Green Future Networks: A Roadmap to Energy Efficient Mobile Networks

v1.0

www.ngmn.org

WE MAKE BETTER CONNECTIONS

GREEN FUTURE NETWORKS: A ROADMAP TO ENERGY EFFICIENT MOBILE NETWORKS

by NGMN Alliance

Version:	1.0
Date:	02 July 2024
Document Type:	Final Deliverable (approved)
Confidentiality Class:	P - Public
Project:	Green Future Networks
Lead Editor / Submitter:	Rishikesh Chakraborty, Vodafone Antonio De Domenico, Huawei
Programme office:	Chris Hogg, representing NGMN
Approved by / Date:	NGMN Board, 24 June 2024

EXECUTIVE SUMMARY

The mobile industry through Next Generation Mobile Network's (NGMN), Green Future Networks programme is working together to increase network energy efficiency (EE). This publication, outlines and prioritises the various options explored within the third phase of the Green Future Networks programme. The previous publication addressed short term solutions deployed by mobile network operators (MNOs) to address a challenging energy landscape. By contrast, this report further expands on the potential technologies that could pave the way for a network EE roadmap to address upcoming environmental and economic challenges faced by MNOs, with each respective potential solution being mapped to a time-horizon indication for its potential future implementation.

First, being able to measure Radio Access Network (RAN) EE is an important part of the Process. There are two approaches explored for assessing Base Station (BS) equipment in the lab; the static measurement procedure for power consumption and the dynamic measurement procedure for EE. Depending on the context, each has their place when testing today's modern BS equipment. However, a BS can be used in many different configurations, purposes and contexts.

In addition, to support network EE optimisation, Artificial Intelligence (AI) could be a key tool to provide Energy Consumption (EC) estimation and prediction while limiting the amount of data collected and transferred throughout the network. This report recommends standards organisations to define methodologies to transfer and update AI models at the nodes where network configuration and parameters are controlled. In this context, with the increasing size of AI models, we advocate the need for solutions able to adjust the model complexity to minimise the EC that arises from: model training, transmission and execution.

We also highlight that, soon, MNOs may seek to integrate novel hardware and software mechanisms to support AI-based network EE modelling and optimisation. At the software level, integrating new intelligent solutions could allow the network to reduce EC by adjusting the available network capacity to the actual traffic load at each given point in time. In indoor deployments, a new energy saving technology is proposed to manage the state of each RU all belonging to a given cell independently to dynamically switch off the Power Amplifier (PA) of any RUs that do not harbour user connection or data transmission at a given point of time. Trials have highlighted that this solution achieves an energy saving gain of 20% relative to an always-on network deployment. Where switching off radio components is not possible, due to non-negligible load, this report shows that a RU implementing an intelligent resource allocation that decreases the transmit power by limiting the transmission spectral efficiency could lead up to 30% reduction in load-dependent EC at the RU, without impacting users' Quality of Service (QoS). We recommend standards organisations to define methodologies to coordinate properly RUs tasked with implementing distinct and potentially competing energy saving mechanisms.

Further EE gains could be realised through RAN solutions that implement different levels of coordination to achieve a more efficient usage of network resources. More specifically, this publication highlights trials related to a novel network optimisation approach that, leveraging heterogeneous QoS requirements in the service area, results in up to 18% of reduction in the average RAN EC. In addition, this report presents network level solutions where multi-carrier coordinated scheduling and spectrum sharing are respectively combined with cell discontinuous transmissions and carrier shutdown to lead up to 26%, EE gain at the RUs.

In addition to coordination mechanisms implemented through new software functionalities, MNOs are encouraged to share part of their wireless infrastructure to reduce component duplications and jointly utilise network resources through RAN sharing to limit EC and carbon emissions.

As reported in the previous publication of the Green Future Networks Programme [1], the usage of renewable energy sources is a crucial solution for reducing carbon emissions and limiting the mobile network dependency on electrical grids. This report also highlights the need for solutions to jointly dimension the power supply and network communication resources. We recommend standards organisations to enhance interworking between mobile networks and the energy suppliers to effectively reduce costs and the respective carbon footprints, while maintaining service availability.

Virtualised and disaggregated mobile networks speed up network deployment and management which can improve operational efficiency. In this context, this publication overviews the current state of the O-RAN ecosystem and the actions envisaged to accelerate the deployment of energy efficient networks.

The report highlights practical applications of AI and its related challenges within mobile networks. We show that AI algorithms could help to make better energy saving decisions, such as controlling the energy saving policy thresholds, by predicting future EE and load states as well as by identifying low-EE sites.

NGMN advises that MNOs carefully assess the solutions to enhance network energy efficiency presented in this publication and analyse the prioritised list of available strategies presented in the Conclusion section.

TABLE OF CONTENTS

01 INTRODUCTION

NGMN's Green Future Networks Programme has the ambition to lead the telecoms industry towards energy efficient operations and a sustainable economy. Following the previous publication presenting the collective MNO response to the energy price challenges confronting the telecoms industry [1], in this publication we outline and prioritise the various options available to build an ambitious EE roadmap. In particular, EE approaches are organised into three broad categories (and time-horizons), and for each solution, information is provided – based on data from live networks and/or simulations - on the size and scope of the potential energy savings (see Section 9).

These solutions are presented in detail in this publication as follows: Section 2 presents measurement methods and related standards that should be used by MNOs to properly measure EC and EE. Section 3 provides an overview of 3GPP New Radio (NR) efforts to enhance network EE and highlights future standardisation challenges. Section 4 describes solutions that could improve the EE of the RUs, which account for the 80% of the BS power consumption. Section 5 introduces mechanisms that could enhance the RAN EE by optimising network resource utilisation and leveraging variations in traffic demand over time and space, without compromising users' QoS. Section 6 presents the challenges of a potential new ecosystem where MNOs and energy suppliers collaborate to meet both environmental and economic sustainability targets, harnessing renewable energy sources, energy storage facilities, and real-time information exchange. Section 7 outlines prospects for boosting EE in virtualised and disaggregated networks, emphasizing key priorities for Open RAN deployment. Section 8 highlights the advantages and challenges related to the integration of Artificial Intelligence (AI) and Machine Learning (ML) to achieve greater net EE in telecoms networks. Finally, Section 9 handles the most recent key learnings from NGMN's Green Future Networks Programme, highlighting the potential EE gains of the solutions presented in this white paper.

02 MEASURING RAN ENERGY EFFICIENCY

Being able to measure energy efficiency in the Radio Access Network (RAN) is an important part of the process of improving it. Depending on the purpose, different types of measurement methods are used by MNOs. In the procurement process, it is common to use lab tests to measure the EC and efficiency of the equipment to be used in the network. During network operation, different performance metrics can be monitored to see how the network EC and network EE evolve over time. In recent years, networks have evolved to become both more advanced and complex. 5G-NR has been introduced, new types of equipment such as Active Antenna Unites (AAUs) capable of massive Multiple Input Multiple Output (MIMO) are being widely deployed, and different features for performance enhancement, network optimisation, and energy saving are becoming more common. This obviously impacts network EC and network EE, and also affects how EC and EE should be measured in an accurate way. This section will highlight some of these impacts, and related measurement aspects and standards.

2.1. EQUIPMENT LEVEL MEASUREMENTS

2.1.1. STATIC AND DYNAMIC MEASUREMENTS

There are two approaches for testing BS equipment in the lab; one for measuring the power consumption in a static procedure, and the other for measuring the EE in a dynamic procedure. In the static measurement procedure, a static Physical Resource Block (PRB) load is put on the equipment under test and the resulting power consumption is measured. This is repeated for three load levels (low, medium and high – the high load reflects a typical traffic load in a busy hour), and by post-processing the average power consumption as well as the daily energy consumption of the equipment under test can be calculated. The main advantages of this test are its simplicity for setting up and conducting the test, but it has limitations in its ability to reflect the behaviour of a BS in a real-world network operation. For example, the effects of varying radio channel environments and some traffic dependent EE features are not captured. Static tests are widely used by MNOs, the most well-known standard is provided by ETSI [2]. ATIS, CCSA and ITU-T have similar standards.

To capture in a more dynamic manner the effects of different EE features, functionalities and radio network characteristics, a so-called dynamic measurement procedure can be used. In this test, a User Equipment (UE) emulator is used to generate traffic from several UEs at different locations to load the BS under test (see Figure 1), and the BS is also enabled to use all possible RAN features and functionalities. Time and even space variations and burstiness of the data traffic are captured, which allows performance enhancing and energy saving features to take effect accordingly. Power consumption is measured during the test, but also the delivered performance in terms of data volume, which allows to calculate both energy consumption as well as EE of the BS equipment. Hence, this test provides information on the dynamic behaviour of the BS equipment in operation but requires on the other hand a more advanced measurement set-up in the lab, e.g., UE as well as channel and fading emulators. ETSI has recently released new versions of the dynamic measurement standards for LTE [3] and NR [4] (conducted tests only), respectively, and also CCSA and ITU-T [5] are in the process of defining/updating dynamic BS measurement standards.

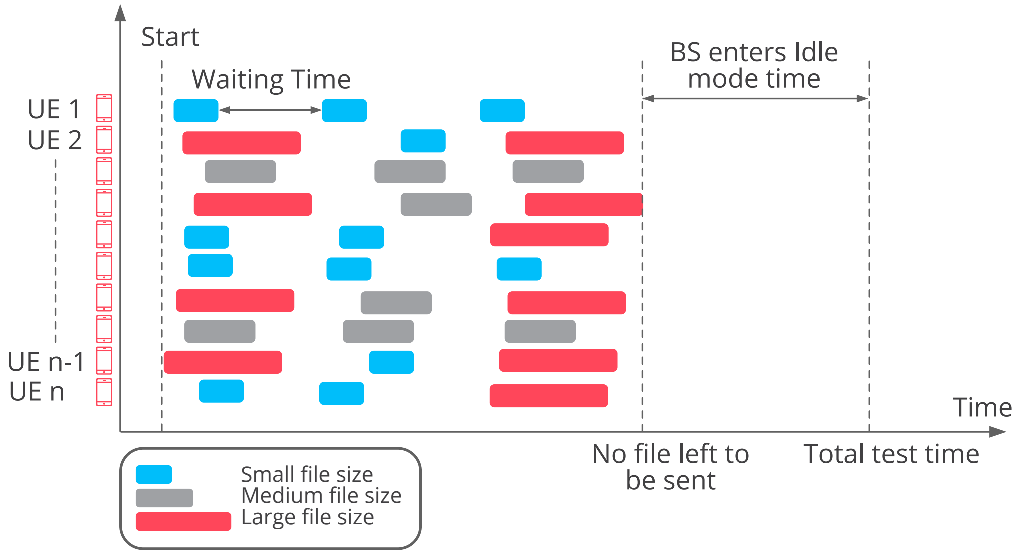

Figure 1. Illustration of traffic model in dynamic test [3], [4].

For today's modern BS equipment, which operate in more complex mobile network environments with many performance enhancing and optimising features and functionalities, both static tests and dynamic tests have their place. They are not mutually exclusive but offer a wide array of laboratory measurement methods depending on the needs and context. However, it should be emphasized that these lab tests only provide snapshots, while the most reliable and accurate measurements of energy consumption and energy efficiency are obtained at network level, which will be discussed in Section 2.2.

2.1.2. MEASUREMENT ON ACTIVE ANTENNA UNITS

When measuring power and energy consumption of RUs, standards [2] mandate that the RUs should be configured to have the same output power. This should be verified by measuring the output power at the test port. For Remote Radio Units (RRUs), this typically means the antenna connectors of the RRU, i.e., before the passive antenna (see Figure 2, left). Once the RRUs are tuned following this guideline, their power consumption can be measured. Thereafter, EC [2] and power efficiency [5] values can also be derived. Since the assumption is that the RRUs under test can be connected to the same passive antenna, fair tests are guaranteed, as in this case, the same output power at the RRU results in the same effective isotropic radiated power (EIRP)[1] from the antenna.

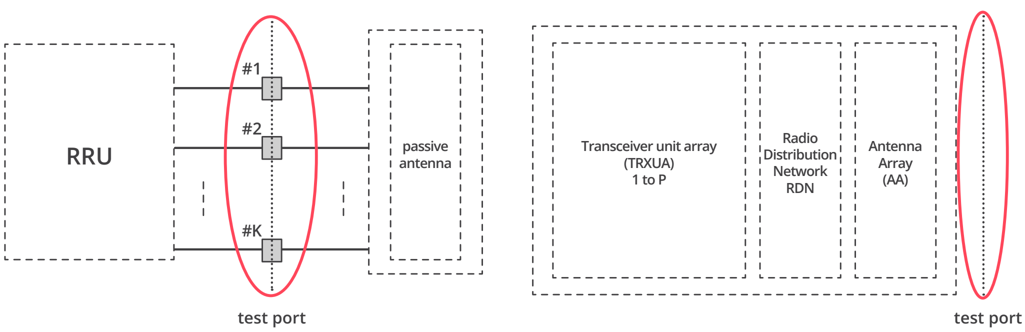

Figure 2. Test ports – antenna connectors of the RRU [2], [6]] (left), and the RIB of the AAU [2], [7] (right).

[1] EIRP [dBm] = TX power [dBm] + Antenna Gain [dB].

However, since AAUs typically have their passive antenna arrays integrated together with the radio chains in one single unit, the same procedure to measure power and energy consumption used for RRUs cannot be applied. The reason is that different AAUs may have different antenna array capabilities, e.g., one has more antenna elements than the other, and/or the directivity of their antenna elements differs. If we would test them using the procedure described above, i.e., configuring them to have the same output power, the different antenna array capabilities of the AAUs would result in different EIRPs, i.e., different operating points. Therefore, for AAUs, the test port is defined as the radiated interface boundary (RIB), i.e., after the antenna (see Figure 2).

Hence, for AAUs an EIRP-based test (see Figure 3) should be used when measuring power and energy consumption. Such a test is developed by 3rd generation partnership project (3GPP) [7] and adopted by ETSI for EC measurements [2]. It is then important that the AAUs to be measured are tuned to provide the same EIRP in order to make sure that the measurements are carried out at the same operational point.

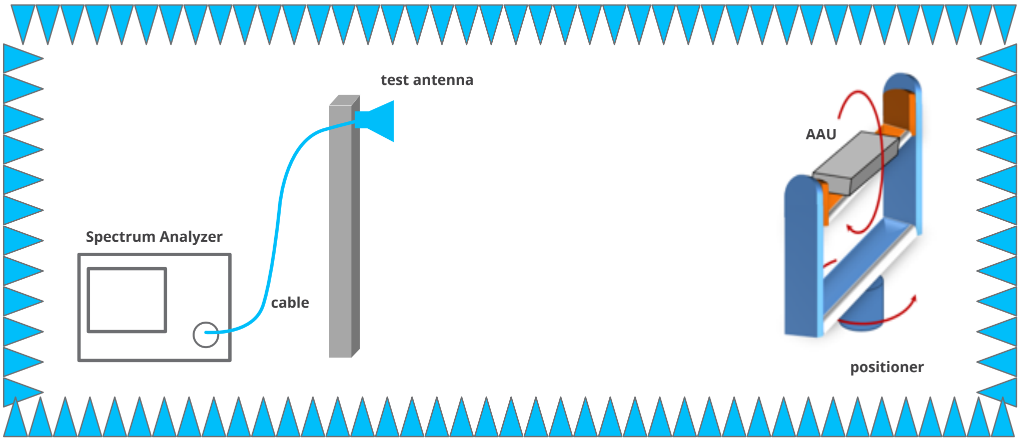

Figure 3. EIRP-based AAU test.

2.2. NETWORK LEVEL MEASUREMENTS

The previous subsection focused on BS equipment measurements in the lab, however, the most reliable and accurate energy efficiency Key Performance Indicators (KPIs) can be obtained at network level. A mobile network can be built in many different ways, and at network level the whole dynamism including interference from other cells, real world radio channel characteristics, deployment and operation aspects, as well as RAN features and functionalities come into play and affect the network EC and EE.

As no two networks are the same, it is difficult to compare network EE of different networks. However, by monitoring different performance metrics and EC, it is possible for MNOs to see how network EC and network EE evolve over time in their respective networks. For a long time, delivered data volume has been the main performance metric to consider when assessing network EE. However, with the roll-out of 5G new services and use cases have been introduced. For these, there might be other performance metrics that are of importance, e.g. number of connections, latency, and/or reliability. 3GPP has defined EE metrics based on these [8], [9], while ETSI [10] and ITU-T [11] define methodology for assessing network EE on a network level using these metrics. An update of these standards is ongoing, where e.g., EC of Virtualised Network Functions (VNFs) is considered. It is recommended that MNOs use metrics and methods in these standards to monitor and assess how their network energy efficiency evolves over time. In addition, one key challenge is to provide EC-related information to monitor, control and optimise the network without increasing the overall EC. AI could be a key tool to provide EC estimation and prediction in the most efficient manner.

03 OVERVIEW OF 3GPP RADIO ACCESS NETWORK ENERGY EFFICIENCY ACTIVITIES

In the context of mobile communications, the EE of the mobile terminals has always been an important design criterion. This is mainly driven by the requirement for efficient usage of the handset battery. However, the EC at the network side is becoming increasingly important due to reasons of economic sustainability, especially regarding the EE of the RAN since the RAN is the most hungry energy consumer in a mobile network [12].

