IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 4, 2016, accepted July 21, 2016, date of publication August 10, 2016, date of current version September 28, 2016.

Digital Object Identifier 10.1109/ACCESS.2016.2598813

Survey of Strategies for Switching Off Base Stations in Heterogeneous Networks for Greener 5G Systems

FENGXIA HAN[1], SHENGJIE ZHAO[2], (Senior Member, IEEE), LU ZHANG[3], AND JINSONG WU[4], (Senior Member, IEEE)

[1]Tongji University, Shanghai 201804, China
[2]Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai 200092, China
[3]Nokia Shanghai Bell Company, Ltd., Shanghai 201206, China
[4]Department of Electrical Engineering, Universidad de Chile, Santiago 8370451, Chile

Corresponding author: S. Zhao (shengjiezhao@tongji.edu.cn).

This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2014CB340404 and in part by the National Science Foundation of China under Grant 61471267.

ABSTRACT For heterogeneous network, which has been viewed as one pioneering technology for making cellular networks be evolved into 5G systems, reducing energy consumption by dynamically switching off base stations (BSs) has attracted increasing attention recently. With aiming at optimization on energy saving only or another energy-related performance tradeoffs, several BS switch-off strategies have been proposed from different design perspectives, such as random, distance-aware, load-aware, and auction-based strategies. Furthermore, work has been done to consider joint design for BS switch-off strategy and another strategies, such as user association, resource allocation, and physical-layer interference cancellation strategies. Finally, there have been research results about this topic in emerging cloud radio access networks. In this paper, we take an overview on these technologies and present the state of the art on each aspect. Some challenges that need to be solved in this research filed for future work are also described.

INDEX TERMS Energy consumption, BS switch-off (or called sleeping) strategy, heterogeneous networks, cloud radio access networks, greener 5G systems.

I. INTRODUCTION

Recently, there has been an explosion in mobile data due to the popularization of smart phones and tablets. Since the traffic load on conventional cellular networks is predicted to be increased by 1000 times in the next 10 years [1], the fifth generation (5G) of mobile communication networks will have to deal with an augmented amount of mobile data traffic, which is expected to reach 24.3 exabytes per month by 2019 [2], [3]. Also, the major technical challenges of 5G systems include $0.1 \sim 1$ Gbps user experienced data rate, tens of Gbps peak data rate, tens of Tbps/km^2 traffic volume density, ms-level end-to-end latency, 1 million connections per squared kilometer, etc [4]. One promising approach to tackle the challenge is to exploit the heterogeneous networks (HetNets), where traditional macro base stations (MBSs) are deployed to provide the coverage for large areas while overlaid lower-powered small base stations (SBSs) are used to cover relatively small areas (e.g., hot spots) [5]–[7]. Being constructed with different tiers of small cells and large cells, HetNets can deliver the increased bandwidths, reduced latencies, and higher uplink and downlink data rate to end users.

On the other hand, processing an explosively increased amount of mobile data traffic in 5G systems will also bring ever-increasing energy consumption and carbon footprint to the mobile communication industry. In particular, the whole information and communication technology (ICT) industry has been estimated to contribute to about 2% of global CO_2 emissions, to which the mobile communication industry contributes 15-20% [8]. With increasing awareness of the potential harmful impact on environment and the depletion of non-renewable energy sources, establishing greener mobile communication networks has become an economic issue and a big challenge for sustainable development [9], [10]. In particular, 100 times energy efficiency improvement has been proposed as another technical challenge in the design of 5G systems [4].

Specifically, according to some surveys on energy consumption (e.g., see [11], [12]), 80% of energy consumption in mobile communication networks is due to the operation of BSs. Further, based on the results from laboratory tests done by China Mobile Communications Corporation, a BS consumes 100% energy at the state with the maximum traffic load and about 50%-60% energy at the state with zero traffic load, while the energy consumption of a BS can be reduced to 40% if it is switched off (i.e., in the sleeping state). Therefore, an effective way to achieve energy saving in mobile communication networks is to dynamically switch off BSs, especially for scenarios with low traffic load where less BSs can meet the traffic needs of all user equipments (UEs) [13].

A traditional BS consists of baseband unit (BBU) for baseband signal processing and remote radio head (RRH) for transmitting/receiving radio signals. When a traditional BS is switched off, BBU and RRH of this BS would be switched off together. In contrast, in cloud radio access networks (CRANs) which would be investigated and pursued in 5G systems, BBUs of several traditional BSs are centralized in a single location and the corresponding BBU resources are sliced via virtualization technologies, while RRHs are left at cell sites. With this kind of system architecture, the switch-off operation for RRHs and virtual BBUs could be done separately, through combination with flexible resource allocation on virtual BBUs.

Many works (e.g., see summary works in [14]–[16]) have reported the switch-off (or called sleeping) mechanism for traditional BSs in HetNets and verified the resulting benefit in terms of energy saving with realistic data. Further, there have also existed research efforts (e.g., see [17], [18]) for BS switch-off mechanisms in heterogeneous CRANs (H-CRANs), where the system architecture of CRAN is applied to HetNets. Motivated by these research works and the goal of building greener 5G systems with HetNets being deployed, in this paper, the state of the art in strategies for design on switching off BSs are introduced. Some challenging issues for future work in the related research areas are also discussed.

The rest of this paper is organized as follows. Section II will take an overview on the choice for optimization objective and constraint of the optimization problem used for designing BS switch-off strategy, and analyze their characteristics. Then, section III will describe specific BS switch-off strategies which are designed for HetNets according to different criteria. In Section IV, with focusing on HetNets, we will discuss joint design for BS switch-off strategy and other strategies which are also important for 5G systems. In Section V, the BS switch-off strategies designed for either CRANs or H-CRANs will be reviewed. Finally, conclusions and future work are presented in Section VI.

II. OPTIMIZATION OBJECTIVES AND CONSTRAINTS FOR DESIGNING BS SWITCH-OFF STRATEGY

Appropriate performance evaluation metrics are of primary importance in the design on BS switch-off strategy, because they are directly related to the choice for the optimization objective and constraint of the corresponding optimization problem.

A. OPTIMIZATION ON ENERGY OR POWER CONSUMPTION

The most popular optimization objective is energy or power consumption. Early research in [19] has applied a general energy model to HetNets and evaluated the energy saving in an urban scenario. However, it was not considered in [19] that the energy consumed by a BS is generally not constant but variable with the traffic load. For example, when a SBS is switched off, the coverage-overlapped MBS will take over the traffic load of this SBS so that its energy consumption is increased. Illustrated by Fig. 1 [20], the energy model used in [20] gets rid of the above limitation by assuming that the energy/power consumption of a MBS is linearly varied with its traffic load while the energy/power consumption of a SBS is constant, thus allowing to perform a more practical analysis for the performance of BS switch-off strategy in HetNets.

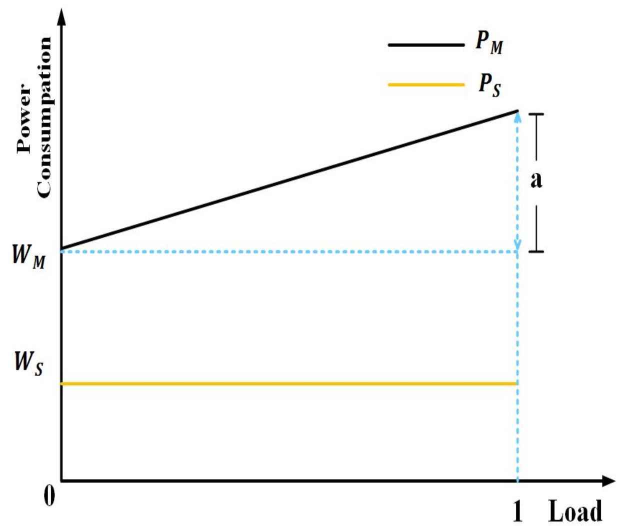

FIGURE 1. Power consumption of MBS (denoted as P_M) and Power consumption of SBS (denoted as P_S) as a function of the traffic load [20]. ($P_M = a \times \text{load} + W_M$ where a is linear coefficient and W_M is the basic circuit power consumption of a MBS when there is no traffic load, $P_S = W_S$ where W_S is a constant independent of traffic load.)

Further, in [21], the power consumption of a SBS is modeled in a more realistic way, which can be found as below in equation (1):

$$P_{\mathrm{SBS}} = \begin{cases} P_0 + \Delta_m P_m^T, & \text{active mode} \\ P_{\mathrm{sleep}}, & \text{sleep mode} \end{cases} \tag{1}$$

where P_0 is the basic circuit power consumption of a SBS in the active mode, Δ_m is the proportionality coefficient of load dependent power consumption, P_m^T is the transmission power of a SBS, and P_{sleep} denotes the power consumption when SBS is in the sleep mode (i.e., being switched off).

