A Distributed Q-Learning Approach for Adaptive Sleep Modes in 5G Networks

Ali El Amine, Mauricio Iturralde, Hussein Al Haj Hassan, Loutfi Nuaymi

A Distributed Q-Learning Approach for Adaptive Sleep Modes in 5G Networks

Ali El-Amine*, Mauricio Iturralde*, Hussein Al Haj Hassan† and Loutfi Nuaymi*
*IMT Atlantique, Rennes, France
†American University of Science and Technology, Beirut, Lebanon
e-mail: *name.lastname@imt-atlantique.fr, †hhajhassan@aust.edu.lb

***Abstract*—In 5G networks, specific requirements are defined on the periodicity of Synchronization Signaling (SS) bursts. This imposes a constraint on the maximum period a Base Station (BS) can be deactivated. On the other hand, BS densification is expected in 5G architecture. This will lead to an energy crunch if kept ignored. In this paper, we propose a distributed algorithm based on Reinforcement Learning (RL) that controls the states of the BSs while respecting the requirements of 5G. By considering different levels of Sleep Modes (SMs), the algorithm chooses how deep a BS can sleep according to the best switch-off SM level policy that maximizes the trade-off between energy savings and system delay. The latter is calculated based on the wake-up time required by the different SM levels. Results show that our algorithm performs better than the case of using only one type of SM. Furthermore, our simulations show a gain in energy savings up to 90% when the users are delay tolerant while respecting the periodicity of the SS bursts in 5G.**

***Index Terms*—Energy consumption, 5G, sleep modes, Q-learning**

I. Introduction

With the explosive increase in the number of mobile subscribers and services, mobile networks will have to support a much higher capacity demand [1]. Nevertheless, the growing increase in traffic demand is proliferating the energy consumption of the Information and Communication Technology (ICT) sector. It is estimated that ICT consumes around 4.7% of the world's electrical energy, releasing into the atmosphere about 1.7% of the global CO_2 emissions [2].

On the other hand, the Fifth Generation of Cellular Mobile Communications (5G) is expected to provide ubiquitous internet access with 1000 times higher data rate compared to present cellular systems. As a result, the innovations in 5G systems are not limited to physical layers techniques, but also introduce new network architectures and application scenarios. In particular, a major trend in 5G networks is the deployment of a large number of small-scale BSs, also known as network densification. Trying to achieve this ambitious goal relying on the above paradigm and architecture is not sustainable since it will lead to an energy crunch with serious economic and environmental concerns.

Owing to the economic concerns of mobile operators as well as the environmental ones and in order to cater for the vision of 5G communication systems, a lot of studies have investigated distinct approaches to reduce the energy consumption in mobile networks. Since the energy consumption of a telecommunication network is dominated by the Radio Access Network (RAN) and in particular the BS consuming $75 - 80\%$ of the network's energy [3], most of the work focus on reducing the power at this level to enhance the energy efficiency of cellular networks. Several techniques attack the aforementioned challenge on different layers: *a)* network planning and deployment, *b)* switching-off techniques, *c)* radio resource management optimization, *d)* component level enhancement, and *e)* the use of renewable energy resources [4]. Among the different layers, sleep mode techniques are considered to be among the most efficient approaches for energy savings.

In [3], 5G BSs sleep model are proposed to reduce the energy consumption of a cellular network. The proposed model describes different levels of SM the BS can switch to, where each SM level is characterized by an activation/deactivation period and a power consumption. Implementing these different stages of SM in 4G networks is challenging due to backwards compatibility and reference signaling requirements. However, in 5G, the Cell Reference Signals (CRS) are removed [5] making it possible for a BS to explore these different SM levels. In contrast to models that apply only one type of SM, in this work, we study the trade-off between energy savings and added service delay, i.e., the time the user has to wait in the buffer until the BS reactivates. We consider a cellular network with unbalanced traffic where each BS chooses its appropriate SM level in order to achieve a desirable balance between energy consumption and latency. We propose an adaptive sleep scheme algorithm based on RL that controls the states of the BSs, where each BS has access only to its local information in order to learn the best energy saving policy.

