Energy-Efficient and Load-Proportional eNodeB for 5G User-Centric Networks

Haris Pervaiz, Oluwakayode Onireti, Abdelrahim Mohamed, Muhammad Ali Imran, Rahim Tafazolli and Qiang Ni

***Abstract*—Nowadays, dense network deployment is being considered as one of the effective strategies to meet capacity and connectivity demands of the fifth generation (5G) cellular system. Among several challenges, energy consumption will be a critical consideration in the 5G era. In this direction, base station on/off operation, i.e., sleep mode, is an effective technique to mitigate the excessive energy consumption in ultra-dense cellular networks. However, current implementation of this technique is unsuitable for dynamic networks with fluctuating traffic profiles due to coverage constraints, quality-of-service requirements and hardware switching latency. In this direction, we propose an energy/load proportional approach for 5G base stations with control/data plane separation. The proposed approach depends on a multi-step sleep mode profiling, and predicts the base station vacation time in advance. Such a prediction enables selecting the best sleep mode strategy whilst minimizing the effect of base station activation/reactivation latency, resulting in significant energy saving gains.**

I. Introduction

ENERGY Energy saving technologies for cellular communications system has recently received a lot of attention in the context of growing energy demand and increasing energy prices. Many international research projects on energy saving have sprang up over the past decade, e.g. GreenTouch, Energy Aware Radio and network technologies (EARTH), mobile VCE Green-Radio, and have reported some energy savings. Most of these projects found that radio access network (RAN) nodes of current systems, such as the long term evolution (LTE), consume high energy even in low traffic conditions. In other words, the base stations (BSs) are characterized by (almost) load-independent power consumption profiles. As an illustrative example, the EARTH power model shows that the LTE pico BS without load consumes 92.9% of the power consumed by a fully loaded pico BS. Such a profile can be traced to the always-on service approach adopted in traditional RANs in addition to the BS component design and power characteristics.

In this direction, more energy can be saved by fully adapting the energy consumption of BSs to the traffic, as the state of the art BSs still consume significant amount of power in low traffic hours and active-idle (sleep) mode. In the sleep mode, the BS does not provide data services but consumes less energy than BSs in the on state. However, energy consumption in the sleep mode is still significant because the sleeping BSs has to be quickly activated when needed. The main power supply, DC-DC power supply and the baseband components, which have a very long reactivation time, must stay on in this mode. Consequently, the baseline power consumption of a small BS in sleep mode remains high and could reach about 50% of the peak values [1]. In the off state, the BS is deactivated completely, i.e., almost all of the components in its circuitry are switched off. Hence, the BS energy consumption in the off state is negligible. Based on the current technology, the time to transmission from the off to the on state is non-negligible, and this may violate the quality of service (QoS) constraints especially when the traffic increases suddenly. As a result, the state of the art energy saving schemes only focus on putting BSs in the sleep mode, which consumes a non-negligible energy even without any traffic.

Achieving a fully energy proportional network requires adopting deep sleep modes, i.e., partial deactivation of the hardware component, and a fast wake-up. The evolution of the 5G RAN with new concepts, such as the dual connectivity RAN with control/data plane separation, can be advantageous in the light of achieving an energy proportional network. In such architecture, connectivity and data transmission are provided by separate nodes, macro BS and small BS respectively. This provides a degree of freedom because the small BSs can be put in deep sleep modes or completely switched off without affecting the basic mobility and connectivity services [2], [3], [4]. As a result, the RAN energy consumption can be significantly reduced by adopting an on-demand approach with deep sleep modes in the small BSs. This in turns allows the small BSs to stay in deep sleep for longer time. Moreover, the control-data separated architecture (CDSA) has inbuilt support for a deeper sleep state at data BSs, due to its mobility management functionality being located at the control plane.

In this paper, we propose an energy/load proportional approach for BSs of the dual connectivity CDSA. This architecture has been adopted in the 5G RAN standard. The proposed approach depends on estimating vacation period of the BS in advance in order to optimally match its vacation time with the sleep depth. The latter compromises of the deactivation duration, actual sleep duration and reactivation duration. We demonstrates that the power consumed in the actual sleep duration is an exponential decaying function of the deactivation duration. Numerical results demonstrate the effectiveness of finding the optimal deactivation duration for each sleep depth resulting in significant level of power saving gains for 5G networks.

H. Pervaiz **(Corresponding Author)**, A. Mohamed and R. Tafazolli are with Institute of Communication Systems, Home of 5GIC, University of Surrey, UK. Emails: {h.pervaiz, abdelrahim.mohamed, r.tafazolli}@surrey.ac.uk.
O. Onireti and M. Imran are with the School of Engineering, University of Glasgow, UK. Email: {oluwakayode.onireti, muhammad.imran, }@glasgow.ac.uk.
Q. Ni is with the School of Computing and Communications, Lancaster University, UK. Email: q.ni@lancaster.ac.uk.