Consequently, the NR standardisation activities in 3GPP targeted the so-called Network Energy Saving (NES) functionalities in Release 18 that improve the EE at the BS side as depicted in Figure 4. Most of the NES functions aim at reducing the BS power consumption for low or zero carried load conditions. The ideal BS power profile would then be a power consumption very close to zero Watt for zero carrier load, which is conceptually represented with the dashed green line in Figure 4. Note, however, that this ideal/desired power profile would also require many hardware optimisations that fall outside of the 3GPP standardisation work.

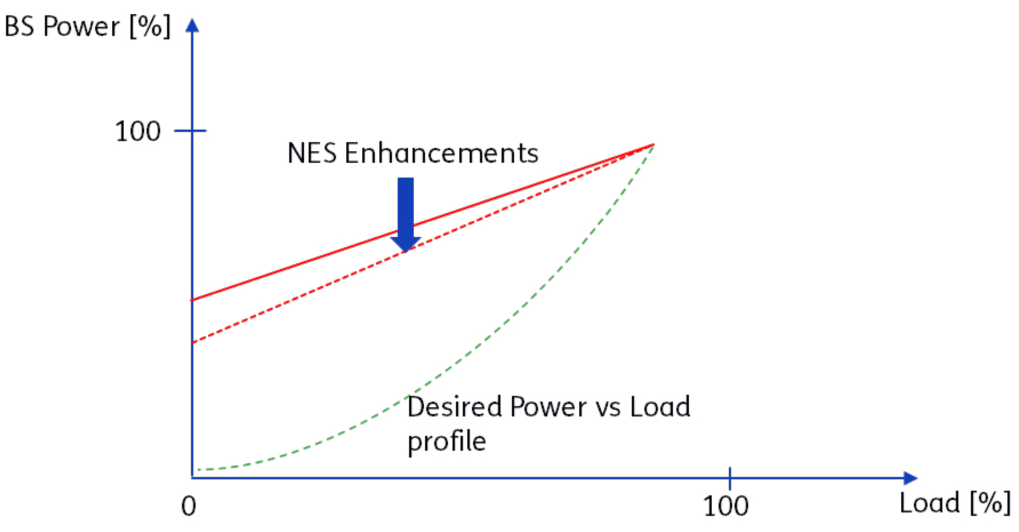

Figure 4. NES enhancements towards improving power vs load BS profile.

The 3GPP NR standardisation work on NES in Release 18 is summarised in [13], where companies have also quantified the achievable power consumption gains of their proposed NES functionalities versus the impact on the service quality, e.g., uplink or downlink throughput, access delay, etc.

Finally, a few promising functionalities from the different proposed NES functions in [13] were selected for the standardisation in Release 18, as briefly described in the following section.

3.1. NETWORK ENERGY SAVING TECHNIQUES IN 3GPP NR

3GPP NR initiated its efforts in energy-saving with a focus on UE power saving in Release 16, spanning from 2019 to 2020 [14]. As indicated in Figure 5, this pursuit continued through Release 17 [15], and by the advent of Release 18, the focus of discourse changed towards NES strategies. The initial phase of Release 18 involved studying and exploring diverse techniques applicable to the RAN for energy reduction on the network side [13].

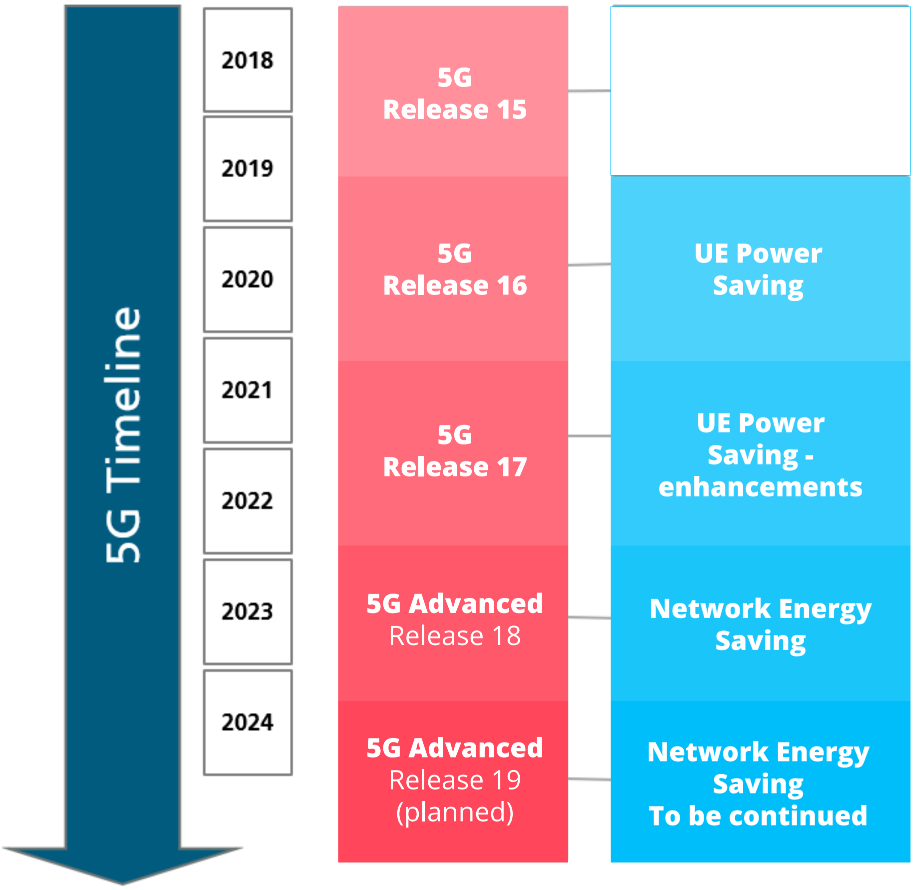

Figure 5. 3GPP standardisation timeline since the start of UE power saving.

Currently, networks' EC does not fully align proportionally with the traffic load. This indicates a high energy-saving potential during low-traffic periods (e.g., nights) without significantly impacting user experience. The NES techniques studied by 3GPP NR and documented in [13] fall into four categories: Time, Frequency, Spatial, and Power domain techniques, each targeting NES in the respective domain.

- The time domain techniques enable component-level shutdown during inactive periods, with no transmission or reception by the BS. Multiple sleep (inactivity) levels were defined, as indicated in Figure 6, with deeper sleep levels involving larger groups of components. Crucially, the depth of sleep modes correlates with the extent of component deactivation. However, a trade-off exists, as deeper sleep modes entail longer delays in transitioning away from that sleep mode. With the current configuration deployed in 5G NR due to periodically broadcasting System Information, the extended transition time of 25 ms will prevent the Next Generation Node B (gNB)[2] from going to deep sleep.

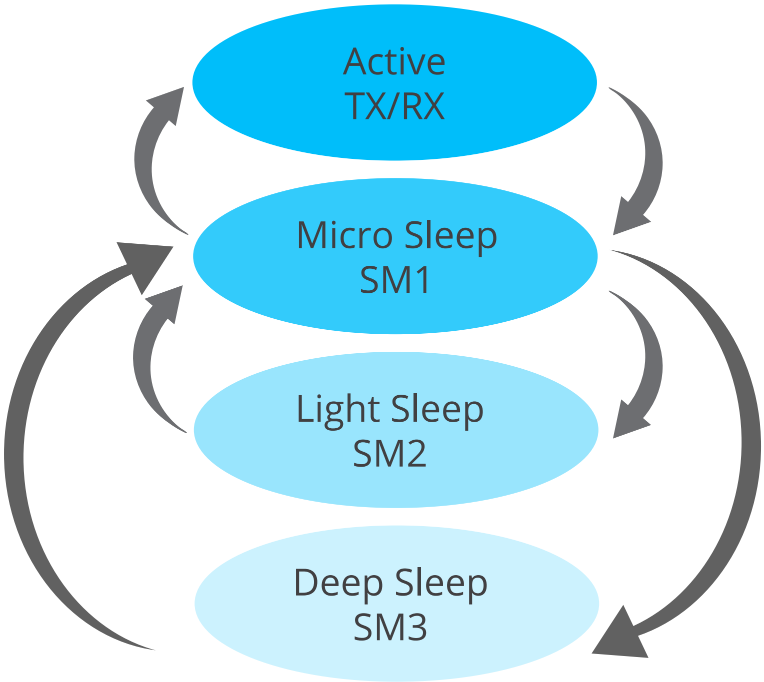

Figure 6. gNB energy saving states.

- The frequency domain techniques explored the adaptation of frequency resources in general, including the shutdown of secondary cells/carriers (SCell) in multi-carrier operation and adjustments in bandwidth segments.

- The spatial domain techniques involved dynamically deactivating spatial elements, i.e., antenna ports/elements and/or TX/RX points (TRPs) in multi-TRP operation.

- The power domain techniques delved into dynamic adaptation of downlink (DL) transmit power and energy-efficient DL transmission through novel approaches for peak-to-average power ratio reduction, pre-distortion, post-distortion, PA power bias adaptation etc.

[2] A BS in the 3GPP NR terminology.

Among the numerous techniques considered during the NES study item, the Release 18 work item introduced the following new features:

Channel State Information (CSI) enhancements for adaptations in spatial and power domains, introducing a new CSI reporting framework, which enables the UE to report N>=1 CSI sub-report(s) in one reporting instance, each sub-report corresponds to multiple TX shutdown patterns and/or DL power reduction. This enhanced CSI reporting enables the gNB to perform more dynamic adaptation in the spatial and/or power domains for energy saving.

Cell Discontinuous Transmission and Discontinuous Reception (Cell DTX/DRX), involves a technique where the gNB can enter a sleep state during periods of inactivity [11]. However, specific signals, such as Synchronization Signal Block (SSB), System Information Block (SIB)-1, paging, and others, need to be periodically transmitted regardless of the UEs presence in the cell, which limits the ability to enter into sleep modes. To facilitate reducing gNB activity time, Release 18 has introduced features to enable the alignment between Cell DTX/DRX and UE C-DRX Connected Mode, such that control channels/signals are not expected to be received/transmitted on the corresponding cell by a given UE.

SSB-less SCell is extended from the intra-band contiguous FR1 co-located scenario (supported in Release-15) to inter-band FR1 co-located scenario. By omitting the SSB transmission, the energy cost in SCell can be reduced. The scenarios and working conditions for inter-band SSB-less SCell, and how to measure and get essential information for a SCell are all specified.

A new barring bit, optionally present in SIB1, was introduced in Release 18 to allow only UEs capable of cell DTX/DRX to access a cell that has enabled this technique, thus preventing legacy UEs camping on cells adopting the Release-18 NES techniques.

Finally, Release 18 has introduced **enhancements on Conditional Handover (CHO) procedure for NES cell(s) to handover a UE to other cells** as fast as possible if the source cell enters NES mode.

The above Release 18 techniques provide more flexibility for the gNB to realise power-saving gains. When these techniques are applied, the BS needs to make appropriate decisions based on feedback from the UE and the network KPIs to strike an appropriate balance of user experience and network energy saving.

In this context, AI/ML algorithms may be used to optimise network energy-saving decisions by leveraging the data collected in the RAN to address inaccurate cell load prediction, individual gNB efficiency, and performance trade-offs. More specifically, AI/ML algorithms can infer future energy saving opportunity or gNB load to enhance decision-making and balance the trade-off between network performance and energy saving.

In 3GPP Release 18, dynamic adaptations of transmission power and antennas stand out as the most effective techniques for improving the EE of BS transmissions. These techniques can achieve RU energy savings of 15-30% under low-to-medium cell load levels [16]. When considering daily average load, the combined use of these techniques offers the optimal balance between energy savings and throughput impact. By implementing careful design and mitigation strategies, it is possible to completely avoid any impact on user throughput.

It is recognised that Release 18 focused primarily on Radio Resource Control (RRC) Connected UEs, user-specific signals and channels, while enhancements related to the UEs in RRC Idle/Inactive state were not considered. Further improvements in network energy saving are expected to continue in 3GPP studies. The Release 19 covers the procedures and signalling methods to support on-demand SSB secondary cell operation for UEs in connected mode configured with intra-band and inter-band carrier aggregation. In addition to this, 3GPP will work on on-demand SIB-1 for UEs in idle/inactive mode and common signal/channel transmissions adaptation.

3.2. FUTURE OPTIMISATIONS IN RAN BEYOND 3GPP STANDARDISATION ACTIVITIES

3GPP agreed in Release 18 not to alter the regular transmission of common signals by the gNB, which limits the potential gain in EE since the gNB cannot enter deeper sleep modes and has to wake up periodically to transmit system information. One viable optimisation strategy could be to increase the periodicity of broadcasting system information. By doing so, the gNB would have extended idle periods, enabling it to enter longer sleep modes and efficiently wake up to transmit the essential common signals. This adjustment could enhance energy-saving capabilities in the network.

As an illustrative example, the impact of the periodicity of SSB transmission is presented in Figure 7. With an increase in SSB periodicity, there is a potential to enter deeper sleep modes, showcasing the correlation between transmission intervals and the achievable level of energy-saving modes. This method has the potential to deliver notable energy saving gains [15].

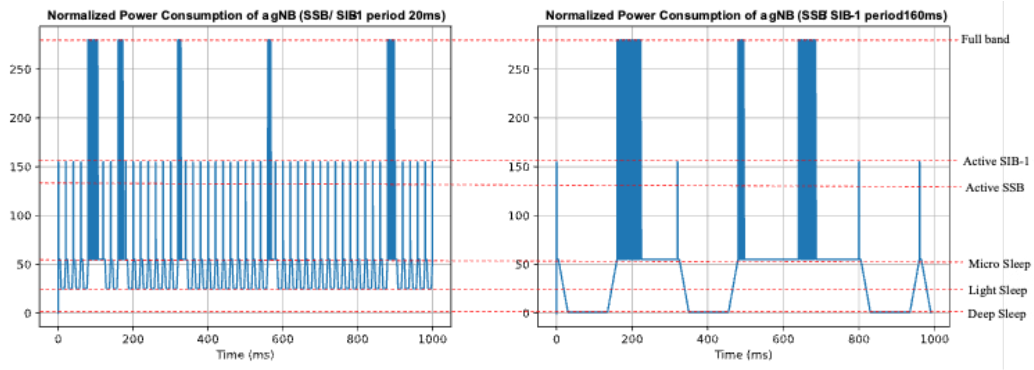

Figure 7. Normalised EC of a gNB in respect to EC of deep sleep mode in low load applying sleep modes.

An alternative approach to address this challenge is to transmit SSB or SIB-1 on demand. In this method, UEs could send a Wake-Up Signal (WUS) to prompt the transmission of SSB or SIB-1 when needed. This mechanism empowers the gNB to enter a more profound sleep mode, leading to energy savings. By relying on demand-driven signalling, the network can strategically manage the transmission of SSB/SIB-1, optimising EE [15]. In addition, AI could be used to estimate and predict EC and KPIs related to users' QoS, thus supporting network optimisation with a limited overhead, without the need to continuously measure and share collected measurements throughout the network. To achieve this, it would be necessary to define methodologies to transfer and update AI models together with the required input at the nodes where network configuration and parameters are controlled. From the standpoint of EC, these domains could significantly influence energy usage.

04 RADIO UNIT ENERGY EFFICIENCY IMPROVEMENTS

Since the RAN accounts for the majority of electricity consumption in the telecoms network, its EE is an unremitting subject to deal in addressing the environment challenge. This section describes solutions to improve EE of RUs, which are responsible, by far, for the largest share of the power consumption in each BS site.

4.1. PASSIVE ANTENNA EFFICIENCY OPTIMISATION

Antennas have a critical role in networks of connecting the BS site equipment with the end user devices. The efficiency of antennas is fundamental to both the performance and the energy efficiency of the RAN.

In the communication process between the BS and user devices energy is lost. Essentially, there are three phases of energy losses, which all impact the overall network EE: energy losses inside BS antennas, energy losses from antenna radiation leaking outside the coverage area, and energy losses from sub-optimal alignment of the projection direction of BS antennas. High efficiency BS antennas aim to minimise these losses.

Improving the end-to-end efficiency of antennae, and defining related antenna efficiency evaluation metrics, will support MNOs to select antennas that enable the highest network EE. Indeed, based on field trials and lab tests conducted in 2022 by NGMN Partner Huawei together with a large European MNO, legacy antennae already deployed have a large margin of improvement considering against the latest antenna technologies, which can improve **Radio Frequency (RF) efficiency, coverage efficiency, and alignment efficiency** (see Figure 8). The potential gains from these three improvement axes are independent and complementary: actual solutions can focus on one or several of them.