B. OPTIMIZATION ON TRADEOFF BETWEEN ENERGY CONSUMPTION AND QoS INDICATORS

Besides considering the optimization on only energy or power consumption for designing BS switch-off strategy, efforts have been made to optimize the tradeoff between energy consumption and data throughput. With focusing on HetNets where SBSs have multi-level sleeping modes, the tradeoff between energy consumption and data throughput is quantified in [22] using an analytical model, where a metric named "energy efficiency (EE)" and with the unit of bits per joule is derived as below:

$$EE = \frac{R}{P} \tag{2}$$

where R (with the unit of bits per second) and P (with the unit of Watt) denote overall throughput and overall power consumption, respectively. In [22] and [23], EE maximization has been considered in HetNets to design the optimal sleeping strategy for MBSs (in [23]) or for SBSs (in [22]). Furthermore, in [24], in order to perform joint design on BS sleeping strategy and power control strategy for self-powered SBSs, an optimization problem is formulated to maximize the channel capacity (i.e., the theoretic upper bound of data throughput) with using energy consumption and energy storage limit in the self-powered battery as optimization constraints.

Inspired by using the tradeoff between energy consumption and data throughput as the optimization objective, people have considered another energy-related tradeoffs in the design on BS switch-off strategy (e.g., see [25]–[29]). Those could be the tradeoffs between energy/power consumption and quality-of-service (QoS) indicators except for data throughput (such as traffic delay, coverage probability, outage probability). For example, when a SBS is turned off, UEs originally served by this SBS have to wait until this SBS wakes up or need to be re-associated to the MBS or another SBSs, leading to longer service delay.

There have been research efforts in considering energy-delay tradeoff (EDT) in the design on BS switch-off strategy (e.g., see [25]–[27]). With setting the average system power consumption as optimization objective and using the mean delay as optimization constraint, the work in [25] performs design for control parameters of switching off SBSs (e.g., SBS operation time, user number in queue). Further, the authors of [26] build a more complicated multi-objective optimization problem where the mean power consumption and the mean traffic delay are two optimization objectives, and exploit the explicit relationship between these two optimization objectives to derive the optimal control parameters of sleeping operation for SBSs (including close-down time, sleeping period, and set-up time). In [27], similar as defining EE to represent the energy-throughput tradeoff in one mathematical expression, the EDT is quantified as below for minimization:

$$\varphi(N) = P_{\text{total}} + \eta D \tag{3}$$

where N is the threshold used to switch off or awaken a SBS in terms of the traffic flows which could be served by the SBS, P_{total} denotes average system power consumption and is a function of N, D denotes average delay and is a function of N, and η is the weight that controls the relative proportion of P_{total} and D. As shown in [27], based on utilizing Markov state transfer diagram to derive the proper expression for traffic delay, Fig. 2 illustrates the energy-delay tradeoff for different average distances between the MBS and SBSs with a given traffic load of SBS.

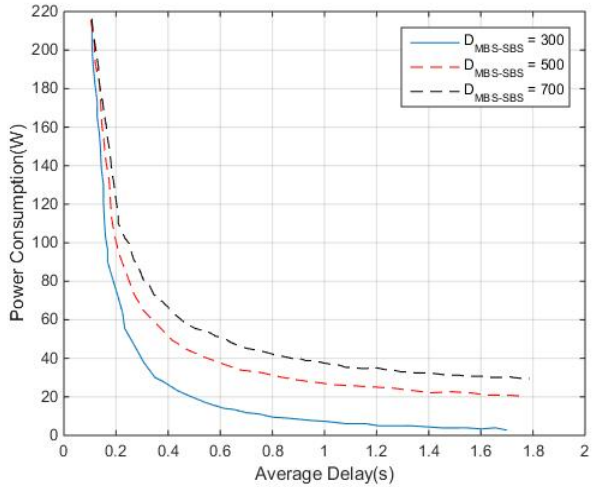

FIGURE 2. Energy-delay tradeoff for different densities of SBS deployment (denoted by average distances between the MBS and SBSs, $D_{\text{MBS-SBS}}$) when the average traffic load of SBSs is given as $\rho = 0.30$, where $\rho = \lambda/\mu$ with λ being the total arrive rate of flows and μ being the serving rate in the modeled Markov process [27].

In addition, to design BS switch-off strategy, some research works also consider the tradeoffs between energy/power consumption and QoS indicators related to successful probability of transmissions over wireless links (e.g., see [28], [29]). With considering a homogeneous network, the work in [28] conducts the design for BS switch-off strategy through considering to reduce the power consumption of BSs and minimize the increment of sum power of UEs, where a constraint on outage probability is used. In [29], with assuming that either MBSs or SBSs in the considered HetNet could be switched off, the QoS indicator like coverage probability is used as the optimization objective, in order to perform joint design on the control parameters for a given BS switch-off strategy and some given user association strategies.

C. OPTIMIZATION ON ANOTHER ENERGY-RELATED TRADEOFF

The tradeoff between energy/power consumption and energy/power cost caused by the transition between sleeping mode and active mode is considered in [30]. Specifically, similar as the way used in equation (3) to quantify the EDT, this tradeoff is represented as below for minimization in [30]:

$$f(v_k^{\text{upp}}, v_k^{\text{low}}) = P^{\text{net}} + \xi P^{\text{cost}} \tag{4}$$

where SBS k will be switched on when the number of its served users reaches v_k^{upp} (i.e., the upper threshold) and will switch off when the user number is less than v_k^{low} (i.e., the lower threshold), P^{net} denotes network power consumption, P^{cost} denotes mode transition power consumption, and ξ is the

TABLE 1. Optimization objectives and constraints for designing BS switch-off strategy.

Literature	Main Tradeoff	Optimization Objective	Meaningful Opt. Constraints
[19-21]	Energy/Power consumption only	Power consumption	\
[22-24]	Energy/Power consumption -Data throughput	EE maximization [22]	Coverage probability and average delay for BSs to wake up [22]
		EE maximization [23]	\
		Channel capacity maximization [24]	Energy consumption and energy storage limit in the self-powered battery [24]
[25-27]	Energy/Power consumption -Delay	Average system power consumption [25]	Mean delay [25]
		Mean power consumption and mean traffic delay (which is multi-objective opt.) [26]	\
		Minimize linear combination of average system power consumption and average delay [27]	\
[28-29]	Energy/Power consumption-Successful probability of transmissions	Power consumption [28]	Outage probability (and implicitly consider the increment of sum power of UEs) [28]
		Coverage probability [29]	\
[30]	Energy/Power consumption -Mode transition cost	Power consumption	Average number and average data rate of served users
[31-33]	Energy/Power consumption-Financial cost or benefit	Total financial cost of the MNO and all third parties [31-32]	\
		Profit of the third party and profit of each MNO (either separate opt. or multi-objective opt.) [33]	

Note: Not all the literature listed here is based on system model with HetNets (system models in [28] do not consider HetNets).

parameter that balances the tradeoff between P^{net} and P^{cost}. Further, in [30], the average number of users served by active SBSs and the average data rate for users served by active SBSs are used as optimization constraints.

Also, a novel tradeoff between energy/power consumption and some certain financial cost or benefit could be considered (e.g., see [31]–[33]), which is applicable for the emerging situation that operator(s) lease capacity and bandwidth resources from one or multiple third parties who deploy lower-powered SBSs to enable offloading traffic from BSs of operators to SBSs of the third parties during peak hours. When considering one mobile network operator (MNO) and multiple third parties who can rent out their SBS resource, the works in [31] and [32] perform design on BS switch-off strategy by setting the total financial cost of the MNO and all third parties as the optimization objective. In [33], under the situation that multiple MNOs and one single third party are considered, besides traditional energy consumption, another two finance-related optimization objectives are considered, including the profit of the third party and profit of each MNO.

D. SUMMARY AND FUTURE WORK

The comparative summary for the literature reviewed in this section can be seen in Table 1.

Summary: Further, the four main findings from the above discussions are:

1) Energy/power consumption is the most popular component when choosing optimization objective and constraint of the optimization problem used for designing BS switch-off strategy.

2) Considering the tradeoffs between energy/power consumption and QoS indicators (such as data throughput, service delay, coverage probability, outage probability) will generate more comprehensive and flexible choices on optimization objective and constraint, which will make the resulted design on BS switch-off strategy be more effective.

3) The transition between sleeping mode and active mode for a BS would increase the operational cost, for example, additional energy/power consumption is required to switch on or off the hardware. It is a more realistic way to consider this kind of mode transition cost in the design on BS switch-off strategy.

4) The involvement of various entities which have different financial goals generates a new tradeoff between energy/power consumption and some certain financial cost or profit.

Future Work: In the future, besides the energy-related tradeoffs already investigated, the infrastructure cost or backhaul overhead can be further taken into account in the design on BS switch-off strategy in HetNets.