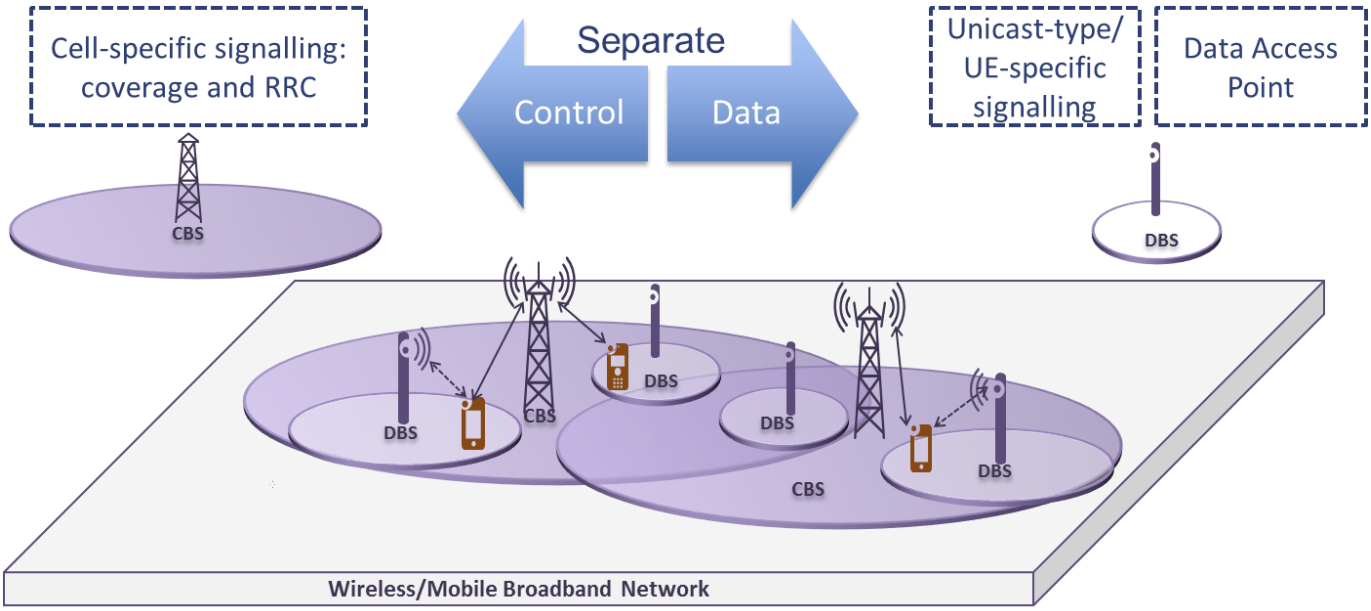

Figure 1: High Level Overview of Control/Data Separation Architecture [3]

The reminder of this paper is structured as follows. Section II provides an overview of the CDSA. Section III presents the sleep mode strategies and the proposed load/energy proportional approach. Section IV presents and discusses performance results, while Section V draws key conclusions.

II. RAN with Control/Data Plane Separation

BS on/off operation via sleep modes is one of the most effective techniques to mitigate the excessive energy consumption in ultra-dense cellular network. However, current implementation of this technique is unsuitable for dynamic networks with fluctuating traffic patterns [5]. Coverage constraints limit the achievable gains of BS sleep mode in the conventional RAN architecture. This can be traced to the coupling between connectivity and data access points which require an always-on RAN regardless of the actual traffic profile. Expressed differently, most of the BS sleep mode challenges and limitations originate from the control plane (CP) and data plane (DP) coupling approach adopted in the conventional RAN. In this direction, separating the connectivity services (provided by the CP) and the data services (provided by the DP) produces a framework with relaxed constraints for BS sleep modes.

The basic concept behind the CDSA is the fact that ubiquitous connectivity does not imply high data rate transmission. The latter is needed on demand which suggests an adaptive on/off data layer complemented by an always-on coverage layer. In this direction, the CDSA consists of two layers as can be seen in Fig. 1.

1) Control base station (CBS) layer: Provides ubiquitous connectivity, and consists of macro BSs operate in low frequency bands.
2) Data base station (DBS) layer: Efficient small BSs deployed within the CBS footprint to provide data transmission and on-demand services.

As shown in Fig. 1, idle users are anchored to the CBS only. On the other hand, active users establish a link with the DBS (for data transmission) in addition to the connectivity link with the CBS [3]. In other words, the DBS connection is only needed for active users as the connectivity services are provided by the CBS layer. As a result, the DBS can be dynamically switched on and off depending on the traffic profile. At the hardware level, the CDSA allows independent design for the DBS and the CBS components. For instance, the CBS power amplifier can be independently designed to operate near the saturation region for low rate coverage services with low order constant envelope modulations.