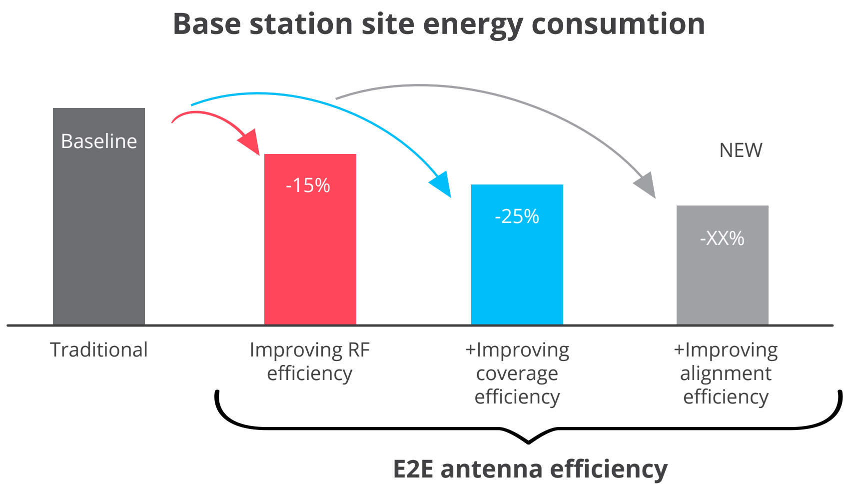

Figure 8. Reducing BS EC by improving end to end antenna efficiency[3].

[3] NOTE: The study excluded improving the alignment of the antenna hence why the figure indicates an unknown number – XX%.

RF efficiency refers to the proportion of RF energy provided from a RU to an antenna, that gets converted by the antenna into energy radiated into the air. Increasing the RF efficiency of an antenna reduces the power required from the radio unit, to achieve the same total radiated power, while at the same time improving the uplink performance as it also increases the useful uplink signal strength at the receiver side.

Coverage efficiency refers to the proportion of energy radiated in the useful coverage area, to the total radiated energy[4]. Ideally, radiated energy should be focused in the sector required for coverage, while minimising the radiation in unwanted directions such as towards adjacent sectors, causing interference that degrades network performance, and towards the sky causing energy waste (see Figure 9).

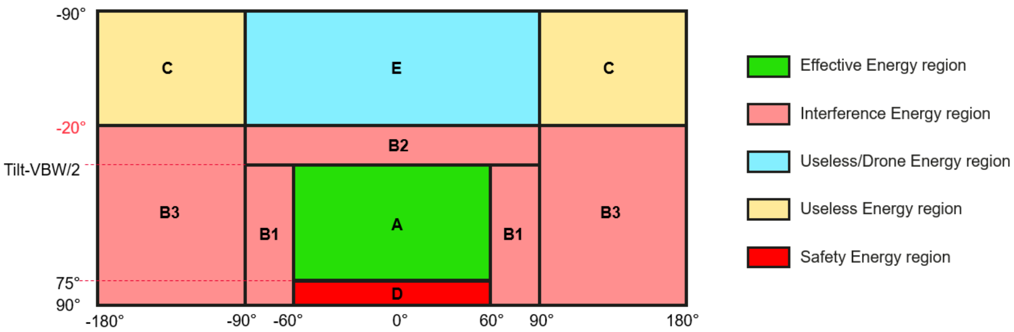

Figure 9. 2D antenna pattern, showing effective energy projection and unwanted projections.

Increasing coverage efficiency results in both BS site energy savings, by limiting waste radiated energy, and network performance improvements, by reducing interference.

Alignment efficiency reflects the favourability of the radiation projection direction of the BS antenna: For the best network performance, the projection directions of BS antennas need to be carefully optimised, to maximise coverage and reduce interference. Identification of antenna alignment directions, both horizontally and vertically, is typically handled as part of network planning and optimisation.

There are several reasons why antennas in real networks might point in wrong directions, such as installation errors, and changes in network topology since original site installations, resulting in antennas overlapping and interfering in their coverage. Moreover, dynamic network conditions, such as user locations shifting over time would require different orientations of antennas for best network performance.

An antenna with high alignment efficiency, enables the antenna orientation to be remotely controlled, which is much faster and cheaper as it does not require the manpower and equipment needed at BS sites for manual adjustments. Remote Azimuth Steering (RAS) requires two-dimensional adjustment capabilities, both vertical and horizontal. To maximise the network optimisation potential, both the horizontal range and the vertical range should be large enough to cover the typical adjustment cases, for example 20 degrees vertically and 60 degrees horizontally.

Antennae with remote control for both tilt and azimuth enable efficient AI-based network optimisation. Realising this vision using classic optimisation tools is challenging due to 1) the large number of parameters to control in a network composed by thousands of BS,

[4] This metric is studied in the current BASTA project and will be presented in its next publication.

where each individual cell's performance depends on the configuration of neighbouring BS and 2) it is not possible to clearly define mathematically a straightforward relation between the tunable parameters and the global optimisation goal. Therefore, stochastic optimisation can be efficiently used to improve network performance by tuning RF parameters [17]. The numerical results shown in Figure 10 shows the improvement achievable in a 5G network composed by 308 cells using the so called zeroth order (ZO) stochastic optimisation. The performance of ZO is compared with those provided by deploying a Tabu Search (TS), a classic optimisation algorithm in terms of global optimisation score, a metric providing the fraction of cells whose Signal-to-Interference-plus-Noise Ratio (SINR) and Reference Signal Received Power (RSRP) requirements are satisfied.

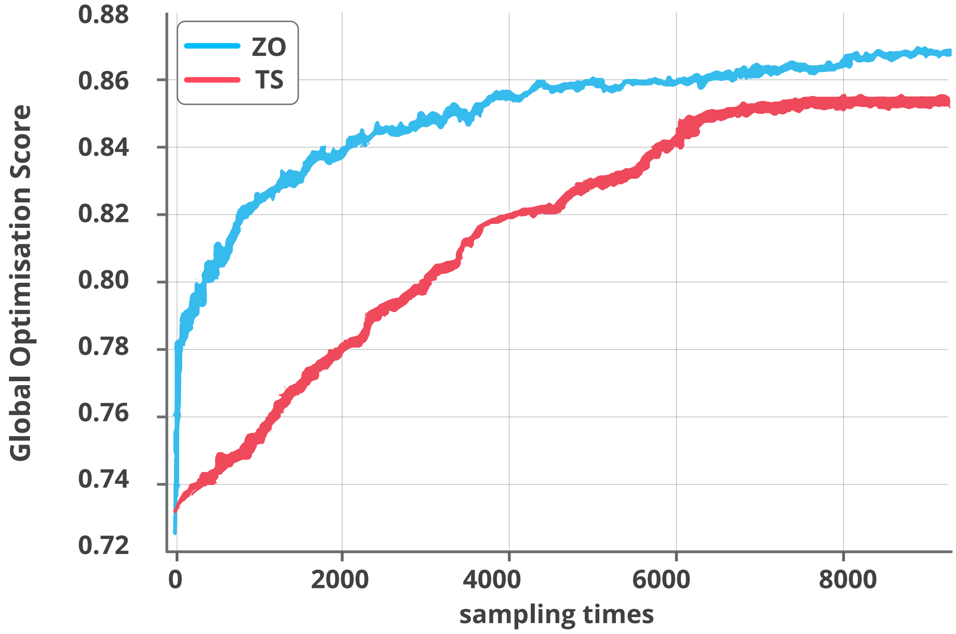

Figure 10. Network performance improvements by tuning horizontal and vertical antenna orientation: at each sample time the network RF parameters are updated.

As shown in Figure 10, ZO allows to improve the initial network performance of nearly 25% in terms of a global optimisation score. Indeed, ZO provides much larger gain with respect to TS even after a few numbers of iterations, and converge to better performance for the same transmit power. Equivalently ZO can lead to notable energy saving by allowing to reduce the transmit power required to deliver the target network performance.

4.2. ENERGY SAVING OF INDOOR PRRU

A pico RRU (pRRU) is a very small Radio Unit (RU) usually deployed for indoor coverage and characterised by low power consumption. At present, the energy saving feature of pRRUs is coordinated on the cell level [5] such that all pRRUs within a given cell maintain the same synchronised energy saving states. For example, in a given cell, when a pRRU operates with high load, all other pRRUs in that entire cell together with the reference pRRU are prevented from entering the energy saving mode, which leads to energy waste.

[5] Several pRRUs are used in this context for the transmission of one logical cell, usually deployed for indoor coverage in a similar architecture as an intelligent Distributed Antenna System (iDAS, as already discussed in previous NGMN publication from 2020: https://www.ngmn.org/wp-content/uploads/Publications/2020/Small-cell-economics-external-full-report-v1_1-clean.pdf).

Intelligent pRRU energy saving technology is proposed as an enhanced energy saving method. It works as follows. A UE receives and measures the reference signals transmitted by each pRRU, the Baseband Unit (BBU) monitors which pRRU each UE is connected to, the corresponding traffic load and/or the number of users connected to each pRRU. For those pRRUs without user connection and ongoing data transmission, the PA is turned off in a slot-specific manner to realise appropriate management of energy saving states rendering separate modes for each separate pRRU, as shown in Figure 11.

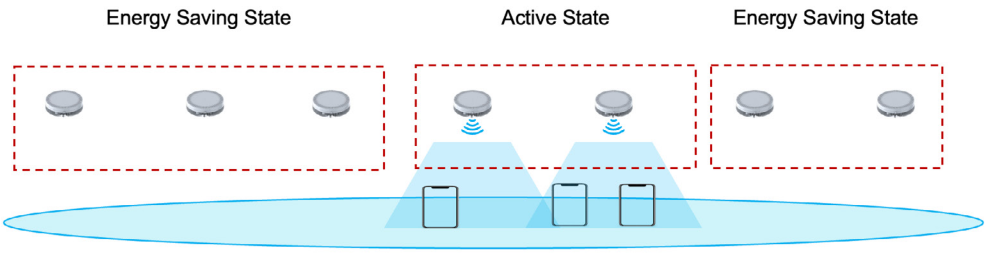

Figure 11. Energy saving at pRRU level.

Compared with existing energy saving technologies, the enhanced technology can achieve targeted signal transmission according to the location information of the UE, and shut down the power amplifier of the vacant pRRU to achieve energy saving. The mechanism consists of three main steps as follows:

STEP 1: The pRRU transmits reference signals to UEs in the cell and determine whether the UE is mainly served by the pRRU or not via a measurement report from the UE.

STEP 2: Based on the mapping relationship between the UE and the pRRU, the BBU is able to identify which pRRU is vacant in the cell.

STEP 3: Based on the service status of the pRRU, the BBU gives an instruction to the pRRUs with data transmission to maintain a “normal working” state, and an instruction to the pRRUs without data transmission to enter an energy saving state.

Test results were conducted by China Mobile in an office building with 6 pRRUs within a real network at low load. During the test, the intelligent pRRU energy saving feature was turned off for 3 days and then activated for 3 days. The obtained results show that this feature leads to energy saving gain of 20% with respect to the baseline where pRRUs are kept awake. Additionally, shutting down pRRUs can effectively reduce the electromagnetic interference to neighboring cells.

4.3. OPTIMISING THE RU LOAD-DEPENDENT POWER CONSUMPTION

The historical dominance of load-independent EC in RUs has prompted the development of various mechanisms aimed at powering down hardware components to optimise the amount of energy consumed. However, current projections anticipate a continued decrease in load-independent properties of RUs towards load-sensitive or load-dependent behavior, thanks to the advancements in power-efficient hardware, particularly in RRUs. As the share of load-dependent EC becomes more relevant, new strategies for reducing the overall EC of a BS concentrate on decreasing energy consumption in RUs already endowed with load-dependent properties.

These methods are particularly advantageous as they can also be employed in BSs providing continuous coverage, which generally cannot be switched off to save energy.

Load-dependent RU EC is primarily driven by the transmit power, which scales with the traffic handled by BS. Classically, the total transmit power at the BSs is evenly distributed among the total available PRBs, resulting in a linear growth of load-dependent power consumption with the serviced traffic. Emerging load-dependent energy-saving schemes aim to optimally adjust the allocated transmit power per PRB to reduce the overall EC.

Specifically, with load-dependent energy-saving schemes, when a user experiences good channel quality, indicated by a high SINR, the scheduler decreases the transmit power per PRB, while increasing the number of PRBs assigned to the user. In this way, the overall transmit power can be reduced, leading to a decrease of the load-dependent EC while maintaining the target data rate. Significantly, this approach distinguishes itself from traditional energy saving methods, which primarily focus on shutting down certain components of the BS, since it does not affect network coverage or QoS.

It is essential to note that, since this method involves increasing the allocated PRBs to the user, it is typically implemented during Time Transmission Intervals (TTIs) with low load, where it is possible to increase the PRBs to be allocated to each active UE. However, the impact from neighboring BSs, which employ shutdown strategies, must be considered. Typically, the shutdown of a BS leads to an increased load in neighboring BSs, which must now serve users from the deactivated BS. This increase in PRB usage can subsequently deplete the pool of available PRBs necessary for the successful execution of load-dependent RU energy saving. To mitigate this issue, exchange of energy saving information between BSs using shutdown and load-dependent energy saving methods is required.

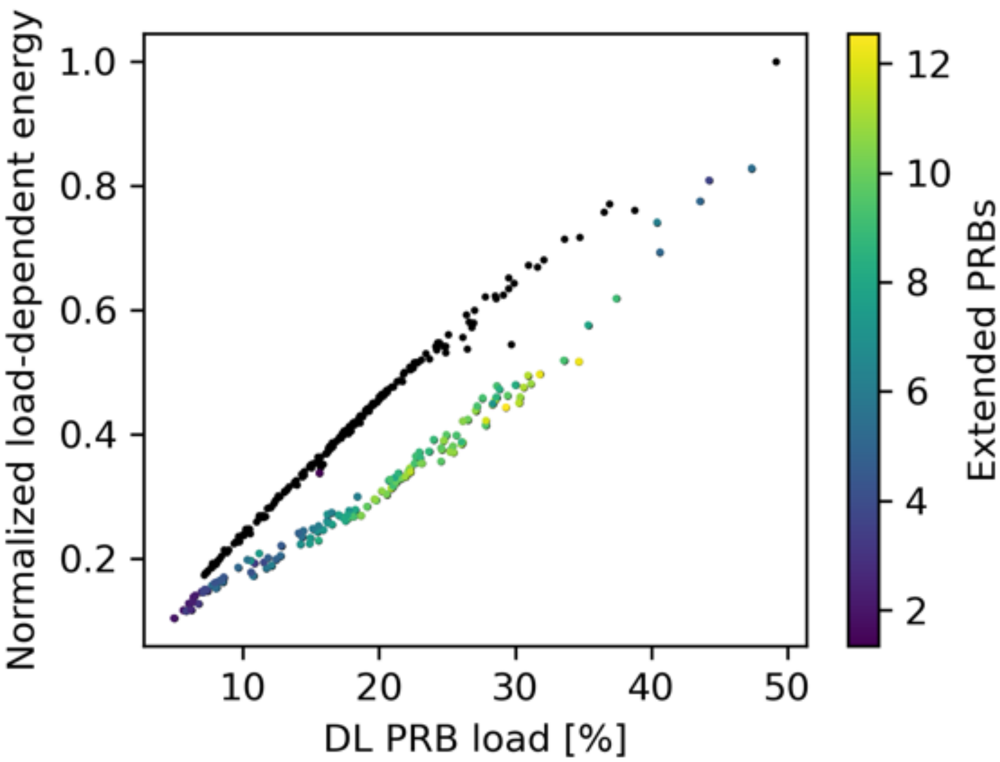

Figure 12. Normalised load-dependent EC achieved at different DL PRB load in a test base station.

The positive impact on EC is visually represented in Figure 12 depicting hourly EC samples in a BS implementing the described method. It is important to note that the energy values have been normalised to the total energy to address privacy concerns.

The graph displays, in black, the load-dependent energy consumed in a baseline scenario where the total transmit power at the BS is evenly distributed among the available PRBs. As anticipated, there is a linear increase in load-dependent energy consumption with the DL PRB load.

Furthermore, for a given DL PRB load, Figure 12 illustrates the load-dependent energy consumed when expanding the number of PRBs allocated to UEs while concurrently reducing the transmit power per PRB. The reported samples reveal that in this particular BS, a notable 30% reduction in the RU load-dependent EC can be achieved through this method.

It is essential to emphasize that the degree of transmit power reduction highly depends on the channel quality experienced by the users and the BS load. In fact, users facing poor channel conditions cannot tolerate a decrease in SINR. Additionally, BSs with high traffic loads typically have limited available PRBs to additionally allocate to UEs.

4.4. FRONT-END ADAPTIVITY FOR INCREASED ENERGY EFFICIENCY

4.4.1. THE ANALOG-TO-DIGITAL CONVERTER CHALLENGE

Were higher data rates to be delivered to users in future networks, the analog-to-digital converter (ADC) could form a power consumption bottleneck. The energy per conversion step of the ADC is constant over the Nyquist sampling frequency, and thus bandwidth, up to a threshold frequency of approximately 300 MHz [18], see Figure 13. For signal bandwidths below this threshold frequency, the ADC power consumption scales linearly with the bandwidth. However, for higher bandwidths above this threshold, a quadratic increase in ADC power consumption is expected. A 10x bandwidth increase will thus lead to a 100x power increase. Even with potential semiconductor technology improvements for ADCs, which increased efficiency by 20x between 2009 and 2020 [19] , such a substantial increase in power consumption is unlikely to be fully offset, leading to a potentially problematic surge in overall power consumption [19].

