KICS The Korean Institute of Communications and Information Sciences

Available online at www.sciencedirect.com

ScienceDirect

ICT Express 8 (2022) 525–529

www.elsevier.com/locate/icte

Multi-agent Q-learning based cell breathing considering SBS collaboration for maximizing energy efficiency in B5G heterogeneous networks

Howon Lee[a,b], Eunjin Kim[a,b], Hyungsub Kim[c], JeeHyeon Na[c], Hyun-Ho Choi[d,b,*]

[a] School of Electronic and Electrical Engineering, Hankyong National University, Anseong, Republic of Korea
[b] IITC, Hankyong National University, Anseong, Republic of Korea
[c] Electronics and Telecommunications Research Institute, Daejeon, Republic of Korea
[d] School of ICT, Robotic, and Mechanical Engineering, Hankyong National University, Anseong, Republic of Korea

Received 14 June 2021; received in revised form 1 September 2021; accepted 8 September 2021
Available online 22 September 2021

Abstract

In B5G heterogeneous cellular networks, a rapid increase in the number of small cell base stations (SBSs) to support a massive number of devices tends to waste a considerable amount of energy. Therefore, intelligent management of SBSs' power consumption is one of the most important research issues. We herein propose quasi-distributed Q-learning-based cell breathing (QD-QCB) considering full and partial SBS collaborations for maximizing network energy efficiency. Also, the concept of an aggregated active SBS set based on regional user distributions is proposed for computing- and energy-efficient operation. Through intensive simulations, we show that the proposed QD-QCB algorithm can achieve optimal energy efficiency, and improve the network energy efficiency significantly compared with conventional algorithms such as no transmit power control, random cell breathing, and greedy cell breathing algorithms.

Keywords: Multi-agent Q-learning; Cell breathing; SBS collaboration; Energy efficiency; Heterogeneous cellular network

1. Introduction

Beyond fifth-generation (B5G) cellular networks aim at supporting tremendous mobile data traffic and a massive number of mobile devices while improving the entire network energy efficiency compared to the previous generation mobile networks [1,2]. In particular, with the explosively increasing mobile data traffic, small cell networks are emerging as a promising solution for B5G cellular networks [3–5]. In addition, it is worth noting that most of the energy in current cellular networks is consumed by BSs, which is approximately 58% of the total power consumption [6,7]. Thus, to minimize this severe network energy consumption, several greening algorithms have been proposed [8–12]. On the other hand, to support the massive amount of data traffic generated in B5G networks, more BSs may be deployed more densely. In other words, a lot of small cell BSs (SBSs) would be deployed with macro cell BSs (MBSs). Accordingly, the average inter-site distance between BSs (SBSs and MBSs) and users is exponentially decreasing, and the link quality and network capacity could be improved significantly. However, this may cause severe interference between neighboring SBSs and MBSs, also ushering in a vast amount of energy consumption in the entire network [4,13–15]. Therefore, saving this network energy consumption is one of the most important challenges for B5G heterogeneous cellular networks in practice.

In [8], Z. Hasan et al. proposed the scheme for BS mode adjustment to minimize the network energy consumption. Also, potential gains and limitations of the ultra-dense networks (UDNs) were investigated, which considered the impact of idle-mode operation of BSs, transmission power control, user density, and user distribution on the network energy efficiency in [9]. In [12], energy-aware user association and power allocation algorithms were proposed for millimeter-wave (mmWave) based UDNs with energy-harvesting BSs. Furthermore, Z. Jian

* Correspondence to: School of ICT, Robotics & Mechanical Engineering, Hankyong National University, 327 Chungang-no, Anseong-si, Kyonggi-do, 17579, Republic of Korea.
E-mail address: hhchoi@hknu.ac.kr (H.-H. Choi).
Peer review under responsibility of The Korean Institute of Communications and Information Sciences (KICS).

