The Value of BS Flexibility for QoS-Aware Sleep Modes in Cellular Access Networks

Gianluca Rizzo*, Balaji Rengarajan†, and Marco Ajmone Marsan‡
gianluca.rizzo@hevs.ch, balaji.rengarajan@acceleramb.com, ajmone@polito.it
*HES-SO Valais, Switzerland
†Accelera Mobile Broadband, USA
‡Politecnico di Torino, Italy and Institute IMDEA Networks, Spain

***Abstract*—Sleep modes are one of the most widely investigated techniques to decrease energy consumption in cellular access networks. However, the application of such algorithms on the base station (BS) equipment of today presents several challenges. Indeed, currently installed BSs are unfit for frequent on/off cycles. This may lead to increased failure rates and malfunctioning, ultimately resulting in significant CAPEX and OPEX increases for mobile network operators (MNOs). This situation calls for a new generation of flexible BSs endowed with a "hot standby" mode, which guarantees quick activation times without affecting BS availability. However, when such new BS models become available, MNOs will need to determine a migration path to a new network deployment with progressive replacement of old BS equipment. In this paper, we propose an approach to quantify the benefits obtained by MNOs with the deployment of flexible BSs, in terms of maximum energy efficiency achievable with a given fraction of flexible BSs in their network. More specifically, we propose a method for estimating, for a given percentage of flexible BSs, the energy optimal density of static and flexible BSs, which is sufficient to serve a given set of active users with predefined performance guarantees. We show how to apply our method to derive bounds on the maximum energy saving achievable through sleep modes, as a function of the fraction of flexible BSs. We determine the effect of uncertainty in traffic predictions on sleep modes performance, and we derive indications for optimal network planning strategies.**

I. Introduction

Sleep modes help alleviating the problem of raising OPEX (operational expenditures) in wireless access networks, due to the ever increasing energy costs and to the explosive growth of mobile network traffic. In cellular networks, sleep modes (or standby modes, or low power idle modes) are of particular interest, since energy costs account for a large fraction (up to 50%) of the total OPEX of mobile operators [1]. By putting to sleep base stations (BSs) in period of low traffic load, standby modes help reducing the waste of energy and making the whole network more energy proportional (the energy consumption becomes more proportional to the actual traffic load). Indeed, with the presently installed BS models, energy consumption exhibits only a weak dependency on the amount of traffic served, going at most up to 40% of the maximum energy consumed by a BS, in most recent equipment. In such a scenario, it has been shown [2] that sleep modes are the main instrument to decrease power consumption in wireless cellular networks.

In the present networking scenario, the problem is further aggravated by the fact that mobile network operators (MNOs) tend to overprovision their networks even with respect to peak traffic, trying to anticipate the forecasted traffic growth, so as not to have to update their networks too often.

Over the years, MNOs have deployed different generations of BSs. Most of the currently deployed BSs do not allow a 'standby mode', i.e., a low-power consumption mode that is designed to save energy (like for TV sets and PCs). However, these BSs can be turned off, as is done when they have to be upgraded, repaired, etc. When turning such BSs back on, a significant delay is incurred before they are fully operational. Such a slow responsiveness constitutes a constraint on how closely networks consisting of such base stations can be adapted to follow time-varying loads. And this, of course, has potentially a significant impact on the actual energy saving which can be achieved through sleep modes. Moreover, currently deployed BSs have not been designed to withstand on/off cycles with the frequency required by sleep modes. Applying sleep modes to static BSs might lead to high rates of equipment failures and of events of malfunctioning, resulting in lower BSs availability and shorter BSs lifetime. Ultimately, such problems bring to an increase of the MNO CAPEX and, through maintenance costs, also of OPEX.

In order to address these issues, new BS designs are currently being proposed, which better support sleep modes by implementing a standby state, minimizing in this way also the effects which frequent state changes can have on their lifetime [3]. In standby mode, a BS consumes more energy than in the off state, but less than in the fully operational mode. The amount of energy consumed, and the time needed to wake up, depend on the "depth" of the sleep mode, i.e. on which (and on how many) components of the BS are put to sleep. In any case, we assume that such flexible BSs can transition from the sleep mode to the fully operational state very quickly (on the order of seconds [4]), enabling fast adaptation to changes in traffic.