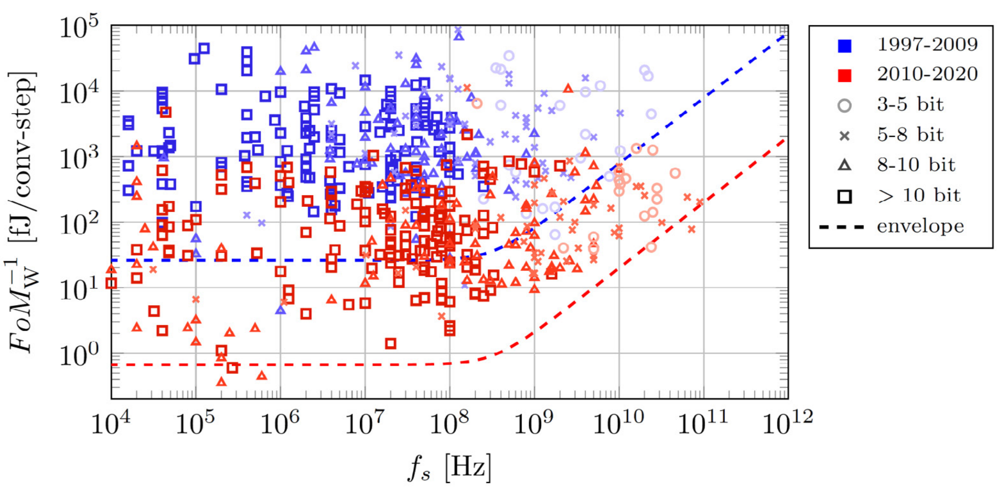

Figure 13. ADC energy per conversion step over sampling frequency [19].

4.4.2. THE WAY FORWARD? ADAPTIVE FRONT-ENDS

The EC for peak spectral efficiencies and peak data rates is expected to rise. However, in real-life scenarios lower data rates and lower spectral efficiencies are frequently sufficient. For these requirements, the EC of the RU can be significantly reduced by selecting modulation schemes that can work with an energy-efficient radio front-end [19], e.g., ZXM [20], which allows receiver designs with 1-bit ADCs at the price of decreased spectral efficiency. Analogous to a car's gearbox, the Gearbox-PHY selects a suitable modulation scheme and corresponding transceiver front-end based on data rate requirements and spectral availability, cf. Table 1. This way the high EE at low rates and low spectral efficiencies can potentially compensate the increase in EC for peak rates, that are required for certain applications.

Table 1: A possible Gearbox-PHY scenario [19].

Spectral Availability	**Abundant**	**High**	**Medium**	**Low**
PHY/ "Gear"	e.g., Impulse Radio	e.g., ZXM [20] or CPM	e.g., MIMO-OFDM	e.g., MIMO-OFDM or OTFS [21]

05 NETWORK SOLUTIONS TO OPTIMISE POWER CONSUMPTION

This section presents solutions to enhance network EE without introducing QoS degradation for users. These solutions leverage the variations of traffic demand in time and space together with a more efficient usage of network resource to improve network EE. More specifically, Section 5.1 introduces the fundamental steps for online optimisation of network EE. Section 5.2 describes a novel optimisation solution that improves network EE leveraging heterogeneous QoS requirements in the service area, resulting in up to 18% of reduction in the average network power consumption. To meet the rate requirements of 5G services, today's RRUs/AAUs aggregate the capacity across distinct carriers, which has increased the RAN power consumption. Section 5.3 and Section 5.4 introduce novel solutions where multi-carrier coordinated scheduling and spectrum sharing are respectively combined with energy saving schemes leading up to 26% of EE gain. Finally, Section 5.5 introduces inter-MNO RAN sharing solutions and highlights the benefits in terms of network EE improvement and carbon emissions reduction.

5.1. NETWORK ENERGY SAVING AND NETWORK QOS DEGRADATION

Conceptually, when describing the network solutions to achieve energy savings, we can distinguish the essential steps in this continuous optimisation process broken down as: monitoring, analyse-and-decide and execute, as outlined in Figure 14. The analyse-and-decide step includes the intrinsic trade-off between network EC versus communication service quality as guided by operator policies, e.g., expert rules, intent, service level agreements.

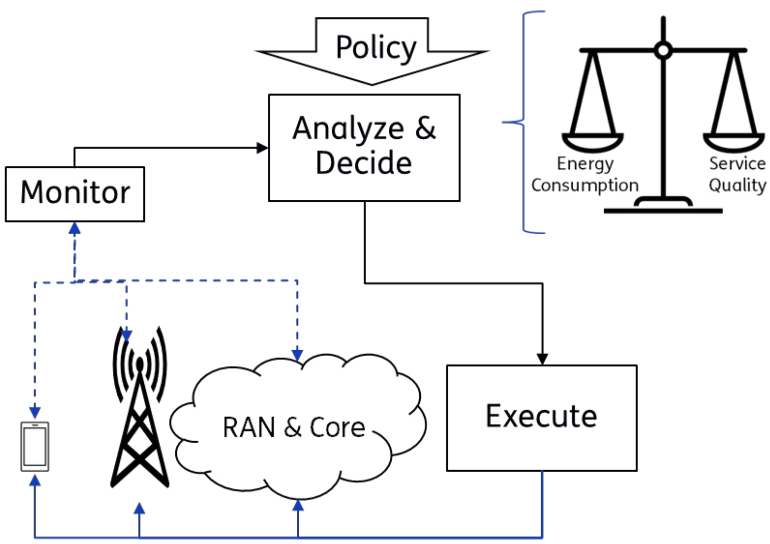

Figure 14. Energy saving optimisation loop and its trade-off with service quality.

MNOs use traditional rule-based network automation methods in the analyse-and-decide step to minimise the impact on communication service quality levels. Such rules are manually defined or combined with AI/ML-enhanced functionality, such as using traffic pattern predictions to adjust parameters such as the rule-based automation load thresholds, the number of active antenna branches, etc.

In the future, it is possible that with the introduction of EE service criteria, see Section 6.1, the policies that control the trade-off between energy consumption and service quality will get more extensive and complex. With the assistance of AI/ML functions, it would be preferable to translate the 'intent,' for example, 'reduce energy consumption while the performance is not degraded more than 5%', into concrete/quantifiable targets for the optimisation process. The performance goal and optimisation target can simultaneously cover different services and technologies for MNOs. It's like considering VoLTE and circuit-switched services' quality priority higher than packet-switched services. If AI/ML functions were to be more widely deployed, it would be beneficial for these functions to also adjust the optimisation targets whenever there are significant changes in the network resources and topology, or for example, when changes in traffic load trends, either unexpected or predictable, are experienced.

5.2. HIERARCHICAL NETWORK ENERGY EFFICIENCY OPTIMISATION

In the Phase II white paper [12], we presented a network EE optimisation methodology to find the optimum network configuration settings that would save energy while maintaining a pre-defined average service quality of a typical metropolitan area. However, in practice, the demand for service quality in diverse scenarios is different. Then, a hierarchical methodology with three types of policies for achieving the optimum energy saving while simultaneously fulfilling different service quality requirements of different zones is proposed (see Figure 15):

- **Policy I:** Experience first is set for zones with high data rate requirements. A power reduction solution is adopted to decrease the co-coverage area among cells to save energy while the percentage of users in good coverage is guaranteed.
- **Policy II:** Experience balance is set for zones with the requirement of the balance between power saving and user experience. A shutdown solution is adopted to migrate traffic towards the most energy-efficient cells.
- **Policy III:** Energy saving first is set for zones with low rate requirements by applying both power adjustment and shutdown solutions.

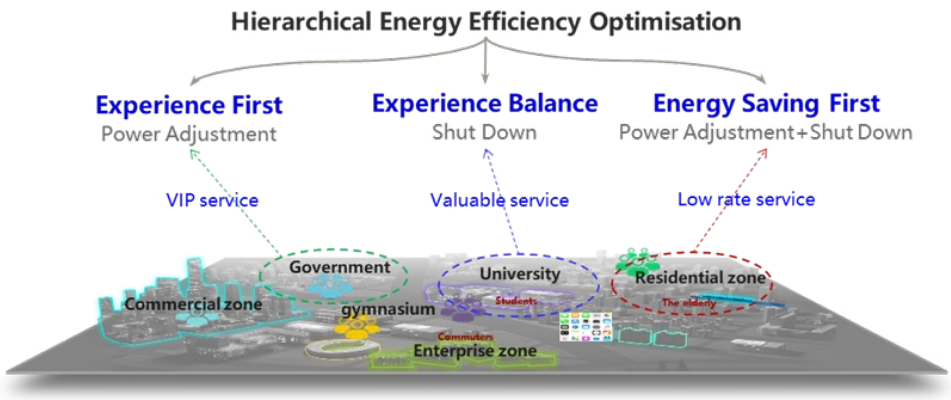

Figure 15. Hierarchical Energy Efficiency Optimisation.

5.2.1. APPLICATION OF HIERARCHICAL ENERGY EFFICIENCY OPTIMISATION

To demonstrate the benefit of the hierarchical EE optimisation, we apply it in a real network comprising 134 sites including 400 LTE cells (operating in multiple frequency bands) and 155 5G cells (using 2.6GHz and 700MHz). Four phases were conducted in this experiment. The first three phases are: all energy saving schemes deactivated; expert-based configuration of energy saving schemes; network-level optimisation of energy saving schemes. During the fourth phase, the whole metropolitan area was divided into multiple zones based on the geographical location, throughput of BSs, and service type. A specific energy-saving policy was set for each zone. Then, we computed the optimum configurations of the parameters in the energy saving policies by applying the hierarchical EE optimisation, including shutdown solution thresholds and duration, power adjustment, as well as handover parameters.

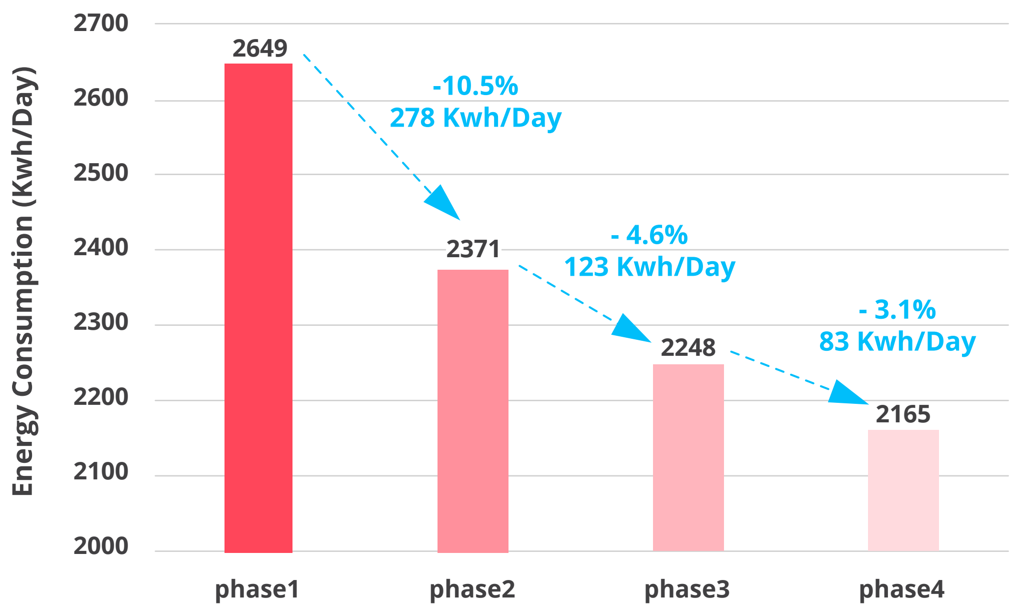

Figure 16. Energy-saving in each phase of the optimisation process for the overall sites under test.

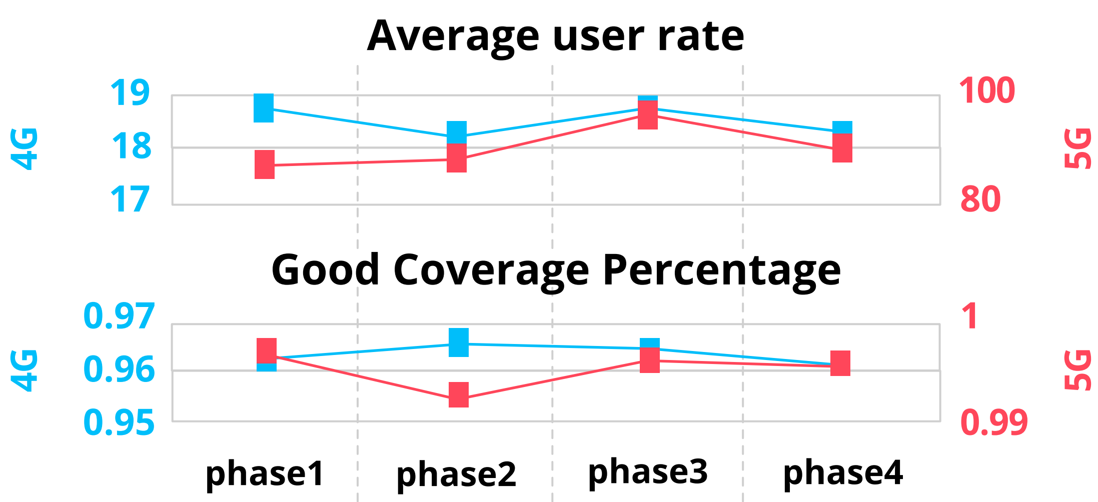

Figure 17. KPI performance in each phase of the optimisation process.

The experiment results are shown in Figure 16 and Figure 17. When applying the proposed hierarchical energy efficiency optimisation algorithm in phase 4, an additional 83 kWh was saved per day compared to phase 3, i.e., 3.1% energy saving gain. As shown in Figure 17, the average DL user rate and the percentage of users in good coverage in Phase 4 were both lower than each of their respective values in Phase 3 since the hierarchical EE optimisation yielded data rate reduction in zones with low data rate requirements in order to increase the energy-saving gains.

5.3. IMPROVING ENERGY EFFICIENCY THROUGH COORDINATED RESOURCE USAGE IN MULTI-CARRIER SYSTEMS

When frequency resource utilisation is not high, to achieve further energy saving, sub-frame shutdown can be implemented by collecting data packets together and delivering them on fewer symbols during downlink scheduling. Also, it's appropriate to send the data packets on continuous symbols to increase the probability of shutting down the PA and to increase the number of continuous deactivated symbols to achieve further energy saving gains (see Figure 18).

Since the sub-frame shutdown enhancement technique may lead to increased user plane delay , it is necessary to adopt converged scheduling considering the specific 5g QoS identifier (5QI) value in any given scenario. Importantly, the increased delay of data packets should not adversely affect user experience.

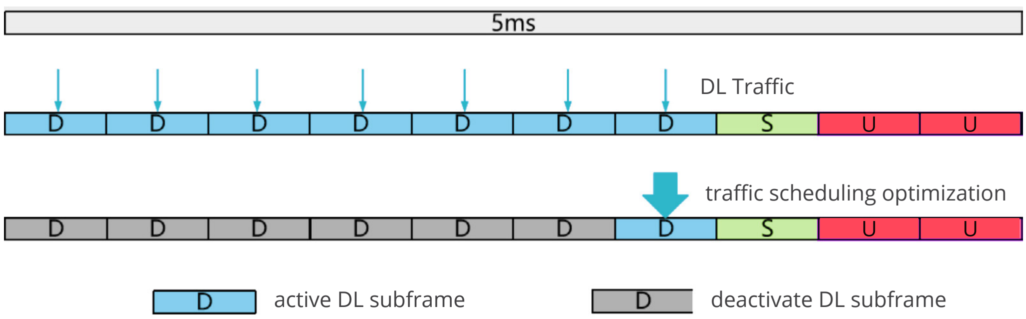

Figure 18. Enhancement of sub-frame level shutdown.

In scenarios of multiple carriers using common RF channels, if the BS detects that no carrier has data to transmit, the sub-frame shutdown is enabled to shut off the PA in the empty symbols to reduce energy consumption. When the BS detects that any carrier has data to transmit in this symbol, the sub-frame shutdown function is disabled to ensure the integrity of data transmission. However, due to the randomness of the traffic burst, the benefit of the sub-frame shutdown enhancement technique applied for separate carriers is limited. Therefore, the concentration of activated symbols among multiple-carriers should be considered during downlink scheduling to increase the probability of the PA shutting off.

Specifically, the data packets of each single carrier are scheduled in a centralised manner while not exceeding the delay constraints of different services. At the same time, the scheduling of symbols in multiple carriers can be aligned in the time domain to improve the proportion of deactivated symbols or sub-frames and save energy (see Figure 19).