https://doi.org/10.1016/j.icte.2021.09.006

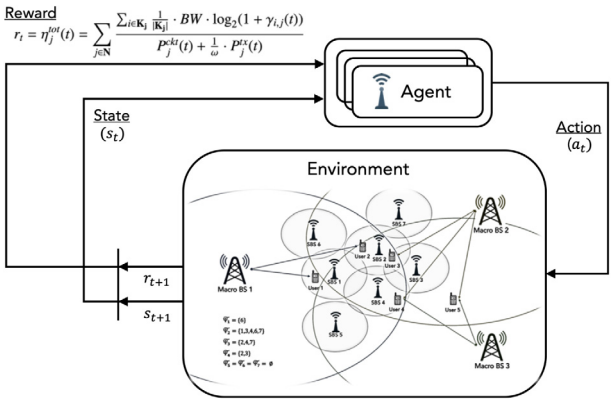

Fig. 1. System model for our proposed QD-QCB algorithm in B5G heterogeneous cellular networks.

et al. proposed a joint optimization framework for an energy-efficient switching on/off strategy and user association policy in UDNs with partial conventional BSs in [16].

Machine learning is able to identify patterns in observed data, create models that explain the digital and physical worlds, and predict things without having any related experiences, explicit pre-programmed rules, and models [17]. In particular, reinforcement learning (RL) tries to follow the fundamental way in which humans learn. RL enables an agent to interact with its environment and to learn from its previous experiences without a training data-set [18,19]. Therefore, RL-based approaches can be utilized to improve network energy efficiency in the B5G heterogeneous cellular networks. Especially, because tabular Q-learning does not use deep neural networks for functional approximation, it can significantly reduce computation overhead caused when performing neural network training in deep reinforcement learning. As mentioned before, the problem of unnecessary power waste of SBSs is very severe in these B5G heterogeneous cellular networks due to spatially and temporally varying traffic loads, regional imbalance of user density, and dynamic user mobility. So, in this paper, we propose an energy- and computing-efficient quasi-distributed Q-learning based cell breathing (QD-QCB) algorithm considering an aggregated active SBS set to maximize the network energy efficiency.

The rest of this paper is organized as follows. The system model for our proposed algorithm is described in Section 2. Also, we propose the QD-QCB algorithm with the aggregated active SBS set in Section 3. In Section 4, we show the simulation results in terms of the energy efficiency of our proposed QD-QCB algorithm compared with conventional algorithms: No TPC, random cell breathing, and greedy cell breathing. Finally, the conclusions are drawn in Section 5.

2. System model

In this section, we describe the system model for our proposed algorithm and assumptions used in this paper. We consider a downlink heterogeneous network configured with several MBSs and SBSs, as shown in Fig. 1. We assume that MBSs as interferers to users associated with SBSs. In addition, the SBS is the agent of the proposed multi-agent Q-learning framework.

2.1. Creation of aggregated active SBS set

In the proposed QD-QCB algorithm, users estimate the channel quality of the serving cells and neighboring cells through the reference signal received power (RSRP) measurement report. This RSRP value is a very common parameter to determine the association of the user in cellular networks. To consider wireless channel fluctuation by small-scale fading and noise, we use an infinite impulse response (IIR) averaging scheme [20]. That is, we can obtain the reliable and stable RSRP value by exploiting the IIR based averaging scheme. By using the IIR averaging scheme, the averaged user's RSRP measurement ($\bar{\zeta}_{ij}(t)$) of user i from cell j at the tth time step is calculated as

$$\bar{\zeta}_{ij}(t) = (1-\kappa)\cdot\bar{\zeta}_{ij}(t-1) + \kappa\cdot\zeta_{ij}(t). \tag{1}$$

Here, $\zeta_{ij}(t)$ is the instantaneous measured RSRP value of user i form cell j, and κ is a filter coefficient parameter configured by the small cell networks.

Then, user i constitutes its active SBS set (Ψ_i) with this averaged RSRP value. If $\bar{\zeta}_{ij}(t)$ is greater than the RSRP threshold (ζ_i^{th}), the SBS j is added to Ψ_i. Among elements in Ψ_i, the SBS which provides the best RSRP (or the best signal-to-interference-plus-noise ratio (SINR)) becomes the serving SBS of user i. Then, user i periodically transmits its Ψ_i information to its serving SBS. Namely, all users send their active SBS set information to their serving SBSs, respectively. After receiving active SBS set information from users, SBS j forms an aggregated active SBS set ($\tilde{\Psi}_j$) through a simple set operation such as intersection or union. By using $\tilde{\Psi}_j$, the SBSs, which are the agents of our proposed QD-QCB algorithm, can manage and utilize their Q-tables for their energy- and computing-efficient cell breathing operations. That is, in each episode, each agent calculates and updates its Q-table corresponding to its $\tilde{\Psi}$ to maximize the reward of our proposed QD-QCB algorithm where the reward is the sum of network energy efficiency.