The rest of the paper is organized as follows. Section II presents an overview of the existing work. In Section III, we detail the system model along with the 5G sleep mode model. We give an overview of the distributed Q-learning algorithm in Section IV, whereas the proposed algorithm is presented in Section V. Finally, we present the simulation results in Section VI before concluding in Section VII.

II. Related Work

The literature on Energy Efficiency (EE) cellular networks is huge [6], [7], and the references therein. BS ON-OFF switching is considered among the best methods to save energy since it does not require changes to current network architecture, and it is easy to implement. This well known

method initially proposed in IEEE 802.11b [8] has attracted a lot of attention on the research community. For example, the authors in [9] studied the energy savings problem by switching-off macro BSs under coverage constraints using stochastic geometry. Their results achieved an EE gain of around 1.6 compared to the case where all the BSs are active. In [10], the authors opted to maximize the EE of a heterogeneous network using ON-OFF switching subject to traffic load constraints. In [11], the authors proposed an optimization mechanism based on delay-constrained energy-optimal BS sleeping policies. In [12], dynamic programming is applied to find the optimal BS ON-OFF policy given the on-grid energy price in order to minimize the on-grid energy cost purchased by the operator while assuring the downlink transmission quality at the same time. However, the above mentioned work focused only on one level of SM, and do not take into account the activation/deactivation times required for the BS to switch between states ON and OFF.

In contrast to binary BS models (active and sleep), recent models have split the SMs state into several levels [13]–[16]. Considering that a BS can switch to different SM levels gives it more flexibility to adjust with the type and pattern of traffic to further enhance the system performance. In [13], the EE of a heterogeneous network is studied by switching small cells to different SM levels while preserving the Quality of Service (QoS). Using stochastic geometry, the authors determined the optimal operating probability for each SM level of each BS. However, the proposed solution is offline, and the SM levels discussed are not suitable for 5G networks. In [14]–[16], the authors proposed the concept of Advanced Sleep Mode (ASM) which corresponds to gradual deactivation of the BSs components in order to decrease the energy consumption of the BS. Even though the used sleep model do not violate the 5G requirements, [14], [15] focused on the energy savings potential and not the best SM policy to be applied, while [16] is limited to only one BS.

Recently, artificial intelligence has received significant attention as a highly effective alternative to conventional methods [17]. In particular, machine learning has found wide-ranging applications in wireless networks [18]. In [19], a cognitive engine with reinforcement learning is implemented at each BS to improve the system capacity and QoS. Furthermore, a BS switching operation algorithm is proposed to dynamically switch between sleep and active modes. In [20], the authors presented a distributed Q-learning-based algorithm that learns energy inflow and traffic demand patterns in a heterogeneous network. By interacting with the environment, each agent (i.e. BS) decides its optimal policy (ON or OFF) to improve the system performance. The authors in [21] proposed a method for optimizing ON-OFF policies for ultra dense networks. The mechanism is based on Deep Q-Learning (DQL) to solve the dynamic optimization problem. Using DQL, the authors looked for maximizing the EE while respecting the QoS of the network.

The above mentioned studies take advantage of machine learning in order to find the best ON-OFF policy to reduce the energy consumption of the BS. However, none of them respect the requirement of 5G on the SS bursts that need to be transmitted periodically (ranging from 5 to 160ms); hence, limiting the duration time a 5G BS can switch to SM. For example, [13] applies SMs that requires several seconds for the BS to wake up from. The work in [16] use Q-learning to apply 5G adapted SMs; however, the work is limited to only one BS.

In this paper, in contrast to most prior work we focus on 5G SM proposed in [3]. We do not limit the work to only one type of SM, rather we consider a network where each BS can choose different SM states to switch to. We propose a distributed Q-learning approach that finds the best policy for each BS in order to reduce the energy savings while maintaining the QoS of the users under the 5G requirements.