Several of the sleep mode algorithms which have been recently proposed are indeed characterized by frequent changes in the number of active base stations [5]–[7]. The expectation is that by adapting closely to traffic variations over time, further improvements in energy efficiency are achievable with respect to traditional, slowly adapting, sleep mode algorithms. However, being tied to a specific algorithm and network setup, such works do not give a clear idea of the impact of the adoption of flexible BS models on the potential energy savings of sleep modes. Moreover, several of the proposed algorithms do not directly consider user-perceived QoS, or they only do it a posteriori, assuming some amount of performance degradation as acceptable, in return for a decrease in the energy consumed.

In this paper, we propose an approach to quantify the benefit gained by a MNO through the deployment of flexible BSs. Our contributions are:

- For a given BS topology, and for a given pattern of user density variation over time, we propose a method for estimating the energy-optimal density of static and flexible BSs, which is sufficient to serve a given set of active users, with fixed performance guarantees.
- Through numerical evaluation, we compute bounds on the maximum energy saving as a function of the amount of flexible BSs, and illustrate the impact of various system parameters. We determine the effect of uncertainty in traffic predictions on the performance of sleep modes, and we derive indications for an optimal network planning strategy.

These results are particularly interesting in light of the practical impossibility of replacing all of a network BSs at once. MNOs need to identify effective migration paths from current networks to new, energy-efficient networks. This requires the estimation of the benefits obtained with the introduction of a fraction of new BS equipment in their network. Our work offers a tool for doing just this.

The paper is organized as follows. In Section II we present our model for the distribution of users and of BSs, and we state the main assumptions underlying our approach. In Sec.III, we use the results of the previous sections to compute the energy-optimal BS density for a given user density, and to estimate the achievable energy savings. In Sec. IV, we present numerical results, and we conclude the paper in Section V.

II. Model and Assumptions

We consider the downlink information transfer in a cellular access network. At any time instant, users form a homogeneous planar Poisson point process with intensity $\lambda_u(t)$ users per km^2. We focus on a geographical area where we assume BSs to be distributed according to a two-dimensional homogeneous Poisson point process, with density λ_b BSs per km^2. Such distribution has been shown to approximate reasonably well some real BS deployments [2].

We consider two classes of BSs, *static* and *flexible*, and three possible operational states: *on*, *sleep*, and *off*. We assume that static BSs can be either on or off, while flexible BSs can also be in the sleep state. BSs of both types can only serve users when in the on state.

We assume that the density of users fluctuates slowly over time. Typically, transitions between off and on states require a substantial amount of time (up to 15 minutes in most currently deployed BS equipment [4]). Therefore, we model a policy that follows variations in user density over time by dividing the time of a day into N periods, within each of which the set of BSs in the off state is not modified. The duration of such periods (which is not necessarily the same for all periods) is much greater than the BS switch-on time. We assume that at the beginning of each period, such policy turns off BSs based on a predicted spatial load distribution for the period, allowing for prediction errors and load fluctuations within the period. In order to exploit flexible BSs to follow the load fluctuations within the period, the considered policy divides each of the N periods into M subperiods, within each of which no BS goes into or out of the "off" state, but in which flexible BSs can change from sleep to on, or viceversa. The duration of each subperiod is much greater than the time required to switch from sleep to on states (which we assume to be of the order of seconds).

More specifically, the density of users in the n-th period, λ_n^u, is a random variable with pdf $f_{\lambda_n^u}$. We assume that at the beginning of each period, this pdf is known exactly (e.g. from historical data). The density of users in subperiod m within time period n, is therefore a random variable with pdf $f_{\lambda_n^u}$. Note that we assume the user density in a subperiod to be constant over the subperiod. At the beginning of each subperiod, we assume that the forecasted value of user density in that subperiod is predicted exactly. In what follows, we assume $f_{\lambda_n^u}$ to be a Gaussian centered at the value of λ_n^u, with a variance reflecting the extent of the prediction error for the n-th period.