2.2. Power consumption model for SBS

In this paper, we consider two types of power consumption: static power consumption at tth time step ($P_{j,mode}^{ckt}(t)$) caused by baseband signal processing, battery backup, and site cooling, and transmit power consumption at tth time step ($P_j^{tx}(t)$) [21]. Also, we assume that the static power consumption is independent of the transmit power consumption of SBS. Accordingly, the total power consumption at tth time step ($P_j^{tot}(t)$) of the SBS j can be represented as

$$P_j^{tot}(t) = P_{j,mode}^{ckt}(t) + \frac{1}{\omega}\cdot P_j^{tx}(t), \tag{2}$$

where ω is power amplifier efficiency, and $P_{j,mode}^{ckt}(t)$ and $P_j^{tx}(t)$ are the amounts of static and transmit power consumption of SBS j, respectively. Here, the mode denotes the current

state of SBS: active or sleep. According to the mode of the SBS, the amount of power consumption could be different.

2.3. SINR and network energy efficiency

The SINR of user i associated with SBS j at the tth time step ($\gamma_{i,j}(t)$) is represented as

$$\gamma_{i,j}(t) = \frac{P^S_{i,j}(t) h_{i,j}(t) d^{-\beta_S}_{i,j}}{I_i(t) + \sigma^2_i}. \tag{3}$$

$$I_i(t) = \sum_{n \neq j, n \in \mathbf{N}} P^S_{i,n}(t) h_{i,n}(t) d^{-\beta_S}_{i,n} + \sum_{m \in \mathbf{M}} P^M_{i,m}(t) h_{i,m}(t) d^{-\beta_M}_{i,m}. \tag{4}$$

where $I_i(t)$ is the total interference experienced by the user i at tth time step, and $\mathbf{N}$ and $\mathbf{M}$ are the sets of SBSs and MBSs, respectively. $P^S_{i,x}(t)$ is the transmit power of user i from SBS x, and $P^M_{i,m}(t)$ is the transmit power of user i from MBS m. Also, σ_i is the thermal noise power at user i, $h_{x,y}$ denotes small scale fading (e.g., Rayleigh fading) between the user x and the BS y, and $d_{x,y}$ is the distance between the user x and the BS y. Furthermore, β_S and β_M mean the path loss exponents for SBS and MBS, respectively.

With Eqs. (2)–(4), the energy efficiency of SBS j at tth time step ($\eta_j(t)$) can be calculated as

$$\eta_j(t) = \frac{\sum_{i \in \mathbf{K_j}} \frac{1}{|\mathbf{K_j}|} \cdot BW \cdot \log_2(1 + \gamma_{i,j}(t))}{P^{ckt}_j(t) + \frac{1}{\omega} \cdot P^{tx}_j(t)}. \tag{5}$$

Here, $\mathbf{K_j}$ denotes the set of users associated with SBS j and BW is the entire bandwidth available at SBS j.

3. Quasi-Distributed Q-learning based Cell Breathing (QD-QCB)

In the proposed QD-QCB algorithm, the SBS, which is the agent of the proposed reinforcement learning framework, adjusts its transmit power to maximize the entire network energy efficiency. The reward of our proposed QD-QCB algorithm (r_t), which is the network energy efficiency ($\eta^{tot}_j(t)$), can be obtained as

$$r_t = \eta^{tot}_j(t) = \sum_{j \in \mathbf{N}} \frac{\sum_{i \in \mathbf{K_j}} \frac{1}{|\mathbf{K_j}|} \cdot BW \cdot \log_2(1 + \gamma_{i,j}(t))}{P^{ckt}_j(t) + \frac{1}{\omega} \cdot P^{tx}_j(t)}. \tag{6}$$

Accordingly, the value function of our proposed reinforcement learning framework ($Q(s,a)$) is described as

$$Q(s,a) = E[r_t + \mu \max_{a'} Q(s', a') | s, a], \tag{7}$$

where μ is the discount factor of our Q-learning framework. This Q-value at each state is calculated by using the following iterative procedure [18,19].

$$Q_{t+1}(s_t, a_t) = (1 - \alpha_t) \cdot Q_t(s_t, a_t) + \alpha_t [r_{t+1} + \mu \cdot \max_{a_t} Q_t(s_{t+1}, a_t)]. \tag{8}$$

Here, α_t is the learning rate at tth time step, r_{t+1} represents the reward at the current time step, and '$\mu \times \max_{a_t} Q_t(s_{t+1}, a_t)$' is the maximum expected future reward. At the very beginning, the basic transmission power of the agent is set as the default value. Also, all Q values are initialized to zero and each agent randomly selects its transmission power level.

