This is the author's accepted version of the article. The final version published by IEEE is O. E. Martínez-Durive, J. Suárez-Varela, J. Omaña Iglesias, A. Lutu and M. Fiore, "An Evaluation of RAN Sustainability Strategies in Production Networks," IEEE International Conference on Computer Communications, 2025, pp. TBP, doi: TBP.

An Evaluation of RAN Sustainability Strategies in Production Networks

Orlando E. Martínez-Durive*†, José Suárez-Varela ‡, Jesús Omaña Iglesias ‡, Andra Lutu ‡ and Marco Fiore*
*IMDEA Networks Institute, Spain, †Universidad Carlos III de Madrid, Spain, ‡ Telefónica Research, Spain
{orlando.martinez, marco.fiore}@imdea.org, {jose.suarez-varela, jesusalberto.omana, andra.lutu}@telefonica.com

***Abstract*—Reducing energy consumption is a primary goal for the mobile telecommunication industry, with strong environmental and economic implications. The main target for savings is the Radio Access Network (RAN), which is responsible for more than 70% of the total energy costs incurred by operators. Lowering energy costs at the RAN is possible by reducing the number of active carriers at off-peak locations and times where the demand can be served with a lower capacity than deployed. While the scientific community has been proposing a plethora of complex solutions to switch-off underutilized carriers, production networks largely rely nowadays on threshold-based strategies that run at individual RAN equipment and are typically enabled only overnight. Moreover, there are no real-world evaluations of the effectiveness of carrier switch-off approaches in reducing energy consumption or their impact on the end users. In this paper, we benchmark five fixed threshold-based cell sleep policies deployed in a production network serving large geographical regions. The study provides unprecedented insights on industry-grade RAN sustainability at scale, in terms of actual energy savings and trade-offs with user experience. Our insights suggest that the capability of the tested policies in reducing the energy costs hits a clear barrier if no degradation is admissible for any user, and provides a strong empirical basis in support of more flexible approaches to save energy at the RAN.**

I. Introduction

Energy efficiency is one of the major challenges that mobile network operators (MNOs) face today, as operating very large and power-hungry mobile network infrastructures is becoming increasingly complicated by rapidly growing energy costs and global climate emergency awareness. There is therefore a strong drive towards making mobile communication systems more sustainable both in economic terms, *i.e.*, reducing operating expenses linked with the power consumption of network equipment, and environmental terms, *i.e.*, improving the social acceptability of the cost that pervasive digital services entail for the planet. As a result, all major MNOs and equipment vendors are nowadays making substantial efforts to investigate strategies to reduce their energy and carbon footprint [1] and meet the goals set by global policymakers [2], [3].

The bulk of the energy consumption in mobile networks is associated to Radio Access Network (RAN) operation, which accounts for 73% of the total energy costs incurred by a typical operator according to GSMA [4]. Indeed, RAN equipment performs power-intensive functions for signal processing and transmission at tens of thousands of base stations that ensure coverage across territories of thousands of square kilometers. Moreover, the co-existence of overlaid Radio Access Technologies (RATs) belonging to different generations like 2G, 3G, 4G and 5G, coupled with challenges in sunsetting the older ones, exacerbates the problem of high RAN energy footprint.

On the bright side, RANs are also highly redundant systems, since they have to cope with strong spatial and temporal fluctuations in the user demands. For instance, mobile traffic loads are dramatically different in urban and rural areas [5], are known to feature circadian rhythms with high volumes during daylight hours that drop overnight [6], [7], and show prominent peaks occurring at specific moments of the day that change depending on the location considered [8]. To efficiently serve such time-varying demands, modern RANs have a layered design with multiple antennas that cover a same geographical area at different frequency bands and that jointly provide the capacity needed to serve the local traffic peaks.

This layered architecture paves the way for a dynamic management of the RAN where equipment that is not needed to sustain the demand in a specific time period can be turned off, hence reducing its associated energy cost. Options supported by present RAN hardware include shutting down individual symbols, radio channels or whole carriers, which offer diverse levels of flexibility and savings [9], [10]. However, the problem of deciding when to switch-off or switch-on RAN elements is extremely complex since sustainability goals inherently conflict with the strict requirements that mobile networks are expected to meet in terms of availability, reliability, throughput, or latency. While a plethora of proposals exist in the scientific literature, they mostly are, as later expounded in Section VI, theoretical or simulation studies that do not consider: (i) the limitations imposed by industry-grade RAN hardware that cannot be instantaneously and perfectly controlled, (ii) the striking heterogeneity of real-world user demands that are instead modeled with simplistic measurement data [9], [10], or (iii) the exact impact of RAN configuration changes on users' experience that are arduous to establish [11]–[13].

These aspects are crucial for production mobile networks serving actual users, where even minimal and localized degradation of the users' Quality of Experience (QoE) is unacceptable, since it would entail, *e.g.*, bursts of calls to customer service, the backlash on social media and ultimately churn in a very competitive market. These very real and significant penalties cause players in the mobile ecosystem to be highly conservative in RAN sustainability. Even leading MNOs currently adopt basic and cautious strategies that only switch-off a minimum portion of the RAN equipment overnight [14].

In this context, it is difficult to understand if more ambitious targets for RAN sustainability can be achieved in production. In this paper, we shed some light on this subject, by presenting the setup, measurements and results of trials executed at scale in a production network where diverse energy saving

TABLE I: Energy saving strategies tested by the MNO.

Strategy	Deployment time window	Load Thresholds	
		Policy ON	Policy OFF
Night-loose	23h - 6h (night)	5%	10%
Night-strict	23h - 6h (night)	10%	20%
Full-loose	24h (all day)	5%	10%
Full-moderate	24h (all day)	7%	12%
Full-strict	24h (all day)	10%	15%

strategies were tested for several weeks. Our work sets forth the following main contributions.

- We present an unprecedented benchmarking of five different cell sleep strategies for RAN energy savings that are based on utilization thresholds and that were tested in the RAN of a top-tier MNO in a Western European country, covering two regions with diverse urbanization levels.
- We characterize the switch-on/off process that each strategy induces on the real-world RAN using several Key Performance Indicators (KPIs) for measuring cell downtime, energy consumption, and impact on users' throughput.
- We evaluate the gains of the different strategies in terms of energy savings for the MNO, using a realistic energy consumption model, and show that the reduction in power consumption varies dramatically not only across strategies but also among regions and individual carriers.
- We capture for each strategy the trade-off between the achieved energy savings and the impact on end-user performance, using user throughput as a reference. The results reveal that throughput variations are remarkably heterogeneous across cells in different locations.

Ultimately, our work provides a first-of-its-kind deep dive into the operation of energy saving policies in production, and points at clear directions for the design of RAN sustainability strategies that are both effective and viable in practice.

II. Methodology & Dataset

Our study focuses on the cellular network deployment of a top-tier MNO in a Western European country with over 40 million connected devices. During February - March 2023, the MNO tested five different energy efficiency strategies for the RAN, each one over a week, using commercial solutions from equipment vendors (see Table I). We focus here on strategies that rely on cell sleep energy efficiency policies, where specific cells within base stations are put into a lower-power sleep mode during times of low traffic load.

We collect fine-grained KPIs at the cell level, and leverage this data to evaluate the different strategies in terms of the trade-off between estimated energy savings and the impact on end-user performance.

