Power Modeling of the O-RAN O-RU & Application of Advanced Sleep Modes for Enhanced Energy Efficiency

Muhammad Qamar Usman*[a], Cormac J. Sreenan*[b], Marcin Dryjanski†[c], Aisling O'Driscoll*[d],
* nasc research.
School of Computer Science & Information Technology, University College Cork, Ireland.
† Rimedo Labs Poznan, Poland.
[a] mq.usman@cs.ucc.ie, [b] cormac.sreenan@ucc.ie,
[c] marcin.dryjanski@rimedolabs.com, [d] aisling.odriscoll@ucc.ie

2025 IEEE 22nd Consumer Communications & Networking Conference (CCNC) | 979-8-3315-0805-0/25/$31.00 ©2025 IEEE | DOI: 10.1109/CCNC54725.2025.10975933

***Abstract*—Open Radio Access Networks (O-RAN) enhance RAN flexibility, interoperability, and intelligence by introducing open interfaces, disaggregation of RAN components, and data-driven control. This flexibility has the potential to significantly reduce capital and operating expenses (CAPEX/OPEX) by enhancing efficiencies in network energy consumption. Advanced Sleep Modes (ASMs) are an advantageous energy-saving technique for 5G base stations (BS), with a set of sleep depths that progressively deactivate different RAN components. The deployment and effectiveness of ASMs in O-RAN raises a number of challenges that have yet to be investigated. This paper surveys and critically reviews RAN ASMs, as well as other energy-enhancing techniques applied in O-RAN. It presents the first granular, per-component O-RAN Radio Unit (O-RU) power model, that will enable researchers to quantify the power savings gained from deploying ASMs. A calculation of the energy savings achieved at each ASM level per O-RU component for different 3GPP functional splits is provided. The power consumption of O-RU sub-components for functional splits 6, 7.2x, and 8 is shown. The default 3GPP functional split 7.2x as adopted by O-RAN shows that substantial power savings can be achieved by applying ASMs to the O-RU, achieving a minimum of 80% savings depending on the choice of SM as compared to power consumption at full load. Finally, a discussion of the open research challenges and opportunities in deploying ASMs in O-RAN is provided.**

Index Terms*—Open Radio Access Network (O-RAN), O-RAN Radio Unit (O-RU), Advanced Sleep Modes (ASMs), power model, energy efficiency.

I. Introduction

Mobile Network Operators (MNOs) account for an estimated 2%-3% of global energy consumption [1] with mobile network energy expenditure accounting for approximately 5% of their annual operating expenditure. When profiling the overall energy consumption of mobile network components [2], it has been shown that the Radio Access Network (RAN) consumes as much as 73%. Half of this is consumed by the radio unit, baseband processing, and main control, with the second half consumed by air conditioning. Thus, the energy consumption of the RAN poses a serious challenge for MNOs. Potential energy savings (ES) can be achieved by employing next-generation RAN architectures. More recently, the O-RAN architecture shown in Figure 1 is being investigated for its many benefits including open interfaces that break vendor lock-in, disaggregation of RAN components, commercial off-the-shelf hardware (COTS) hardware, and intelligent controllers that run Machine Learning (ML) algorithms for near real-time network optimization. These characteristics are especially applicable for network ES, with the disaggregation benefits of architectural components, as well as the utilization of technology improvements such as power amplifier (PA) enhancements, spectral efficiency, and energy harvesting techniques. It is expected that this will be further extended for 6G-related energy enhancements such as zero-touch networks, rate splitting, and leveraging the benefits of Intelligent Reflecting Surfaces (IRS) [3]. However, such technological improvements also have their unique challenges. PA improvements are not cost-efficient as they require high-performance amplifiers and additional cooling systems. Spectral efficiency requires advanced technologies, which can be expensive and complex, and some energy harvesting techniques are weather-dependent. In contrast, O-RAN can utilize the Non-Real-Time and the Near-Real-time RAN Intelligent Controllers (RIC) to implement more intelligent and adaptive ES measures. These can include SMs [4], network virtualization, and steering traffic to energy-efficient base stations. Sleep mode techniques in particular can save significant energy [2] and do not require any additional hardware. Furthermore, they require less effort to implement in the network. However SMs need to be carefully implemented as they can increase network latency, and the activation time required to wake up RAN components can degrade the customer experience. The flexibility of RICs may allow for this latency to be minimized by using AI traffic predictions. The remainder of this paper is organized as follows: Section II discusses the implications of the O-RAN architecture for ES with a particular emphasis on deployment scenarios and functional splits. Section III gives an overview of SMs and reviews O-RAN literature related to ES. Section IV presents a per-component O-RU power model that can be used to evaluate the impact of any proposed RAN ES measure at a granular level. In Section V, this model is used to calculate the O-RU ES that can be gained by utilizing various