In ultra-dense deployment scenarios, the DBS density will be high while the CBS density could be low to moderate. Thus the linear relationship between the energy consumption and the network density suggests focusing on the DBS layer. Fortunately, most of the network adaptation techniques can be used with relaxed constraints in the DBSs. The run time energy saving approaches and the sleep modes could benefit from the flexible DBS adaptation opportunities when mobility management and network connectivity are delegated to the CBS. This in turns increases the energy saving gains [6], [7] and improves the energy/load scaling profile [8]. In particular, by using the conventional single-level sleep strategy in [9], the CDSA can save more than one-third of the overall network energy consumed in the conventional cellular architecture [8]. Further energy savings can also be achieved with the CDSA by the utilization of optimized sleep mode techniques to dynamically change the DBS sleep level. Moreover, the DBS

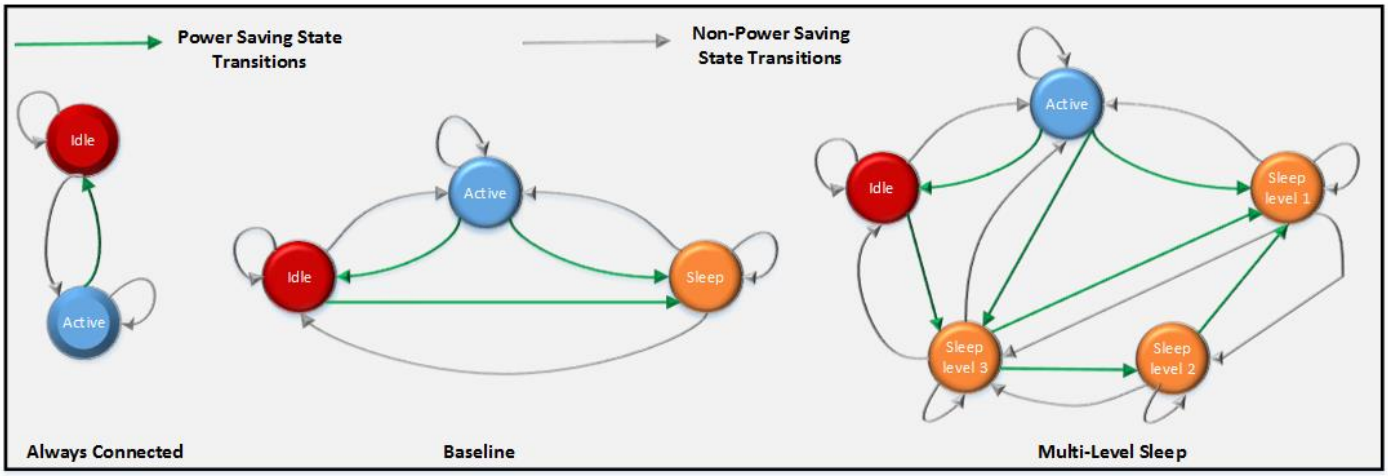

Figure 2: Possible DBS State Transitions for Different Sleep Strategies

also provides one-to-one data transmission with user-specific signals rather than cell-specific signals [3]. This reduction in signalling overhead can as well be translated into an increased energy saving gain.

The CDSA allows network-driven DBS-user equipment (UE) association strategies. Typically, idle UE request resources from the CBS when they start a data session. The CBS chooses the best serving DBS, switches it on, and associates the UE with the chosen DBS. From a delay perspective, [10] estimates that the dual connection feature allows reducing the time to use a just turned on DBS from 1100 ms (in the conventional RAN) to 240 ms (in the CDSA). In addition, system level simulation results reported in [11] show that the BS sleep modes provide throughput gains of 10%–20% when the small BSs are switched on whenever a UE is associated with them (even if the UE is idle as in the conventional RAN). When the small BSs are switched on only when there are active UE (as in the CDSA), the throughput gain reaches 30%–110%. Thus it can be said that the CDSA has a built in feature to support the network-driven sleep mode methods with a lower delay, a lower on/off oscillations, a higher energy efficiency and a higher QoS.

III. A Tractable and Analytical DBS Power Consumption Model for 5G User-Centric Networks

One of the requirements for 5G is the increase in energy efficiency, i.e., the bit-per-joule capacity by at least 10-100$\times$. Technologies such as network densification, massive MIMO and millimeter wave are all aimed towards achieving a high bit rate in 5G networks. However, network densification in particular could result in an increase in the networks' power consumption and consequently, a reduction in the energy efficiency if not well managed. Nevertheless, the 5G energy efficiency target can be achieved by implementing architectures such as the CDSA, a user-centric network and an efficient BS switching on and off strategy. In this section, we introduce the state of the art BS sleep strategy for the CDSA architecture and we propose an optimized user-centric and multi-level sleep strategy which can achieve the 5G energy efficiency target.