A. Channel and Service Model

We do not consider the effects of shadowing or interference, and only take into account distance-dependent path loss. We assume that users are served by the BS that is closest to them, i.e., by the BS that corresponds to the strongest received signal, as it normally happens in reality. We denote the capacity to a user located at distance r from the BS by $C(r)$ bit/s per Hertz. The capacity can be modeled, for example, using Shannon's capacity law, or other models such as a quantized set of achievable rates. We assume that the network only serves best-effort traffic. We assume that BSs use a processor sharing mechanism to divide capacity among all the connected best-effort users. By doing so, a notion of fairness is imposed, since all best effort users associated with a particular BS are served for an identical fraction of time.

We use the per-bit delay seen by the typical user (τ^0) as the performance metric. It is defined as the inverse of the short-term user throughput, i.e., the actual rate at which the user is served, averaged over a short time with respect to the time constant of the variations in user density, and taking into account the capacity to the user as well as the sharing of the BS time across all associated users. Here, the interpretation of a typical user is that provided by Palm theory [8], and $\bar{\tau}$ is computed as the expectation of τ with respect to the Palm distribution P^0 associated with the point process of users.

We assume that the performance target of any sleep mode policy in the network is to keep the average per-bit delay in each period above a given threshold τ_0 with probability p_{th}. Note that we do not consider the effect of interference.

B. Energy Consumption Model

We assume that BSs always transmit at a fixed transmit power. When the BS density is higher than that required to achieve the threshold expected per-bit delay $\bar{\tau}^0$, we assume that BSs only serve users for the fraction of time required to satisfy the performance constraint, and remain idle (i.e., not transmitting to any user) for the rest. We denote with U the utilization of BSs, i.e., U is the average fraction of time in which the BS is transmitting. When a BS is on, we model its power consumption as a constant plus a component proportional to BS utilization. The energy consumption of a BS with utilization U in on state is modeled as: $p_{on} + q_{on}U$. In such an energy model, p_{on} is the power consumed by keeping a BS turned on with no traffic, while q_{on} is the rate at which the power consumed by the BS increases with the BS utilization. We assume that BSs in the off state consume power p_{off}, and flexible BSs consume power p_s in sleep mode, with $0 = p_{off} \leq p_s \leq p_{on}$.

C. Sleep Modes Strategy

We assume that all BS densities are feasible. In the homogeneous Poisson process layout of BSs, if each BS independently makes a decision to either turn off, or stay on, according to some probability, the resulting point process of active BSs is a thinned homogeneous Poisson process, and all BS densities are indeed achievable. We therefore restrict ourselves to strategies that turn each BS off (or set it to sleep mode) independently of the other BSs in the network, with a certain probability. Thus, we can specify a strategy by only specifying the densities of BSs that are on, off, and in sleep mode. Since, in actual BS deployments, not all densities are feasible due to coverage constraints, limitations due to law, etc., the results we obtain can be considered bounds to the maximum energy saving achievable with sleep mode strategies.

III. A Method for Evaluating the Impact of Flexibility on Energy Efficiency

In this section we present our method for estimating, for every time instant, the energy optimal density of active BSs (static and flexible), and its associated energy consumption. Our method consists in two steps. In the first step, for each period, we derive the total density of BSs which satisfies the performance target with the given probability. In the second step, within each subperiod, we show how to determine the density of flexible BS which should be active in order to satisfy the QoS constraint in terms of target per bit delay, in response to fluctuations in user density.

A. Derivation of the Energy-Optimal BSs Density

In this subsection we present the method we use for the derivation, for a homogeneous Poisson distribution of users and of BSs, with density λ_u and λ_b respectively, the minimum density of BSs which allows serving users while meeting the performance target in terms of maximum expected per-bit delay. This method was originally derived in [2].

The central result of this method is an expression which characterizes the expected per-bit delay seen by the typical user who is just beginning service, as a function of the density of users and BSs.