To determine execution action at tth time step (a_t), we apply the decayed ϵ-greedy policy. In brief, with $1 - \epsilon(t)$ probability, the agent chooses the action with the highest Q^* value, $Q^* = \max_a Q(s,a)$, and with probability $\epsilon(t)$, the agent chooses a random action. To achieve the optimal reward, the appropriate adjustment of exploitation and exploration is needed because the agent does not have enough information about the environments in general. Initially, the agent chooses relatively more random actions to find the best action to achieve the optimal reward. As the episode progresses, the agent gradually reduces the ratio of random actions by controlling the $\epsilon(t)$ value.

$$\epsilon(t) = \epsilon_{init}(1 - \epsilon_{init})^{\frac{i}{\delta \times |\mathbf{A}|}}, \tag{9}$$

where ϵ_{init} is the initial ϵ value, i is the episode index, and δ is an exploration parameter. Also, $|\mathbf{A}|$ is the action set size of the proposed QD-QCB algorithm. When considering the aggregated active SBSs set, the total state set of agent j ($\mathbf{S}_j$) can be described as a Cartesian product space, $\mathbf{S}_j = \otimes \mathbf{s}_i$, $i \in \tilde{\Psi}_j$ where $\mathbf{s}_i$ is the state set size of agent i and $\otimes$ represents a set product. Therefore, using $\mathbf{S}_j$ and $\mathbf{A}$, the agent can build its Q-table.

As mentioned in Section 2, the agents of the proposed QD-QCB algorithm can manage their Q-tables with the aggregated active SBS set ($\tilde{\Psi}_j$) obtained from the active SBS set information (Ψ_j) of each user. Therefore, in QD-QCB, the agent does not need to consider the status information of all base stations: $|\mathbf{A}|^{|\mathbf{N}|}$. That is, each agent calculates and updates its Q-table corresponding to its $\tilde{\Psi}$ to maximize the reward of QD-QCB where the reward is the sum of energy efficiency. By using $\tilde{\Psi}_j$, our proposed QD-QCB algorithm can perform energy- and computing-efficient cell breathing operations. The Q-table size of the centralized Q-learning algorithm augments exponentially according to the increase in the number of SBSs. However, because the proposed algorithm only considers the neighbor SBSs included in the aggregated active SBS set, the proposed QD-QCB algorithm has a constant Q-table size corresponding to the aggregated active SBS set size. In this paper, we consider two kinds of SBS collaboration: partial and full. In full SBS collaboration (FSC), the agent uses both $\tilde{\Psi}_j$ and r_t, however in partial SBS collaboration (PSC), the agent uses only r_t when performing QD-QCB.

4. Simulation results

As shown in Table 1, we consider a heterogeneous network environment consisting of 3 MBSs and 4 SBSs. Cell radii for deploying users are 300 m and 400 m, respectively. To show optimal cell breathing performance, we included the energy-efficiency (EE)-optimal cell breathing curve, and we can find that our proposed QD-QCB can efficiently follow the optimal solution compared to a random cell breathing algorithm. Here, the EE-optimal curve is obtained through an exhaustive search method. In the no transmit power control (TPC) algorithm, the

Table 1
Simulation parameters.