III. OVERVIEW OF SPECIFIC BS SWITCH-OFF STRATEGIES IN HETNETS

In HetNets, BS switch-off strategies could be designed either for SBSs or for the coverage-overlapped MBS. Of course, switching off MBS may have non-negligible negative impact on the network coverage. Moreover, when compared with the power consumption of a MBS which significantly increases with its traffic load, the power consumption of a SBS is relatively flat and is not very closely related to its load, so that the aggregated power consumption of SBSs could be larger than the MBS. Thus, more efforts have been put by people into the design on switch-off strategies for SBSs.

A. RANDOM STRATEGY

As we all know, the easiest strategy for switching off BSs is to switch off each BS independently at a certain probability, which has been investigated without considering HetNet deployment (e.g., see [34]–[36]).

The work in [37] focuses on random BS switch-off strategy in HetNets, where the optimal switch-off probability for MBS is derived based on minimizing BS energy consumption. In contrast, in [22], random switch-off strategy for SBSs is designed based on maximizing EE with the constraint for coverage probability. Simulation results in [22] indicate the improvement of approximately 30% in EE with using random sleeping strategy.

Although the random switch-off strategy takes less operational cost and low computational complexity; however, this strategy lack adaptability to varying conditions in realistic mobile networks (such as traffic load, MBS-SBS distance or BS-UE distance).

B. DISTANCE-AWARE STRATEGY

Inspired by the work in [38] which designs more intelligent BS switch-off strategy based on the distance between UEs and their associated BS for a network without HetNet deployment, an optimal distance-based switch-off algorithm for SBSs is proposed in [39] for HetNets. In [39], the operation modes of the SBSs (i.e. switch-off or active) are decided according to their distances to the MBS; specifically, the SBSs closer to the MBS are turned off gradually to minimize the total power consumption in HetNets.

Based on firstly utilizing one well-designed random switch-off strategy for SBSs (as described in Section III-A) to determine the desired number of SBSs in different modes (such as active and sleeping), a distance-aware switch-off strategy for SBSs is further proposed in [22] for HetNets. For this strategy, the number of UEs in the coverage of each SBS and the distance from the nearest UE to each inactive SBS are considered for a static traffic model, while the number of UEs in the coverage of each SBS and the dynamics of the distance from the nearest UE to each inactive SBS (in terms of velocity and moving direction) are considered for a dynamic traffic model. Simulation results in [22] show that the improvement of approximately 45% in EE can be achieved by using this distance-aware sleeping strategy.

It needs be realized that, due to some certain reasons (such as technology limits or designing principles), the information about BS-UE distances may not be able to be accurately or easily obtained. Therefore, it is meaningful to seek for more feasible switching methods.

C. LOAD-AWARE STRATEGY

The distribution of traffic load at different geographical areas would not be uniform in realistic networks. Fortunately, such spatial traffic load fluctuation creates the opportunity to save energy significantly, by switching off the underutilized SBSs. In [40], a load-aware approach is used to switch off a SBS when its traffic load is below a certain threshold for a certain period. It is observed in [40] that, by incorporating load balancing between different types of BSs into the design on switch-off strategies for SBSs, EE improvements up to 68% for low traffic load and up to 33% for medium traffic load can be achieved in HetNets. In [41], a heuristic algorithm is developed for determining which BS to be switched on or off, based on the number of UEs that each BS in the considered HetNet can serve. However, it requires the knowledge on the load and location of all UEs, making this strategy infeasible in realistic networks. In [42], with taking user mobility into account, an optimal load-aware strategy is derived as below:

when the load is not high, the MBS can alone handle the traffic and all the SBSs could be switched off; as the load increases, one or more SBS(s) will be switched on depending on the estimated load and localization info of traffic. Then, the behavior of this strategy is investigated for the case with complete traffic localization information and the case with partially known or delayed traffic localization information.

The traffic load will also have fluctuation over time. As for time-varying traffic load, if we consider all UEs in a given urban area, the distribution of these UEs might fluctuate during one day, resulting in various traffic states. The simplest BS switch-off strategy is proposed in [43] based on only traffic statistics over time, where a SBS is switched off based on a fixed timer and this timer is manually configured for a statistical cycle when the traffic load is very low (e.g., during some time periods at night). More adaptive sleeping strategies based on dynamic traffic monitoring have been discussed in [44] and [45] for HetNets. In [44], based on minimizing the cumulative energy consumption which is obtained by considering time-varying traffic load, the switch-off strategy for a SBS is derived for complete information case and incomplete information case, respectively. With the purpose of adapting the necessary network resource to the actual traffic demand at the moment, a heuristic algorithm and a progressive algorithm are introduced in [45] to dynamically switch off the unnecessary SBSs in HetNets. Both algorithms can well track the traffic variation over time. In [45], it is assumed that the daily traffic profile of the whole HetNet is the same and repeats periodically, which can be approximated by a sinusoidal-like periodic behavior. As shown in [45], Fig. 3 illustrates the traffic curve when assuming the statistic traffic information is available and making the service arrival be modeled as a Poisson process. The corresponding periodic sinusoidal traffic profile could be further obtained easily.

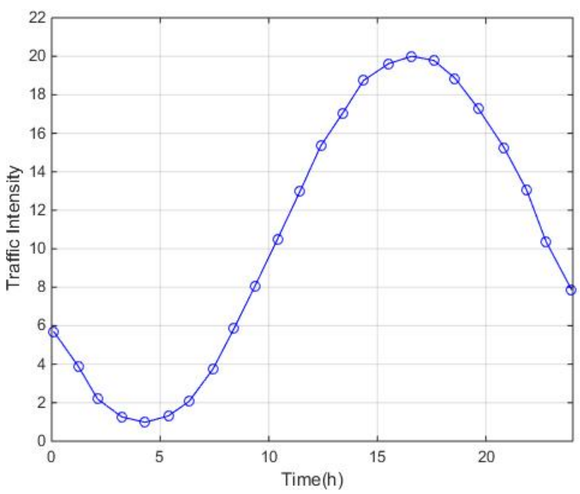

FIGURE 3. Average daily network traffic variations of UE service arrival rate [45]. Here, the service arrival is modeled as a Poisson process with intensity $\rho(t) = \frac{A}{2^b}[1 + sin(\frac{\pi t}{12} + B\pi)]^b + C$ where A denotes the parameter that controls the amplitude of the traffic profile, B means the phase of traffic curve which regulates the position of the peak, C denotes the minimum traffic intensity in the network, $b \in \{1, 3\}$ is used to modulate the gradient of traffic profile curve.

The coexistence of multiple MNOs in the same geographic area, which has not been considered in another works, motivates the research work in [46], where a traffic load based switch-off strategy is proposed for BSs (including MBSs and SBSs) with the help of cooperation between MNOs, and the QoS of re-associated UEs is guaranteed by incorporating the roaming-based infrastructure sharing among the rival MNOs. Furthermore, many aforementioned works have assumed that there exists a centralized entity in the network. This centralized entity will be used to collect the required information over the whole network for implementing the designed BS switch-off strategy, thus causing tremendous communication overhead. With focusing on decentralized implementation, the authors of [47] propose a practical BS switch-off strategy, in which the exchanged information for coordination between neighboring BSs (including the load management information) complies with the X2 interface defined in the 3GPP LTE-Advanced standard. In particular, the always active MBS (called coverage BS in [47]) will be in charge of switching off SBSs (called capacity BSs in [47]) in the order of their traffic load.

D. AUCTION-BASED STRATEGY

To avoid significantly increased capital and operational expenditures (CapEx and OpEx) resulted by densely deployed SBS while meeting the traffic demands in an economical way, MNOs may consider leasing SBS resources from one or multiple third parties. Focusing on the case with one MNO and multiple third parties, authors in [31] and [32] propose an auction-based offloading scheme for implementing the switch-off of the MNO's BSs. In particular, the corresponding reverse combinatorial auction problem is formulated using the integer linear programming model, which provides the optimal allocation for the auction, namely the SBSs to be rented out and the mobile data traffic that can be offloaded. Of course, the works in [31] and [32] assume the capacity of SBSs rented by third parties is sufficient to serve the offloading demand of the MNO. However, it is shown in [48] that the presence of multiple MNOs in the same area will make the situation become more complicated, since the SBSs may not be able to fully support the offloading demand of all MNOs.

For the case with multiple MNOs and one single third party, when each MNO proposes a single bidding value for its requested capacity (which is based on the maximum prediction about the traffic that it is willing to offload), this prediction may not always correspond to the actual traffic volume, which will lead to decreased benefit for the operators and inefficient use of the third party's SBS resources. Motivated by this issue, the authors of [33] propose to let each of multiple MNOs submit a set of bidding values to the third

TABLE 2. Overview of specific BS switch-off strategies in HetNets.

