, X. Liu, and Z. Niu, "Toward dynamic energy-efficient operation of cellular network infrastructure," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 56–61, Jun. 2011.
[4] Z. Hasan, H. Boostanimehr, and V. Bhargava, "Green cellular networks: A survey, some research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 4, pp. 524–540, Nov. 2011.
[5] D. Feng, C. Jiang, G. Lim, J. Cimini, L.J., G. Feng, and G. Li, "A survey of energy-efficient wireless communications," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 167–178, Feb. 2013.
[6] I. Ashraf, F. Boccardi, and L. Ho, "Sleep mode techniques for small cell deployments," *IEEE Commun. Mag.*, vol. 49, no. 8, pp. 72–79, Aug. 2011.
[7] Y. Han, D. Grace, and P. Mitchell, "Energy efficient topology management for beyond next generation mobile broadband systems," in *Proc. Int. Symp. Wireless Commun. Syst.*, 2012, Conference Proceedings, pp. 331–335.
[8] A. S. Alam, L. S. Dooley, and A. S. Poulton, "Traffic-and-interference aware base station switching for green cellular networks," in *IEEE 18th Int. Workshop Comput. Aided Modeling and Design of Commun. Links and Networks*, 2013, Conference Proceedings, pp. 63–67.

[9] K. Samdanis, D. Kutscher, and M. Brunner, "Self-organized energy efficient cellular networks," in *Proc. IEEE 21st Int. Symp. Personal, Indoor and Mobile Radio Commun.*, 2010, Conference Proceedings, pp. 1665–1670.

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

[11] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[12] L. C. Godara, "Application of antenna arrays to mobile communications, part II: Beam-forming and direction-of-arrival considerations," *Proc. IEEE*, vol. 85, no. 8, pp. 1195–1245, 1997.

[13] A. Hirata, E. Taillefer, H. Yamada, and T. Ohira, "Handheld direction of arrival finder with electronically steerable parasitic array radiator using the reactancedomain MUltiple SIgnal Classification algorithm," *IET Microw., Antennas & Propag.*, vol. 1, no. 4, pp. 815–821, Aug. 2007.

[14] Z. Wang and S. A. (Reza) Zekavat, "Omnidirectional mobile NLOS identification and localization via multiple cooperative nodes," *IEEE Trans. Mobile Comput.*, vol. 11, no. 12, pp. 2047–2059, Dec. 2012.

[15] V. Zhang and A.-S. Wong, "Combined AOA and TOA NLOS localization with nonlinear programming in severe multipath environments," in *Proc. IEEE Wireless Commun. and Netw. Conf.*, 2009, pp. 1–6.

[16] Z. Li, D. Grace, and P. Mitchell, "Cell division, migration and death for energy efficient 5G ultra-small cell networks," in *Proc. Globecom Workshops*, 2014, pp. 942–947.

[17] H. Wang, X. Zhou, and M. Reed, "Coverage and throughput analysis with a non-uniform small cell deployment," *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 2047–2059, Apr. 2014.

[18] T. Novlan, R. Ganti, A. Ghosh, and J. Andrews, "Analytical evaluation of fractional frequency reuse for OFDMA cellular networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 12, pp. 4294–4305, Dec. 2011.

[19] S. rae Cho and W. Choi, "Coverage and load balancing in heterogeneous cellular networks with minimum cell separation," *IEEE Trans. Mobile Comput.*, vol. 13, no. 9, pp. 1955–1966, Sep. 2014.

[20] M. Haenggi, J. Andrews, F. Baccelli, O. Dousse, and M. Franceschetti, "Stochastic geometry and random graphs for the analysis and design of wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 7, pp. 1029–1046, Sep. 2009.

[21] M. Haenggi, "Mean interference in hard-core wireless networks," *IEEE Commun. Lett.*, vol. 15, no. 8, pp. 792–794, Aug. 2011.

[22] V. M. Nguyen and H. Claussen, "Efficient self-optimization of neighbour cell lists in macrocellular networks," in *Proc. IEEE 21st Int. Symp. Personal, Indoor and Mobile Radio Commun.*, 2010, pp. 1923–1928.

[23] D. Kim, B. Shin, D. Hong, and J. Lim, "Self-configuration of neighbor cell list utilizing E-UTRAN NodeB scanning in LTE systems," in *Proc. 7th IEEE Consum. Commun. and Networking Conf.*, 2010, pp. 1–5.

[24] J. Lim and D. Hong, "Management of neighbor cell lists and physical cell identifiers in self-organizing heterogeneous networks," *J. Commun. and Netw.*, vol. 13, no. 4, pp. 367–376, Aug. 2011.