Parameter	Value	Parameter	Value
N	4	M	3
K	20	R	300 m or 400 m
$P^{ckt}_{j,active}$	0.25 W	$P^{ckt}_{j,sleep}$	0.025 W
$P^{S}_{i,j}$	$0 \sim 2.0$ W	σ_i	-174 dBm/Hz
β_S &β_M	3	ω	1
α_t	0.1	μ	0.9
ϵ_{init}	0.99	δ	330

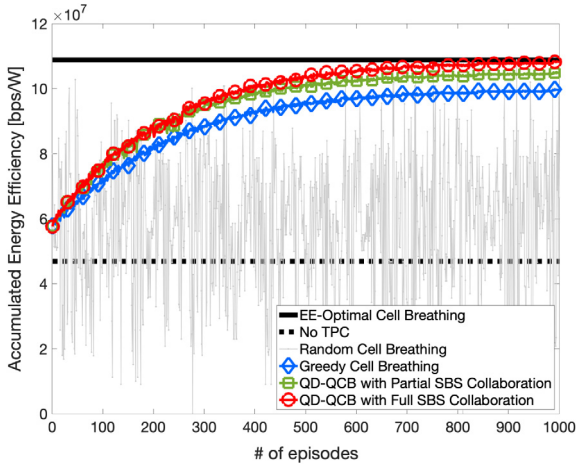

Fig. 2. Accumulated energy efficiency when $N = 4$, $M = 3$, $K = 20$, and $R = 300$ m.

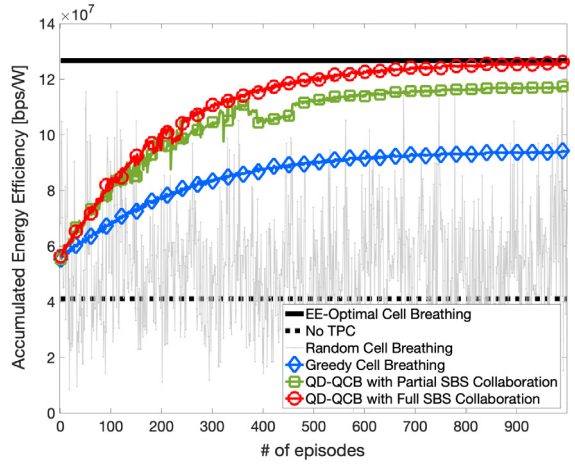

Fig. 3. Accumulated energy efficiency when $N = 4$, $M = 3$, $K = 20$, and $R = 400$ m.

SBS decides its mode considering the initial user deployment. In the No TPC algorithm, if a certain SBS has no users associated with it, this SBS becomes sleep. Furthermore, in the greedy cell breathing algorithm, the SBS controls its transmit power based on Q-learning, however only considers its state and its reward, not the entire network energy efficiency.

As shown in Fig. 2, when $N = 4$, $M = 3$, $K = 20$, and $R = 300$ m, we can show that our proposed QD-QCB with FSC rapidly converges to the EE-optimal value. When the agent applies PSC, the saturated energy efficiency is less than that of the QD-QCB with FSC. Although QD-QCB with PSC has a smaller energy efficiency compared with that of QD-QCB with FSC, QD-QCB with PSC uses relatively less status information to determine the next action. So, according to the computing capability of the SBS, the amount of SBSs' collaboration can be adaptively adjusted. The greedy cell breathing where the agent uses only its state and reward has relatively smaller energy efficiency compared with QD-QCB algorithms. In addition, QD-QCB with FSC and QD-QCB with PSC have better network energy efficiency compared with random cell breathing and No TPC.

The overall trend of Fig. 3 is similar to Fig. 2. However, because the cell radius in Fig. 3 is larger than that in Fig. 2, the performance difference is relatively larger. Also, we can show that our proposed QD-QCB rapidly converges to the EE-optimal value. In particular, when the number of episodes is $350 \sim 450$, we can see that the fluctuation of the accumulated energy efficiency occurs because of imperfect Q-table building and high probability of exploration.

5. Conclusions

To improve the network energy efficiency in B5G heterogeneous networks, we proposed QD-QCB algorithms considering full and partial SBS collaborations. For computing- and energy-efficient operation for QD-QCB, we proposed the concept of the aggregated active SBS set based on the regional user distributions. In addition, for the more computing-efficient operation, we can determine the amount of SBSs' collaboration corresponding to the computing capability of SBSs: FSC and PSC. Through intensive simulations, we showed that the proposed QD-QCB algorithm could achieve the optimal solution and improve the network energy efficiency significantly compared to No TPC, random cell breathing, and greedy cell breathing.

CRediT authorship contribution statement

Howon Lee: Supervision, Conceptualization, Methodology, Writing – original draft. **Eunjin Kim:** Data curation, Software, Visualization. **Hyungsub Kim:** Investigation, Resources. **Jee-Hyeon Na:** Conceptualization, Methodology. **Hyun-Ho Choi:** Supervision, Conceptualization, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2018-0-01659, 5G Open Intelligence-Defined RAN (ID-RAN) Technique based on 5G New Radio).