Literature	Strategy	Contributions
[37]	Random sleeping	Derive the optimal switch-off probability for MBS in HetNets based on minimizing energy consumption.
[22]	Random sleeping	Design random switch-off strategy for SBSs based on maximizing EE with the constraint for coverage probability.
[22]	Distance-aware	Design switch-off strategy for SBSs by utilizing the following key metrics: the number of UEs in the coverage of each SBS and the distance from the nearest UE to each inactive SBS.
[39]	Distance-aware	Design switch-off strategy for SBSs based on their distances to the MBS.
[40]	Load-aware	Switch off a SBS when its traffic load is below a certain threshold for a certain period.
[41]	Load-aware	Design switch-off strategy for BSs (including MBSs and SBSs) based on the number of UEs that each BS can serve.
[42]	Load-aware	All BSs are switched off when the MBS can alone handle the traffic; one or more SBS(s) will be switched on depending on the estimated load and localization info of traffic.
[43]	Load-aware	Design switch-off strategy for SBSs based on only traffic statistics over time (a fixed timer is configured based on the load).
[44]	Load-aware	Design switch-off strategy for SBSs based on minimizing time-varying energy consumption caused by time-varying traffic load.
[45]	Load-aware	Design switch-off strategy for SBSs based on optimizing a utility function which is related to load, rate, and interference strength.
[46]	Load-aware	Focusing on the case with multiple MNOs, design switch-off strategy for SBSs (including MBSs and SBSs) with the help of cooperation between MNOs.
[47]	Load-aware	Switch off SBSs in the order of their traffic load, with focusing on decentralized implementation.
[31-32]	Auction-based	Focusing on the case with one MNO and multiple third parties, design switch-off strategy for the MNO's BSs.
[33]	Auction-based	Focusing on the case with multiple MNOs and one single third party, design switch-off strategy for BSs of MNOs, where each of multiple MNOs submits a set of bidding values to the third party.

party, which request different capacity resources provided by SBSs of the third party. Further, in [33], an auction-based switch-off strategy is designed for BSs of MNOs, where each of multiple MNOs will quantify its tolerance about requesting less SBS resources with different bidding values. Conflicting interests of the involved parties are also considered in [33], so as to the optimal solution, which maximizes the financial profits of both the third party and the MNOs and minimizes the network energy consumption simultaneously, is provided by solving a multi-objective optimization problem.

E. SUMMARY AND FUTURE WORK

The comparative summary for the literature reviewed in this section can be seen in Table 2.

Summary: Further, the main findings of this section are described as follows:

1) The design of specific BS switch-off strategy in HetNets can be done based on different criteria/perspectives. However, the design which can better adapt to the dynamics of network conditions would achieve more energy saving and better network performance.

2) The design could be performed with focusing on applying the switch-off strategy to only SBSs, or only MBSs, or both types of BSs in HetNets.

Future Work: Although lots of works have been done for the design on specific BS switch-off strategy in HetNets, enhancement and challenges remain to be investigated in the future. For instance, it is meaningful to investigate the designs done by appropriately exploiting the combination of different criteria/perspectives. For example, combination of random and load criteria, combination of load and distance criteria, etc.

IV. OVERVIEW OF JOINT DESIGN FOR BS SWITCH-OFF STRATEGY AND OTHER STRATEGIES IN HETNETS

In this section, we will focus on the joint design for BS switch-off strategy and another strategies which are also important for 5G systems, such as user association strategy, resource allocation strategy, physical-layer (PHY-layer) interference cancellation strategy.

A. JOINT DESIGN FOR BS SWITCH-OFF AND USER ASSOCIATION

In the aforementioned works, the user association strategy is designed based on the given design of BS switch-off strategy. Here, we consider joint design for BS switch-off strategy and user association strategy. The work in [49] formulates the following combinatorial optimization problem to perform this joint design:

$$\begin{aligned}
&min\ P_{\text{total}} = \sum_{0\leq j\leq N} b_j \sum_{1\leq i\leq M} a_{i,j}p_{i,j} \\
&s.t.\ \sum_{i=1}^{M} n_{i,j}a_{i,j} \leq N_{RB}^{\text{total}}, \quad \forall j \\
&\qquad \sum_{j=0}^{N} a_{i,j} = 1, \quad \forall i \\
&\qquad b_j = \begin{cases} 1, & \sum_i a_{i,j} \neq 0 \\ 0, & \sum_i a_{i,j} = 0 \end{cases}
\end{aligned} \tag{5}$$

In equation (5), $a_{i,j} = 1$ denotes that the i-th user is attached to the j-th BS while $a_{i,j} = 0$ denotes the opposite case, $b_j = 1$ denotes the "switch-on" status of the j-th BS while $b_j = 0$ denotes the "switch-off" status of the j-th BS, $n_{i,j}$ denotes the minimum number of resource blocks for the i-th user to meet the QoS constraint of minimum data rate; moreover, the first constraint represents that the load of each BS should be no more than the maximum number of available resource, and the second constraint indicates that every user can only access one BS at one time. Solving the above optimization problem requires high computational complexity of $o(2^{N+M})$ for completing exhaustive searching. In [49], based on transforming this optimization problem to be a classical 0/1 knapsack problem, a dynamic programming algorithm is used to get the solution. In [50], also aiming at minimizing the total power consumption in HetNets, the joint design for BS switch-off strategy and user association strategy is formulated as a form of the minimum set cover problem. One characteristic of the work in [50] is that the simulations are conducted based on the system-level simulation platform of LTE (which is a commercial mobile communication system); this make the feasibility of the solution be verified more strictly.

The joint design considered in this subsection has also been investigated when taking the maximization on EE as the optimization objective (e.g., see [51]–[53]). In [51], the optimization problem is transformed to be a linear programming problem, and a series of Lagrangian dual methods are used to obtain the solution when assuming a central controller is available to collect the information about the whole network. In addition, to get a more feasible solution without the need of a central controller, a user repeated bidding game is utilized in [51] to find out the solution of the original EE maximization problem. In contrast, in [52], the joint design is done based on providing a heuristic solution for the formulated EE maximization problem. The key design principle is to consider the impact of switching off of a SBS on the network, in terms of ensuring additional load increments on its neighboring BSs (including the MBS and neighboring SBSs) can be handled well with meeting the required QoS. Specifically, for any one given load state, switch off SBSs one by one and update user association iteratively till the outage constraint cannot be satisfied. In [53], the EE maximization problem is also solved by using a heuristic algorithm. In particular, the SBS with the lowest EE value is the most possible candidate for being switched off; if the system EE is improved after this SBS is deactivated and its served users are re-associated, the algorithm is repeated to successively switch off the SBS with lowest EE value at each round, till there is no gain in EE anymore.

Unlike other works which perform joint design in terms of both methodology and related control parameters, the work in [29] aims at the joint design on only control parameters, when the methodologies for both BS switch-off strategy and user association strategy have been given.

B. JOINT DESIGN FOR BS SWITCH-OFF AND RESOURCE ALLOCATION

Due to aggressive frequency reuse, serious inter-tier interference observed in HetNets has raised a major challenge to the optimization on system capacity. Both of wireless resource allocation and BS switch-off have been

TABLE 3. Overview of joint design for BS switch-off strategy and other strategies in HetNets.

Literature	Strategy	Contributions
[49-50]	Joint design for BS Switch-off and User Association	Aim at minimizing the total power consumption:
		[49]: Formulate the joint design as a combinatorial optimization problem.
		[50]: Formulate the joint design as a form of the minimum set cover problem.
[51-53]	Joint design for BS Switch-off and User Association	Aim at maximizing EE:
		[51]: Provide solutions for the case with a central controller (via Lagrangian dual methods) and the case wihout a central controller (via using a user repeated bidding game).
		[52]: Key point of heuristic solution: considering the impact of switching off of a SBS on the network, in terms of ensuring additional load increments on its neighboring BSs can be handled well with meeting the required QoS.
		[53]: Key point of heuristic solution: the SBS with the lowest EE value is the most possible candidate for being switched off.
[29]	Joint design for BS Switch-off and User Association	Aim at the joint design on only control parameters when the methodologies for both BS switch-off strategy and user association strategy have been given.
[54]	Joint design for BS Switch-off and Resource Allocation	Spectrum resource allocation is considered in the joint design and user association is also coupled in the joint design.
[55]	Joint design for BS Switch-off and Resource Allocation	Spectrum resource allocation is considered in the joint design.
[56]	Joint design for BS Switch-off and Resource Allocation	Partial spectrum reuse is considered in the joint design.
[24]	Joint design for BS Switch-off and Resource Allocation	Transmit power allocation is considered in the joint design.
[57]	Joint design for BS Switch-off and PHY-layer Interference Cancellation	A SNR maximization problem is formulated to find the optimal beamforming pattern of SBSs with setting the maximal number of active SBSs as one opt. constraint.
[58]	Joint design for BS Switch-off and PHY-layer Interference Cancellation	Choosing which BSs to be turned on is done by either minimizing the transmission power consumption resulted by beamforming transmissions or maximizing the sum rate achieved by beamforming transmissions.
[59]	Joint design for BS Switch-off and PHY-layer Interference Cancellation	The design of pre-coding used in beamforming will determine the amount of cross-tier interference, while switching on or off a SBS is decided by the amount of its experienced interference generated to or by the MBS.

regarded as useful mechanisms for achieving interference reduction and/or EE improvement. Performing joint design for these two kinds of strategies will be beneficial to further enhancing the performance of future HetNets. Many relevant works have been done in the literature (e.g., see [24], [54]–[56]).