III. System description

A. 5G Energy Savings Sleep Modes

In [3], GreenTouch identified four distinct SM levels by grouping sub-components with similar transition latency when being activated or deactivated. The presented model enables to quantify the power consumption of the BS in each of the four SMs. These are:

- SM 1: It considers the shortest time unit of one OFDM symbol (i.e. $71\mu s$) comprising both deactivation and reactivation times. In this mode only the power amplifier and some processing components are deactivated.
- SM 2: It corresponds to the case of sub-frame or Transmission Time Interval (TTI) (i.e. 1 ms). In this SM, more components enter the sleep state.
- SM 3: It corresponds to the frame unit of 10 ms. Most of the components are deactivated in this mode.
- SM 4: This is the deepest sleep level. Its unit corresponds to the whole radio frame of 1s. It is the standby mode where the BS is out of operation but retains wake-up functionality.

Higher energy savings can be achieved when switching BSs to a deeper SM, since more components will be deactivated. However, this will be associated with longer transition latency which may impact the QoS of the system. In Table I, we present the SM levels characteristics.

Along with SM and users' dynamics, the BS has to wake up periodically to send signalling bursts. In contrast to Long Term Evolution (LTE) systems where each antenna must transmit every 0.2 ms a unique CRS for channel quality estimates and mobility measurements among other SS, no CRS is required for 5G [5]. Instead, SS and Physical Broadcast CHannel (PBCH) are transmitted in SS/PBCH block periodically. It has been agreed in Third Generation Partnership Project (3GPP) [5] that this periodicity can be set to any value among [5, 10, 20, 40, 80, 160 ms]. With these values, SM 4 cannot be used. Hence, we limit our work to the first three SM levels.

B. Network Description

We consider a large-scale wireless downlink cellular network composed of M BSs serving K users. We further

TABLE I: BS Sleep Modes Characteristics [3].

Sleep level	Deactivation duration	Minimum sleep duration	Activation duration
SM 1	35.5 μs	71 μs	35.5 μs
SM 2	0.5 ms	1 ms	0.5 ms
SM 3	5 ms	10 ms	5 ms
SM 4	0.5 s	1 s	0.5 s

consider that each BS can switch to one of the following SM levels: SM 1, SM 2 or SM 3.

Whenever a user requests a service from a BS in SM, it triggers the activation mode and the user is buffered until the BS wakes up. This has an impact on the latency added to the system. The deeper the SM is, the more time the user will have to wait until the BS reactivates. When the user is served and if the BS is in idle mode, it goes back to its SM level to save energy until the next users arrival.

We further consider BSs with different users arrival rates. Thus, we divide the BS into two sets. One corresponds to the BSs with high users arrival rate denoted by $\mathcal{B}^{\text{Dense}}$, and the other grouping the BSs with low users arrival rate denoted by $\mathcal{B}^{\text{Light}}$, such that $|\mathcal{B}| = |\mathcal{B}^{\text{Dense}}| + |\mathcal{B}^{\text{Light}}| = M$. Note that the users are uniformly distributed within the cell of each BS.

C. Downlink Transmission Model

We measure the downlink transmission quality between the serving BS m and a user k based on the Signal-to-Interference-plus-Noise Ratio (SINR) as follows:

$$\text{SINR}_m(k) = \frac{P_m^{\text{Tx}} h_m(k)}{\sigma^2 + \sum\limits_{m' \in \mathcal{S}, m' \neq m} P_{m'} h_{m'}(k)} \tag{1}$$

where P_m^{Tx} is the transmitted power of BS m, $h_m(k)$ is the channel gain from BS m to user k, which accounts for the path loss and shadowing effect, and σ^2 is the additive white Gaussian noise power density.

We can express the peak rate offered to user k and served by BS m using Shannon-Hartley theorem as follows:

$$R_m(k) = \alpha \times W \times \log_2(1 + \text{SINR}_m(k)) \tag{2}$$

where W represents the bandwidth and α the fraction of bandwidth used for the data transmission.

IV. Distributed Q-Learning

Distributed Q-learning is an online optimization technique that aims at controlling multi-agent systems, i.e., a system featuring M BSs which take decisions (select the appropriate SM level) in an uncoordinated fashion. Each BS has to learn independently a policy (SM 1, SM 2 or SM 3) through real-time interactions with the environment. Q-learning finds the optimal policy in the sense that it maximizes the expected value of the total reward (Q-value) over all successive episodes. The agents (i.e., BSs) have a partial view of the overall system, and their actions may differ since the users are unevenly distributed over the network. In particular, the decision of a BS to choose a SM level is affected by how many users it has to serve, and on how delay-tolerant these users are.