References

[1] ITU-R, IMT Vision - Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond, Recommendation ITU-R M.2083-0, 2015, pp. 1–19.

[2] H. Yu, et al., What is 5G? Emerging 5G mobile services and network requirements, MDPI Sustain. 9 (10) (2017) 1–22.

[3] X. Ge, et al., Energy efficiency challenges of 5G small cell networks, IEEE Commun. Mag. 55 (5) (2017) 184–191.

[4] W. Lee, et al., DeCoNet: Density clustering-based base station control for energy-efficient cellular IoT networks, IEEE Access 8 (2020) 120881–120891.

[5] S. Zhang, et al., Fundamental green tradeoffs: Progresses, challenges, and impacts on 5G networks, IEEE Commun. Surv. Tutor. 19 (1) (2017) 33–56.

[6] C. Han, T. Harrold, S. Armour, I. Krikidis, Green radio: Radio techniques to enable energy-efficient wireless networks, IEEE Commun. Mag. 49 (6) (2011) 46–54.

[7] J. Wu, Y. Zhang, M. Zukerman, E.K.N. Yung, Energy-efficient base-stations sleep-mode techniques in green cellular networks: A survey, IEEE Commun. Surv. Tutor. 17 (2) (2015) 803–826.

[8] Z. Hasan, et al., Green cellular networks: A survey, some research issues and challenges, IEEE Commun. Surv. Tutor. 13 (4) (2011) 524–540.

[9] D. López-Pérez, M. Ding, H. Claussen, A.H. Jafari, Towards 1 Gbps/UE in cellular systems: understanding ultra-dense small cell deployments, IEEE Commun. Surv. Tutor. 17 (4) (2015) 2078–2101.

[10] G. Koudouridis, H. Gao, P. Legg, A centralised approach to power on-off optimisation for heterogeneous networks, in: Proc. IEEE VTC, 2012, pp. 1–5.

[11] L. Saker, et al., Optimal control of wake up mechanisms of femtocells in heterogeneous networks, IEEE J. Sel. Areas Commun. 30 (3) (2012) 664–672.

[12] H. Zhang, S. Huang, C. Jiang, et al., Energy efficient user association and power allocation in millimeter wave based ultra dense networks with energy harvesting based stations, IEEE J. Sel. Areas Commun. 35 (9) (2017) 1936–1947.

[13] W. Lee, B.C. Jung, H. Lee, ACEnet: Approximate thinning-based judicious network control for energy-efficient ultra-dense networks, MDPI Energ. 11 (5) (2018) 1–11.

[14] Q.C. Li, G. Wu, R.Q. Hu, Analytical study on network spectrum efficiency of ultra dense networks, in: Proc. IEEE PIMRC, 2013, pp. 2764–2768.

[15] Q. Ren, J. Fan, X. Luo, Z. Xu, Y. Chen, Analysis of spectral and energy efficiency in ultra-dense network, in: Proc. IEEE ICCW, 2015, pp. 2812–2817.

[16] Z. Jian, et al., Energy-efficient switching on/off strategies analysis for dense cellular networks with partial conventional base-stations, IEEE Access 8 (2020) 9133–9145.

[17] F. Hussain, et al., Machine learning for resource management in cellular and IoT networks: potentials, current solutions, and open challenges, IEEE Commun. Surv. Tutor. 22 (2) (2020) 1251–1275.

[18] S.K. Sharma, et al., Collaborative distributed Q-learning for RACH congestion minimization in cellular IoT networks, IEEE Commun. Lett. 23 (4) (2019) 600–603.

[19] Z. Chen, et al., Heterogeneous machine-type communications in cellular networks: random access optimization by deep reinforcement learning, in: Proc. IEEE ICC, 2018, pp. 1–6.

[20] F.B. Tesema, et al., Evaluation of adaptive active set management for multi-connectivity in intra-frequency 5G networks, in: Proc. IEEE WCNC, 2016, pp. 1–6.

[21] F. Richter, A. Fehske, G. Fettweis, Energy efficiency aspects of base station deployment strategies for cellular networks, in: Proc. IEEE VTC, 2009, pp. 1–5.