In this trend, the authors in [5] presented four different sleep modes[1] with each mode achieving different power savings, which is related to the BS's hardware capability. The sleep mode 1 corresponds to the shortest BS sleeping time which is of the OFDM symbol duration, i.e. 71μs. However, the BS remains fully operational and is able to receive data. Hence, the sleep mode 1 can be used when the BS is not transmitting actively. The sleep mode 2, on the other hand, is the medium sleep state where more components go into the sleeping state that corresponds to the LTE subframe duration, i.e., 1ms. In sleep mode 3, most of the BS's components are deactivated and it can be referred to as a slow sleep mode which corresponds to the duration of 1 LTE radio frame, i.e. 10ms. The last sleep mode is sleep mode 4, which also corresponds to the BS standby mode. In this mode, the BS is out of operation but it can be woken up. The time unit of the sleep mode 4 is defined by its minimum duration, which is 1s. In [12], using the four sleep modes defined in [5][1], the authors evaluated the impact of sleeping BS on the overall BS energy consumption. Their results showed a gain of about 22% in energy saving gains. In [13], the authors proposed an advanced sleep modes (ASMs) where BS are gradually deactivated in order to achieve a reduction in energy consumption. The proposed scheme allows for the management of users whose service request happens when the BS is in sleep mode. In [14], the authors utilized a curve fitting function to approximate the sleep mode power consumption as a function of the sleep depth. Their evaluation allows for selecting the optimal sleep depth for a given BS idle period while considering the BS deactivation and reactivation power consumption. In [15], the authors have formulated the EE resource allocation optimization problem in the downlink transmission scheme of a sparsely deployed

[1] IMEC, "Power model for today's and future base stations," Available: http://www.imec.be/powermodel

phantom cellular network with non-perfect CSI available at the BSs transmitters. It is also shown that the achievable EE deteriorates with an increase in the number of phantom cells.

A. Always Connected DBS Strategy

In the "Always On Always Connected" paradigm as shown in Fig. 2, it is assumed that the DBSs are always switched on, i.e., they are either in active or idle state. This kind of paradigm is more driven towards the QoS or spectral efficiency rather than the power efficiency or savings as the only possibility for the power saving occurs if the state transition from an initial state to a new state results in a lower power consumption. For example, if the DBS transits from an initial state of active with the power consumption of P_A to a new state idle with the power consumption P_I and stay in this new state for some duration will result in power saving gains if and only if $P_A > P_I$.

B. Conventional DBS Single-Level Sleep Strategy

It is important to highlight that the DBS in the "Baseline" paradigm or single-level sleep strategy as shown in Fig. 2 can operate in three states namely the sleep, idle or active states. The components of DBSs are considered to be completely active in both idle and active states whereas in the latter it is assumed to be receiving or transmitting data and in the idle state it is waiting for the arrival of the new user requests. If the DBS does not receive any user request for the hysteresis time interval T, then it will enter a sleep state and otherwise, it will immediately start its service remaining in an active state. The power consumption of a DBS in the sleep state (or unable to serve the user requests) denoted by P_S is lower than the active or idle states with their respective power levels of P_A and P_I. As the power consumption in the active state is far more than the other two states, the power savings can be achieved for all the state transitions originating from the active state. Similarly, the state transition from idle state to the sleep state will result in power savings gain due to the fact that the power consumption in the sleep state is always less than P_I. This kind of paradigm is more focused towards the power savings and can switch off (or switch to sleep state) DBS. The sleep state consist of the deactivation, actual sleep level and reactivation phases. After the DBS actual sleep duration expires, it requires setup time for the wake-up and after that if there is still no user request then it will enter an idle state.

C. Optimized Multi-Level Sleep Strategy for DBSs

In the proposed multi-level sleep strategy, the vacation period and operational time of a DBS can be estimated or predicted in advance using some intelligence in the network via self-organizing network (SON) concepts by applying the machine learning techniques such as support vector machine (SVM) regression model on the historical network traffic profile. Based on memory and history of the network, a sleep mode profile is defined for each DBS. This profile captures the statistics of the idle time duration and defines the DBS sleeping depth. The DBS-based sleep mode profile is considered as a location-based approach that implicitly takes into account the spatial variation of traffic demand and idle time duration. That's why a time varying DBS sleep mode profile is suggested as the pure DBS-based profile does not take into account the temporal variation of idle time duration.

In contrast to the aforementioned conventional single-level sleep strategy, the DBS can enter another vacation period if there is no user request for the DBS after the vacation period and the DBS vacation period can be rightly matched with the sleep depth that minimizes the average power consumption of a DBS. The objective is to provide the optimal matching of the sleep depth (i.e., optimized deactivation, actual sleep level and reactivation transition latency) with the DBS vacation period to maximize the power savings. Firstly, we present a tractable and analytical power consumption model for the various phases of the sleep depth of DBS for a given vacation period.