To perform joint design for BS switch-off strategy and resource allocation strategy, in [54], an optimization problem is formulated in terms of minimizing total energy consumption of all BSs in HetNets subject to some certain user QoS constraint. This problem is then decoupled into the following two sub-problems: switch-off strategy for MBSs in the considered HetNet, user association and resource allocation; then, the algorithm for solving each sub-problem is developed.

A threshold-based optimization problem is formulated in [55] to jointly optimize switch-off strategy for SBSs and spectrum resource allocation. It is found in [55] that the lightly-loaded SBSs should be switched off first when their sleeping probabilities are low, while the most heavily-loaded ones should go into sleep when their sleeping probabilities are high. The work in [56] focuses on jointly designing partial spectrum reuse (PSR) and switch-off strategy for BSs (including MBSs and SBSs). With assuming the sleeping/active probability of each MBS is the same and the sleeping/active probability of each SBS is the same, in [56], the first decomposed sub-problem of maximizing the coverage probability of all BSs is solved by deriving the relationship between the optimal PSR factor and the ratio of MBSs' active probability to SBSs' active probability, and the solving for the second sub-problem of minimizing the total network energy cost is then followed.

Furthermore, if transmit power is also viewed as one kind of wireless resource, the joint design for BS switch-off strategy and transmit power allocation strategy is investigated in [24] for self-powered SBSs.

C. JOINT DESIGN FOR BS SWITCH-OFF AND PHY-LAYER INTERFERENCE CANCELLATION

Due to their potential for achieving higher link reliability and data rates, multiple-input multiple-output (MIMO) techniques utilizing multiple antennas at the transmitter and/or receiver have emerged as a milestone of modern wireless communications. As one of MIMO techniques, beamforming is the most popular PHY-layer interference cancellation technique used in commercial mobile communication systems, for mitigating intra-cell multi-user and/or multi-stream interference and inter-cell multi-BS interference. When beamforming is exploited to combat inter-cell multi-BS interference, the resulted transmission scheme is also called coordinated multi-point (CoMP) transmission. Some works have been done to investigate joint design for BS switch-off strategy and beamforming (e.g., see [57]–[59]).

The joint deign considered in this subsection is done in [57] by formulating a signal-to-noise ratio (SNR) maximization problem, which tries to find the optimal beamforming pattern of SBSs with setting the maximal number of SBSs which can be active as one optimization constraint. In [58], an optimized solution about choosing which BSs (including MBSs and SBSs) to be turned on is obtained by either minimizing the transmission power consumption resulted by beamforming transmissions or maximizing the sum rate achieved by beamforming transmissions. When utilizing transmission power minimization to do the joint design, a QoS constraint is considered in terms of effective signal-to-interference-plus-noise ratio (SINR) at the receiver side. Furthermore, in [59], the amount of cross-tier interference is utilized to perform the joint deign considered in this subsection. Specifically, the design of pre-coding used in beamforming technology will determine the amount of cross-tier interference, while switching on or off a SBS is decided by the amount of the interference generated by or to the MBS.

D. SUMMARY AND FUTURE WORK

The comparative summary for the literature reviewed in this section can be seen in Table 3.

Summary: The main finding is that the joint design for BS switch-off strategy and another strategies in HetNets will simultaneously improve the energy saving and enhance the overall network performance, to get greener and also more robust 5G systems.

Future Work: In the future, the effort could be put into the joint design for BS switch-off strategy and another advanced HetNet-related strategies, such as inter-site carrier aggregation (CA) between MBS and SBS [60], dual connectivity (DC) between MBS and SBS [61], downlink-uplink decoupled biased user association [62], [63], etc.

V. OVERVIEW OF BS SWITCH-OFF STRATEGIES IN CLOUD RADIO ACCESS NETWORKS

When CRAN is considered, the amount of research works for BS switch-off strategies has not been large and most of them focus on CRAN without HetNet deployment. Thus, in this section, we will firstly review BS switch-off strategies in CRAN without HetNet deployment; then, the emerging research direction of designing BS switch-off strategies for H-CRAN will be discussed.

A. BS SWITCH-OFF STRATEGIES IN CLOUD RAN WITHOUT HETNET DEPLOYMENT

In CRAN, baseband signal processing for multiple distributed RRHs is performed in the BBU pool (or called cloud), where the computing and storage resources are configured as several virtual BBUs. Depending on the computing and storage capability given to each virtual BBU and the maximum amount of baseband processing tasks needed for each RRH, either multiple RRHs could be supported by one virtual BBU or one RRH at most needs to be served by multiple virtual BBUs.

In the following, for the sake of simplicity, we just use the term "BBU" to represent "virtual BBU".

1) STRATEGY VIA SWITCHING OFF BBUs

In the BBU pool, the number of used BBUs can be reduced by matching a right amount of BBU resources with the corresponding RRHs according to the traffic load. Some works have been done to investigate how to appropriately perform the "BBU-to-RRH" mapping and switch off underutilized BBUs (e.g., see [64]–[67]).

With the objective of minimizing network power consumption, in [64], the "BBU-to-RRH" mapping problem is modelled as the typical bin packing problem when assuming the traffic load has been given. In particular, each BBU is modeled as a bin and the baseband signal processing tasks which come from the corresponding RRHs are modeled as items to be packed. After properly packing these items into some BBUs, the number of used BBUs needs to be minimized. Since the modeled optimization problem is NP-hard, three heuristic algorithms are proposed in [64] to generate good but not necessarily optimal "BBU-to-RRH" mapping. In [65], based on more practically modeling energy consumption for housing facilities (e.g., cooling facility), internal switch in central office at which the BBU pool is located, centralized BBUs, optical line terminal in central office, RRHs, and optical network units at cell sites, a network energy consumption minimization problem is formulated to minimize the number of active BBUs. Then, the problem is solved for static traffic load and dynamic traffic load, respectively. For the case of static (i.e., given) traffic load, the problem is more difficult to solve when compared with the problem modeled in [64], because each BBU is modeled as two sub-bins to pack two kinds of items including user processing and cell processing.

Focusing on the case that multiple RRHs could be supported by one BBU in the BBU pool, a traffic load balancing based strategy is designed in [66], to assign a minimum number of active BBUs to the RRHs so that power saving is maximized. In this strategy, if the resource usage of one BBU reaches the upper limit, partial traffic of this BBU will be offloaded to another light-loaded or sleeping BBU; in contrast, an underutilized BBU will be switched off after its traffic is completely offloaded to another suitable BBU. Similar as [66], through proper scheduling and timely offloading/consolidation for traffic load required to be processed by active BBUs, one heuristic strategy is proposed in [67] to dynamically minimize the number of active BBUs. In addition, in [67], a testbed is constructed for concept verification and practical performance evaluation.

2) STRATEGY VIA SWITCHING OFF RRHs AND FRONTHAUL LINKS

In CRAN, with the centralized processing at the BBU pool, the transport links between RRHs and the BBU pool (i.e., fronthaul links) need to provide high-capacity connections, so that the power consumption over fronthaul links becomes enormous and cannot be ignored. Thus, to reduce power consumption of the entire network for CRAN, it is also crucial to switch off some RRHs and the corresponding fronthaul links based on the data traffic requirements. There have been some relevant works in the literature (e.g., see [68]–[70]).