In Q-learning, each agent takes an action a_m^t from an action set $\mathcal{A}$, then moves to a new state s_m^{t+1} while receiving a reward r_m^t. This reward is then used to update the Q-value locally, $Q(s_m^t, a_m^t)$, indicating the level of convenience of selecting action a_m^t when in state s_m^t. The Q-value is updated following the update rule:

$$\begin{aligned} Q(s_m^t, a_m^t) \leftarrow Q(s_m^t, a_m^t) + \\ \alpha[r_m^t + \gamma \max_{a \in A} Q(s_m^{t+1}, a_m^{t+1}) - Q(s_m^t, a_m^t)] \end{aligned} \tag{3}$$

where α is the learning rate that represents the speed of convergence, and $\gamma \in [0, 1]$ is the discount factor that determines the current value of the future state costs.

During the learning phase, each agent selects the corresponding action based on the ϵ-greedy policy, i.e., it selects with probability $1-\epsilon$ the action associated with the maximum Q-value, and with probability ϵ selects a random action:

$$a_m^t = \begin{cases} \underset{a \in \mathcal{A}}{\text{argmax}}\, Q(s_m^t, a_m^t), & \text{if } y > \epsilon \\ rand(\mathcal{A}), & \text{otherwise} \end{cases} \quad m = 1, \ldots, M. \tag{4}$$

By implementing the ϵ-greedy policy, the BS would have explored all possible actions and avoided local minima. For more details on RL and Q-learning the reader is referred to, e.g., [22].

V. Adaptive Partial Scheme Algorithm

In this section, we present the proposed Q-learning-based algorithm that controls the states of the BSs. Consider the state representation $\mathcal{S}$ of the BSs. Each BS can be active (serving users), in sleep mode (SM 1, SM 2 or SM 3) or idle (active but not serving any user).

$$\mathcal{S} = \{\text{Active}, \text{Idle}, \text{SM 1}, \text{SM 2}, \text{SM 3}\}$$

We define the set of possible actions $\mathcal{A}$ the state to which the BS can switch to. In contrast to schemes that implement only one type of SM (i.e., SM 1, SM 2 or SM 3), in this work we emphasis the potential of traffic-aware scheduling where each group of BSs in a given network can switch to different SM levels. This will fully utilize the advantages of BSs in the different states for both energy savings and delay reduction. For instance, the group of BSs switching to SM 1 will serve rate sensitive users (i.e. VoIP), whereas the other group of BSs in SM 3 level will serve delay tolerant users (i.e. HTTP requests). We call such a schedule an adaptive partial scheme. The final policy is then to determine how many BSs chose to switch to SM 1, SM 2 and SM 3.

An episode starts when all the BSs are in idle mode, and it finishes when all the users are served. During each episode, a BS chooses an action, then stores a quality-value linking the states $s \in \mathcal{S}$ to the chosen action $a_m \in \mathcal{A}$ following Eq. (3). The action consists of choosing the best sleep mode level having the highest Q-value.

Our goal is to find the best combination of SMs levels each group of BSs can switch to in order to maximizes the reward r_m^t. We define the reward as the weighted-sum of the energy gain G and the added delay D, both resulting from the sleep mode level chosen during an episode.

$$r = (1-\eta)G - \eta D \tag{5}$$

where $\eta \in [0,1]$ is a parameter which controls the trade-off between energy gain and the delay performance. Note that for $\eta = 0$, the problem reduces to maximizing the energy gain without considering the added delay, whereas for $\eta = 1$ the problem becomes minimizing the delay. We note the delay measured in this work is the time a user has to wait in the buffer until the BS reactivates from its sleep mode state.

VI. Numerical Analysis

A. Simulation Parameters

We consider a network of 25 BSs with an inter-site distance of 500 m. We model the arrival of the users following a Log-normal distribution with mean λ_a and variance ν. We consider that a user request to download a file with mean size 500 MBytes. The file size follows a Weibull distribution having a CDF of $F(x) = 1 - e^{-(x/\lambda)^k}$, where $\lambda > 0$ is the scale parameter, and $k > 0$ is the shape parameter. Table II summarizes the assumptions and parameters used in our simulation.