Theorem 3.1: The average per-bit delay $\bar{\tau}$ perceived by a typical best-effort user joining the system is given by:

$$\bar{\tau}(\lambda_b, \lambda_u) = \int_0^{\infty} \left(\int_0^{\infty} \int_0^{2\pi} e^{-\lambda_b A(r,x,\theta)} \lambda_u x \, \mathrm{d}\theta \, \mathrm{d}x \right) \frac{e^{-\lambda_b \pi r^2} \lambda_b 2\pi r}{C(r)} \, \mathrm{d}r. \quad (1)$$

where $A(r, x, \theta)$ is the area of the circle centered at (x, θ) with radius x that is not overlapped by the circle centered at $(0, -r)$ with radius r.

For a given density of BSs and of users, the performance constraint that is enforced is as follows: if the per-bit delay experienced by a *typical* user, $\bar{\tau}$, is less than a predefined threshold $\bar{\tau}^0$ seconds, then users are said to perceive satisfactory performance, and the corresponding BS distribution is feasible. Given the expression of the expected per bit delay, the energy optimal BSs density is derived by solving the following optimization problem:

OPTIMIZE(λ_u,$\bar{\tau}_0$)

$$\begin{aligned} \underset{\lambda_b}{\text{minimize}} \quad & \lambda_b \left(p_{on} + q_{on} \frac{\bar{\tau}(\lambda_b, \lambda_u)}{\bar{\tau}_0} \right) \\ \text{subject to} \quad & \bar{\tau}(\lambda_b, \lambda_u) \leq \bar{\tau}_0 \\ & \lambda_b \geq 0 \end{aligned} \quad (2)$$

OPTIMIZE can easily be solved by exhaustive search, being a problem with only one variable. For energy models which are completely insensitive to traffic (i.e. $q_{on} = 0$) this problem boils down to finding the BS density which satisfies the constraint on $\bar{\tau}$ with equality, i.e. the minimum feasible density for a given user density.

B. Characterization of Optimal Strategies

Any sleep mode strategy has to adapt to variations in traffic load by periodically changing the configuration of the network. How frequently this is done depends on the period of the day, as well as on traffic characteristics, and is constrained by the time it takes to switch on/off a static BS, by the time it takes to collect statistics about the traffic, to elaborate a new configuration for the network, and to apply the new configuration, migrating users away from BSs which are to be turned off in a way which does not impact their perceived QoS. Since the typical time for turning on a static BS can be assumed to be around 15 minutes, a reasonable *minimum* period between two changes in BSs configurations can be of the order of one hour.

In networks where a given amount of flexible BSs is available, sleep modes have to take advantage of their fast response times while coordinating flexible BSs management with the periodic switching on and off of static BS. Moreover, as flexible BSs still consume a non-negligible amount of energy while in standby mode, it might still be necessary to periodically put them into cold off state in periods of low traffic load. Since the speed at which the set of flexible BSs can adapt to traffic variations is much higher than the speed at which static BS can be adapted, sleep modes in such a mixed environment are generally composed by a *slow sleep* algorithm, and by a fast adaptation algorithm. While the slow sleep mode acts on all base stations, the fast adaptation algorithm only manages those BSs which can be put in on state within seconds. Slow sleep modes, being characterized by relatively long response times due to the time taken by switching on/off a device, act mainly based on predictions of average user density, while fast sleep modes are mainly based on real time measurements of traffic load.

An optimal slow sleep strategy needs to determine, for each period, the energy-optimal maximum density of active BSs (static and flexible) such that the target performance is achieved with the given probability within that period. That is, it has to guarantee an average per-bit delay which is above the target value with a probability p_{th}. Since we assume that at the beginning of each period the pdf of the user density is known, the minimum value of user density λ_{th} such that the probability of the user density being larger than λ_{th} be less p_{th} can be derived as $\int_{-\infty}^{\lambda_{th}} f_{\lambda}(l)dl \geq p_{th}$. Then the maximum density of active BSs (flexible and static) sufficient to satisfy the performance target can be derived by solving OPTIMIZE(λ_u^{th},$\bar{\tau}_0$). In this way, we determine the energy-optimal fraction of BSs which should be active during the period, as well as the maximum power consumption over the period (worst case, in which all BSs which are not in off state are active during the whole period). Of course, the larger the prediction error, the more conservative will be the estimation of such quantities.