In [68], the joint design on selection of active RRHs and coordinated beamforming among active RRHs is done, with the objective of minimizing the total power consumption of RRHs and the corresponding fronthaul links. This optimization problem is firstly solved by one greedy algorithm, which iteratively chooses one RRH to switch off at each step while

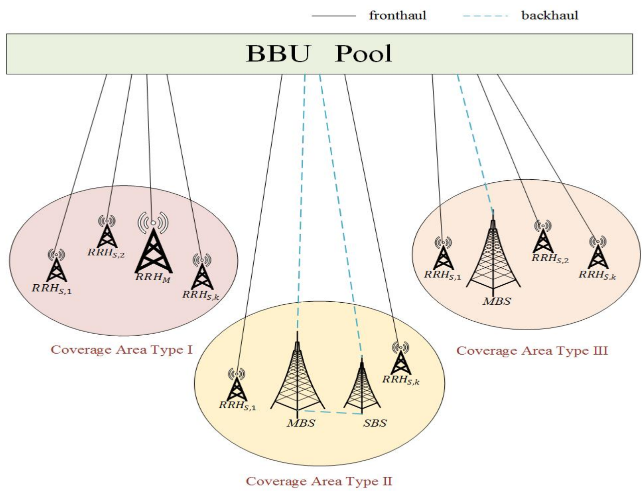

FIGURE 4. Flexible constructions for the network architecture of H-CRANs.

re-optimizing the coordinated beamforming for the remaining active RRHs. Further, to reduce the complexity, an equivalent sparsity-based representation of the focused optimization problem is proposed, where the group sparsity pattern of the aggregated beamforming vector indicates those RRHs which can be switched off. The work in [69] is an extension of [68], where the considered set of UEs is divided into multiple non-overlapping multicast subsets so that the formulated sparsity-based optimization problem is more difficult to solve. With QoS constraints on potentially a large number of served UEs, the focused power consumption minimization problem may be unable to find feasible solutions. While this practical concern is ignored in [68], the work in [69] proposes to reformulate the joint design problem to be maximizing the number of supported users through user admission control.

The work in [70] considers the same joint design as that focused in [68] and [69] and also formulates an optimization problem in terms of minimizing the total power consumption of RRHs and the corresponding fronthaul links. However, unlike [68] and [69] where SINRs of all the served UEs are used as QoS constraints, traffic delay is used in [70] for setting optimization constraint on QoS. More importantly, in contrast to [68] and [69], in [70], the power consumption minimization is formulated as a stochastic optimization problem, based on considering random traffic arrivals and time-varying channel conditions.

B. BS SWITCH-OFF STRATEGIES IN HETEROGENEOUS CLOUD RAN

When transforming a traditional HetNet into a H-CRAN, in theory, an extreme situation is to make all of traditional SBSs and MBSs in the considered HetNet be transformed into one or more pools of virtual BBUs plus the corresponding RRHs. Of course, this kind of cloudification could be flexibly done for only part of traditional SBSs and part of traditional MBSs. One situation, which is preferred by most of the exiting literature, is to do the cloudification for all of

TABLE 4. Overview of BS switch-off strategies in cloud radio access networks.

Literature	Strategy	Contributions
[64-65]	Switch off BBUs in CRAN without HetNet deployment	Formulate a network power/energy consumption minimization problem to minimize the number of active BBUs:
		[64]: The problem is solved for static traffic load.
		[65]: Each BBU is modeled as having two independent capacities including user processing capacity and cell processing capacity. The problem is solved for static and dynamic traffic load, respectively.
[66-67]	Switch off BBUs in CRAN without HetNet deployment	Design traffic load balancing based strategies to assign a minimum number of active BBUs to the RRHs:
		[66]: Two resource usage thresholds (including upper and lower thresholds) are used in its heuristic solution.
		[67]: One workload threshold (set for triggering BBU sleeping) is used in its heuristic solution; further, construct a testbed for concept verification and practical performance evaluation.
[68-70]	Switch off RRHs in CRAN without HetNet deployment	Joint design for RRH switch-off strategy and coordinated beamforming among active RRHs, with the objective of minimizing the total power consumption of RRHs and the corresponding fronthaul links:
		[68]: Besides providing one greedy algorithm, design lower-complexity algorithms for an equivalent sparsity-based representation of the focused optimizaton problem. Moreover, downlink SINR is used for QoS constraint.
		[69]: Foucs on the case that the considered set of UEs is divided into multiple non-overlapping multicast subsets. Further, reformulate the problem to be maximizing the number of supported users via user admission control in case that the origianlly focused power consumption minimization problem is unable to find feasible solutions. Moreover, downlink SINR is used for QoS constraint.
		[70]: The focused power consumption minimization is formulated as a stochastic optimization problem, based on considering random traffic arrivals and time-varying channel conditions. Moreover, traffic delay is used for QoS constraint.
[17]	Switch off RRHs in H-CRAN	Consider a H-CRAN where all of traditional SBSs are cloudified but none of traditional MBSs is cloudified, and mention that EE could be improved if the number of activated RRHs is adaptive to traffic volume.
[18]	Switch off RRHs in H-CRAN	Consider a H-CRAN where only part of traditional SBSs are cloudified and none of traditional MBSs is cloudified, and mention that energy saving could be achieved via minimizing the number of active RRHs and traditional MBSs/SBSs.

traditional SBSs but for none of traditional MBSs. To the best of the authors' knowledge, none of the existing literature has clearly represented the network architecture of H-CRANs like this. Further, the authors of this survey paper use Fig. 4 to illustrate flexible constructions for the network architecture of H-CRANs.

It can be seen from Fig. 4 that, if every large-scale coverage area in a H-CRAN is constructed like Coverage

Area Type I, the extreme situation with full cloudification occurs. Of course, the mixed deployment of Coverage Area Types I, II, and III is also possible for a H-CRAN. Further, one currently preferred situation is to make each large-scale coverage area be constructed like Coverage Area Type III. In this case, while being used to guarantee the backward compatibility with the existing cellular systems, traditional MBSs could also be used to deliver all the control signaling in a H-CRAN so that the capacity and delay constraints on fronthaul links are alleviated.

For H-CRANs, there has been little literature to discuss the switch-off strategy for "virtual BBUs in the BBU pool" and/or "RRHs plus the corresponding fronthaul links". Further, in quite a few related works which we can find (e.g., see [17], [18]), the switch-off strategy is just simply mentioned at a concept level. In [17], with focusing on a H-CRAN constructed by deploying all large-scale coverage areas with only the way of Coverage Area Type III shown in Fig. 4, it is briefly mentioned that the number of activated RRHs should be adaptive to traffic volume so that the EE performance of the focused H-CRAN could be improved. In [18], for a H-CRAN which is constructed by deploying all large-scale coverage areas with only the way of Coverage Area Type II shown in Fig. 4, with using the concept-level statement, it is pointed out that the dynamic energy saving could be achieved via minimizing the number of active RRHs and traditional MBSs/SBSs while ensuring the desired QoS.

C. SUMMARY AND FUTURE WORK

The comparative summary for the literature reviewed in this section can be seen in Table 4.

Summary: The main findings of this section can be found as below:

1) In CRANs or H-CRANs, for the cloudified BSs, the design of switch-off strategy would be done through making either "virtual BBUs in the BBU pool" or "RRHs plus the corresponding fronthaul links" enter into the sleeping mode.

2) With the help of centralized baseband signal processing in the BBU pool, CoMP transmission among RRHs would be implemented more easily; therefore, it is meaningful to perform joint design for RRH switch-off strategy and coordinated beamforming among active RRHs.

Future Work: Much more research efforts would be needed to investigate the switch-off strategy for cloudified BSs in H-CRANs, although some switch-off strategies already designed in CRANs could be directly applied to H-CRANs. Further, for the cloudified BSs, it will be quite attractive to investigate how to simultaneously minimize the number of active virtual BBUs and the number of active RRHs.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we discussed the state-of-the-art research on BS switch-off design for greener 5G systems with the deployment of HetNets. Optimization objectives and constraints for designing BS switch-off strategy, specific BS switch-off strategies, joint design for BS switch-off strategy and another strategies which are important for 5G systems, and research results about this topic in emerging CRANs/H-CRANs were introduced. For readers to easily find interesting references which have been presented in this paper, Tables 1, 2, 3, 4 give the comparative summary for each of the above four aspects, respectively.

Existing research works show that well-designed BS switch-off strategies (either with joint design considering also another strategies or not) could significantly reduce the energy consumption of the entire network and improve overall network performance. Nevertheless, current research results are still quite preliminary; enhancement and challenges remain to be investigated in the future.

As for the specific research directions in the future, besides what have been presented in Section II-D, Section III-E, Section IV-D, and Section V-C, the design under more advanced HetNet architectures in 5G systems need to be considered too. Specifically, with the introduction of Millimeter wave (mmWave) frequencies in 5G systems, the desired deployment is a mixed sub-6GHz and mmWave HetNet deployment where mmWave BSs act as small cells and sub-6GHz BSs act as MBSs [71], [72]; a further inclusion of sub-6GHz SBSs also needs to be considered. Clearly, new design issues would exist when applying the BS switch-off mechanism to this kind of novel 5G HetNets.

REFERENCES

[1] *The 1000x Data challenge*, Qualcomm, San Diego, CA, USA, 2013.

[2] *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013–2018*, Cisco, San Jose, CA, USA, 2014.

[3] *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2014–2019*, Cisco, San Jose, CA, USA,, 2015.

[4] "5G wireless technology architecture," IMT-2020(5G) Promotion Group, White Paper, 2015.

[5] A. Cimmino *et al.*, "The role of small cell technology in future smart city applications," *Trans. Emerg. Telecommun. Technol.*, vol. 25, no. 1, pp. 11–20, Jan. 2014.