TABLE II: Simulation Parameters.

Parameter	value
Antenna height	30 m
BS Tx Power	45 dBm
Bandwidth	20 MHz
Thermal noise	-174 dBm/Hz
Pathloss	128.1 +37.6 $\log_{10}$(d) dB
Shadowing	Log-normal (6 dBm)
User's arrival	Log-normal, λ_a = 1, υ= $\lambda_a/10$
Service type	file with mean=4 Mbps
Scale parameter	$\lambda = 441.305$
Shape parameter	$k = 0.8$

In order to show the impact of SM levels on the energy consumption, we consider in this study a low load traffic with mean arrival rate $\lambda_a = 1$ user/s/Km^2. We run 500 independent simulations on Matlab in order to acquire the average energy gain and the average added delay on the performance of the network. From [23], we found the power figures for the different states of the BS. These values are given in Table III.

TABLE III: Power Consumption of a 2×2 MIMO BS.

Active	Idle	SM 1	SM 2	SM 3
250 W	109 W	52.3 W	14.3 W	9.51 W

B. Convergence Analysis

The Q-learning algorithm requires a training phase in which the agent explores the state-action couples to converge towards the optimal policy. Fig. 1 shows the convergence of the momentary Q-values (initialized to zero) corresponding to the best action over the different states summed over all the BSs with $\eta = 0.7$. Then, the best policy is stored in a look-up table that will be used during the exploitation process. We can observe that after few iterations, the algorithm converges for all BSs and over all the state-action pairs.

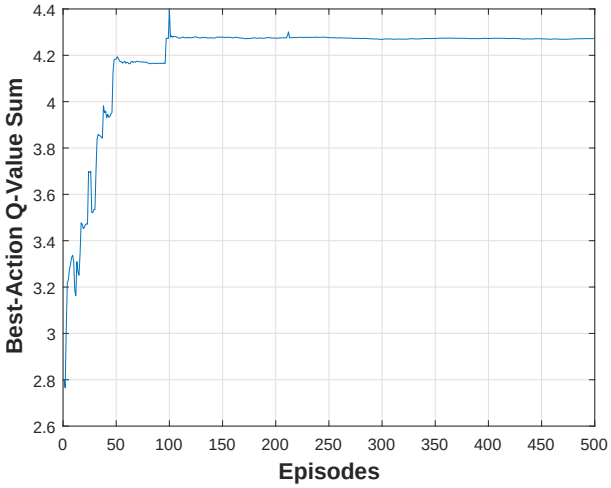

Fig. 1: Convergence of the Q-learning algorithm.

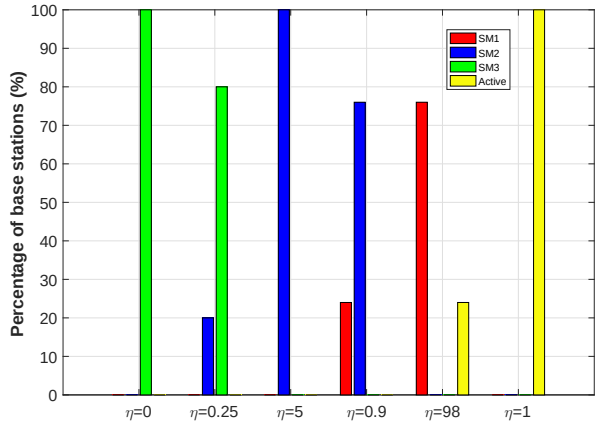

Fig. 2: States of the BSs for different different values of η.

C. Numerical Results

In Fig. 2, we present the distribution of the states of the BSs for different values of η, where 25% of the sites $\in \mathcal{B}^{Dense}$. We observe that when energy saving is prioritized over delay ($\eta = 0$), all BSs choose SM 3 that saves energy the most, whereas staying idle policy is preferred when the users are delay sensitive ($\eta = 1$). The states combination is observed for other values of η. In these cases, we observe a gradual shift from SM 3 to the other states. The incentive behind this shift is in the non-uniformity distribution of the traffic resulting in having some BSs with no traffic (since the overall traffic is low), thus able to switch to a deeper sleep mode level than the others.