A. RAN energy saving strategies

Figure 1 illustrates the general behavior of the commercial energy saving solutions tested by the MNO. All solutions rely on continuous monitoring of the Physical Resource Block (PRB) utilization, *i.e.*, the portion of fundamental radio resource units available in a cell that are allocated to users for data transmission or reception. PRB utilization is tracked at the level of each *power group*, *i.e.*, the set of all cells covering a same geographical area over different frequency bands.

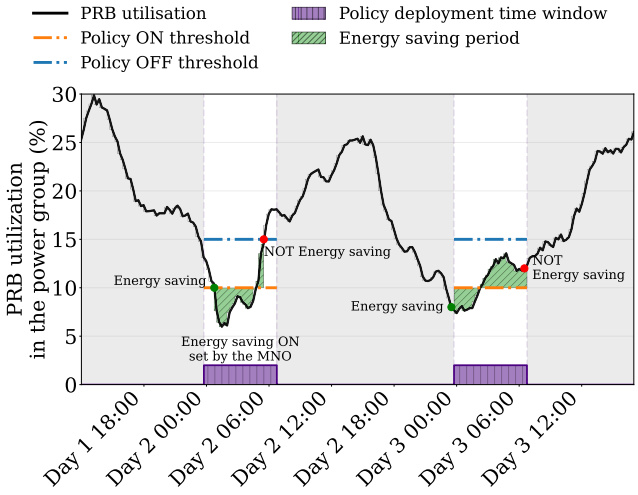


Fig. 1: Example of energy saving strategy. During the policy deployment time window (purple intervals on the x-axis), a PRB utilization (on the y-axis) dropping below the Policy ON threshold puts the cell in sleep mode. A PRB utilization rising above the Policy OFF threshold wakes the cell from sleep mode. The strategy is disabled outside the policy deployment time window.

The tested solutions apply dynamic cell sleeping strategies based on the time evolution of the average PRB utilization in cells of a power group. With reference to Figure 1, the following parameters define an energy efficiency strategy.

Policy deployment time window: the time period in which the energy saving policy is enabled. In the strategies that we analyze (Table I), *Night-loose* and *Night-strict* are active during night time, *i.e.*, the deployment time window spans 23:00 to 06:00. The other policies operate all day.

ON/OFF load thresholds: These thresholds set the PRB utilization levels that trigger the activation (ON) or deactivation (OFF) of energy-saving policies. When the average PRB utilization in a power group falls below the ON threshold, the energy efficiency solution identifies the cell to send in sleep mode.[1] Conversely, if PRB utilization exceeds the OFF threshold and cells in the power group are in sleep mode, then one of the sleeping cells is woken up. In the MNO trials (Table I), each strategy sets different ON/OFF load thresholds that represent various levels of aggressiveness: higher thresholds entail higher energy saving, while they may impact more the end-user performance. This characterization of such a trade-off is part of the focus of our study.

Time to trigger: refers to the delay between exceeding the ON/OFF load thresholds and the activation or deactivation of the energy-saving mechanism, designed to prevent erratic switching due to rapid load fluctuations. In our case, the MNO has set this delay to 10 minutes for the cell sleep process and 5 minutes for reactivation. To minimize user impact, the cell power is gradually reduced before a full shutdown, ensuring a smooth transition for users from the cell edge to its center. All tested solutions adopt the same time to trigger implementation.

[1]Due to non-disclosure agreements with the MNO, we cannot detail the strategy adopted by the tested solutions to select the cell put to sleep.

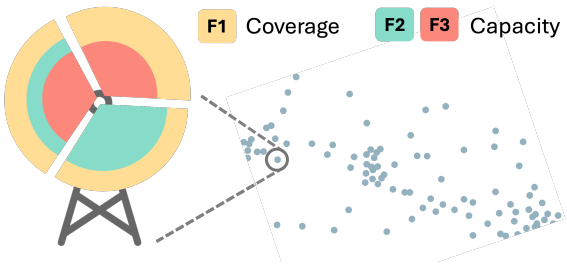


Fig. 2: Example of a three-sector site with cells deployed on three different frequency bands, for coverage and capacity. Blue dots represent base stations included in a specific trial.

B. Energy saving trials

In line with standard deployment strategies, the MNO catalogs its licensed frequency bands into two classes: low-frequency bands that are used to optimize *coverage*, and mid- to high-frequency bands that are dedicated to enhancing *capacity*. From these, capacity cells are the only ones that implement energy saving policies, while coverage cells typically remain untouched to guarantee service availability over the entire geography. Accordingly, each power group comprises a set of capacity cells and one or more coverage cells. Figure 2 shows an example of a cell site located in one of the regions the MNO used for the pilot. A three-sectored site uses cells operating in three frequency bands. Cells in the F1 band are used for coverage, while F2 and F3 bands are used for capacity and are dynamically put to sleep mode and woken up as dictated by the deployed energy saving strategy.

The MNO performed trials of five energy saving strategies, each over a one-week period (see Table I), and over two geographic regions delimited by specific Tracking Area Codes (TAC). Table II summarizes the RAN infrastructure features in two trial regions, which include diverse population and cell deployment densities (cells/sq. km.). Hereafter, we refer to these regions as *Dense* and *Sparse*, respectively. Energy saving strategies were only tested on LTE cells, which account for 71.51% and 54.14% of the cells deployed in the two regions, and all other RATs, including 2G, 3G and 5G, were excluded from the trial. Capacity cells, which are affected by the energy saving policy, account for 23.05% of the tested cells in Dense regions and 27.43% in Sparse regions.

C. Measurement data feeds

For our evaluation, we use cell-level measurements from LTE cells in the two study regions (Table II), collected throughout the whole trial period. Below, we describe the data feeds that were processed and combined for our analysis.

Radio access network deployment: We collect an inventory of all the cells deployed in the two regions of study. This includes cell-level information such as the location (lat/lon coordinates), the eNodeB ID, the orientation (azimuth), tilt, antenna manufacturer and model, and the carrier (*i.e.*, frequency channels). This inventory is updated daily to reflect changes in the network topology, such as the deployment of new cells and the decommissioning of older technologies.

Radio access network KPIs: The MNO monitoring infrastructure collects more than 160 KPIs at the cell-level for all the RATs deployed, including from GSM (2G) to the newest 5G NR cells. Measurements are collected every 15 minutes and averaged over hourly intervals. In our study, we evaluate the different energy-saving strategies using three main KPIs: (i) *cell availability*, which measures the amount of time a cell was woken up within a specific hour, with granularity down to seconds; (ii) *cell load*, expressed as the PRB utilization in downlink[2]; and (iii) *average user throughput*, which indicates the average data rate served by the cell.

Mobility management signaling: We collect control-plane signaling messages from the Mobility Management Entity (MME) of the 4G Evolved Packet Core (EPC). This data captures all mobility-related events (e.g., attach, detach, handover) at the LTE cells involved in the MNO trials, with timestamps recorded to the millisecond. We use this data to analyze user transitions during cell sleep or wake up events and identify the *fallback cells*, *i.e.*, the cells that absorb the load and UEs.

III. Energy Saving Policies in Action

Optimizing energy saving in the RAN presents a significant challenge: evaluating the energy saved and its impact on users' performance [14]. In this section, we focus on analyzing the energy saving aspect by examining (i) the cell downtime (*i.e.*, the amount of time that cells were in sleep mode), and (ii) the energy consumption. We present the impact of energy saving on user performance later in Section IV.