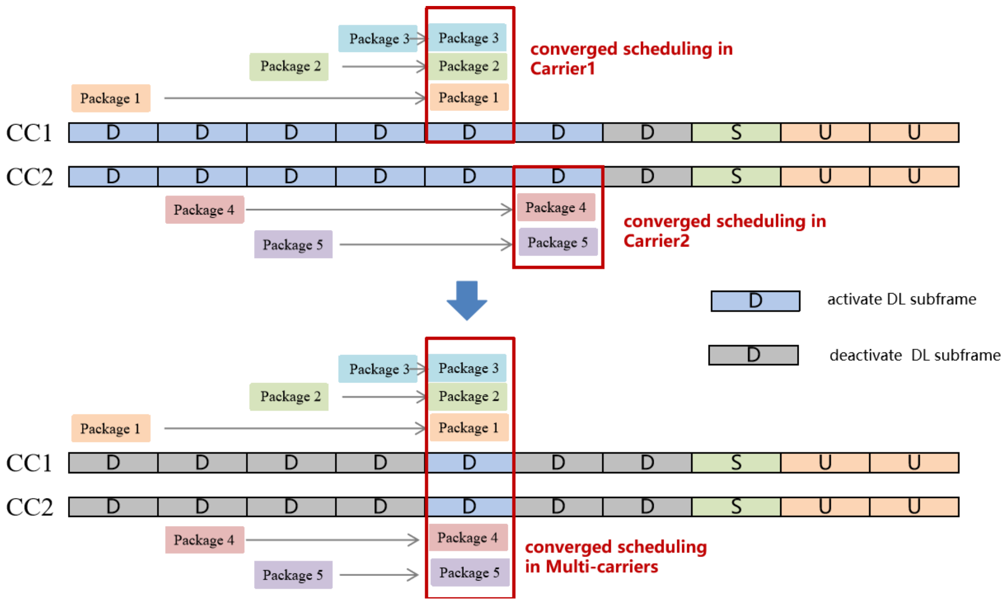

Figure 19. Multi-carrier converged scheduling for enhancement of sub-frame level shutdown.

5.4. ENERGY SAVING AND SPECTRUM SHARING IN MULTI-CARRIER SYSTEMS

To enable a carrier shutdown mechanism without adversely impacting coverage and data rates, 3GPP has introduced the concept of co-coverage relationships where capacity booster cells and coverage cells are paired, and they coordinate during cell shutdown and activation mechanisms such that the users' QoS is not reduced.

In general, paired coverage and capacity cells are deployed on distinct AAUs to maximise energy saving through carrier shutdown, since if the AAUs are distinct, the paired coverage and capacity cells do not share the same PA. Although the above technique can theoretically reduce the EC of cellular networks, its current implementation has some limitations: in particular, the network spectrum efficiency is limited due to the momentary deactivation of capacity cells and the temporarily unused spectrum of these shut-down cells.

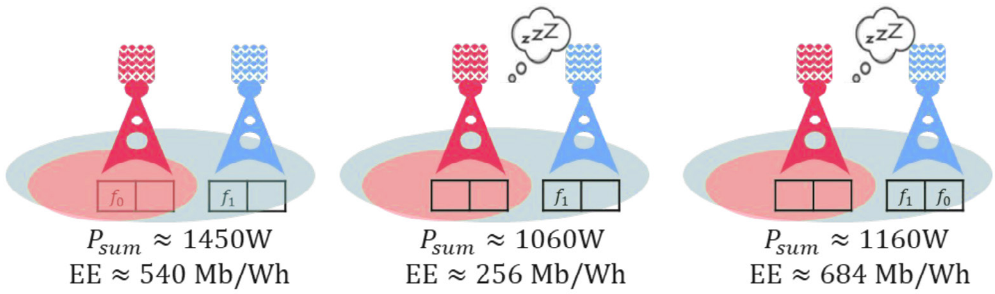

Figure 20. Power consumption and EE assessment of joint spectrum sharing and carrier shutdown using the power model in [12].

By implementing spectrum sharing between a paired shutdown capacity booster cell and its coverage cell MNOs would allow the latter to handle larger traffic volumes relative to its baseline configuration without spectrum sharing. This would in turn reduce the need for activating the switched off capacity cell, when the coverage cell's traffic volume were to increase, which would further increase the network EE with respect to standard carrier shutdown.

Figure 20 provides a qualitative example of the power saving and EE gains brought by combining spectrum sharing and carrier shutdown, evaluated using the power model introduced in [12]. In this example, we consider a capacity booster cell operating at 2.6 GHz (f_0) with 20 MHz bandwidth, and a coverage cell operating at 1.8 GHz (f_1) with 10 MHz bandwidth. When both are active the maximum power consumption of the two cells is 1450W and the maximum EE is 540 Mb/Wh (see the left side of Figure 20). When the load of the capacity booster cell decreases, it can shutdown saving around 450 W; however, the overall EE decreases as the capacity of the shutdown cell is lost (see the middle of Figure 20). By enabling the coverage cell to borrow f_0, the overall power consumption slightly increases with respect to standard carrier shutdown due to the additional power required at the coverage cell to transmit data on f_0; however, the EE can increase up to 684 Mb/Wh (see the right side of Figure 20). To summarise, spectrum sharing between capacity booster cells and coverage cells can lead to network energy efficiency enhancements by slightly increasing the network energy consumption to service an even greater increase in traffic volume, relative to the baseline energy consumption when only carrier shutdown is implemented.

5.5. INTER-MNOS RAN SHARING

RAN sharing is a strategic collaboration where two or more MNOs share various components of their wireless infrastructure, thereby enhancing network coverage, capacity, and quality, while reducing costs, energy consumption, and carbon emissions. This collaboration can encompass a wide array of resources, including cell towers, antennae, BS equipment, and even carrier frequencies. There are two primary types of sharing:

- **PASSIVE SHARING:** involves the sharing of passive infrastructure elements. In this model, operators share common physical structures, such as cell towers, masts, and site facilities e.g., cooling and direct current (DC) power systems. While MNOs maintain separate radio equipment, the shared infrastructure significantly reduces the environmental impact by eliminating the need for duplicating tower construction. Passive sharing is particularly effective in reducing the visual and environmental impact of cell towers, making it a more sustainable option for network expansion.

- **ACTIVE SHARING:** involves operators jointly deploying and operating network equipment. In this model, multiple operators share both the physical infrastructure and the active radio products. This includes the deployment of shared BSs, antennae, and transmission equipment. Active sharing allows operators to maximise resource utilisation, thereby reducing the number of duplicate network elements and lowering energy consumption. It also facilitates a more efficient use of spectrum resources.

Moreover, RAN sharing brings about several key benefits in terms of network EE and carbon emissions reduction. These benefits include:

- **REDUCED ENERGY CONSUMPTION:** by eliminating the duplication of under-utilised resources and sharing network components, RAN sharing leads to a significant reduction in EC. Fewer BSs and antennae result in lower power requirements, cooling needs, and maintenance costs.

- **OPTIMISED NETWORK UTILISATION:** network elements are used more efficiently, reducing the overall energy required to provide the same level of service. Furthermore, MNOs can implement energy-saving technologies and practices more effectively in a shared network environment.

- **ECONOMIES OF SCALE:** collaborative efforts in RAN sharing enable operators to leverage economies of scale when it comes to network upgrades and technology investments. This, in turn, makes it more affordable for MNOs to implement energy-efficient technologies, such as renewable energy sources.

A case study was implemented for a symmetric agreement between 2 MNOs with similar market shares and RAN resources such that:

- 50% of the shared RAN resources are owned (and operationally controlled) by one MNO, while the other 50% are owned (controlled) by the other one.
- Any shared RAN resource may be used to process traffic to / from any of the 2 MNOs' subscribers.

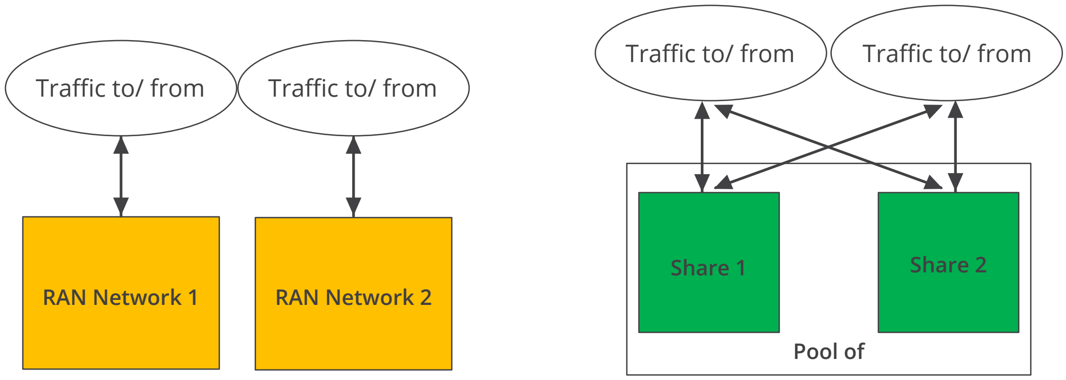

Figure 21. Non-Active RAN Sharing vs Active RAN sharing.

This case study is compared with standard dedicated RAN architecture in Figure 21.

This collaborative approach to RAN sharing not only promotes efficiency and flexibility but also results in significant energy savings: in the active RAN sharing case study, each MNO benefits from approximately 30% energy OPEX savings.

It is important to note that quality of the network is not only based on the radio access part but on the end-to-end link, inducing an optimisation process of the full chain, from the gateways, to the backhaul and the caches. Network sharing does not preclude MNOs from differentiating their network services and proposing "customer fit" services, thanks end to end optimisation.

06 INTERWORKING BETWEEN MOBILE NETWORKS AND ENERGY SUPPLIERS

This section introduces a near-future ecosystem where MNOs and energy suppliers cooperate to respond to environmental as well as economic sustainability challenges, leveraging renewable energy sources, energy storage facilities, and real-time information exchange.

6.1. ENERGY EFFICIENCY AS A SERVICE CRITERION

3GPP is studying the definition of "Energy Efficiency as a Service Criteria" [22] for Release 19. This criterion enables the delivery of services with diverse EE and EC policies. In this context, MNOs can define subscription policies that cover the aggregate quantity of EC of the network elements and functions used by the subscriber. These policies may include a maximum EC rate to limit energy consumption within a given time interval or a maximum energy credit to restrict the total amount of EC. Subscription policies can be applied either to the subscriber (all services) or to particular services. Energy credits are a quantity of credits related to the energy associated with network resources usage by the subscriber (e.g., EC) that can be used for credit control by the 5G system. In some context energy credit can be related to carbon intensity/emissions as well. The EC/charging rate defines "how fast" energy credits can be consumed via some mapping rule or conversion algorithm that can take several other metrics into account, e.g., time, location.

Since enforcing EE may negatively impact the QoS delivered to users, the network should, at least, support EC rate or energy credit enforcement for services without QoS criteria, such as best-effort traffic, which has no performance guarantees. If a user agrees to subscribe to "green communication" for some services with QoS criteria, the operator can provide a replaceable SLA with slightly reduced quality guarantees for the service when enforcing EC policies in agreement. Additionally, the MNO may change dynamically the charging rate, in terms of 3GPP "policy and charging framework", of the service provided to the user. Users inclined towards eco-friendly practices may indicate their preferences on a more granular level, i.e., during each service use. Implementing mechanisms that enable users to directly communicate their willingness to adjust QoS — either by downgrading for energy savings or upgrading for enhanced experience — could allow the network to dynamically adapt resource allocation and energy saving strategy in alignment with the user's current expectations. This approach may not only facilitate network cells entering in "energy state" but also potentially empower users by providing them the capability to impact their network environmental footprint.

The term "energy state" refers to the state of a cell, network element, and/or network function with respect to energy. By shutting down certain equipment, such as symbol shutdown, carrier shutdown, or channel shutdown, the cell can enter an energy state and provide a certain capacity. In line with the cell load, a predefined list of energy states supported by the cell can be configured for verticals and operators. Each state is associated with a respective cell capacity and energy efficiency. Different energy states that vary with the working status of telecommunications equipment could be beneficial for verticals and operators to help them save energy.

6.2. COOPERATION BETWEEN COMMUNICATIONS NETWORK AND ELECTRIC-POWER GRID

In recent years, the growth of the Information and Communication Technology (ICT) infrastructure, led by 5G BSs and data centers, has highlighted the energy consumption challenge within the communication industry. In this context, the usage of renewable energy sources such as solar and wind energy has been regarded as a crucial solution for energy conservation and carbon reduction. However, due to the unpredictability and instability of the renewable energy sources, large-scale renewable EC poses a significant challenge for the construction and operation of power grids.

The EC of communication networks needs to be characterised by flexibility and controllability. Hence, cooperation between communications and energy networks could enable flexible interactions between energy sources and traffic loads, which could be a key enabler for the low-carbon transformation of energy, achieving peak carbon dioxide emissions and carbon neutrality goals. There are two primary methods to introduce renewable energy to communications networks:

1. With suitable climates, small-scale wind turbines and rooftop photovoltaic systems can be employed to directly power BS equipment.
2. Communications networks collaborate with power grids, utilising grid-supplied renewable energy to power BSs.

Since renewable energy sources are volatile and unpredictable, in order to adopt renewable energy in mobile networks, three technical directions are presented in this publication (see Figure 22):

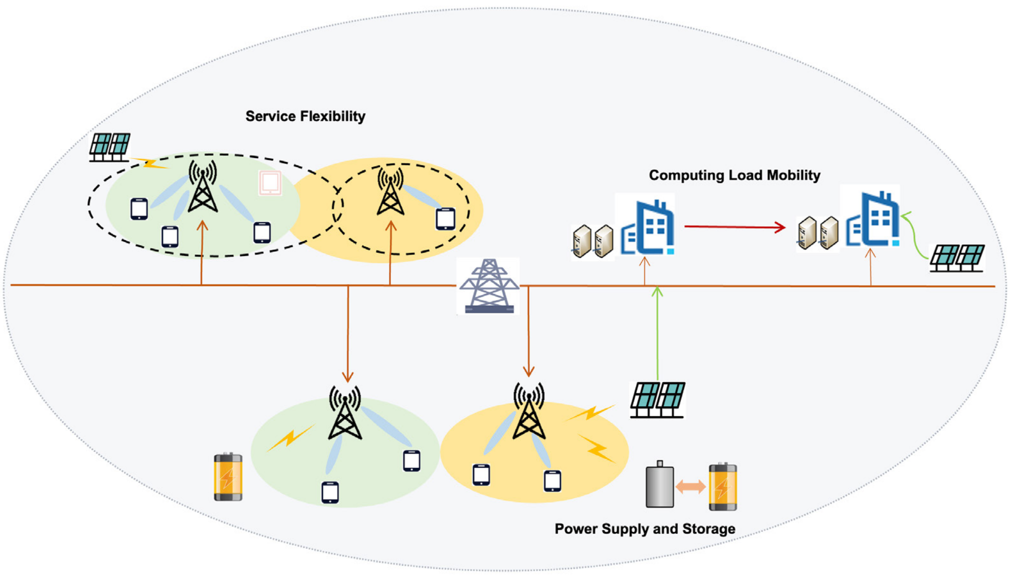

Figure 22. Technical directions for integrating renewable energy in mobile networks.

- **SERVICE FLEXIBILITY:** This involves dynamically adjusting the energy saving states, scheduling strategy, user access management, and transmit power of the BS to meet the requirements of renewable energy consumption in the current power supply area, achieving the balance between energy demand of service and supply of renewable energy.
- **COMPUTING LOAD MOBILITY:** This involves offloading computing tasks to edge computing nodes with sufficient renewable energy supply. It also leads to improving the ability of monitoring energy and power consumption so that computing routing technology can be used to transfer computing load between data centers more accurately, enabling flexible scheduling of EC and assisting power grids in consuming renewable energy locally.
- **POWER SUPPLY AND STORAGE:** This involves ensuring that communications equipment rooms have sufficient and abundant energy storage resources to stabilise power supply, as well as meeting computing requirements, maintaining storage space for batteries, sustaining appropriate cooling conditions, providing advantages in participating in the power grid market with auxiliary services. The power supply systems in the equipment room can be improved to support the capabilities of both power supply and storage, enabling flexible power scheduling. By battery charging and discharging, dynamically responding to renewable EC requirements, could help to strike the right balance between power supply and consumption in the power grids.

6.3. BALANCING ENERGY CONSUMPTION AND ENERGY SUPPLY FOR TELECOMS SYSTEM

Telecoms networks rely on diverse sources of energy supply in different countries and regions. Some networks are supported by reliable power grids that can deliver sufficient and continuous power enabling carbon reduction solutions to be sought with renewable energy, while in other deployments BSs are still equipped with diesel generators. In both scenarios, the joint optimisation of EC and energy supply, as presented in this publication and depicted in Figure 23, has great potential to deal with the growing sustainability challenges faced by the industry.