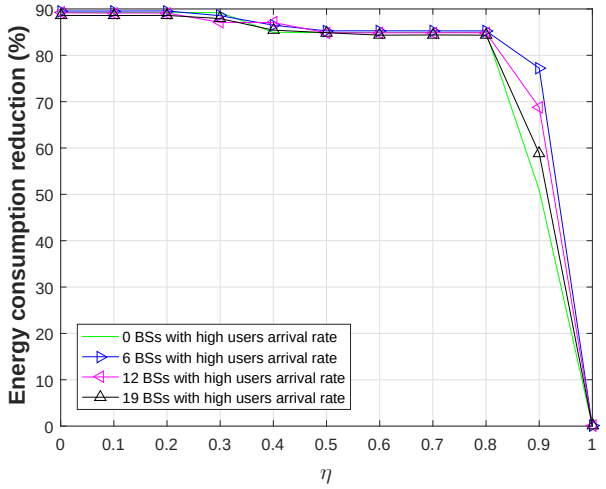

Fig. 3: Average network energy savings of our proposed algorithm for different users distribution.

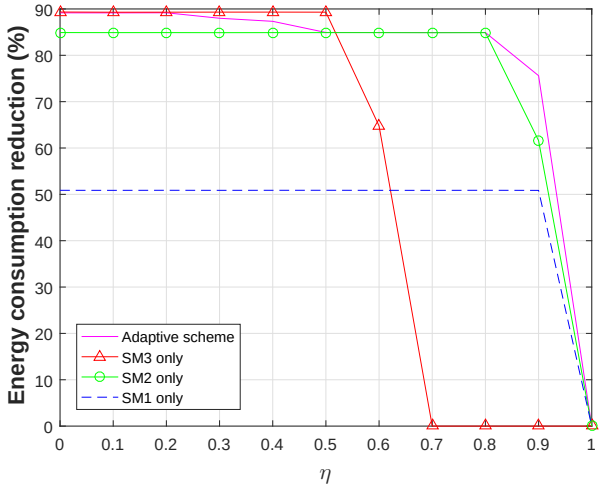

Fig. 5: Energy savings comparison between our proposed algorithm and the cases of using only one type of sleep mode.

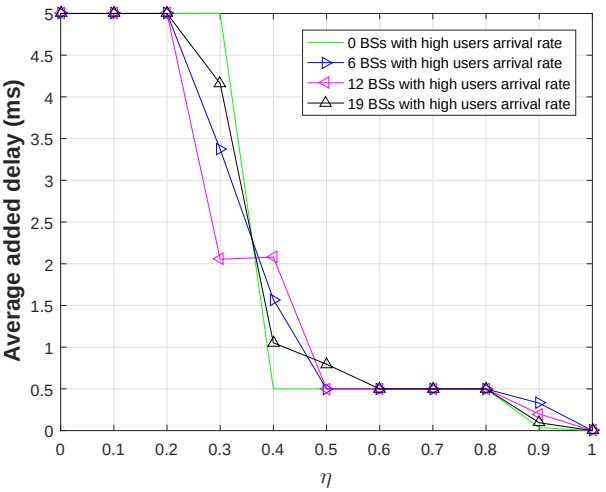

Fig. 4: Average network added delay of our proposed algorithm for different users distribution.

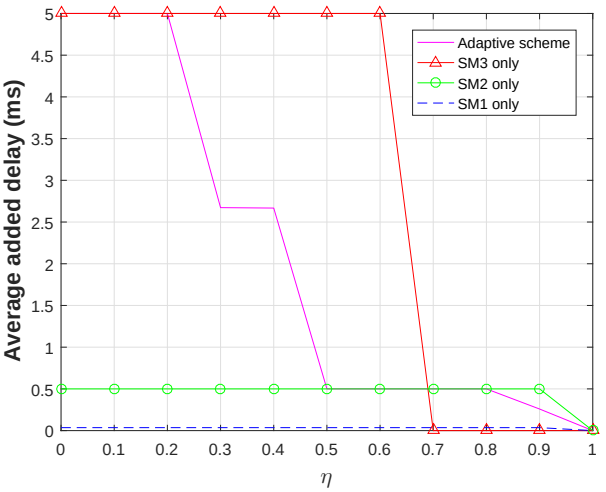

Fig. 6: Added delay comparison between our proposed algorithm and the cases of using only one type of sleep mode.