A. Cell downtime

As mentioned earlier in Section II-B the MNO applies cell sleep strategies only on capacity cells, while coverage cells remain always active. Hence, we measure the impact of the different strategies taking as a reference the *cell downtime*, *i.e.*, the amount of time that capacity cells were in sleep mode. For this, we use the cell availability measurements collected from the RAN monitoring infrastructure of the MNO, as described in Section II-C. These values allow us to measure the cell downtime with a granularity of seconds.

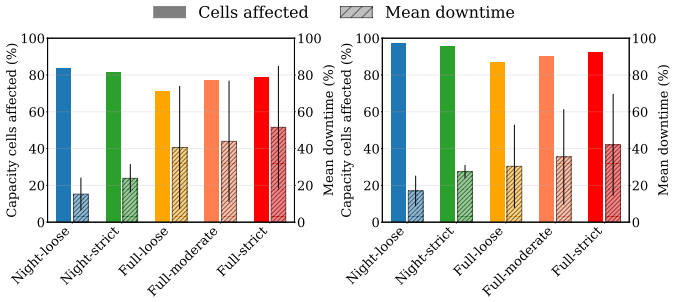
Figure 3 shows the percentage of capacity cells affected by each of the strategies, *i.e.*, cells that entered at least once into sleep mode over the weekly trial periods, and the average percent downtime of these cells. Overall, the strategies that operate only over the nighttime (Night-loose and Night-strict) exhibit significantly less downtime, while the number of cells affected (at least once) is slightly larger. Full-day strategies (24h) affect more cells as the strategy is more aggressive (*i.e.*, higher ON/OFF thresholds are higher).

The figure also reports the average and standard deviation of the total percent of the time during which cells remain in sleep mode for each strategy. This statistic grows with the aggressiveness of the strategy. For example, in the Dense region, we observe a rise in the mean downtime from 40.6% with the Full-loose strategy (5/10 for ON/OFF load thresholds) to 51.5% with the Full-strict strategy (10/15% for ON/OFF thresholds). It is interesting to note that although the Full-loose strategy has the same ON/OFF load thresholds as the

[2]As most frequency bands use FDD, downlink channels get congested earlier than uplink ones, hence downlink PRB utilization is a better indicator.

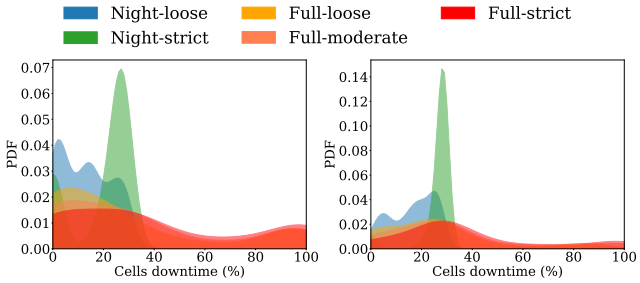
TABLE II: Description of the regions under study

Region name	Area km^2	LTE share	LTE Density (cells/sq. km.)	LTE cells / site	Capacity cells (%)
Dense region	189.28	71.51%	5.41	11.0 ± 6.10	23.05%
Sparse region	4986.82	54.14%	0.21	8.5 ± 5.98	27.43%



(b) Dense region (c) Sparse region

Fig. 3: Percentage of capacity cells affected by the energy saving strategies and their corresponding mean downtime (*i.e.*, the percentage of time that cells were in sleep mode). Error bars show the standard deviation across all the measured capacity cells.



(a) Dense region (b) Sparse region

Fig. 4: Distribution of cell-level downtime (in %) of capacity cells for the energy saving strategies tested by the MNO.

Night-loose, the fraction of capacity cells affected is slightly larger in the latter. Upon verifying with the operation team of the MNO, this was, in fact, a symptom of misbehavior by the nighttime energy saving strategies, where the cells were failing to wake up from cell sleep, even when the load was reaching levels above the OFF threshold. Nevertheless, the results are still consistent when we look at the average cell downtime. The mean downtime increased by 25.0% and 36.3% in the Dense and Sparse regions, respectively, when comparing the Full-strict with the Night-loose strategies.

To further dig into the behavior of the different strategies, Figure 4 shows the distribution of cell-level downtime. We observe that policies operating only during nighttime exhibit some inconsistency in their distribution across the two regions, likely due to the issue mentioned above. In contrast, full-day strategies show more consistent behavior across cells. For these strategies, a main takeaway is that as the policy becomes more aggressive (*e.g.*, Full-strict), cells display a more heterogeneous behavior, observing sparser distributions as the strategy becomes stricter, *e.g.*, more cells experience considerably higher downtime, suggesting that these strategies even keep a small portion of cells inactive for most of the time (see right part of the distributions). While these less-conservative behaviors can help save more energy, they may also have some impact on end-user performance –and aspect that we further investigate in Section IV.

We complete the characterization of the downtime by looking at the temporal evolution of the amount of time that cells spend in sleep mode over different moments of the day. Figure 5 shows the hourly evolution of the mean downtime of capacity cells across all the different strategies, each tested for a one-week period. Overall, we observe clear day-night patterns, where all strategies benefit from the low traffic load over the night period, as shown by downtime peaks occurring during the nighttime for all strategies. Likewise, full-day strategies consistently show increased downtime as they become stricter, both during the day (valleys of the curve) and at night (peaks of the curve). In addition, the Night-strict strategy shows longer sustained peak periods during the night, this is mainly due to the higher OFF loading threshold (20% PRB utilization) which forces the cells to remain in sleep mode for longer periods of time while the policy is active.

B. Energy consumption

In this section, we translate the cell downtime into actual energy savings. For this, we focus on the energy consumed at the cell level, and rely on an operational energy consumption model that estimates the energy usage of a full base station [15]. The energy consumption model is expressed in Eq. 1 as:

$$E_{BS} = E_{BBU} + \sum_{i} E^{i}_{RRU}, \tag{1}$$

where E_{BBU} accounts for the energy consumed by the processing of baseband signals and remains relatively constant and not affected by traffic load, while E^{i}_{RRU} is the cost associated with the radio unit and expressed as follows:

$$E_{RRU} = \alpha \cdot R_{PRB} \cdot P_m + \gamma, \tag{2}$$

where R_{PRB} represents the fraction of PRBs used (*i.e.*, the load), and P_m denotes the maximum transmit power of the antenna (set by the operator). α and γ account for the power amplifier efficiency and fixed circuit power, respectively, and are specific to the antenna type, frequency, and number of transmitters.[3]

In the analyzed MNO trials, coverage cells remain always active, hence base stations never turn off. This means that the E_{BBU} element of the energy consumption model remains always active for all base stations. Consequently, energy savings are only reflected on the E_{RRU} component: when a cell enters in sleep mode the consumption related to the radio unit drops to zero according to the energy consumption model (*i.e.*, E_{RRU} = 0 kWh).

[3]For confidentiality reasons we cannot disclose the values of the model parameters used in the two formulas, which were empirically obtained.