The transition of a DBS in the lth vacation period from the active state to the sleep state is termed as *deactivation* phase with the power level P_{D_l}. In the sleep state, the DBS can switch off only those components whose transition latency[2] is shorter than its vacation period v which is termed as *actual sleep* phase with the power level P_{S_l}. The transition of a DBS from the sleep state to the active state is defined as the *reactivation* phase with the power level P_{R_l}. After the DBS wake-up, it requires some further reactivation time to warm up and afterwards it can start serving the user requests. It should also be noted that the power consumption in this reactivation time captures the cost for the DBS state transition and is mostly higher than P_{I_l}. The DBS l^{th} vacation period v_l compromises of the component deactivation latency v_{D_l}, the actual sleep duration v_{S_l} and the component reactivation latency v_{R_l} as shown in Fig. 3[3]. From Fig. 3, we can also observe that the actual sleep duration for the DBS is less than the DBS vacation period. Further, the measurements in the EARTH project have shown that the reactivation transition latency is always higher than the deactivation transition latency. Similarly, the power consumed during the deactivation and reactivation phase is more than the power consumed in the actual sleep state.

The components are deactivated in a manner such that the components with the shorter deactivation transition latency[4] are the first to be deactivated. This process continues until no further component deactivation effect the power consumption of the DBS. The power consumption in the actual sleep duration P_S decreases with an increase in the DBS deactivation transition latency allowing the deactivation of more components as shown in Fig. 3. Similarly, for the given DBS vacation period v, the power consumed in the actual sleep level, P_S, is modelled as an exponential decay function of the component deactivation latency v_D, decay constant ω and idle state power consumption P_I. Since the deactivation phase arises due to the DBS transition from the idle to the sleep state, the consumed power is somehow dependent on both P_I and P_S. Hence, the

[2] It comprises the time interval duration for both deactivation and reactivation phases.

[3] From now onwards, we have dropped the subscript l for the convenience.

[4] It should be noted that each component deactivation latency is matched to its corresponding reactivation latency.