In each period, we assume that those static BSs that cannot access sleep modes, are turned off first, and flexible BSs are turned off only after all the static BSs have been turned off. This in order to always have the maximum flexibility in the network. Note that this is an optimal behavior for sleep modes only because the energy model of flexible BSs assumes the same power consumption in the on state as for static BSs. If this would not be the case, then the method we have presented can be easily adapted in order to optimize, over the whole day, also over the share of flexible BSs.

As for the "fast" sleep modes, for each subperiod within the considered period, an optimal sleep strategy should determine the energy optimal density of flexible BSs which should be in on state, given the traffic measurements and user density forecasts up to that time, and set to sleep/on state a given amount of flexible BSs accordingly. We assume the duration of the subperiod to be short with respect to the speed at which user density varies over time, so that we can assume the value of user density in the subperiod to be known, possibly with some error margin δ. Then, the energy optimal amount of flexible BSs to be activated in the m-th subperiod of the n-th period is obtained by solving OPTIMIZE($\lambda_{u,m}^{n}+\delta$,$\bar{\tau}_0$), and by subtracting from the resulting BS density the density of static BSs which are active during that period.

IV. Numerical Evaluation

In this section we apply our analysis method to a simple network scenario, in order to estimate the increase in energy efficiency associated with the introduction of flexible BSs, for a specified target quality of service.

We assume in our analysis that the BS transmit power p is 10 W. BSs work at a frequency of 1 GHz, and use a bandwidth of 10 MHz. We use a log distance path loss model, with path loss at a reference distance of one meter calculated using Friis equation, and with a path loss exponent $\alpha = 3$. We assume that the data rate perceived by users is given by Shannon's capacity law. Thus, the capacity to a user located at a distance r from the base station is given by $C(r) = 10^7 \log_2 \left(1 + \frac{pr^{-\alpha}}{N_0}\right)$ bit/s, where $N_0 = -174$ dBm/Hz is the power spectral density of the additive white Gaussian noise. In accordance with typical values found in the literature for currently deployed BSs, we assume the maximum power consumption of a BS (i.e., the one at 100% utilization) to be 1500 W. Moreover, we choose the fraction of power consumed by a BS at zero utilization to be 60% of the maximum, in accordance with typical values found in the literature [4]. We assume that, in the on state, static and flexible BS have the same energy model. In the sleep state, unless otherwise stated, we assume flexible BSs consume 30% of the maximum, and we assume no energy is consumed in the off state by both types of BSs.

TABLE I. Parameter values used in the numerical evaluation.

Parameter	Value
BS transmit power p	10 W
Bandwidth	10 MHz
Path loss exponent α	3
Maximum BS power consumption	1500 W
Power in idle state	900 W
Power consumed in standby state	450 W
Target per-bit delay τ_0	1 μs, 10 μs
Peak user density	100 users/km^2
Minimal duration of a period	15 minutes
Threshold violation probability p_{th}	1%, 2%

Unless otherwise stated, we assume a 2% probability for the user density to be larger than the threshold value in any period. The target value of expected per bit delay is set to 10 μs.

In order to model the variation of the average user density over 24 hours, we use a classical m-shaped curve (see for instance [9]) where the peak user density is set to 100 users per km^2. The curve, shown in Fig. 1, is derived from data measured in the cellular networks of a large Italian mobile operator, and it is relative to a working day in the city of Pisa. We assume the user density prediction error to be Gaussian distributed.

In a first set of evaluations, we consider equally sized periods of 3 hours, starting at midnight. For each period, the threshold user distribution is computed over a Gaussian, whose mean is equal to the maximum average user density within the 3 hour interval. Therefore, in each period, the threshold user density is determined by both the prediction error σ and the variation of the average user density during that interval.