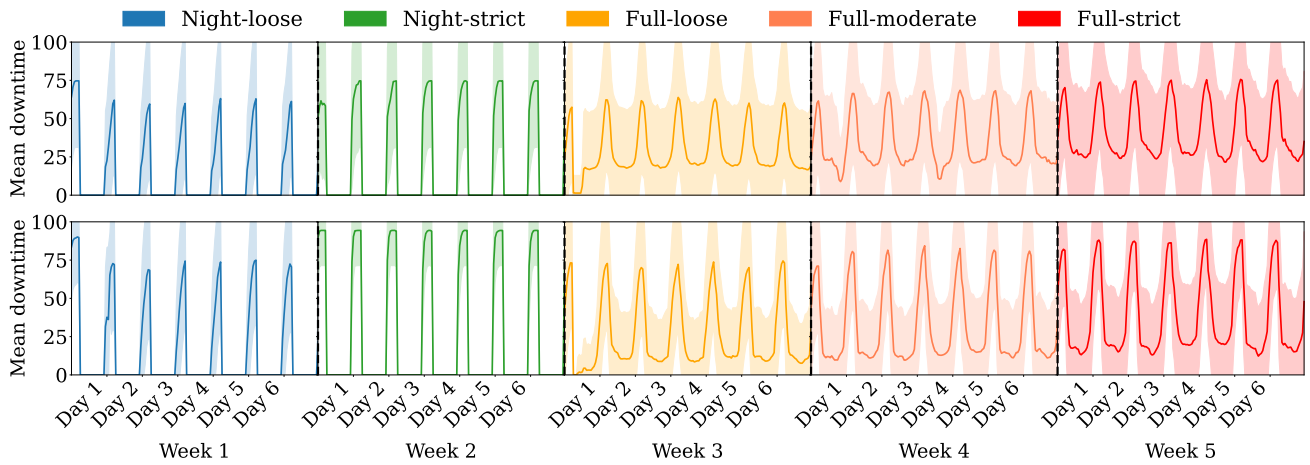
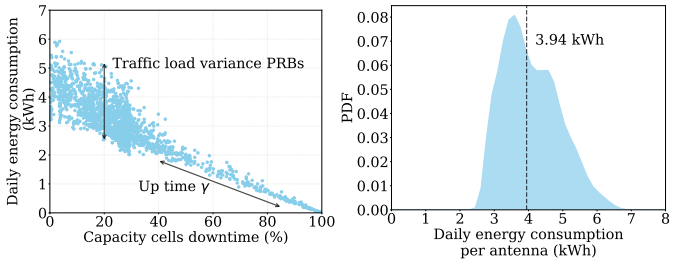


Fig. 5: Hourly time series of the mean cell downtime; Top: Dense region, and bottom: Sparse region. Shades indicate the standard deviation of the downtime distribution across all capacity cells.



(a) Capacity cells (b) Coverage cells

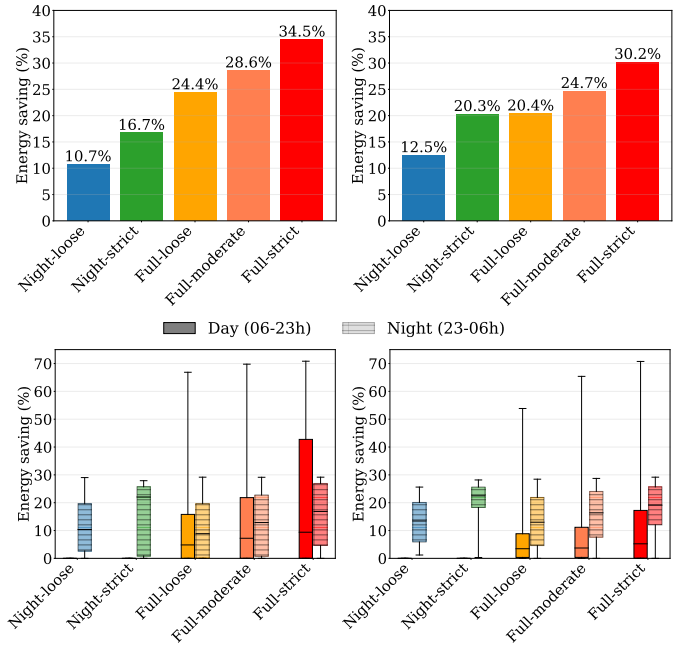
Fig. 6: Energy consumption of Radio units (E_{RRU}).

Figure 6a shows the relationship between the cell downtime and the corresponding cell-level energy consumption on daily intervals for all capacity cells in the two studied regions. Since measurements are at the cell level, we show only energy consumption due to the E_{RRU} component; note that E_{BBU} applies at the base station level and does not depend on cell downtime. Here we observe that γ produces a constant contribution to the energy consumption (see Eq. 2), which is proportional to the time that the cell is active (*i.e.*, inversely proportional to the downtime). Likewise, traffic load plays a crucial role in energy consumption, which affects the R_{PRB} component in Eq. 2. We would expect that capacity cells that go into cell sleep more often (*i.e.*, higher downtime) consistently handle low daily traffic loads. This is reflected in the results, where we see that energy consumption experiences greater variability as cell uptime becomes longer (left part). This is explained by the increasing heterogeneity of daily traffic loads on those antennas that remain active for a longer time period. To complement our analysis, we show in Figure 6b the average daily energy consumption per antenna for all coverage cells – which remain active all day. The distribution shows that energy consumption also varies significantly among cells of this type, with a median consumption of 3.94 kWh.

During the trial periods, we observed some variability in the aggregated traffic load in the two studied regions. These changes in the user traffic demand typically depend on external factors, such as local holidays and social or sports events. Specifically, we see a traffic variability of 3.2% for the Dense region and 3.7% for the Sparse region over the different weeks of the trials. Note that all strategies were tested for an entire week, which mitigates variability due to the difference in traffic between weekdays and weekends across the different strategies. To make a fair comparison, we take as a reference a model that represents a reverse trial, *i.e.*, *what would have been the energy consumption of cells in the absence of the energy saving strategies*. In this case capacity cells do not enter into cell sleep mode, hence the energy consumption model always accounts for the fixed circuit power γ on all radio units, even if the cell load drops to zero. Conversely, in the tested strategies the model considers E_{RRU}=0 (*i.e.*, γ=0; R_{PRB}=0) when cells enter into sleep mode. This no-sleep configuration allows us to compare the various energy saving strategies in relative terms, reporting their percent saving with respect to a no-sleep scenario so that the results are robust to possible traffic load fluctuations over the trial periods.

Figure 7 (top) shows the energy-saving improvement with respect to the benchmark model (in %) in the two regions. Overall, all strategies exhibit significant energy savings, which are clearly correlated with the policy time windows (night vs full-day) and the ON/OFF thresholds that each strategy implements. The most aggressive Full-strict strategy exhibits an energy saving of 34.5% and 30.2% in the Dense and Sparse regions, respectively. To further investigate the behavior of these strategies along the day, we differentiate between energy savings during daytime (06-23h) and nighttime (23-06h). The nighttime coincides with the time window where Night-loose and Night-strict strategies are deployed. Figure 7 (bottom) shows the distributions of energy savings per cell over the two periods of time. The results reveal that most cells experience greater savings during the night (see median values), which is expected due to the lower traffic load during this time period.