ASM mechanisms while considering multiple functional splits. Finally, Section VI describes open research challenges and opportunities relating to ES in O-RAN.

II. Implications of the O-RAN architecture for energy savings

The O-RAN Alliance is a worldwide community of MNOs, vendors, and research/academic institutions endeavoring to transform RAN to be more open, intelligent, and interoperable. There are 11 dedicated Working Groups (WGs), each responsible for a specific O-RAN component or functionality. The activities of WG1 [5] are the most relevant to this paper, as they describe ES use cases, slicing capabilities, as well as an overview of the general O-RAN architecture. The O-RAN Alliance introduced open interfaces and virtualization that enable the operators to mix and match different hardware and software components from different vendors. Such open interfaces ensure interoperability between different equipment providers and enable the integration of diverse components within the network. Virtualization decouples software from hardware and enables network functions to run on COTS rather than proprietary hardware.

The O-RAN architecture disaggregates the BBU into an O-RAN Central Unit (O-CU) and Distributed Unit (O-DU), as shown in Figure 1. These run on COTS hardware in a cloud environment. The O-DU manages higher-layer functionalities such as the Radio Link Control (RLC) layer, Medium Access Control (MAC), and the higher part of the physical layer functionalities. The O-CU manages the control plane including Radio Resource Control (RRC), Packet Data Convergence Protocol (PDCP) and the Service Data Adaptation Protocol (SDAP). The O-RU is usually located at the cell site and performs low layer PHY and radio frequency (RF) operations. The concept of the RIC is an essential tenet of O-RAN that provides intelligence and enhances the efficiency of radio network operations. The RIC is comprised of the Near-Real-Time RIC (Near-RT RIC) and Non-Real-Time (Non-RT) RIC. The Near-RT RIC runs time-sensitive micro-services with time frames between 10 ms and 1s and orchestrates near real-time supervision. The Near-RT RIC serves as a platform, enabling multiple applications called xApps to support various network functions. xApps are independent of the Near-RT RIC and may incorporate ML models. They perform real-time optimization tasks such as traffic steering (TS), interference management, and cell sleeping management. The Near-RT RIC interacts with the O-DU and O-CU via the E2 interface to manage radio resources and ensure efficient network performance. The Non-RT RIC, with a control loop of more than one second, usually performs long-term RAN optimization management and it sends control policy guidance to the Near-RT RIC through the A1 interface. The Non-RT RIC supports rApps that can be provided by third parties and could be provided by the RIC platform provider itself to offer services that support RAN optimization and execute AI/ML functionalities for policy management and orchestration of RAN resources. rApps perform tasks that include network slicing or energy-

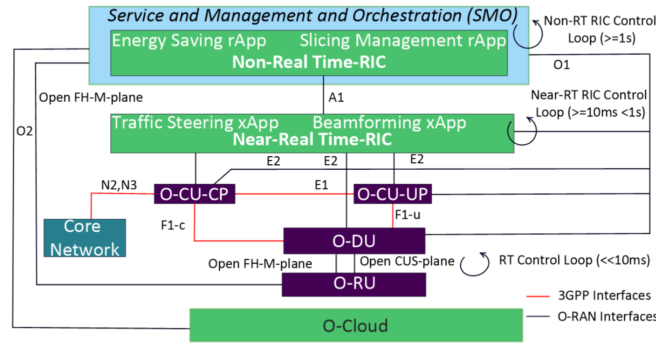

Fig. 1: O-RAN Architecture.

saving management. The Service Management and Orchestration (SMO) is responsible for the deployment, configuration, and scaling of RAN components. The O1 interface connects the RAN components with the SMO framework.