In Fig. 2 we analyze the impact of the prediction error on the energy saving, as a function of the percentage of flexible BSs in the network. The percentage of flexible BSs is computed with respect to the maximum, over the day, of the total density of BSs required to serve the users with the given performance target. The daily energy consumption is normalized with respect to the ideal case of perfect forecast, with no prediction error. The curves show the effect of the

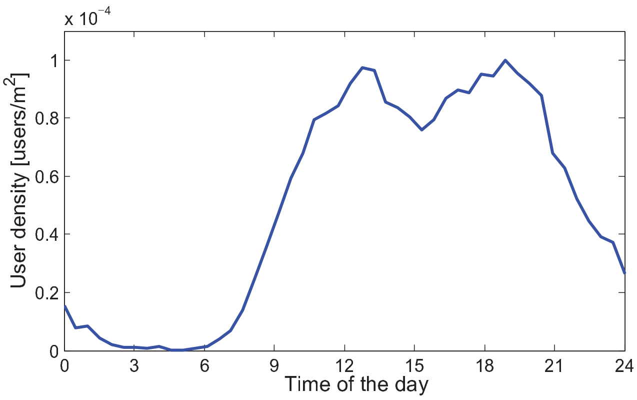

Fig. 1. Average density of active users over 24 hours.

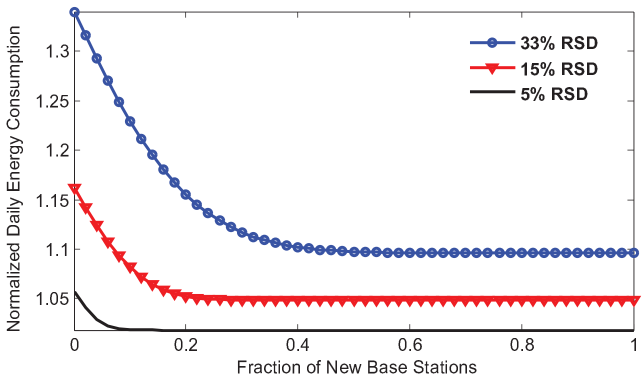

Fig. 2. Average energy consumed daily per km^2, as a function of the fraction of flexible BS, normalized w.r.t. the energy consumed in the case of zero prediction error.

prediction error, in terms of standard deviation over the mean (relative standard deviation – RSD), on the efficiency of the sleep modes. We can see how a prediction error with a 33% RSD increases the energy consumed by an optimal sleep mode of 34% with respect to the ideal case of no prediction error. As expected, we see that the use of flexible BSs helps recovering part of the loss in energy efficiency due to uncertainty in traffic forecasts. From the plots we see that, as a rule of thumb, for a standard deviation of the prediction error of x%, substituting x% of the static BSs with flexible ones is enough to get all the possible benefits out of flexibility. Indeed, through the energy consumed in sleep mode, flexible BSs end up contributing to the OPEX of the network too, and thus prevent from recovering in full the energy cost of uncertainty in traffic prediction.

In Fig. 3 we plot the percentage of energy saved, in one day, by an optimal sleep mode strategy with respect to the configuration in which all BSs are always active, as a function of the percentage of flexible BSs. We see that the higher the prediction error (measured in terms of relative standard deviation, or RSD), the higher the savings due to sleep modes. Indeed, as the prediction error affects the optimal maximum BS density, sleep modes allow to save between 25% and 33% of the energy consumed in one day, as a function on the prediction error, even with no flexible BS. Moreover, such plots show that by introducing flexible BSs, the sleep modes savings increase of about 12%.

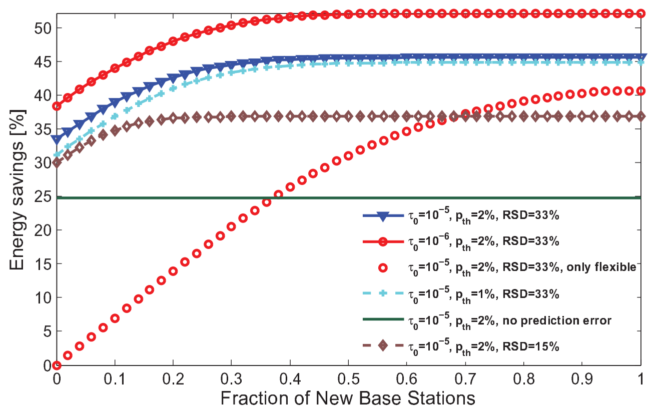

Fig. 3. Sleep modes energy savings, as a function of the fraction of flexible BS.