Figures 3 and 4 illustrate the average network energy savings and added delay for different values of η and for different percentage of BSs with higher users arrival rate. We observe that the performance of the system depends on the reward function and in particular η. An energy saving gain of around 90% is achieved but at a cost of increased delay (5ms). It is then important to carefully choose η in order to satisfy the requirements of the different 5G use cases. We also note that having an unbalanced traffic diversifies the best policy for each BS, and thus we observe better performance when having different arrival rates for each BS.

In Fig. 5 and Fig. 6, we compare our algorithm with a benchmark that uses only one type of sleep mode. We observe that the adaptive scheme outperforms all three benchmarks in terms of energy savings. Even though using SM 3 only has slightly better energy savings for some range of η, it adds more delay to the user as shown in Fig. 6. Thus, for applications when there is a hard restriction on delay, using SM 3 only may not be tolerated, rather an adaptive scheme is required. Similarly, in applications where both delay and energy consumption are considered, the adaptive scheme algorithm finds its best policy to combine the different sleep modes to achieve the necessary requirements.

In order to show the performance of our proposed algorithm, we compare the reward that evaluates both the energy savings and added delay. We notice that the adaptive algorithm combines the advantages of the different sleep mode levels in order to maximize the reward. For instance, the adaptive algorithm prefers SM 3 when the users are rate sensitive (small epsilon) and SM 1 when the users are delay sensitive. The combination

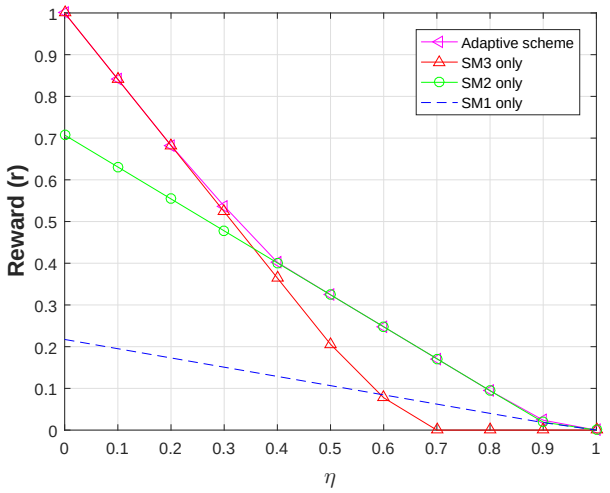

Fig. 7: Performance comparison between our proposed algorithm and the cases of using only one type of sleep mode.

of these states gives a higher reward than if these states where solely chosen.

VII. Conclusions

In this work, we investigated the problem of energy savings and delay associated with 5G networks. While it is critical to reduce the energy consumption of these networks, hard constraints are agreed on in the 5G requirements to periodically activate 5G cells for signaling synchronization. In this work, we proposed an adaptive partial sleep scheme algorithm based on reinforced learning that controls the state of the base stations in order to maintain a desirable trade-off between energy consumption and delay. We focused on 5G sleep modes that respect the 5G requirements. We showed that having a combination of different states of sleep modes in a network can achieve better performance than having one state only. Simulation results revealed an energy gain ranging between 90% and 50% for rate and delay tolerant users, respectively.

This work opens the door for accessing the performance of dense networks, where a huge number of cells are deployed, without violating the 5G requirements. The effect of signaling bursts on the energy consumption is not studied in this work. This might affect the optimal policies that might be used for different signaling periodicities allowed in 5G networks. Another interesting topic is to study the optimal policies in the context of renewable energy and smart grid where the amount of energy harvested varies from one location to another, and the price of electricity varies from one retailer to another.

References

[1] *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016-2021*. Cisco, Tech. Rep., February 2017.

[2] Erol Gelenbe and Yves Caseau. The impact of information technology on energy consumption and carbon emissions. *Ubiquity*, 2015(June):1:1–1:15, June 2015.