[6] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[7] N. Bhushan *et al.*, "Network densification: The dominant theme for wireless evolution into 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 82–89, Feb. 2014.

[8] J. B. Rao and A. O. Fapojuwo, "A survey of energy efficient resource management techniques for multicell cellular networks," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 154–180, 1st Quart., 2014.

[9] I. Chih-Lin, C. Rowell, S. Han, Z. Xu, G. Li, and Z. Pan, "Toward green and soft: A 5G perspective," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 66–73, Feb. 2014.

[10] C.-X. Wang *et al.*, "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 122–130, Feb. 2014.

[11] A. De Domenico, E. C. Strinati, and A. Capone, "Enabling green cellular networks: A survey and outlook," *Comput. Commun.*, vol. 37, pp. 5–24, Jan. 2014.

[12] Y. Li, H. Celebi, M. Daneshmand, C. Wang, and W. Zhao, "Energy-efficient femtocell networks: Challenges and opportunities," *IEEE Wireless Commun.*, vol. 20, no. 6, pp. 99–105, Dec. 2013.

[13] D. Feng, C. Jiang, G. Lim, L. J. Cimini, Jr., G. Feng, and G. Y. Li, "A survey of energy-efficient wireless communications," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 167–178, 1st Quart., 2013.

[14] X. Chen, J. Wu, Y. Cai, H. Zhang, and T. Chen, "Energy-efficiency oriented traffic offloading in wireless networks: A brief survey and a learning approach for heterogeneous cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 4, pp. 627–640, Apr. 2015.

[15] R. L. Cavalcante, S. Stanczak, M. Schubert, A. Eisenblaetter, and U. Türke, "Toward energy-efficient 5G wireless communications technologies: Tools for decoupling the scaling of networks from the growth of operating power," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 24–34, Nov. 2014.
[16] G. Wu, C. Yang, S. Li, and G. Y. Li, "Recent advances in energy-efficient networks and their application in 5G systems," *IEEE Wireless Commun.*, vol. 22, no. 2, pp. 145–151, Apr. 2015.
[17] M. Peng, Y. Li, J. Jiang, J. Li, and C. Wang, "Heterogeneous cloud radio access networks: A new perspective for enhancing spectral and energy efficiencies," *IEEE Wireless Commun.*, vol. 21, no. 6, pp. 126–135, Dec. 2014.
[18] M. Peng, Y. Li, Z. Zhao, and C. Wang, "System architecture and key technologies for 5G heterogeneous cloud radio access networks," *IEEE Netw.*, vol. 29, no. 2, pp. 6–14, Mar./Apr. 2015.
[19] M. A. Marsan, L. Chiaraviglio, D. Ciullo, and M. Meo, "Multiple daily base station switch-offs in cellular networks," in *Proc. IEEE Int. Conf. Commun. Electron. (ICCE)*, Aug. 2012, pp. 245–250.
[20] P. Dini, M. Miozzo, N. Bui, and N. Baldo, "A model to analyze the energy savings of base station sleep mode in LTE HetNets," in *Proc. IEEE Int. Conf. Green Comput. Commun. (GreenCom)*, Aug. 2013, pp. 1375–1380.
[21] G. Auer *et al.*, "How much energy is needed to run a wireless network?" *IEEE Trans. Wireless Commun.*, vol. 18, no. 5, pp. 40–49, Oct. 2011.
[22] C. Liu, B. Natarajan, and H. Xia, "Small cell base station sleep strategies for energy efficiency," *IEEE Commun. Veh. Technol.*, vol. 65, no. 3, pp. 1652–1661, Mar. 2016.
[23] Y. S. Soh, T. Q. S. Quek, M. Kountouris, and H. Shin, "Energy efficient heterogeneous cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 5, pp. 840–850, May 2013.
[24] C.-Y. Chang, K.-L. Ho, W. Liao, and D.-S. Shiu, "Capacity maximization of energy-harvesting small cells with dynamic sleep mode operation in heterogeneous networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2014, pp. 2690–2694.
[25] X. Guo, Z. Niu, S. Zhou, and P. R. Kumar, "Delay-constrained energy-optimal base station sleeping control," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1073–1085, May 2016.
[26] Z. Niu, X. Guo, S. Zhou, and P. R. Kumar, "Characterizing energy-delay tradeoff in hyper-cellular networks with base station sleeping control," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 4, pp. 641–650, Apr. 2015.
[27] B. Liu, M. Zhao, W. Zhou, J. Zhu, and P. Dong, "Flow-level-delay constraint small cell sleeping with macro base station cooperation for energy saving in HetNet," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, Sep. 2015, pp. 1–5.
[28] I. Aydin, H. Yanikomeroglu, and U. Aygolu, "User-aware cell switch-off algorithms," in *Proc. Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Aug. 2015, pp. 1236–1241.
[29] S. Song, Y. Chang, X. Wang, and D. Yang, "Coverage and energy modeling of HetNet under base station on-off model," *ETRI J.*, vol. 37, no. 3, pp. 450–459, Jun. 2015.
[30] G. Yu, Q. Chen, and R. Yin, "Dual-threshold sleep mode control scheme for small cells," *IET Commun.*, vol. 8, no. 11, pp. 2008–2016, Jul. 2014.
[31] S. Paris, F. Martignon, I. Filippini, and L. Chen, "A bandwidth trading marketplace for mobile data offloading," in *Proc. IEEE INFOCOM*, Apr. 2013, pp. 430–434.
[32] W. Dong *et al.*, "iDEAL: Incentivized dynamic cellular offloading via auctions," *IEEE/ACM Trans. Netw.*, vol. 22, no. 4, pp. 1271–1284, Aug. 2014.
[33] A. Bousia, E. Kartsakli, A. Antonopoulos, L. Alonso, and C. Verikoukis, "Multiobjective auction-based switching off scheme in heterogeneous networks to bid or not to bid?" *IEEE Trans. Veh. Technol.*, to be published.
[34] E. Altman *et al.*, "Stochastic geometric models for green networking," *IEEE Access*, vol. 3, pp. 2465–2474, Nov. 2015.
[35] D. Tsilimantos, J.-M. Gorce, and E. Altman, "Stochastic analysis of energy savings with sleep mode in OFDMA wireless networks," in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, Apr. 2013, pp. 1097–1105.
[36] Y. S. Soh, T. Q. Quek, and M. Kountouris, "Dynamic sleep mode strategies in energy efficient cellular networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2013, pp. 3131–3136.
[37] J. Peng, P. Hong, and K. Xue, "Stochastic analysis of optimal base station energy saving in cellular networks with sleep mode," *IEEE Commun. Lett.*, vol. 18, no. 4, pp. 612–615, Apr. 2014.
[38] A. Bousia, A. Antonopoulos, L. Alonso, and C. Verikoukis, "'Green' distance-aware base station sleeping algorithm in LTE-advanced," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2012, pp. 1347–1351.
[39] S. Cai, Y. Che, L. Duan, J. Wang, S. Zhou, and R. Zhang, "Green 5G heterogeneous networks through dynamic small-cell operation," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1103–1115, May 2016.
[40] A. A. Abdulkafi, T. S. Kiong, D. Chieng, A. Ting, and J. Koh, "Energy efficiency improvements in heterogeneous network through traffic load balancing and sleep mode mechanisms," *Wireless Pers. Commun.*, vol. 75, no. 4, pp. 2151–2164, Apr. 2014.
[41] W. Vereecken *et al.*, "Evaluation of the potential for energy saving in macrocell and femtocell networks using a heuristic introducing sleep modes in base stations," *EURASIP J. Wireless Commun. Netw.*, vol. 2012, no. 1, pp. 1–14, Dec. 2012.
[42] L. Saker, S.-E. Elayoubi, R. Combes, and T. Chahed, "Optimal control of wake up mechanisms of femtocells in heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 664–672, Apr. 2012.
[43] S. Navaratnarajah, A. Saeed, M. Dianati, and M. A. Imran, "Energy efficiency in heterogeneous wireless access networks," *IEEE Wireless Commun.*, vol. 20, no. 5, pp. 37–43, Oct. 2013.
[44] X. Gan *et al.*, "Energy efficient switch policy for small cells," *China Commun.*, vol. 12, no. 1, pp. 78–88, Jan. 2015.
[45] J. Wu, S. Jin, L. Jiang, and G. Wang, "Dynamic switching off algorithms for pico base stations in heterogeneous cellular networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2015, no. 1, pp. 1–18, Dec. 2015.
[46] M. Oikonomakou, A. Antonopoulos, L. Alonso, and C. Verikoukis, "Cooperative base station switching off in multi-operator shared heterogeneous network," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–6.
[47] E. Yaacoub, "A practical approach for base station on/off switching in green LTE-A HetNets," in *Proc. IEEE Int. Conf. Wireless Mobile Comput., Netw. Commun. (WiMob)*, Oct. 2014, pp. 159–164.
[48] M. A. Marsan and M. Meo, "Network sharing and its energy benefits: A study of European mobile network operators," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2013, pp. 2561–2567.
[49] G. Wu, L. Dong, Z. Qin, and Z. Xu, "Dynamic programming-based pico base station sleep mode control in heterogeneous networks," *Int. J. Commun. Syst.*, to be published, doi: 10.1002/dac.2967.
[50] J. He *et al.*, "Energy efficient BSs switching in heterogeneous networks: An operators perspective," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2016.
[51] M. Feng, S. Mao, and T. Jiang, "BOOST: Base station on-off switching strategy for energy efficient massive MIMO HetNets," in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, Apr. 2016, pp. 1–9.
[52] L. Zhou *et al.*, "Green small cell planning in smart cities under dynamic traffic demand," in *Proc. IEEE Int. Conf. Comput. Commun. Workshops (INFOCOM)*, Apr./May 2015, pp. 618–623.
[53] G. K. Tran, H. Shimodaira, R. E. Rezagah, K. Sakaguchi, and K. Araki, "Dynamic cell activation and user association for green 5G heterogeneous cellular networks," in *Proc. IEEE Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Aug./Sep. 2015, pp. 2364–2368.
[54] P. Ren and M. Tao, "A decentralized sleep mechanism in heterogeneous cellular networks with QoS constraints," *IEEE Wireless Commun. Lett.*, vol. 3, no. 5, pp. 509–512, Oct. 2014.
[55] S. Zhang, J. Wu, J. Gong, S. Zhou, and Z. Niu, "Energy-optimal probabilistic base station sleeping under a separation network architecture," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2014, pp. 4239–4244.
[56] X. Chai, Y. Li, Y. Lv, and Z. Zhang, "Joint spectrum-sharing and base-station-sleep model for improving energy efficiency of HetNet," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 1851–1856.
[57] G. Bartoli, L. Carla, R. Fantacci, and D. Marabissi, "Efficient multicast beamforming in HetNet with an opportunistic cells activation," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2015, pp. 1–6.
[58] W.-C. Liao, M. Hong, Y.-F. Liu, and Z.-Q. Luo, "Base station activation and linear transceiver design for optimal resource management in heterogeneous networks," *IEEE Trans. Signal Process.*, vol. 62, no. 15, pp. 3939–3952, Aug. 2014.
[59] A. Adhikary, H. S. Dhillon, and G. Caire, "Massive-MIMO meets HetNet: Interference coordination through spatial blanking," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 6, pp. 1171–1186, Jun. 2015.
[60] "LTE-advanced carrier aggregation optimization," Nokia, Espoo, Finland, White Paper, 2015.
[61] H. Wang, C. Rosa, and K. I. Pedersen, "Dual connectivity for LTE-Advanced heterogeneous networks," *Wireless Netw.*, vol. 22, no. 4, pp. 1315–1328, May 2016.
[62] H. Elshaer, F. Boccardi, M. Dohler, and R. Irmer, "Downlink and uplink decoupling: A disruptive architectural design for 5G networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2014, pp. 1798–1803.