As already discussed, one of the main problems related to the introduction of sleep modes is that presently available BSs are not built for being frequently turned on and off. As a consequence, applying sleep modes to static BSs might bring to much higher failure rates and to malfunctioning, thus shortening the BSs lifetime. Ultimately, such problems lead to an increase of the CAPEX and, through maintenance costs, of the OPEX of the MNO. We assume that future BSs will be designed to better support sleep modes by implementing a standby mode, minimizing in this way also the effects of on/off cycles on their lifetime. For this reason, we analyzed the case in which only flexible BSs implement sleep modes, while static BS are never turned off. In Fig. 3 we plot the energy saving relative to this configuration. We see how replacing 30% of the static base stations with flexible BSs brings energy savings between 10% and 20% over a whole day. These results show the importance for MNOs of an accurate prediction of their daily traffic, in order to achieve the best performance with sleep modes. From the plots we can also see how decreasing the target mean per-bit delay brings to higher achievable energy efficiency. This is a consequence of the fact that tighter QoS constraints require a higher density of base stations, and a more conservative dimensioning of the network configuration, thus leaving more room for optimization through dynamic network management algorithms. Finally, we see that bringing the threshold violation probability p_{th} (defined in Section III-B) from 2% to 1% does not affect significantly the achievable energy savings.

A. Impact of the Traffic Adaptation Strategy

One of the most important aspects of a sleep mode strategy is the determination of the frequency and the time at which the set of active BSs should be changed in order to adapt to traffic variations. In settings with a mix of flexible and static BSs, this is still an open problem. In particular, it is not clear if the presence of flexible BSs and fast adaptation algorithms makes it still useful to also adapt frequently to traffic variations, given the impact that on/off cycles may have on the BS availability and lifetime.

In order to clarify this, we consider different possible ways of defining the periods during which the set of BSs in off state cannot be modified, and evaluate their impact on potential

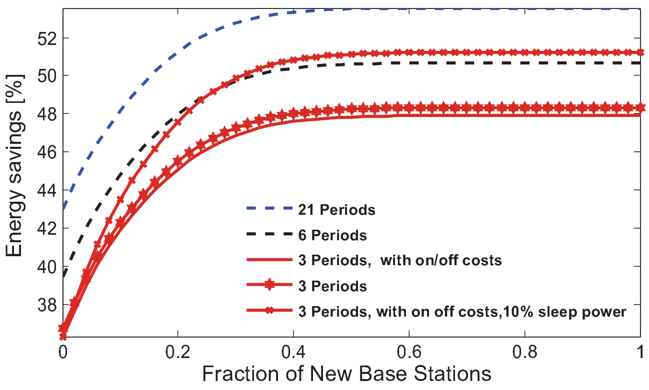

Fig. 4. Sleep modes energy savings, as a function of the fraction of flexible BSs, for different numbers of periods.

energy savings of sleep modes. More specifically, we assume that the 24 hours are divided into the minimum number of periods such that, within each period, the variation of average user density is less than $Y\%$ of the maximum user density over the 24 hours. Moreover, we impose a minimal duration of a period of 15 minutes. In Fig. 4 we plot the achievable energy savings with respect to the absence of sleep modes, as a function of the fraction of flexible BSs, for three different values of the maximum variability Y. By increasing this last parameter, of course, not only the number of periods changes, but also their position in the day and their durations. As expected, by decreasing Y we allow "slow" sleep modes to follow more closely variations in user density over the day, improving the achievable energy saving. However, we also note how these improvements are only of the order of few percentage points, so that a strategy with only three periods does not lose much with respect to one with 21 periods. Rather, we see how the introduction of 30% flexible BSs has a much stronger impact on energy saving.

We also evaluated what happens when the power consumed in standby mode goes from 30% to a more optimistic 10% of the maximum nominal power (a setting which is still realistic if we consider, for instance, the amount of power consumed by PCs in standby mode [10]). As expected, this improves both the marginal increase in energy efficiency, and the maximum energy efficiency achievable when flexible BSs are introduced in the network.