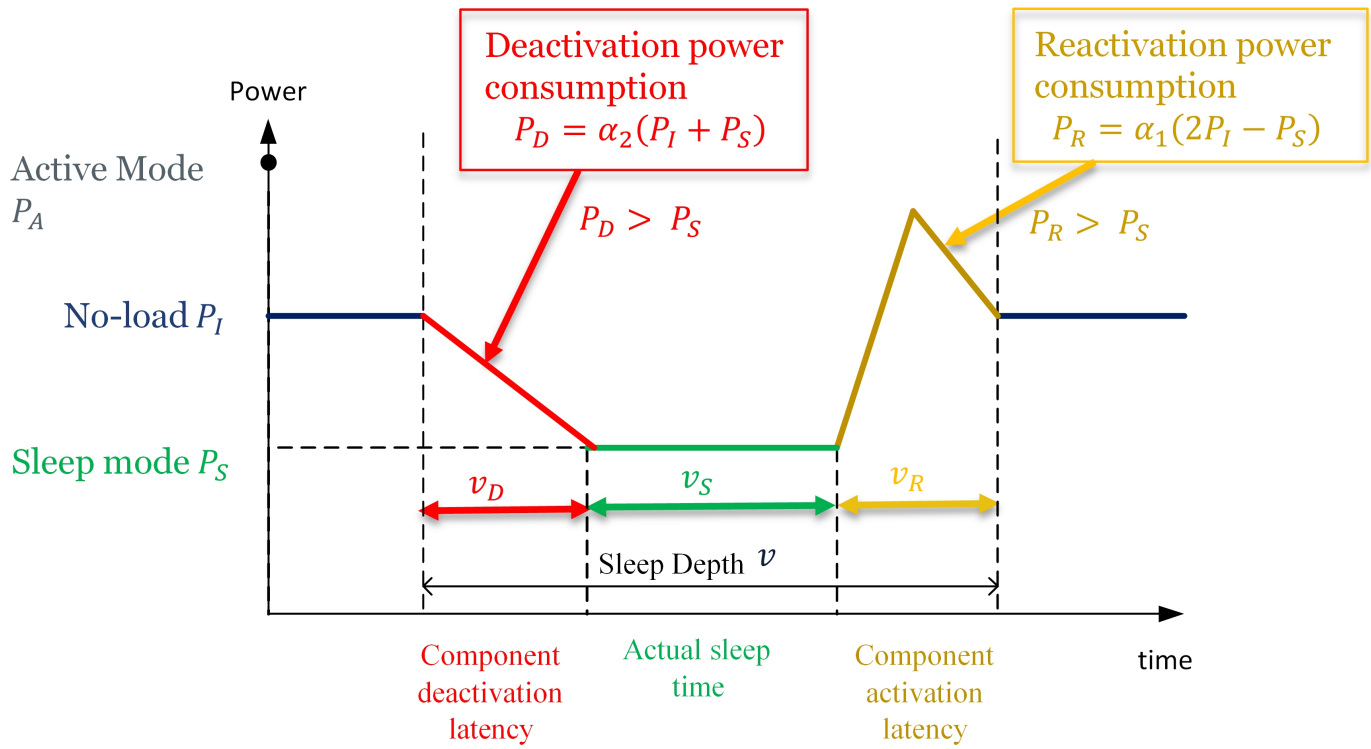

Figure 3: Average Power Consumption of DBS in the Sleep Mode

power consumption in the deactivation phase P_D is defined as the linear function of the mean consumed power over the transition period from idle state till when the actual sleep level is achieved. P_D decreases with an increase in the deactivation latency v_D which will result in decreasing the actual sleep level (or reduced sleep power consumption) and increasing the reactivation transition latency (or increasing the reactivation power consumption) due to the fact that more components need to be reactivated. Finally, the power consumption in the reactivation phase P_R is interpreted as the function of the power consumption in the actual sleep level and increases with an increase in the deactivation transition latency. It is also assumed that $\frac{v_R}{v_D} \geq N$, wherein $N \geq 1$ and due to this fact P_D, P_S and P_R can be represented in terms of the deactivation transition latency v_D. Increasing N implies an increase in the component reactivation latency will also cause a decrease in the optimal component deactivation latency as more power is consumed in the reactivation phase in comparison to the deactivation phase.

D. Discrete Multi-Level Sleep Strategy for DBSs

In this multi-level sleep strategy, the power consumption for deactivation and reactivation phases are computed in a similar manner as outlined in Section IV-C. The major difference is in the computation of the power consumption for the actual sleep level and different deactivation transition latency v_D dependent on the vacation period v for each discrete sleep level is predefined as shown in Table I. As the deactivation transition latency is predefined, we can utilize this information to compute the reactivation transition latency which is usually a scalar multiplier N times the deactivation transition latency and also compute the actual sleep duration in terms of the deactivation transition latency. The component deactivation, actual sleep level and component reactivation latency for this strategy are fixed in contrast to the previous strategy wherein the respective latency of the three phases were computed in an optimal manner to minimize the total power consumption.

Table I: Discrete Sleep level

v		$v < 0.5$	$0.5 < v < 1$	$1 < v < 2$	$v > 2$
1-Sleep level	v_D (s)	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
	Power (W)	P_I	$P_I e^{-\frac{v}{8}}$	$P_I e^{-\frac{v}{8}}$	$P_I e^{-\frac{v}{8}}$
2-Sleep levels	v_D (s)	0	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{4}$
	Power (W)	P_I	$P_I e^{-\frac{v}{8}}$	$P_I e^{-\frac{v}{4}}$	$P_I e^{-\frac{v}{4}}$
3-Sleep levels	v_D (s)	0	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$
	Power (W)	P_I	$P_I e^{-\frac{v}{8}}$	$P_I e^{-\frac{v}{4}}$	$P_I e^{-\frac{v}{2}}$

IV. Performance Evaluation

The DBS is made-up of several components/ subcomponents. Each DBS component/ subcomponent is characterized by its activation and deactivation latency. Let the set **B** with cardinality $|B|$ denote the set of all the DBS components/ subcomponents. The reactivation latency v_R^i of the ith DBS component/subcomponent is the period required for it to go