However, we observe an interesting behavior over daytime on full-day strategies: as the thresholds become higher (*i.e.*, more aggressiveness) some cells start to experience significant energy savings during this period. Although the



(a) Dense region (b) Sparse region

Fig. 7: Top - Percentage of energy saving obtained by capacity cells in the two regions; Bottom - Distribution of energy savings over the capacity cells during day and night periods.

median cells achieve similar energy savings across strategies over daytime, a portion of the cells experience considerably greater savings with more aggressive thresholds. Particularly, if we look at the saving of top-25% cells during daytime (*i.e.*, pct-75), these cells achieve savings of 15.8% (8.8%), 21.8% (11.2%), and 42.7% (17.2%) for the Full-loose, Full-moderate, and Full-strict strategies, respectively, in the Dense (Sparse) regions. These results reveal that a considerable portion of the cells can achieve more energy savings during daytime than during the nighttime. Note that these results are subject to the definition of night period, which is 7 hours, and daytime, which extends over the remaining 17 hours in our case.

IV. Impact of Energy Savings on User Performance

The main goal of energy saving strategies is to minimize the energy consumption while ensuring no impact on the Quality of Experience of end-users. MNOs employ several metrics and approaches to quantify service quality, ranging from end-users reporting anomalies to user equipment testing devices deployed across the country and server-end KPIs. We focus here on an objective metric –the average user throughput at the cell level– which is an average estimation of the throughput served to UEs by a specific cell. The question we aim to answer in this section is: when a capacity cell enters into sleep mode, are the *fallback cells* (i.e., the active cells that receive the load previously accommodated by the now sleeping capacity cell) able to deliver a good service performance?

A. Characterizing post-sleep handovers

Related work [16]–[18] assumes that when the capacity cell enters into the energy saving mode and slowly decreases, the emitting power causes the UE to trigger a detach and attach to a specific cell that also fully overlaps the area of the capacity cell. In practice, this is a far more complex phenomenon, and UEs attach to a variety of nearby cells.

We propose a data-driven approach to gain insights into these transitions, and rely on the detach and attach events using logs from the MME that we captured during the trials. We use the following methodology:

- Given the set of capacity cells and the specific time with 1-second resolution, when the cell sleep mode is activated: $(C_i, [t_1, t_2, ..., t_n])$.
- For each capacity cell C_i and for each time instance t_i, we obtain the set of UEs $[\text{UE}_1, \text{UE}_2, ..., \text{UE}_m]$ whose last service request was served by C_i inside the time windows $(t_i - \delta_{start}, t_i]$. This time window is in place to ensure that the interaction was recent.
- Next, we identify the cells F where the UEs are attached to during the time windows $(t_i, t_i + \delta_{end}]$.
- Finally, we obtain a set of tuple $(C_i, t_i, \text{UE}_k, F_l)$ that represents that the UE_k was attached to capacity cell C_i and when this enters into sleep mode, reattached from it, and attach to a *fallback cell* F_l.

We tested different values for δ_{start} and δ_{end} for one day of experimentation. Empirical results showed that using a symmetric window $\delta_{start} = \delta_{end} = 30$min provided a good trade-off to ensure that the transfer was due to the activation of the sleeping mode and that the number of samples collected is sufficient. With this, we collect the final measurements in the two trial regions for more than one week, and obtained more than $81,000$ transfer tuples.

These results show that 40% of the transfers in the Dense region and 44% in the Sparse region occur to relatively nearby cells located in other cell sites (within 1.8km and 3.9km, respectively; note that these distances vary in function of the cell deployment and density). Conversely, 60% and 56% of the attachments happen to active cells in the same site as the capacity cell C_i, of which around 80% of these happen to cells in the same sector. Although more than 40% of the transfers happen to other base stations, they are geographically sparse, meaning that the UEs spread across these cells and their density decays with the increase in distance. Meanwhile, the cells in the same sector as the capacity cell provide service to a more cohesive group of UEs, which may affect network performance. Therefore, we focus on these cells, which we refer to as the *fallback cells* in the rest of the paper.

B. Quantifying critical cell sleep decisions

Our objective is to identify *fallback cells* that were adversely affected by the UEs transfer after cell sleep events, *i.e.*, by a critical decrease in their offered user average throughput. In order to ensure that the user average throughput decrease was related to an increase in demand, we also analyzed the *fallback cell* utilization to disambiguate the cases where the throughput can decay because of low resource utilization (*e.g.*, by users with little traffic demands) or due to the impact of the new UEs absorbed by the cell.

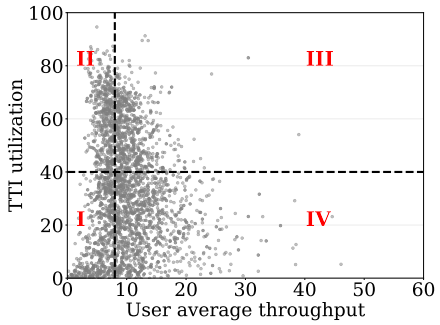


Fig. 8: TTI utilization and user average throughput of fallback cells instants before the corresponding capacity cell enters into sleep mode.

Figure 8 shows how the Transmission Time Interval (TTI) utilization –which, similarly to PRB utilization, is a proxy for the level of saturation of a cell– and user average throughput are related; each sample corresponds to one *fallback cell*, for which we report the two metrics above in the aftermath of a sleep event of a capacity cell in the same site and sector. We then define two thresholds that allow distinguishing between normal/low user throughput and normal/high cell utilization. For TTI utilization the threshold is set at 45% (horizontal line in Figure 8) based on the third quartile of the TTI utilization distribution over the full period of the trials; the average user throughput is considered low below 8Mbps (vertical line in the figure), based on literature analysis, private communications with the operations teams, and statistical analysis of the empirical throughput distribution. Note that these thresholds do indicate non-operational conditions that would trigger an urgent intervention by the MNO but are symptomatic of a significant degradation of the end-user experience.

Using these thresholds, we can define four quadrants in Figure 8: I) low user average throughput and low utilization, II) low average user throughput and high utilization, III) within range user throughput and high utilization, and IV) within range user throughput and low utilization. Based on the transition between states on the *fallback cells* before and after the corresponding capacity cells enter into sleep mode, they can be classified into these four transition classes:

- **Critical**: The *fallback cell* average user throughput was within the expected values, but after the capacity cells sleep, it is lower than the 8Mbps threshold while the utilization increases (*i.e.*, quadrants I, III, and IV $\rightarrow$ quadrant II).
- **Saturate**: The *fallback cell* had below the threshold user throughput and high utilization and remained in this state after the capacity cells went to sleep (*i.e.*, quadrant II $\rightarrow$ quadrant II).
- **Recovery**: The *fallback cell* experiences below the threshold user average throughput and high utilization, but after the capacity cell go to sleep, the user average throughput is within the expected range (*i.e.*, quadrant II $\rightarrow$ quadrants I, III or IV).
- **Normal**: All the cases where the *fallback cell* user's average throughput remains within the expected range before and after the capacity cells sleep, indicating no impact.

With this, Figure 9 shows the percentage of times that we categorized the *fallback cells* in a transaction class. We find that (i) all policies generate all four situations in similar fractions and (ii) the fractions of 'critical' and 'saturated' cells do not grow significantly with the aggressiveness of the policy.

When breaking down energy savings at the level of individual capacity cells, we find a very heterogeneous behavior. Figure 10 shows the energy saving distribution of the capacity cells: there is high variability among the savings, especially with more aggressive policies, in all four cases. This implies that *the same energy saving strategies may be extremely good or bad, depending on the considered cell*. For instance, even a Full-loose policy can be very bad for some cell that saves just 5% of energy, and cause severe saturation every time they enter into sleep mode. Conversely, a Full-strict strategy can work extremely well at some cells, with savings of 60%, that have no impact on the end-user.