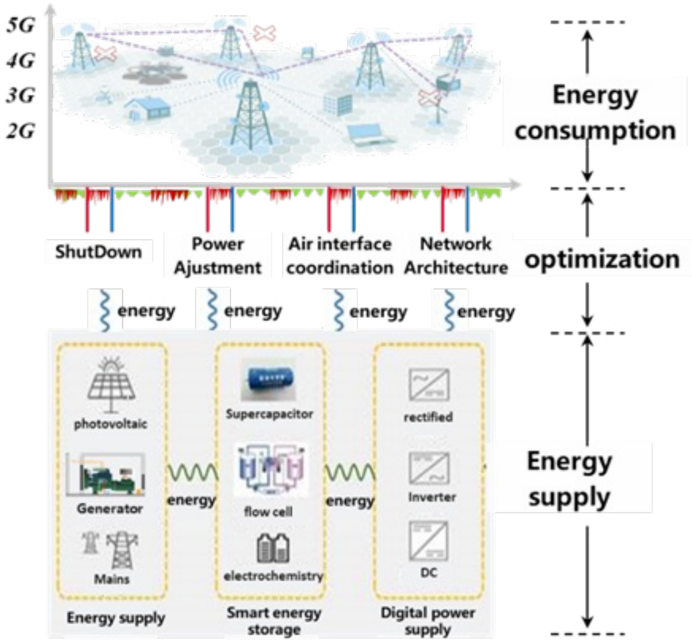

Figure 23. End to end energy optimisation framework in a telecom system.

In Figure 23, the EC layer describes the network characteristics in terms of EC, while the energy supply layer models the energy sources, i.e., mains power, renewable energy, diesel generator, and each associated energy storage system. Classic network EC optimisation is based on various software and hardware network solutions, such those described in Section 3, which are designed to balance the power demand and the energy supply in the network. In this context, the potential for cooperation between the EC and energy supply layers stem from two aspects:

- Optimising power demand distribution to match the energy supply by adjusting the network configuration. For regions where mains power is scarce, an optimised system transfers power demand to BSs with diesel generators to guarantee sufficient energy supply and maintain network availability and user experience. For regions with more abundant renewable energy, power demand is transferred to BSs with renewable energy to decrease the energy supply cost and environmental footprint.

- Deploying energy storage facilities next to BSs with high power demand and stringent service requirements, together with renewable energy sources, can improve energy efficiency and network availability.

By optimising jointly power demand and energy supply, renewable energy can be exploited more efficiently, and the usage of diesel generator and mains power can be reduced, without affecting network availability and user experience.

07 ENERGY EFFICIENCY IN DISAGGREGATED NETWORKS

In this section, we present opportunities to increase EE for virtualised and disaggregated network and an overall EE requirement that Open RAN memorandum of understanding (MoU) operators consider to be priorities for Open RAN deployment.

7.1. EE OPPORTUNITIES IN VIRTUALISED AND DISAGGREGATED NETWORKS

A virtualised and disaggregated mobile network offers MNO the flexibility in hardware, software and systems integration which drives innovation and agility in cloud services to efficiently deliver a myriad services with different performance requirements. Discussion of power measurements, the power measurements model, and algorithms to optimise energy consumption of VNFs based on commercial off-the-shelf hardware and how to increase energy efficiency developed by the industry is covered by the NGMN Virtualised Network Infrastructure Metering Publication, Phase 3 [23].

In a disaggregated open RAN, multiple RAN vendors can contribute various components, offering substantial opportunities to enhance EE. Unlike a traditional RAN where a single vendor provides solutions, an open and disaggregated RAN facilitates rapid implementation of innovative technologies in radios, processors, servers, hardware, software, AI/ML features, and RIC applications from competing vendors. One example is of an architecture where a single RU can be connected to several Distributed Units (DUs) over optical paths. This approach brings the benefits of resource pooling and added reliability through redundancy. By dynamically reducing the number of DUs to which a RU is connected based on traffic load, it's possible to achieve reduced power consumption. Dynamic mobile fronthaul optical path switching saves power usage by leveraging the well-documented fluctuation in the number of mobile users in an area between day and night, especially in urban areas where the variation can typically be more than a factor of ten. As the number of mobile users and associated traffic in a particular area varies, so too does the required network resources in a proportional manner. The process of moving mobile services from one DU to another is carried out directly at the optical transport layer, utilising wavelength switching. Power usage saving can be realised by switching optical paths in the fronthaul to move traffic to smaller overall number of consolidated DUs when network traffic load is low, freeing unused DUs to go into low power mode or be turned off completely.

7.2. EE REQUIREMENTS IN OPEN RAN MOU

7.2.1. OPEN RAN MOU SCOPE

Open RAN MoU is a Memorandum of Understanding (MoU) signed by a group of 5 European MNOs (Deutsche Telekom, Orange, TIM, Telefónica and Vodafone Group). The goal of the MoU is to promote the timely deployment of Open RAN technologies and ensure that a strong ecosystem of companies emerges in Europe.

MoU communication is made mainly via the delivery of white papers, progress reports and via the annual release of the "Open RAN Technical Priorities Document" [24] which provides a comprehensive list of technical requirements that the MNO signatories to the Open RAN MoU consider priorities for Open RAN deployment. Three technical releases have been delivered so far. Energy efficiency topics have been mainly analysed in Rel. 2 and Rel. 3 of the Open RAN Technical Priorities Document. EE will also continue to be analysed in Rel.4, planned for being finalised in Q2 2024. A summary of the EE requirements highlighted by the MoU in Rel.2 and Rel.3 is provided in Section 7.2.2. The MoU has also provided its vision on the status of EE in Open RAN in a white paper entitled "Open RAN MoU progress update on maturity, security and energy efficiency" [25]. The key messages related to EE are summarised in Section 7.2.3.

7.2.2. OVERVIEW OF MOU EE TECHNICAL REQUIREMENTS

EE is an end-to-end requirement involving all domains of the Open RAN architecture. The overall objective for Open RAN networks is to gradually become more energy efficient than traditional RAN without sacrificing Open RAN concepts such as cloudification and disaggregation.

EE for Open RAN networks should therefore rely on the following pillars:

- Power efficient hardware.
- Report of EE KPIs at different hardware and software levels.
- Open RAN features to improve EE on a functional level compared to traditional RAN.
- Intelligence and orchestration to automate EE features.

The Open RAN MoU has provided EE priorities across the three releases of the MoU Technical Priority Document, encompassing different network components: O-RU, O-CU/O-DU SW & HW, O-Cloud SW platform, RIC, RAN features and SMO. The main requirements identified by the MoU are summarised in Table 2.

Table 2: Summary of MoU EE Technical requirements.

Domain	**EE Requirements**
O-RU	Achieve recommended power consumption targets
O-Cloud infra	Vendor should provide: • Energy Efficiency KPI & Monitoring • Power, energy and environmental (PEE) parameters and measurement data at the workload level (e.g., pod, CNF, etc.)
O-CU O-DU	Vendor should provide energy efficiency counters/KPI (CNF)
RAN features	Vendor should support the following features: • Base station Sleep mode mechanisms • Battery consumption saving • Enhanced Radio Deep Sleep Modes • ON/OFF DL MIMO adaptation • LTE/NR Layer switch-off in multi-layer sites • Artificial RF load generation

SMO	Vendor should support the following features and metrics: • Activating/deactivating energy saving features • Cluster/ CNF/ PNF Energy efficiency KPI (availability of data for HW, CaaS, PaaS, NF instance and availability of dashboard)
RIC use cases: O-RAN EE use cases	Vendor should provide the following use cases: • Carrier and Cell Switch Off/On: algorithm to trigger carrier/cell off/on switching may be hosted in the Non-RT RIC/SMO or in the Near-RT RIC • RF Channel Reconfiguration Off/On: algorithm to trigger RF Channel Reconfiguration may be hosted in the Non-RT RIC or in the Near-RT RIC • Advanced Sleep Mode: impact on Non-RT RIC and in the Near-RT RIC • O-Cloud Resource Energy Saving Mode: enable optimisation of O-Cloud resources, e.g., O-Cloud node(s) shutdown, modify P-State and C-state via SMO / Non-RT RIC

7.2.3. KEY MESSAGES FROM MOU PROGRESS REPORT ON EE

In [25] the MoU presents its vision of the current state of the O-RAN ecosystem and the actions envisaged to accelerate the deployment of energy efficient O-RAN networks.

Progress is being made in the industry to increase the EE of all Open RAN building blocks, with particular focus on radio transmitters, cloud infrastructure, energy monitoring and intelligence-based mechanisms:

- **RADIO TRANSMITTERS:** O-RUs contribute the greatest part of the total power consumption of the RAN (around 80%), so it is essential that Open RAN O-RUs are at least as energy efficient as their counterparts in the traditional RAN. Enhancements are being considered in the O-RAN ALLIANCE to enable EE features available to traditional vendors, e.g., allowing control of the O-RU switch-over and shutdown modes by the O-DU through the Open Fronthaul interface.

To this end, the Open RAN MoU MNOs have defined EE targets for Open RAN O-RU in both loaded and unloaded conditions, which may be used for benchmarking Open RAN O-RU with traditional RAN.

- **CLOUD INFRASTRUCTURE:** EE for cloud infrastructure is improving in the industry with progress in the efficiency of the CPU, in accelerator technologies, and more integrated chipsets natively optimised for lower power consumption.

Open RAN MoU MNOs have performed a first review of all the cloud infrastructure hardware elements to be optimised, considering not only processors but also sub-components such as memory storage, NIC cards, fans, power supply. Ultimately, the goal is to identify relevant monitoring and energy-saving features, and automation mechanisms.

- **ENERGY MONITORING:** While the O-RAN ALLIANCE has primarily focused on CPU optimisation, Open RAN MoU MNOs wish to extend monitoring to a variety of server components including accelerators, memory storage, etc. To this end, the MoU promotes open Application Programming Interfaces (APIs) at all levels of the Open RAN system to monitor the power consumption of all possible hardware elements: in particular, for the monitoring of O-RUs and of most sub-components of the cloud infrastructure.
- **INTELLIGENCE:** EE can be improved by using open standardised interfaces together with intelligence provided by O-RAN architecture. Intelligence will eventually allow Open RAN to be more energy efficient than traditional RAN, thanks to real-time monitoring, intelligent switch-off, adaptation to traffic, and native AI/ML offering more granularity in terms of the energy management of RAN elements. Open RAN MoU MNOs are willing to develop an AI/ML architecture framework, which would be capable of tackling the optimisation of EE.

08 ENERGY AWARE AI/ML

Machine Learning (ML) and Artificial Intelligence (AI) increasingly have practical applications within mobile networks. This section provides insights on the following AI/ML applications to improve the EE of mobile networks:

1. Identification of the low-EE BS sites in a network.
2. Assistance to the optimisation of network energy saving feature management.
3. Aligning the network capacity and energy consumption under performance constraints.
4. Optimising the AI/ML model size and hence its energy transmission cost.

8.1. IDENTIFICATION OF LOW ENERGY-EFFICIENCY BSS BASED ON AI/ML

A possible low-hanging fruit in the journey to optimise and improve the EE of the RAN, is to identify the low-EE BS sites among many BS sites. This allows adjustment of the configuration of these sites so that the overall network EE can be improved. Due to the large number of varying BS equipment specifications and different parameter configurations, the EE values (bits per energy) of different BSs varies a lot. For a relevant comparison of different sites and to implement such an EE evaluation accordingly, the BS sites can be consolidated into different categories, and low-EE ones can be identified within a one given category of BS sites.

In this scenario, AI algorithms are introduced to deal with the complex and diverse parameters that influence EE, and can automatically identify and classify different types of BS. The specific procedure is represented in Figure 24 and includes the following steps:

1. The BS sites with different static parameters such as BS type, number of channels, carrier frequency, bandwidth, number of cells, maximum transmit power etc. are divided into different groups.
2. For each group, a relationship model is established between EE and the dynamic parameters, e.g. traffic volume, PRB usage, or user numbers. with an appropriate fitting function trained using the Support Vector Regression (SVR) algorithm.
3. Due to the similarity of feature samples of different fitting functions, K-Means or DBSCAN clustering algorithms are employed to cluster the feature set, achieving intelligent classification of BS sites.

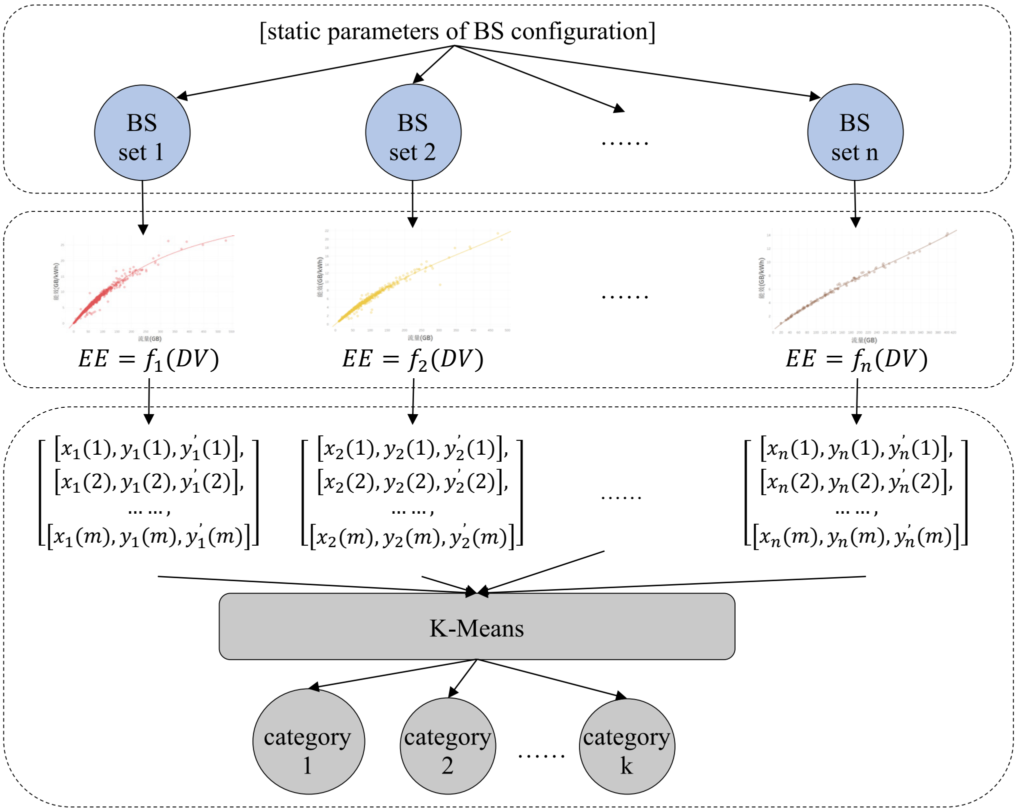

Figure 24. Classification of BS sites in terms of EE.

Even within the same category of BS sites, EE values can vary significantly due to different levels of traffic load. Performing landscape comparisons of EE among BSs of the same type is a challenge.

One method is to use as a baseline the best fit describing the relation between EE and traffic values (see Figure 25). From the trend of the EE curve with respect to the traffic, the EE of each BS site can be assessed by comparing the EE value against the corresponding EE baseline with the same traffic values.

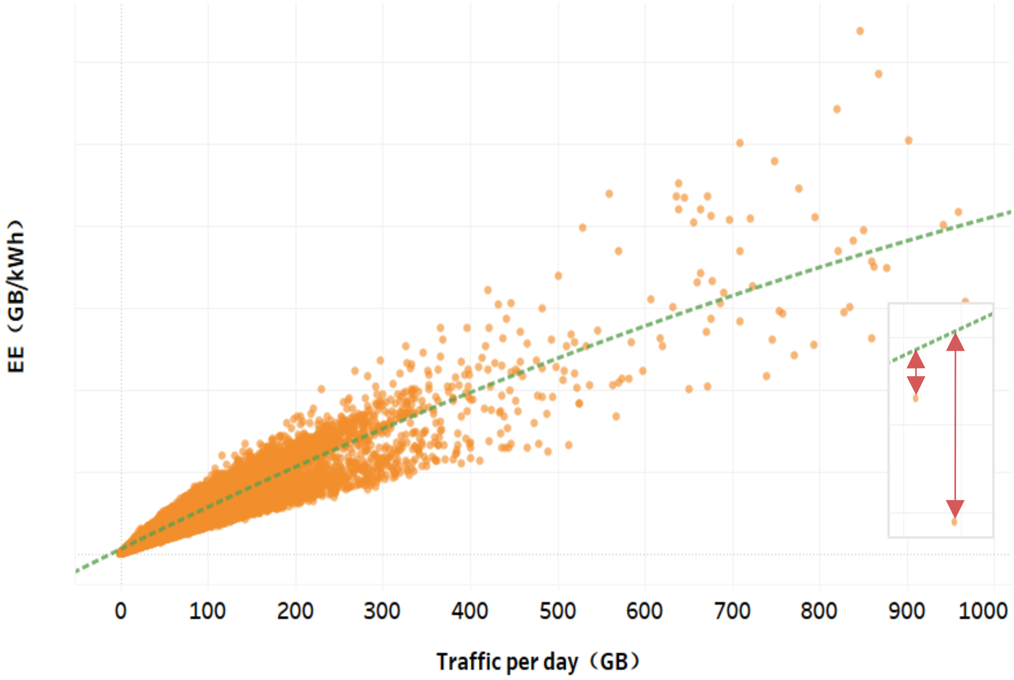

Figure 25. EE baseline calculation based on real network data.