[25] Z. Becvar, P. Mach, and M. Vondra, "Optimization of SINR-based neighbor cell list for networks with small cells," in *Proc. IEEE 24th Int. Symp. Personal Indoor and Mobile Radio Commun.*, 2013, pp. 2346–2351.

[26] P. Kyösti, J. Meinilä, L. Hentilä, X. Zhao, T. Jämsä, C. Schneider, M. Narandzić, M. Milojević, A. Hong, J. Ylitalo, V. Holappa, M. Alatossava, R. Bultitude, Y. Jong, and T. Rautiainen, "IST-4-027756 WINNER II D1.1.2 V1.2 WINNER II channel models," Report, February 2008.

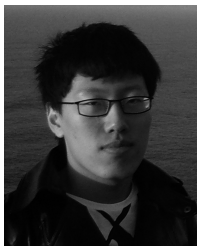
[27] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?" *IEEE Wireless Commun. Mag.*, vol. 18, no. 5, pp. 40–49, Oct. 2011.

[28] X. Guo, S. Zhou, Z. Niu, and P. Kumar, "Optimal wake-up mechanism for single base station with sleep mode," in *Proc. 25th Int. Teletraffic Congr. (ITC)*, Sep. 2013, pp. 1–8.

[29] 3GPP, *Evolved Universal Terrestrial Radio Access (E-UTRA): Physical layer procedures*, 3GPP Standard TS 36.213, Rev. V12.0.0, December 2013.

[30] 3GPP, *Evolved Universal Terrestrial Radio Access (E-UTRA): Radio Frequency (RF) system scenarios*, 3GPP Standard TR 36.942, Rev. V11.0.0, September 2012.

[31] 3GPP, *Evolved Universal Terrestrial Radio Access (E-UTRA): Further advancements for E-UTRA physical layer aspects*, 3GPP Standard TR 36.814, Rev. V9.0.0, March 2010.



Zhehan Li received the B.Eng. degree in Electronic and Electrical Engineering from University of Birmingham, U.K. and the B.Eng. degree in Electronic Science and Technology from Huazhong University of Science and Technology, China, in 2013. He is currently a Ph.D. student in the Department of Electronics at University of York. His current research interests include green networking, 5G small cell networks and self-organised networks.

David Grace (S'95-A'99-M'00-SM'13) received his PhD from University of York in 1999, with the subject of his thesis being 'Distributed Dynamic Channel Assignment for the Wireless Environment'. Since 1994 he has been a member of the Department of Electronics at York, where he is now Professor (Research) and Head of Communications and Signal Processing Research Group. He is also a Co-Director of the York - Zhejiang Lab on Cognitive Radio and Green Communications, and a Guest Professor at Zhejiang University. Current research interests include aerial platform based communications, cognitive green radio, particularly applying distributed artificial intelligence to resource and topology management to improve overall energy efficiency; 5G system architectures; dynamic spectrum access and interference management. He is currently a lead investigator on H2020 MCSA 5G-AURA. He was a one of the lead investigators on FP7 ABSOLUTE and focussed on extending LTE-A for emergency/temporary events through application of cognitive techniques. He was technical lead on the 14-partner FP6 CAPANINA project that dealt with broadband communications from high altitude platforms. He is an author of over 220 papers, and author/editor of 2 books. He is the former chair of IEEE Technical Committee on Cognitive Networks for the period 2013/4. He is a founding member of the IEEE Technical Committee on Green Communications and Computing. In 2000, he jointly founded SkyLARC Technologies Ltd, and was one of its directors. He is currently a Non-Executive Director of a technology start-up company.

Paul Mitchell (M'00-SM'09) received the M.Eng. and Ph.D. degrees from the University of York, York, U.K., in 1999 and 2003, respectively. His Ph.D. research was on medium access control for satellite systems, which was supported by British Telecom. He has been a member of the Department of Electronics at York since 2002, and is currently Senior Lecturer. He has gained industrial experience at BT and QinetiQ. Research interests include medium access control and routing, wireless sensor networks, underwater communications, cognitive radio, traffic modelling, queuing theory, satellite and mobile communication systems. Dr Mitchell is an author of over 90 refereed journal and conference papers and he has served on numerous international conference programme committees. He was general chair of the International Symposium on Wireless Communications Systems which was held in York in 2010, and a track chair for IEEE VTC in 2014. He is an Associate Editor of the *IET Wireless Sensor Systems* journal and Sage International Journal of Distributed Sensor Networks. Dr Mitchell is a Senior Member of the IEEE, a member of the IET and a Fellow of the Higher Education Academy.