V. Discussion

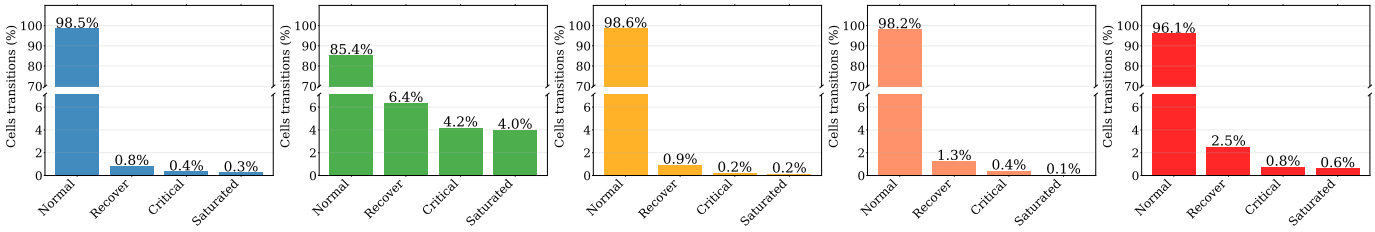
We now discuss the multiple implications of our study for practical energy savings in production mobile networks.

Avenues for improvement in RAN sustainability. Our results clearly show that the use of a single pair of ON/OFF thresholds for all regions is sub-optimal. Fixed thresholds that are indiscriminately applied to a whole RAN deployment, even in the same TAC, risk being overly conservative for some cells and just disruptive for others, as suggested by the diversity seen in Figure 10. A more effective approach could involve setting specific thresholds at the power group level or grouping cells into different clusters with various treatments, which could potentially help yield greater energy savings while keeping an unnoticeable impact on QoE. These tailor-made thresholds could also be dynamically updated using machine learning techniques. Finally, incorporating information about the load of neighboring cells and sites into the decision process could enhance the effectiveness of cell sleep strategies.

Limited energy savings overall. Throughout the paper, we have focused on energy savings at the capacity cells that are allowed to go in sleep mode. We now take a broader view and put the energy savings presented in previous sections in the context of the whole RAN deployments in the target areas.

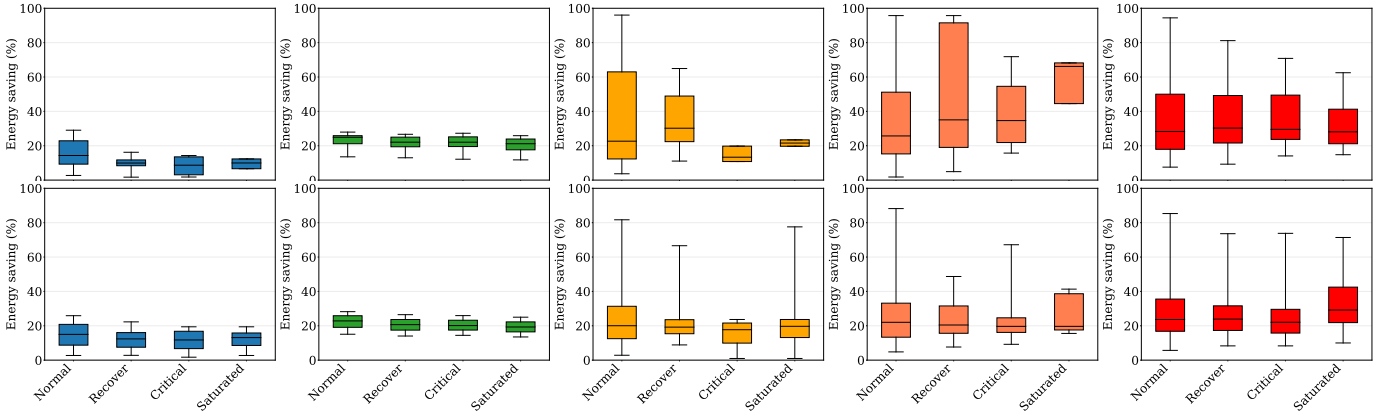
Figure 11 shows (i) the percentage of available cells for each strategy along the day (hourly intervals), including capacity and coverage cells, (ii) the theoretical percentage of cells required to meet the ongoing traffic demand in the region, (iii) the percentage of capacity cells subject to energy saving strategies, and (iv) the percentage of coverage cells, which ensure service along the region. There is a notable gap between available cells, the fraction needed to meet demand, and coverage cells. The Night-strict strategy comes closest to putting all capacity cells to sleep at night, while during the day, the Full-strict strategy is more effective, displaying a similar behavior to other full-day strategies. These results indicate that while these strategies reduce energy consumption, they are far from optimal and could be significantly improved.

Furthermore, Figure 12 shows the total energy saving (in % of improvement over the benchmark model), considering



(a) Night-loose (b) Night-strict (c) Full-loose (d) Full-moderate (e) Full-strict

Fig. 9: Cells transition percentage, for each strategy. Regions were aggregated due to their similar values.



(a) Night-loose (b) Night-strict (c) Full-loose (d) Full-moderate (e) Full-strict

Fig. 10: Energy consumption saving in each transition class for each energy saving strategy. Top - Dense region, bottom - Sparse region.

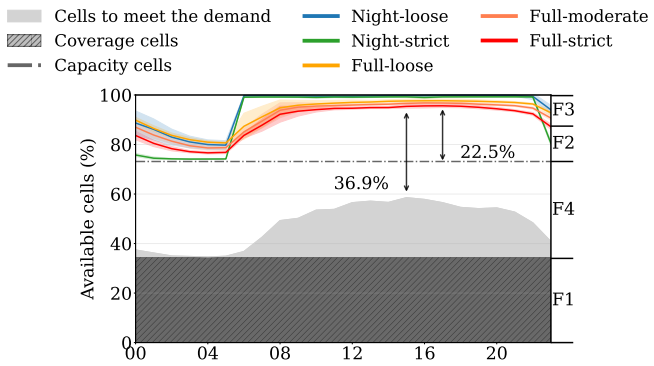
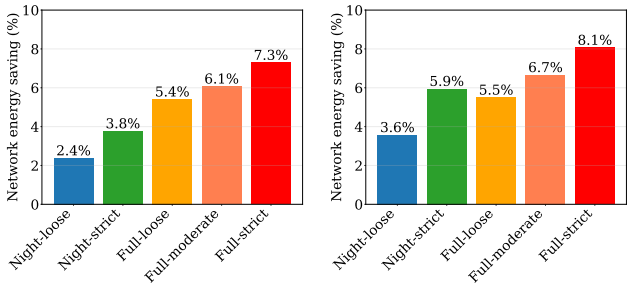


Fig. 11: Percentage of available cells on an hourly basis for each strategy, compared to the fraction of capacity cells, coverage cells, and the cells theoretically needed to meet the current demand. The curves refer to the Sparse region.



(a) Dense region (b) Sparse region

Fig. 12: Total energy saving in the two regions. Although there is a gain in savings as strategies become more aggressive, the total savings are significantly reduced w.r.t. those observed on capacity cells.

all cells deployed in each region – including both coverage and capacity cells. Note that the previous results in Figure 7 considered only savings on capacity cells, which are the ones affected by the cell sleep policies tested by the MNO. These results of total savings reveal more reduced savings at the RAN level, which evidences the conservative nature of current strategies deployed nowadays in production networks.