AI algorithms could also be employed to identify the low-EE BSs by establishing anomaly site models. Specifically, for each type of BSs, an anomaly detection algorithm such as Isolation Forest or Local Outlier Factor can be used for model training based on the distribution of EE values of the same category of BS sites along with the traffic load. An anomaly detection model is then obtained for each type of BS sites to accurately identify the EE outliers, i.e. the low-energy-efficiency BS sites.

8.2. AI/ML ASSISTED NETWORK ENERGY SAVING

Energy optimisation is fundamentally about maximising energy efficiency under performance constraints. Therefore, performance constraints are the core of evaluation and the foundational guarantee for achieving maximum energy efficiency. In addition to traditional KPIs like handover success rate, RRC (re)establishment success rate, the system performance also includes QoS (Quality of Service) KPIs such as scheduling rate, coverage, and latency. Proper energy decision-making strives for a balance between system performance and energy efficiency. AI/ML techniques can help to make better energy saving decisions, e.g., cell activation/deactivation, by predicting future energy efficiency and load states.

In this publication, we present AI/ML-assisted Energy Saving tested in Panyu District, Guangzhou, China, with three components: all-day symbol shutdown, all-day AI/ML-based channel shutdown, and AI/ML-based deep sleep. Load prediction, based on a Long Short-Term Memory (LSTM) ML model leveraging collected historical data, was evaluated across three-time ranges: T0 represents the time without energy-saving strategies, T1 represents the time with traditional energy-saving strategies, and T2 represents the time when AI/ML technology is used. The testing configuration is detailed in Table 3.

Table 3: Configuration Information of Evaluation.

Time	Range	Type	Energy Saving Strategy
T0	2022.06.09 ~ 2022.06.15	None	W/O channel shutdown W/O symbol shutdown W/O deep sleep
T1	2022.05.27 ~ 2022.06.22	Tradition	Channel shutdown Symbol shutdown Deep sleep
T2	2022.06.25 ~ 2022.07.01	AI/ML assisted	AI/ML channel shutdown AI/ML symbol shutdown Deep sleep

Results show that the use of AI/ML techniques has extended the duration of shutdown for these strategies, resulting in a 2.48% reduction of power consumption, and a 16.85% improvement in RAN EE compared with traditional methods. Figure 26(a) shows the average power supply, and the use of AI/ML techniques resulting in a 23.87% reduction in power consumption. Figure 26 (b) shows an energy efficiency improvement of 23.40% achieved by using AI/ML techniques.

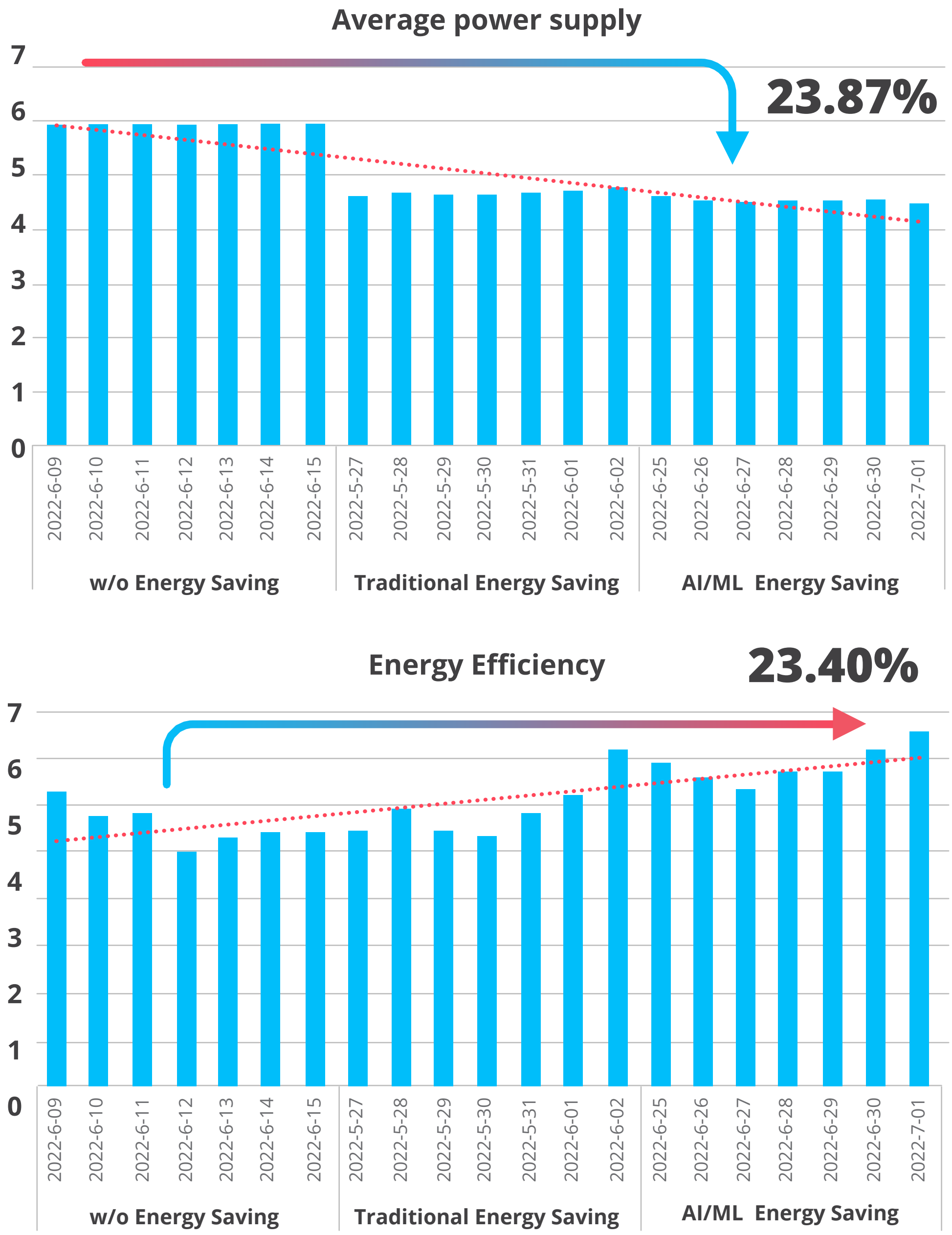

Figure 26. (a). Average power supply (Unit: W); (b). Energy efficiency (Unit: GB/kWh).

8.3. ENERGY SAVINGS VIA CARRIER SHUTDOWN UNDER PERFORMANCE CONSTRAINTS

Today's BSs have typically more than one frequency carrier, as illustrated in Figure 27. By shutting down frequency carriers, the energy consumed by the BS decreases significantly as the PA are responsible for the bulk of the energy consumed by a BS.

But, as user traffic transfers to the remaining active carriers and their load increases, the traffic performance may typically degrade. Here lies the need for the right trade-off so that the energy increment due to the load increase on active carriers is over-compensated by the PA switch-off, to deliver energy savings.

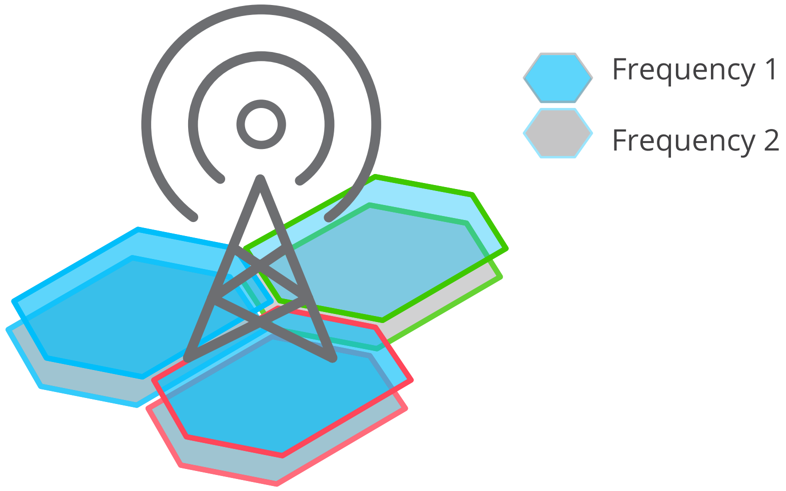

Figure 27. Illustration of a base station with two carriers.

To decide when carriers should be shut down, they are first ranked according to a given criterion (e.g., in increasing coverage radius). Then, each successive carrier is turned off when the traffic load in the given sector dips below a certain minimum-threshold, and switched back on when the load exceeds a certain maximum-threshold. This induces a hysteresis process on the traffic load, that tends to remain within the minimum/maximum-threshold boundaries, as illustrated in the Figure 28.

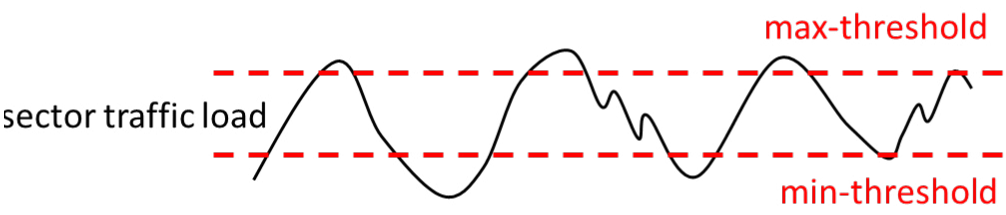

Figure 28. Illustration of the traffic load hysteresis for a given base station sector.

Different sectors may experience very different traffic conditions, in terms of average SINR offered to the users. Hence, different sectors can afford different levels of traffic load (hence, of min/max-threshold) on their cells to offer similar QoS to their connected users. Typically, the better the radio conditions are in a sector, the higher the load (hence, the higher the thresholds) that can be accommodated, leading to higher potential energy savings.

The optimisation goal is to reduce the BS energy consumption while maintaining "acceptable" user QoS with high confidence. One definition of "acceptable" QoS at a given time is the output of a Boolean formula that takes as input the recent values of certain KPIs of interest, i.e., DL throughout, traffic volume, cell load, etc., and produces the value 1 if the KPIs are higher than predefined target values, and 0 otherwise.

The optimisation leads to energy savings by properly tuning the min/max-thresholds mentioned above on a sector basis and across time. In fact, traffic conditions vary within the same sector, following the classic busy/non-busy periodical cycles, which hints at the fact that different hours of the day deserve different threshold values.

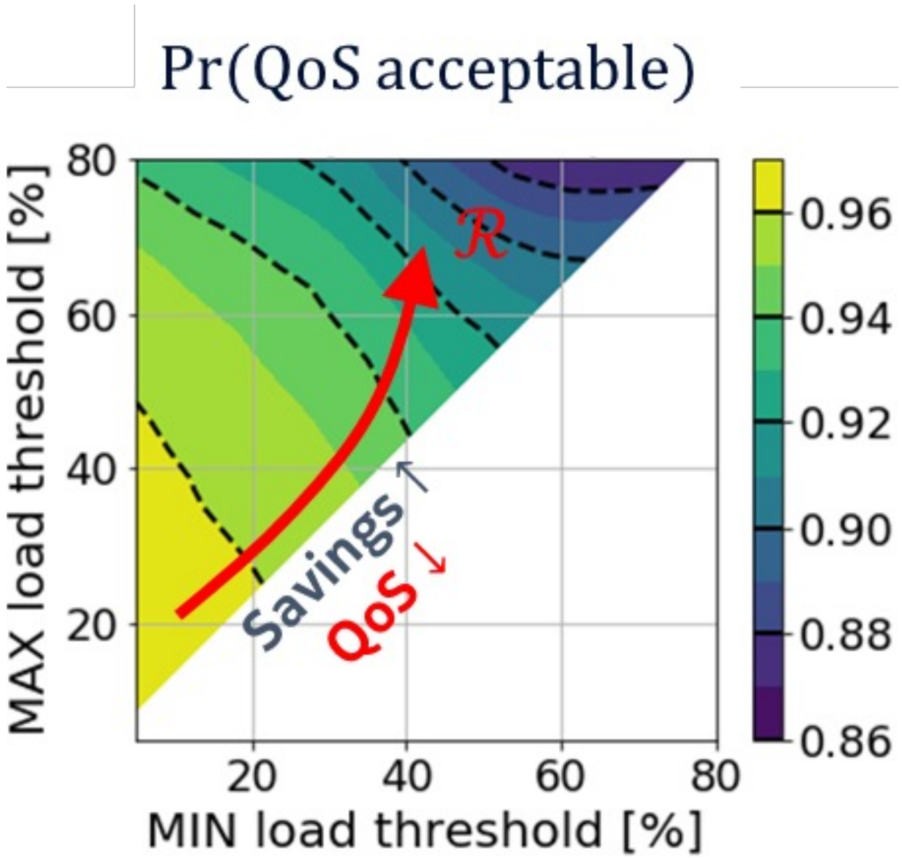

Figure 29. Impact of the min/max-thresholds on the energy savings QoS tradeoff.

As illustrated in Figure 29, there then exists an optimal pair of thresholds (depending on the sector and on the time of the day) such that savings are maximised while QoS remains within the predefined boundaries.

To find and track over time such a pair of optimal thresholds, one approach is to employ Bayesian optimisation (BO) techniques, for the following reasons:

- BO avoids any cold start when the learning begins, as it can naturally incorporate the learnings from historical data.
- BO allows for fast convergence.
- BO takes informed and non-random steps at each time, given the statistical knowledge it has acquired, and hence, it does not incur any performance drop during the learning process.

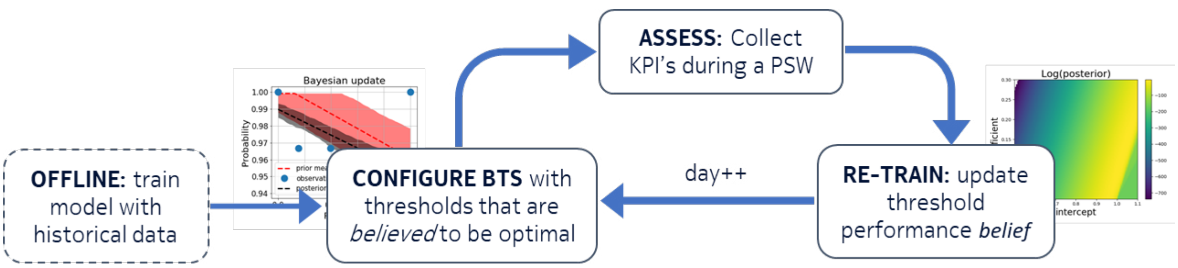

Figure 30 Illustration of a possible approach to track over time the optimal thresholds.

Such a procedure can demonstrably decrease the energy consumption of a real network by more than 10% without degrading the QoS perceived by the user. For example, it was shown during live trials that the energy consumed by BSs could be reduced by 11%, if the system was bound by constraints such that the 4G downlink throughput exceeds Y=5 Mbps for at least X=89% of the time, where such values of X, Y could be predefined. In other scenarios, were any of the value's X or Y to decrease, then higher energy savings would be achieved.

Indeed, depending on the MNO's appetite for acceptable QoS degradation, the level of aggressiveness of the procedure can be tuned by appropriately, as illustrated in Figure 30, defining the formula that defines whether current KPI values are "acceptable", as well as the level of confidence with which we ensure that KPI values are "acceptable".

8.4. A RESOURCE-AWARE MACHINE LEARNING FRAMEWORK

With the increasing size of deep learning models AI/ML-based applications are increasingly becoming computation-intensive, power-hungry, and memory-demanding.

The large size of AI/ML models hinders their deployment on resource-constrained network devices due to insufficient computational power, storage, or memory. Also, this results in long inference times and high energy consumption when running or transferring AI/ML models in the network.

Moreover, a network device may not be able to store all the scenario-dependent AI/ML models due to memory constraints. Consequently, it may request a fresh model to the network endpoint to execute a new task and/or operate under new conditions.

Several optimisation methods have been proposed to address these challenges, including model compression and quantisation. However, these solutions are inefficient from a communication point-of-view [26].

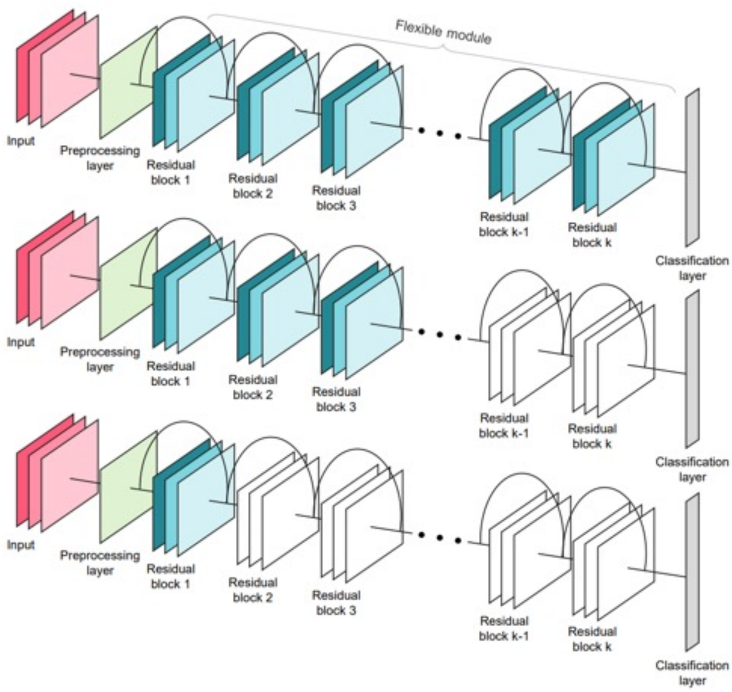

Figure 31. Resource efficient AI/ML model training [27].