As we already discussed, for "slow" sleep modes, the cost of switching BSs more frequently during the day is not only due to the possible decrease in availability of BSs. It also entails an energy cost, due to the time it takes to bring on a BS and to redistribute traffic among neighboring BSs. We can quantify the impact of these costs by assuming a switch-on time of 15 minutes, during which the BS consumes 70% of the power in on state with no traffic. Moreover, we can assume it takes 5 minutes to turn off gracefully a base station, during which the consumption is the same as in the on state. From Fig. 4 we can see that these on/off costs have a marginal impact on the achievable energy savings, and we have also verified that they do not vary significantly with the number of periods, or with the prediction error. Indeed, for a traffic profile such as the one in Fig. 1, for periods with a minimal duration larger than 15 minutes, the vast majority of BSs are turned on and off only once per day, since the traffic profile boils down to one peak and one low per day.

V. Conclusions

In this work, we characterize the value of having flexibility in a fraction of the BSs which are used to implement the access network of a MNO. We call *flexible* those BSs which, in addition to the *on* and *off* operating modes also allow a *standby* mode, which can be used for the implementation of the sleep algorithms which are proposed for the improvement of the proportionality between carried traffic and consumed energy in today's networks. We present a method for the computattion of the benefit in terms of maximum energy efficiency achievable with a given fraction of flexible BS. Our proposed method also helps evaluating the impact of load predictability on the potential energy savings achievable by QoS-aware sleep modes. Our results can be used by MNOs to run a cost versus benefit analysis for sleep modes with flexible BSs, and in particular to determine the marginal benefit of increasing the amount of flexible BS in their network, thus being able to devise an effective migration path to energy-efficient cellular networking.

References

[1] "Mobile networks go green," *ABI Research*, 2012. [Online]. Available: https://www.abiresearch.com

[2] B. Rengarajan, G. Rizzo, and M. Ajmone Marsan, "Bounds on QoS-constrained energy savings in cellular access networks with sleep modes," in *ITC*, sept. 2011, pp. 47 –54.

[3] A. Ambrosy, O. Blume, H. Klessig, and W. Wajda, "Energy saving potential of integrated hardware and resource management solutions for wireless base stations," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2011 IEEE 22nd International Symposium on*, 2011, pp. 2418–2423.

[4] Y. Zhang, L. Budzisz, M. Meo, A. Conte, I. Haratcherev, G. Koutitas, L. Tassiulas, M. Ajmone Marsan, and S. Lambert, "An overview of energy-efficient base station management techniques," in *Digital Communications - Green ICT (TIWDC), 2013 24th Tyrrhenian International Workshop on*, 2013, pp. 1–6.

[5] K. Samdanis, D. Kutscher, and M. Brunner, "Self-organized energy efficient cellular networks," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2010 IEEE 21st International Symposium on*, 2010, pp. 1665–1670.

[6] S. Zhou, J. Gong, Z. Yang, Z. Niu, and P. Yang, "Green mobile access network with dynamic base station energy saving," in *MobiCom '09*, Sep. 2009.

[7] A. Conte, A. Feki, L. Chiaraviglio, D. Ciullo, M. Meo, and M. Ajmone Marsan, "Cell wilting and blossoming for energy efficiency," *Wireless Communications, IEEE*, vol. 18, no. 5, pp. 50–57, 2011.

[8] D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic geometry and its applications*. Wiley, 1987.

[9] O. Blume, H. Eckhardt, S. Klein, E. Kuehn, and W. M. Wajda, "Energy savings in mobile networks based on adaptation to traffic statistics," *Bell Labs Technical Journal*, vol. 15, no. 2, pp. 77–94, 2010. [Online]. Available: http://dx.doi.org/10.1002/bltj.20442

[10] R. Bolla, R. Bruschi, F. Davoli, and F. Cucchietti, "Energy efficiency in the future internet: A survey of existing approaches and trends in energy-aware fixed network infrastructures," *Communications Surveys Tutorials, IEEE*, vol. 13, no. 2, pp. 223–244, 2011.