Energy savings do not imply CO_2 savings. Although this work primarily focuses on energy savings, we also evaluated the impact on CO_2 emissions, specifically Scope 2 emissions as defined by the GHG Protocol [19]. The Greenhouse Gas (GHG)q Protocol categorizes emissions into three scopes: Scope 1 includes direct emissions from company activities like burning fuels; Scope 2 covers indirect emissions from purchased energy, such as electricity from the grid; and Scope 3 encompasses all other indirect emissions across the value chain. Our threshold-based energy-saving policies directly influence Scope 2 emissions, measured in grams of CO_2 equivalent per kilowatt-hour (gCO_2eq/kWh). While reducing energy consumption decreases CO_2 emissions, the alignment between energy savings and emission reductions is not always straightforward. Energy-saving policies focus on reducing energy use without considering the carbon intensity of energy production. In the country under study, carbon fuels constitute a significant part of the energy grid, affecting the overall impact of these policies. Figure 13 shows the carbon emission in the country during the week of the Night-strict trial. Although there is a regular pattern where emissions decrease overnight and increase over daylight hours, there is

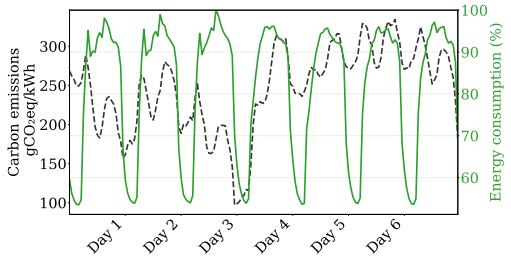


Fig. 13: Estimated carbon intensity in gCO_2eq/kWh and energy consumption in the Sparse region during the week of the second trial. Energy consumption values are normalized with respect to the maximum hourly consumption for confidentiality reasons.

a significant variability along the days that we can exploit to achieve more carbon-efficient policies.

VI. Related work

Energy-aware management of cellular access networks has been extensively studied. A seminal work by Ajmone et al. [20] characterized energy savings by reducing active cells during low-traffic periods. However, this and subsequent studies often assume uniform traffic across cells and complete offloading to a predefined set of fully overlapping neighboring cells. Further research has focused on optimal sleeping policies derived specifically for non-bursty traffic. These studies reveal a two-threshold policy with a wait-and-see feature, where the server waits to see if there are additional arrivals before switching modes [21]. Despite the lack of theoretical results for multiple servers, these findings align with current carrier shutdown solutions that implement distinct conditions for shutdown and reactivation [13].

While numerous studies have explored enhancing energy efficiency in telecommunication networks through *energy consumption measurement* and *policies for optimizing energy consumption* [9], [22], a gap remains in understanding the real-world impact of these policies. To sensibly assess the performance of different policies, it is necessary to conduct regional trials. This approach would provide a real evaluation of the energy gains or savings achieved by each policy.

Measuring energy consumption. The energy consumption of mobile networks is a major concern for the telecom industry. To evaluate the energy-related performance of the RAN, the industry has defined various measurement methods and metrics across different network levels—network [23], site [24], base station (BS) [25], and user equipment [26]—as well as for different scenarios such as dense urban, urban, and rural coverage [23], and services including enhanced mobile broadband, ultra-reliable low-latency communications, and massive machine-type communications [27]. Piovesan et al. [28] propose a model for characterizing the power consumption of 5G multi-carrier BSs, built on extensive data collection. Additionally, Lopes et al. [9] conducted a survey outlining the primary energy efficiency technologies provided by 3GPP NR, detailing their benefits and challenges. The community recognizes the importance of sustainability, with efforts underway to quantify the carbon footprint of communication networks [29] and optimize network routing [30], alongside general efforts for computer systems [31]. In this study, we present utilization and performance data collected at various times throughout the day. This information is instrumental in assessing whether base station utilization can benefit from temporal flexibility, a crucial aspect in achieving greater carbon efficiency.

Optimizing energy saving policies. Cell sleep policies involve the dynamic shutdown of cells during low-traffic periods and their subsequent reactivation as traffic increases. Accurately predicting traffic loads in the near future is crucial for determining the appropriate times to switch base stations on and off. Marsan et al. [32] [33] evaluated various fixed switching-off schemes in cellular networks, finding that optimal configurations vary with traffic profiles. They later showed that cooperative strategies achieve significant savings among operators for user roaming and energy savings [34]. Donevski et al [35] propose using dense Neural Network and Recurrent Neural Network paradigms, although their approach lacks empirical validation through real-world trials. Moreover, Domenico et al. [10] focus on the carrier shutdown approach, enabling base stations to autonomously power down during low-traffic periods by transferring their load to neighboring active base stations. Other researchers have employed reinforcing learning methods [36] [37], but such approaches have been tested only in simulated environments. It is crucial to acknowledge that such optimization involves the sudden removal and addition of carriers to the network. Therefore, there is a need to develop automated and precise methods for generating configuration parameters for newly added carriers in cellular networks (*i.e.*, planning), as suggested by [11]. These types of solutions are orthogonal to previously mentioned works and to the energy-saving strategies analyzed in this paper.

VII. Conclusions

We analyzed five fixed threshold-based cell sleep energy saving strategies deployed in a production network, examining various dimensions such as downtime, energy consumption, and impact on user performance. The unprecedented visibility into practical solutions for RAN sustainability lets us shed light on the performance and current limitations of these strategies, as well as provide recommendations for the design of new and more effective approaches that can still be deployed in real-world production-grade networks.

Acknowledgment

This work was supported by NetSense Talent Attraction grants 2019-T1/TIC-16037 and 2023-5A/TIC-28944, funded with public funds from Comunidad de Madrid; by the ORIGAMI project (Grant no. 101139270) and the INSTINCT project (Grant no. 101139161), funded by the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation programme; by the Spanish Ministry of Economic Affairs and Digital Transformation and the European Union – NextGeneration EU, in the framework of the Recovery Plan, Transformation and Resilience (PRTR) through the UNICO I+D 5G SORUS project; and by EU Horizon Framework grant no. 101060294 (XGain).

REFERENCES

[1] H. Technology, "Green 5g: Building a sustainable world," 2020.

[2] G. Association *et al.*, "5g energy efficiencies: green is the new black," *white Paper*, 2020.

[3] G. Association *et al.*, "Mobile net zero, state of the industry on climate action 2024," *Accessed: Aug*, vol. 8, 2024.

[4] G. Green, "Going green: Benchmarking the energy efficiency of mobile," *London, UK: GSMA*, 2021.

[5] S. Mishra, Z. Smoreda, and M. Fiore, "Second-level digital divide: A longitudinal study of mobile traffic consumption imbalance in france," in *Proceedings of the ACM Web Conference 2022*, WWW '22, (New York, NY, USA), p. 2532–2540, Association for Computing Machinery, 2022.

[6] T. Aledavood, E. López, S. G. Roberts, F. Reed-Tsochas, E. Moro, R. I. Dunbar, and J. Saramäki, "Daily rhythms in mobile telephone communication," *PloS one*, vol. 10, no. 9, p. e0138098, 2015.