To solve these challenges, a recently introduced AI/ML framework can be used which samples various configurations of a given neural network, the same as that used during the training process to create a flexible model with minimal accuracy loss compared to the full model [26].6 The fundamental component of this approach is the AI/ML model training phase, described in Figure 31. In a nutshell, this selects the number of hidden layers to update at each training iteration, e.g., minibatch, according to a given probability distribution. After choosing the number of layers for a given training iteration, the non-selected layers are skipped when computing the error and updating the weights. This approach is designed to a) accommodate the diverse storage/computational resources available on different devices during the training and inference phase and, b) minimise the energy consumption due to the model transmission and execution. Exemplary results are described in Table 4, where our approach is compared with two benchmarks in terms of test accuracy using Visual Transformers (ViTs) architectures on the CIFAR-100 dataset [28].

Table 4: Comparison of test accuracy (%) for visual transformers models on CIFAR-100 dataset.

Model size equivalence	Independent training; Full model transmission	One training and partial transmission	Our proposal
ViT-1	39.3 ± 0.43	5.8 ± 0.3	35.9 ± 0.6
ViT-2	49.7 ± 0.12	27.5 ± 0.83	49.7 ± 0.6
ViT-3	52.8 ± 0.85	52.8 ± 0.85	52.8 ± 0.85

In the independent training benchmark, each ViT architecture is independently trained and fully transmitted to network device where it is run. In the second benchmark, only ViT3 (the full model) is trained but, due to e.g., bandwidth constraints only part of it is available at the network device. We can observe that the accuracy drops by nearly 50% and by nearly 90% when the available network corresponds to the size of ViT2 (2/3 of the full model) and to the size of ViT1 (1/3 of the full model). In contrast, the results show a negligible accuracy loss when using the proposed solution in the same conditions.

Therefore, one could choose to use the proposed approach to transmit smaller models, with considerable network energy saving, without impacting the quality of service. In addition, the proposed approach requires only one training process, instead of 3, which only requires 37% of the FLOPs needed by the training process of ViT3, which results in significant savings in computational time and energy consumption during training.

09 CONCLUSION

NGMN advises that MNOs carefully assess the solutions to enhance network energy efficiency presented in this publication, and analyse the prioritised list of strategies in Table 5, where energy saving approaches are organised into three broad categories and time-horizons: (short-term) process optimisations; (medium-term) engineering optimisations; and (long-term) new technologies.

Table 5: Energy saving potential from technologies described in this report.

Area	Energy Saving Solution	Examples of Energy Saving Potential	Timeframe
Process Optimisation	Measuring RAN Energy Efficiency (Sec. 2)	This framework allows MNOs to correctly measure EC and EE of RAN equipment and take wise optimisation decisions accordingly	Short-term
	Energy saving of indoor pRRU (Sec. 4.2)	This solution achieves energy saving gain of 20% with respect to the always-on network deployment	Short-term
	Optimising load-dependent power consumption (Sec. 4.3)	This method can achieve 30% reduction in load-dependent energy consumption without impacting UEs' QoS	Short-term
	Hierarchical Network Energy Efficiency Optimisation (Sec. 5.2)	This algorithm can be added on top of the standard energy saving policy to achieve 3.1% additional energy saving gain	Short-term
	Improving energy efficiency through coordinated resource usage in multi-carrier systems (Sec 5.3)	The scheduling of multiple carriers can be aligned in the time domain to increase the symbol shutdown duration.	Short-term
	Energy saving and spectrum sharing in multi-carrier systems (Sec. 5.4)	Implementing carrier shutdown together with spectrum sharing between capacity booster cells and coverage cells can lead to 26% network EE gains.	Short-term
	Identification of low energy-efficiency BSs based on AI/ML (Sec. 8.1)	AI-based mechanism to identify low-EE BS sites such that necessary actions can be taken to improve the overall network energy efficiency	Medium-term
	AI/ML assisted Network Energy Saving (Sec. 8.2)	It extends the duration of shutdown, resulting in a 2.48% reduction of power consumption, and a 16.85% improvement in energy efficiency.	Medium-term

Area	Energy Saving Solution	Examples of Energy Saving Potential	Timeframe
	Energy savings via capacity shutdown under performance constraints (Sec. 8.3)	Measuring RAN Energy Efficiency (Sec. 2)	Short-term
Process Optimisation	Energy savings via capacity shutdown under performance constraints (Sec. 8.3)	The BS energy consumed could be reduced by 11% by ensuring that 4G DL throughput exceeds 5 Mbps for at least 89% of the time	Short-term
Engineering Optimisation	Passive antenna efficiency optimisation (Sec. 4.1)	Passive antennas with remote azimuth and tilt control combined with AI leads to a 25% energy saving gain	Medium-term
	Inter-MNOs RAN sharing (Sec. 5.5)	Active RAN sharing case study has shown 30% energy OPEX savings	Medium-term
	Interworking between communication network and -power supply (Sec.s 6.2 and 6.3)	Reducing network OPEX and increasing its availability by jointly dimensioning and controlling communication and power resources	Medium-term
	Dynamic reconfiguring connection between a RU and DUs based on traffic load (Sec 7.1)	This technique allows power saving realisation by switching off optical paths in the fronthaul and unused DUs when network traffic load is low.	Medium-term
	Open RAN (Sec. 7.2)	The Open RAN MoU has defined EE targets for O-RUs in both loaded and unloaded conditions, for benchmarking Open RAN O-RU with traditional RAN.	Medium-term
New technologies	Front-end Adaptivity for Increased Energy Efficiency (Sec. 4.4)	This architecture allows to select suitable modulation scheme and transceiver front-end based on data rate requirements and spectral availability to reduce energy consumption.	Long-term
	A Resource-Aware Machine Learning Framework (Sec. 8.4)	The output AI model requires 37% of the FLOPs needed by the training process of the original one, leading to significant savings in energy consumption during training	Long-term

10 LIST OF ABBREVIATIONS

3GPP	Third Generation Partnership Project
AI	Artificial Intelligence
AAU	Active Antenna Unit
ADC	Analog-to-digital converter
API	Application Programming Interfaces
ATIS	Alliance for Telecommunications Industry Solutions
BBU	Baseband Unit
BO	Bayesian optimisation
BS	Base Station
CCSA	China Communications Standards Association
CSI	Channel State Information
DC	Direct Current
DL	Downlink
DRX	Discontinuous Reception
DTX	Discontinuous Transmit
EC	Energy Consumption
EE	Energy Efficiency
EIRP	Effective Isotropic Radiated Power
ETSI	European Telecommunications Standards Institute
gNB	Next Generation Node B
ICT	Information and Communication Technology
ITU-T	International Telecommunications Union Telecommunication Standardisation Sector
KPI	Key Performance Indicator
LSTM	Long Short-Term Memory
MIMO	Multiple Input Multiple Output
MNO	Mobile Network Operator

MoU	Memorandum of Understanding
NES	Network Energy Savings
NGMN	Next Generation Mobile Network
NR	New Radio
O-RAN	Open RAN
PA	Power Amplifier
PRB	Physical Resource Block
pRRU	pico Remote Radio Unit
QoS	Quality of Service
RAN	Radio Access Network
RAS	Remote Azimuth Steering
RF	Radio Frequency
RIB	Radiated Interface Boundary
RRU	Remote Radio Unit
RU	Radio Unit
SCell	Secondary Cell
SIB	System Information Block
SINR	Signal-to-Interference-plus-Noise Ratio
SSB	Synchronization Signal Block
SVR	Support Vector Regression
TRP	TX/RX point
TS	Tabu Search
TTI	Time Transmission Intervals
UE	User Equipment
ViT	Visual Transformers
VNF	Virtualised Network Function
WUS	Wake Up Signal
ZO	Zeroth order

11 FIGURES

12 TABLES

13 REFERENCES

[1] NGMN Alliance , "Green Future Networks; Effective Energy Saving Strategies and Best Practices for MNOs," Nov. 2023.

[2] ETSI ES 202 706-1, "Environmental Engineering (EE); Metrics and measurement method for energy efficiency of wireless access network equipment; Part 1: Power consumption - static measurement method".

[3] ETSI TS 102 706-2, "Environmental Engineering (EE); Metrics and measurement method for energy efficiency of wireless access network equipment; Part 2: Energy Efficiency - dynamic measurement method".

[4] ETSI TS 103 786, "Environmental Engineering (EE); Measurement method for energy efficiency of wireless access network equipment; Dynamic energy performance measurement method of 5G Base Station (BS)".

[5] ITU-T Recommendation L.1310 (09/2020), "Energy efficiency metrics and measurement methods for telecommunication equipment".

[6] 3GPP TS 38.141-1 V17.6.0 (2022-06), " NR; Base Station (BS) conformance testing Part 1: Conducted conformance testing".

[7] 3GPP TS 38.141-2 V17.6.0 (2022-06), "NR; Base Station (BS) conformance testing Part 2: Radiated conformance testing".

[8] 3GPP TS 28.310, V17.3.0 (2021-12), "Management and orchestration; Energy efficiency of 5G".

[9] 3GPP TS 28.554 V17.7.0 (2022-06), " Management and orchestration; 5G end to end Key Performance Indicators (KPI)".

[10] ETSI ES 203 228, "Environmental Engineering (EE); Assessment of mobile network energy efficiency".

[11] ITU-T Recommendation L.1331 (01/2022), "Assessment of mobile network energy efficiency".

[12] NGMN Alliance, "Green Future Networks: Network Energy Efficiency Phase 2," Oct. 2023. [Online]. Available: https://www.ngmn.org/wp-content/uploads/NGMN_Network_Engergy_Efficiency_Phase2.pdf.

[13] 3GPP TSG RAN, TR 38.864, "Study on Network Energy Savings for NR," Release 18, V.18.1.0, March 2023.

[14] 3GPP TSG RAN, TR 38.840 "NR; Study on User Equipment (UE) power saving in NR," Release 16, V16.0.0, June 2019.

[15] 3GPP TSG RAN, RP-202846, "Work Item Description; NR Side-link Enhancement ," Release 17, December 2020.

[16] Nokia, "Reducing energy use with 5G-Advanced: The essential guide to RAN energy savings in 3GPP Release 18," 2023.

[17] W. Li, D. López-Pérez, X. Geng, H. Bao, Q. Song and X. Chen, "A Zeroth-Order Continuation Method for Antenna Tuning in Wireless Networks," in IEEE International Conference on Communications ICC, Montreal, QC, Canada, 2021, 2021.

[18] B. Murmann, "ADC Performance Survey 1997-2023," [Online]. Available: https://github.com/bmurmann/ADC-survey.

[19] H. Boche and G. P. Fettweis, "6G: The personal tactile internet—and open questions for information theory," IEEE BITS Inf. Theory Mag., pp. 71-82, 2021.

[20] G. Fettweis, M. Dörpinghaus, S. Bender and L. Landau , "Zero crossing modulation for communication with temporally oversampled 1-bit quantisation," in 53rd Asilomar Conf. Signals, Syst. Comput., Pacific Grove, CA, USA, 2019.

[21] R. Hadani, S. Rakib, M. Tsatsanis, A. Monk, A. J. Goldsmith, A. F. Molish and R. Calderbank, "Orthogonal Time Frequency Space Modulation," in IEEE Wireless Commun. Netw. Conf., San Francisco, CA, USA, 2017.

[22] 3GPP TSG SA V19.2.0 (2023-12), "TR 22.882, Study on Energy Efficiency as a service criteria".

[23] NGMN Alliance, "Green Future Networks: Metering in Virtualised RAN infrastructure," NGMN White Paper, April 2024.

[24] Open RAN , "Technical Priority Document, Rel. 3," April 2023.

[25] Open RAN , "MoU progress update on maturity, security and energy efficiency," February 2023.

[26] F. Ayed, A. De Domenico, A. Garcia-Rodriguez and D. Lopez-Perez, "Accordion: A Communication-Aware Machine Learning Framework for Next Generation Networks," IEEE Communications Magazine, vol. 61, no. 6, pp. 104-110, June 2023.

[27] M. Unsal, A. Maatouk, A. De Domenico, N. Piovesan, and F. Ayed, "FlexTrain: A Dynamic Training Framework for Heterogeneous Devices Environments," in Workshop on Advancing Neural Network Training (WANT) at NeurIPS, 2023.

[28] A. Dosovitskiy, "An image is worth 16x16 words: Transformers for image recognition at scale," CoRR, abs/2010.11929, 2020.

[29] ETSI, "ES 203 228 V1.4.1, Environmental Engineering (EE); Assessment of mobile network energy efficiency," 2022.

[30] ITU-T Recommendation L.1350 (10/2016), "Energy efficiency metrics of a base station site".

[31] B. Murmann, "ADC Performance Survey 1997-2023," [Online]. Available: https://github.com/bmurmann/ADC-survey.

[32] L. Deng, G. Li, S. Han, L. Shi and Y. Xie, "Model Compression and Hardware Acceleration for Neural Networks: A Comprehensive Survey," Proceedings of the IEEE, pp. 485-532, April 2020.

[33] J. Huttunen, M. Pärssinen, T. Heikkilä, O. Salmela, J. Manner and E. Pongracz, "Base Station Energy Use in Dense Urban and Suburban Areas," IEEE Access, vol. 11, pp. 2863-2874, 2023.

[34] 3GPP, Release 19, December 202, "RP-234065, Work Item Description; Enhancements of network energy savings for NR".

[35] NGMN Alliance, "Green Future Networks: Network Energy Efficiency," NGMN White Paper, Dec. 2021.

[36] NGMN Alliance, "Green Future Networks: NETWORK ENERGY EFFICIENCY Phase I," 2021. [Online]. Available: https://www.ngmn.org/wp-content/uploads/211009-GFN-Network-Energy-Efficiency-1.0.pdf.

[37] NGMN Alliance, "Green Future Networks: NETWORK ENERGY EFFICIENCY PHASE 2," 2023. [Online]. Available: https://www.ngmn.org/publications/green-future-networks-network-energy-efficiency.html.

14 ACKNOWLEDGEMENTS

China Mobile, Cong Zhang

China Mobile, Jianhua Liu

China Mobile, Xiaoting Wang

Deutsche Telekom, Paul Pfundt

Deutsche Telekom, Saima Ansari

Ericsson, Daniel Dianat

Fraunhofer IIS, Elke Roth-Mandutz

Fraunhofer IIS, Goerge Geordie

Fraunhofer IIS, Gustavo Wagner

Fraunhofer IIS, Nazanin Vatanian

Fraunhofer IIS, Oliveira da Costa

Huawei, Anterow Aleksis

Huawei, Antonio De Domenico

Huawei, Magnus Olsson

Huawei, Nicola Piovesan

Huawei, Peng Yongcheng

Huawei, Xiaowen Li

Huawei, Xiaoyi Zhou

Huawei, Zhenghu Ding

Intel, Gary Li

Nokia, Olivier Guyot

Orange, Salvatore Costanzo

Orange, Yuanyuan Huang

TNO, Ljupco Jorguseski

TUD, Florian Gast

Turkcell, Mehmet İzzet Sağlam

Turkcell, Taylan Öztürk

Vodafone, Rishikesh Chakraborty

ZTE, Jiajun Chen

NEXT GENERATION MOBILE NETWORKS ALLIANCE

NGMN is a forum established in 2006 by world-leading Mobile Network Operators. NGMN is a global operator-led alliance, comprising over 80 companies and organizations across operators, manufacturers, consultancies and academia.

Its objective is to ensure that next generation network infrastructure, service platforms, and devices will fulfil the requirements of operators and, ultimately, meet end-user demands and expectations.

VISION

The vision of NGMN is to provide impactful industry guidance to achieve innovative, sustainable and affordable mobile telecommunication services for the end user with a particular focus on Mastering the Route to Disaggregation, Green Future Networks and 6G, whilst continuing to support 5G's full implementation.

MISSION

The mission of NGMN is:

- To evaluate and drive technology evolution towards the three **Strategic Focus Topics:**
  - **Mastering to the Route to Disaggregation:** Leading in the development of open, disaggregated, virtualised and cloud native solutions with a focus on the E2E Operating Model
  - **Green Future Networks:** Developing sustainable and environmentally conscious solutions
  - **6G:** Anticipating the emergence of 6G by highlighting key technological trends and societal requirements, as well as outlining use cases, requirements, and design considerations to address them.
- To define precise functional and non-functional requirements for the next generation of mobile networks
- To provide guidance to equipment developers, standardisation bodies, and collaborative partners, leading to the implementation of a cost-effective network evolution
- To serve as a platform for information exchange within the industry, addressing urgent concerns, sharing experiences, and learning from technological challenges
- To identify and eliminate obstacles hindering the successful implementation of appealing mobile services.