[63] F. Boccardi *et al.*, "Why to decouple the uplink and downlink in cellular networks and how to do it," *IEEE Commun. Mag.*, vol. 54, no. 3, pp. 110–117, Mar. 2016.

[64] T. Sigwele, A. S. Alam, P. Pillai, and Y. F. Hu, "Evaluating energy-efficient cloud radio access networks for 5G," in *Proc. IEEE Int. Conf. Data Sci. Data Intensive Syst.*, Dec. 2015, pp. 362–367.

[65] X. Wang *et al.*, "Energy-efficient virtual base station formation in optical-access-enabled cloud-RAN," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1130–1139, May 2016.

[66] M. Khan, R. Alhumaima, and H. Al-Raweshidy, "Reducing energy consumption by dynamic resource allocation in C-RAN," in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, Jun./Jul. 2015, pp. 169–174.

[67] Z. Kong, J. Gong, C.-Z. Xu, K. Wang, and J. Rao, "EBase: A baseband unit cluster testbed to improve energy-efficiency for cloud radio access network," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2013, pp. 4222–4227.

[68] Y. Shi, J. Zhang, and K. B. Letaief, "Group sparse beamforming for green cloud-RAN," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2809–2823, May 2014.

[69] Y. Shi, J. Cheng, J. Zhang, B. Bai, W. Chen, and K. B. Letaief, "Smoothed lp-minimization for green cloud-RAN with user admission control," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 1022–1036, Apr. 2016.

[70] J. Li, J. Wu, M. Peng, and P. Zhang, "Queue-aware energy-efficient joint remote radio head activation and beamforming in cloud radio access networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 3880–3894, Jun. 2016.

[71] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[72] J. Park, S.-L. Kim, and J. Zander, "Tractable resource management with uplink decoupled millimeter-wave overlay in ultra-dense cellular networks," *IEEE Trans. Wireless Commun.*, to be published.

FENGXIA HAN is a Ph. D. student with Tongji University since 2015. She holds a bachelors degree from University of Electronics Science and Technology of China in the area of Wireless Communications. Her research interests include wireless big data, massive MIMO and heterogeneous networks.

SHENGJIE ZHAO (SM'09) received the B.S.degree in electrical engineering from the University of Science and Technology of China (USTC), Hefei, China, in 1988; the M.S. degree in electrical and computer engineering from China Aerospace Institute, Beijing, China, in 1991; and the Ph.D. degree in electrical and computer engineering from Texas A & M University, College Station, TX, USA, in 2004. He is a Professor with the College of Electronics and Information Engineering, Tongji University, Shanghai, China. In previous postings, he conducted research at Lucent Technologies, Whippany, NJ, USA, and China Aerospace Science and Industry Corporation, Beijing, China. His research interests include big data, wireless communications, image processing, and signal processing. Dr. Zhao is a PI of 4th Topic of Project of Big Data and the National Basic Research Program of China (973Program), and was a PI of TD-LTE-Advanced System Research and Development, National Science and Technology Specific Projects of China. He is a Fellow of the Thousand Talents Program of China and is a Member of Sigma Xi..

LU ZHANG received the Ph.D degree in 2009 from the Department of Electrical and Computer Engineering, University of Delaware, USA. And, he received the M.E. and B.E. degrees in 2002 and 1999, respectively, from the Department of Electronic Engineering, Beijing Institute of Technology, China. He is currently a Product Line Manager for 5G products at the Nokia Shanghai Bell Co., Ltd. (which is the former Alcatel-Lucent Shanghai Bell Co., Ltd.). Since 2009, he joined Alcatel-Lucent Shanghai Bell Co., Ltd., where he was a Research Scientist at the Bell Labs China from 2009 to 2012 and began to act as a Product Line Manager since 2012. He was awarded as Distinguished Member of Technical Staff (title for life) in 2015, and awarded as Member of the Alcatel-Lucent Technical Academy in 2011. From November 2008 to March 2009, he was a full-time Research Intern at the Mitsubishi Electric Research Laboratories, Cambridge, USA. From January 2004 to August 2004, he was a System Engineer at the Siemens Ltd., China. From April 2002 to January 2004, he was a R&D Engineer at the Beijing Research Institute, Huawei Technologies Co., Ltd. His research interests include emerging techniques in 5G systems, interference management for small cells and/or heterogeneous networks (HetNet), coordinate multi-point (CoMP) transmission technologies in LTE-Advanced systems, carrier aggregation technologies in LTE-Advanced systems, theoretical analyses for multi-cell co-channel interference in MIMO-based cellular networks, decentralized cooperative relaying techniques in adhoc networks, space-time coding in MIMO systems, etc.

JINSONG WU (SM'11) is the Founder and Founding Chair of IEEE Technical Committee on Green Communications and Computing (TCGCC), which was established in 2011 as an official Technical Subcommittee (TSCGCC) and elevated as TCGCC in 2013. He is also the Co-Founder and Founding Vice-Chair of IEEE Technical SubCommittee on Big Data (TSCBD), established in January 2015, He is Series Editor of IEEE Journal on Selected Areas in Communications Series on Green Communications and Networking, Editor of IEEE Communications Surveys & Tutorials, Associate Editor of IEEE Systems Journal, Associate Editor of IEEE Access. He is the founder and Series Editor of IEEE Series on Green Communication and Computing Networks in IEEE Communications Magazine. He has served as a co-leading Guest Editor of Special Issue on Green Communications, Computing, and Systems in IEEE Systems Journal, and Associate Editor of Special Section on Big Data for Green Communications and Computing in IEEE Access. He was the leading Editor and a co-author of the comprehensive book, entitled Green Communications: Theoretical Fundamentals, Algorithms, and Applications, published by CRC Press in September 2012.