[7] D. Gundogdu, O. D. Incel, A. A. Salah, and B. Lepri, "Countrywide arrhythmia: emergency event detection using mobile phone data," *EPJ Data Science*, vol. 5, pp. 1–19, 2016.

[8] A. Furno, M. Fiore, R. Stanica, C. Ziemlicki, and Z. Smoreda, "A tale of ten cities: Characterizing signatures of mobile traffic in urban areas," *IEEE Transactions on Mobile Computing*, vol. 16, no. 10, pp. 2682–2696, 2017.

[9] D. Lopez-Perez, A. D. Domenico, N. Piovesan, G. Xinli, H. Bao, S. Qitao, and M. Debbah, "A survey on 5g radio access network energy efficiency: Massive mimo, lean carrier design, sleep modes, and machine learning," *IEEE Communications Surveys and Tutorials*, vol. 24, pp. 653–697, 2022.

[10] A. D. Domenico, D. Lopez-Perez, W. Li, N. Piovesan, H. Bao, and X. Geng, "Modeling user transfer during dynamic carrier shutdown in green 5g networks," *IEEE Transactions on Wireless Communications*, vol. 22, pp. 5536–5549, 8 2023.

[11] A. Mahimkar, A. Sivakumar, Z. Ge, S. Pathak, and K. Biswas, "Auric: Using data-driven recommendation to automatically generate cellular configuration," pp. 807–820, ACM, 8 2021.

[12] C. Ge, Z. Ge, X. Liu, A. Mahimkar, Y. Shaqalle, Y. Xiang, and S. Pathak, "Chroma: Learning and using network contexts to reinforce performance improving configurations," in *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*, pp. 1–16, 2023.

[13] D. López-Pérez, A. De Domenico, N. Piovesan, and M. Debbah, "Data-driven energy efficiency modelling in large-scale networks: An expert knowledge and ml-based approach," *IEEE Transactions on Machine Learning in Communications and Networking*, 2024.

[14] R. Tan, Y. Shi, Y. Fan, W. Zhu, and T. Wu, "Energy saving technologies and best practices for 5g radio access network," *IEEE Access*, vol. 10, pp. 51747–51756, 2022.

[15] Y. Ma, T. Li, Y. Du, S. Dustdar, Z. Wang, and Y. Li, "Sustainable connections: Exploring energy efficiency in 5g networks," in *Proceedings of the 20th International Conference on Emerging Networking EXperiments and Technologies*, CoNEXT '24, (New York, NY, USA), p. 33–40, Association for Computing Machinery, 2024.

[16] H. Celebi, Y. Yapıcı, I. Güvenç, and H. Schulzrinne, "Load-based on/off scheduling for energy-efficient delay-tolerant 5g networks," *IEEE Transactions on Green Communications and Networking*, vol. 3, no. 4, pp. 955–970, 2019.

[17] S. M. Shahid, Y. T. Seyoum, S. H. Won, and S. Kwon, "Load balancing for 5g integrated satellite-terrestrial networks," *IEEE Access*, vol. 8, pp. 132144–132156, 2020.

[18] W. Teng, M. Sheng, X. Chu, K. Guo, J. Wen, and Z. Qiu, "Joint optimization of base station activation and user association in ultra dense networks under traffic uncertainty," *IEEE Transactions on Communications*, vol. 69, no. 9, pp. 6079–6092, 2021.

[19] GHG Protocol, "Greenhouse gas protocol," 2023.

[20] M. Ajmone Marsan, L. Chiaraviglio, D. Ciullo, and M. Meo, "Optimal energy savings in cellular access networks," in *2009 IEEE International Conference on Communications Workshops*, pp. 1–5, 2009.

[21] B. Leng, X. Guo, X. Zheng, B. Krishnamachari, and Z. Niu, "A wait-and-see two-threshold optimal sleeping policy for a single server with bursty traffic," *IEEE Transactions on Green Communications and Networking*, vol. 1, no. 4, pp. 528–540, 2017.

[22] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, "Green cellular networks: A survey, some research issues and challenges," *IEEE Communications Surveys and Tutorials*, vol. 13, no. 4, pp. 524–540, 2011.

[23] ETSI Environmental Engineering, "Es2023228: Assessment of Mobile Network Energy Efficiency," October 2020.

[24] ITU, "Energy Efficiency Measurement for Telecommunication Equipment," October 2016. Available: https://www.itu.int/rec/T-REC-L.1350-201610-I/en.

[25] ETSI Environmental Engineering, "TS103786: Measurement Method for Energy Efficiency of Wireless Access Network Equipment; Dynamic Energy Efficiency Measurement Method of 5G Base Station (BS) Dynamic Energy Efficiency Measurement Method of 5G BS," December 2020. Available: https://www.etsi.org/deliver/etsi_ts/103700_103799/103786/01.01.01_60/ts_103786v010101p.pdf.

[26] 3GPP, "Technical Specification Group Radio Access Network; NR; Study on User Equipment (UE) Power Saving in NR (Release 16)," Technical Report TR 38.840, V16.0.0, 3GPP, June 2019.

[27] 3GPP, "Technical Specification Group Services and System Aspects; Management and Orchestration; Study on New Aspects of Energy Efficiency (EE) for 5G (Release 17)," Technical Report TR 28.813, V17.0.0, 3GPP, 2021.

[28] N. Piovesan, D. Lopez-Perez, A. D. Domenico, X. Geng, H. Bao, and M. Debbah, "Machine learning and analytical power consumption models for 5g base stations," *IEEE Communications Magazine*, vol. 60, pp. 56–62, 10 2022.

[29] I. E. Agency, "Data centres and data transmission networks," 2024. Accessed: 2024-06-05.

[30] S. El-Zahr, P. Gunning, and N. Zilberman, "Exploring the benefits of carbon-aware routing," vol. 1, nov 2023.

[31] W. A. Hanafy, Q. Liang, N. Bashir, D. Irwin, and P. Shenoy, "Carbonscaler: Leveraging cloud workload elasticity for optimizing carbon-efficiency," *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, vol. 7, pp. 1–28, 12 2023.

[32] M. A. Marsan, L. Chiaraviglio, D. Ciullo, and M. Meo, "Optimal energy savings in cellular access networks," in *Proc. IEEE ICC Workshops*, pp. 1–5, June 2009.

[33] M. Marsan, L. Chiaraviglio, D. Ciullo, and M. Meo, "Multiple daily base station switch-offs in cellular networks," in *Proc. 4th ICCE*, pp. 245–250, August 2012.

[34] M. A. Marsan and M. Meo, "Energy efficient management of two cellular access networks," *ACM SIGMETRICS Performance Evaluation Review*, vol. 37, pp. 69–73, March 2010.

[35] I. Donevski, G. Vallero, and M. A. Marsan, "Neural networks for cellular base station switching," pp. 738–743, IEEE, 4 2019.

[36] F. Meng, P. Chen, L. Wu, and J. Cheng, "Power allocation in multi-user cellular networks: Deep reinforcement learning approaches," *IEEE Transactions on Wireless Communications*, vol. 19, no. 10, pp. 6255–6267, 2020.

[37] J. S. Pujol–Roigl, S. Wu, Y. Wang, M. Choi, and I. Park, "Deep reinforcement learning for cell on/off energy saving on wireless networks," in *2021 IEEE Global Communications Conference (GLOBECOM)*, pp. 01–07, 2021.