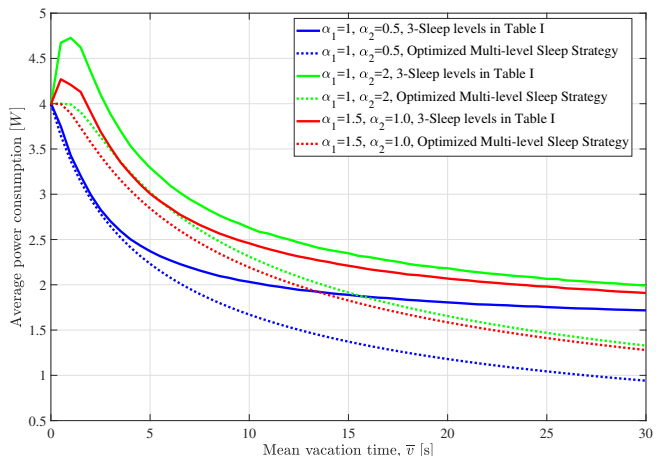

Figure 4: Average Power Consumption versus the DBS mean vacation time $\bar{v}$

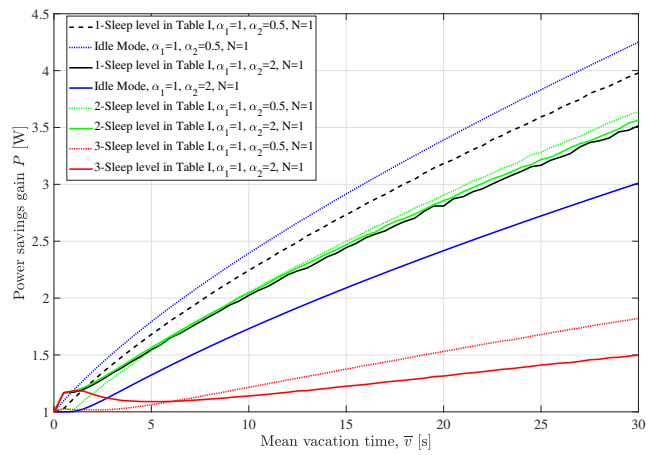

Figure 5: Power Saving gains due to the optimized deactivation latency versus the mean DBS vacation time $\bar{v}$

from sleep (OFF) mode to active (ON) mode, while the deactivation latency v_D^i of the ith DBS component/ subcomponent is the period required for it to go from active to sleep mode. The sum of the ith DBS component/subcomponent's activation and deactivation latency is termed as its transition latency v_i. Each DBS is characterized by Q discrete sleep modes. The qth sleep mode of a DBS is associated with a given duration v_q, where $q \in \{1, 2, \cdots, Q\}$ and $v_{(q+1)} > v_q$. All the DBS components/subcomponents that can enter and exit a sleep mode fast enough are considered sleeping in that mode, i.e., all DBS components/subcomponents with $v_i \geq v_q, \forall i \in \mathbf{B}$ are consider sleeping in that mode such that $\sum_{i \in \mathbf{B}} v_i = v_q$. While the subcomponents having a longer latency are considered to be still active, i.e., all DBS components/subcomponents with $v_i > v_q, i \in \mathbf{B}$ are consider active in that mode. Moving to a higher sleep level i.e. increasing v_q leads to deactivating more DBS components/subcomponents and a reduction in the sleep-mode power consumption.The system parameters are as follows: $P_I = 4$ W, $\omega = 2$, $v_D{}^{\max} = 2$ s, $N = 1$, $v^{\max} = 2\bar{v} - v^{\min}$, $v^{\min} = 71.4\mu$ s and $\tilde{v} = \frac{1}{2}(2\alpha_2 - \alpha_1 + 1)$.

Fig. 4 investigates the average power consumption against the mean vacation time $\bar{v}$, while considering a DBS uniformly distributed vacation time and $v_R = v_D$, for both the cases with the optimized multi-level sleep strategy and the discrete sleep level schemes as defined in Table I. In each discrete sleep level, the actual sleep duration of the vacation period can be obtained as v_S, where the discrete v_D defined in Table I is dependent on the vacation period. Hence, increasing the mean vacation time leads to a reduction in the average power consumption. By increasing the parameters α_1 and α_2 to 1 and 0.5, respectively, and consequently the power consumed during the reactivation and deactivation phases, the parameter $\widetilde{v}$ becomes greater than zero and $v_D^\star = 0, \forall v \leq \widetilde{v}$. Hence, in the optimal case, the DBS continues to operate in no-load for all vacation time which is less than $\widetilde{v}$ since the cost of deactivation and reactivation (increase power during the deactivation and reactivation processes) exceeds the gains from the reduced power consumption due to subcomponent deactivation. For the suboptimal case with 3-sleep levels, we observe that the average power consumption initially increases with $\bar{v}$ to its maximum for $\alpha_1 = 1, \alpha_2 = 2$ and $\alpha_1 = 1.5, \alpha_2 = 1$, before decreasing with further increase in $\bar{v}$. For the suboptimal case, we observe that the average power consumption could even be higher than the no-load power consumption for some mean DBS vacation time.

Fig. 5 depicts the power savings gain versus the mean vacation time $\bar{v}$ in order to evaluate the effectiveness of the optimized multi-level sleep strategy. The power savings gain can be computed as a ratio of the average power consumption according to the discrete sleep levels defined in Table I to the average power consumption based on the optimized multi-level sleep strategy. In this work, the no-sleeping case is considered as the benchmark case wherein DBS consumes the no load power consumption during the vacation period. It is quite evident from Fig. 5 that the optimized multi-level sleep strategy always results in power savings gain greater than 1, $\forall \bar{v}$, showing the effectiveness of the proposed optimized multi-level sleep strategy.