[3] B. Debaillie, C. Desset, and F. Louagie. A flexible and future-proof power model for cellular base stations. In *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, pages 1–7, May 2015.

[4] Luis Suarez, Loutfi Nuaymi, and Jean-Marie Bonnin. An overview and classification of research approaches in green wireless networks. *EURASIP Journal on Wireless Communications and Networking*, 2012(1):142, 2012.

[5] 3GPP. On requirements and design of ss burst set and ss block index indication. In *TS 38.300 Release 15*, 2017.

[6] J. Wu, Y. Zhang, M. Zukerman, and E. K. Yung. Energy-efficient base-stations sleep-mode techniques in green cellular networks: A survey. *IEEE Communications Surveys Tutorials*, 17(2):803–826, Secondquarter 2015.

[7] Mingjie Feng, Shiweb Mao, and Tao. Base station ON-OFF switching in 5G wireless networks: Approaches and challenges. *IEEE Wireless Communications*, 24(4):46–54, August 2017.

[8] IEEE standard for local and metropolitan area networks part 16: Air interface for fixed and mobile broadband wireless access systems amendment 2: Physical and medium access control layers for combined fixed and mobile operation in licensed bands and corrigendum 1. *IEEE Std 802.16e-2005 and IEEE Std 802.16-2004/Cor 1-2005 (Amendment and Corrigendum to IEEE Std 802.16-2004)*, 2006.

[9] J. Peng, P. Hong, and K. Xue. Stochastic analysis of optimal base station energy saving in cellular networks with sleep mode. *IEEE Communications Letters*, 18(4):612–615, April 2014.

[10] M. Feng, S. Mao, and T. Jiang. Boost: Base station ON-OFF switching strategy for energy efficient massive MIMO hetnets. In *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, pages 1–9, April 2016.

[11] X. Guo, Z. Niu, S. Zhou, and P. R. Kumar. Delay-constrained energy-optimal base station sleeping control. *IEEE Journal on Selected Areas in Communications*, 34(5):1073–1085, May 2016.

[12] Y. Che, L. Duan, and R. Zhang. Dynamic base station operation in large-scale green cellular networks. *IEEE Journal on Selected Areas in Communications*, PP(99):1–1, 2016.

[13] C. Liu, B. Natarajan, and H. Xia. Small cell base station sleep strategies for energy efficiency. *IEEE Transactions on Vehicular Technology*, 65(3):1652–1661, March 2016.

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

[15] F. E. Salem, A. Gati, Z. Altman, and T. Chahed. Advanced sleep modes and their impact on flow-level performance of 5G networks. In *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, pages 1–7, July 2017.

[16] Fatma Ezzahra Salem et al. Reinforcement learning approach for advanced sleep modes management in 5G networks. In *2018 IEEE Vehicular Technology Conference (VTC-Fall)*, July 2018.

[17] Y. Bengio, A. Courville, and P. Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, Aug 2013.

[18] C. Jiang, H. Zhang, Y. Ren, Z. Han, K. Chen, and L. Hanzo. Machine learning paradigms for next-generation wireless networks. *IEEE Wireless Communications*, 24(2):98–105, April 2017.

[19] Q. Zhao and D. Grace. Transfer learning for QoS aware topology management in energy efficient 5G cognitive radio networks. In *1st International Conference on 5G for Ubiquitous Connectivity*, pages 152–157, Nov 2014.

[20] M. Miozzo, L. Giupponi, M. Rossi, and P. Dini. Distributed Q-Learning for energy harvesting heterogeneous networks. In *2015 IEEE International Conference on Communication Workshop (ICCW)*, pages 2006–2011, June 2015.

[21] H. Li, H. Gao, T. Lv, and Y. Lu. Deep Q-Learning based dynamic resource allocation for self-powered ultra-dense networks. In *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*, pages 1–6, May 2018.

[22] Richard S. Sutton and Andrew G. Barto. *Introduction to Reinforcement Learning*. MIT Press, Cambridge, MA, USA, 1st edition, 1998.

[23] IMEC power model tool. [Online]. Available: https://www.imec-int.com/powermodel.