V. Conclusion and Future Directions

In this work, we have integrated hardware (components, subcomponents and functional unit) dynamic capability analysis considering the CDSA as a base RAN architecture. Using system level simulations and analytical techniques, the implication of network architectures on the hardware is investigated. The transition delay (i.e. deactivation and reactivation delay), deactivation power consumption and energy consumption at component, subcomponent and functional unit levels is also analyzed. Taking input from the hardware dynamic capability analysis, we have analyzed the deep sleep opportunity and the potential minimum bound of power consumption for sleep mode under CDSA configuration. Multiple discretized sleep levels based on the transition latency are also identified. The network-driven sleep modes with CBS assistance provide the highest potential energy saving gains. Thus, we developed

dynamic energy saving techniques with focus on the CDSA, where the CBS (or a separate entity) assists the DBS sleep mode and wake-up decisions. The indirect measurement and prediction techniques are proposed to avoid the periodic transmission of pilot signals, thus overcoming limitations of the traditional sleep mechanisms whilst improving the prediction outcome.

As a future direction of this work, the investigation of the DBS switching mechanism to improve the energy efficiency in the context of CDSA and the impact of the DBSs switching off on the reliability and low latency needed for the future ultra-reliable low latency communications. In short, there are many aspects relevant to CDSA that needs further investigation to evaluate its suitability as a candidate RAN for 5G networks.

Acknowledgment

This work was supported in part by Huawei under the project: Energy Proportional eNodeB for LTE-Advanced and Beyond, and in part through ESPRC UK Global Challenges Research Fund (GCRF) allocation under grant number EP/P028764/1.

References

[1] E. Ternon, P. K. Agyapong, and A. Dekorsy, "Performance evaluation of macro-assisted small cell energy savings schemes," *EURASIP Journal on Wireless Communications and Networking*, vol. 2015, no. 1, p. 209, 2015. [Online]. Available: http://dx.doi.org/10.1186/s13638-015-0432-0

[2] A. Zakrzewska, D. López-Pérez, S. Kucera, and H. Claussen, "Dual connectivity in lte hetnets with split control-and user-plane," in *Globecom Workshops (GC Wkshps), 2013 IEEE*. IEEE, 2013, pp. 391–396.

[3] A. Mohamed, O. Onireti, M. Imran, A. Imran, and R. Tafazolli, "Control-data separation architecture for cellular radio access networks: A survey and outlook," *IEEE Communications Surveys and Tutorials*, vol. 18, no. 1, pp. 446–465, Firstquarter 2016.

[4] A. Mohamed, O. Onireti, M. A. Imran, A. Imran, and R. Tafazolli, "Predictive and core-network efficient rrc signalling for active state handover in rans with control/data separation," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1423–1436, 2017.

[5] B. Debaillie, C. Desset, and F. Louagie, "A flexible and future-proof power model for cellular base stations," in *IEEE Vehicular Technology Conference (VTC Spring)*, 2015.

[6] A. Capone, I. Filippini, B. Gloss, and U. Barth, "Rethinking cellular system architecture for breaking current energy efficiency limits," in *Proc. of Sustainable Internet and ICT for Sustainability (SustainIT)*, October 2012, pp. 1–5.

[7] E. Ternon, Z. Bharucha, and H. Taoka, "A feasibility study for the detached cell concept," in *Proc. of 9th International ITG Conference on Systems, Communication and Coding*, January 2013, pp. 1–5.

[8] X. Xu, G. He, S. Zhang, Y. Chen, and S. Xu, "On functionality separation for green mobile networks: Concept study over LTE," *IEEE Communications Magazine*, vol. 51, no. 5, pp. 82–90, May 2013.

[9] G. Auer et al., "How much energy is needed to run a wireless network?" *IEEE Wireless Commun. Mag.*, vol. 18, no. 5, pp. 40–49, Oct. 2011.

[10] 3GPP, Panasonic, "Proposal on the time scales for evaluations of small cell ON/OFF," Written Contribution, 3GPP TSG RAN WG1 meeting, Fukuoka, Japan, 20-24 May 2013.

[11] 3GPP, Huawei, HiSilicon, "Network adaptation for small cell operation efficiency enhancement," Written Contribution, 3GPP R1-131854, TSG RAN WG1 meeting, Fukuoka, Japan, 20-24 May 2013.

[12] P. Lahdekorpi, M. Hronec, P. Jolma, and J. Moilanen, "Energy efficiency of 5G mobile networks with base station sleep modes," in *IEEE Conference on Standards for Communications and Networking (CSCN)*, Sept 2017, pp. 163–168.

[13] F. E. Salem, A. Gati, Z. Altman, and T. Chahed, "Advanced sleep modes and their impact on flow-level performance of 5G networks," in *IEEE Vehicular Technology Conference (VTC-Fall)*, Sep 2017.

[14] O. Onireti, A. Mohamed, H. Pervaiz, and M. Imran, "Analytical approach to base station sleep mode power consumption and sleep depth," in *IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Oct.

[15] A. M. Abdelhady, O. Amin, and M. Alouini, "Energy-efficient resource allocation for phantom cellular networks with imperfect csi," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 3799–3813, June 2017.