Based on latency and cost considerations, the O-RAN Alliance has defined 6 deployment scenarios [6] indicating placement options of key network functions, ranging from virtualized network functions (VNFs) deployed at the O-Cloud to physical network functions (PNFs) deployed at the cell site [6]. As such, O-RAN deployment scenarios play an essential role in the energy consumption of the RAN, and given that most of the O-Cloud components are deployed on COTS hardware, this minimizes O-RAN energy consumption. Of the 6 specified scenarios, the O-RAN Alliance has selected scenario B as an initial use case in which the O-CU and O-DU are deployed at the edge cloud and the Near-RT RIC is deployed at the regional cloud. 3GPP functional splits offer the option of placing O-RAN network functions in different Points-of-Presence (PoPs). In functional split 6, all PHY layer and RF functions are performed in the O-RU while the O-CU and O-DU manage the MAC and RLC tasks. In functional split 8, all responsibilities from PHY to RRC layer, except the RF functions, are performed at the O-DU while the RF is performed at the O-RU, but it also requires a high-capacity front haul network. In the 7.2x functional split (recommended as default by the O-RAN Alliance), the low layer PHY and RF tasks are performed at the O-RU, and the higher layer PHY functions are performed at O-DU. 3GPP 7.2x has been shown to offer improved performance because of the low impact on transport bandwidth.

III. Energy saving and sleep modes in RAN and O-RAN

Cellular networks must have the capacity to meet user demand. As the traffic demand fluctuates in both the time and space domains, operating the RAN to handle peak traffic demand can result in significant unnecessary energy expenditure at times when the traffic demand is lower. A key strategy is to turn off RAN components to save energy when low traffic loads are evident. A review of fundamental techniques is now given, with a particular emphasis on 3GPP ASMs.

Furthermore, recent efforts to apply ASMs within the O-RAN architecture are discussed, along with a review of the literature on other O-RAN ES techniques.

A. RAN energy-saving techniques

Numerous RAN ES techniques have been proposed that focus on shutting down selected components. These techniques include symbol shutdown, channel shutdown, carrier shutdown, and ASMs. In symbol shutdown, the PA is muted for a certain number of empty symbols if no information is to be transmitted. It was reported in [2] that this can account for an 8% power reduction in the radio processing unit. In channel shutdown, selected RF channels are muted in low-load scenarios while ensuring that the remaining channels can suffice to serve the current users. In [7], a 20% reduction in power consumption was reported by using this option, with a coverage loss of 6dB. During off-peak hours, higher capacity (frequency) blocks are dynamically shut down exhibiting a commensurate 2% power consumption reduction for the whole BS [2].

B. Advanced sleep modes

The application of the aforementioned RAN ES techniques can have unwanted consequences in terms of customer experience, caused by low data rates from delays in restoring shut-down components. This can be mitigated by utilizing ASMs in which RAN components are gradually turned off, enabling a BS to adjust its power dynamically based on current network conditions, and emerged from papers such as [8]. 3GPP ASMs comprise 4 sleep levels (SM1-SM4) by grouping sub-components with similar transition times that include specified activation/deactivation periods. Their characteristics are summarised in Table I. SM1 is the shortest SM duration with a transition time equal to 74 μs (i.e., one OFDM symbol assuming a sub-carrier spacing (SCS) of 15 kHz). It is worth observing that this will vary depending on the choice of SCS, with 5G NR networks accommodating 15 kHz, 30 kHz, 60 kHz, 120 kHz, and 240 kHz. In this mode, the PA and some components of the digital baseband and analog front-end (FE) are deactivated. SM2 corresponds to a longer sleep duration with a transition time equal to 1 sub-frame of 1 ms. In this mode, more components of the analog FE are deactivated. SM3 requires 10 ms (a frame), and almost all analog FE components are deactivated. SM4 is the longest SM duration with a transition time equal to 1s, where all the components are deactivated, and the base station only retains its wake-up functions. Published ASM-related research focuses on either optimizing the ASM sleep duration [9]–[11] or the decision-making process that triggers the transitions between ASM levels [12]–[15].

While it has been shown that ASM levels and durations allow the network to be optimised based on the current network conditions, the existing literature is limited in that it focuses on base stations acting independently without adopting a network-wide strategy. There is an absence of studies investigating the implementation and evaluation of ASMs in O-RAN in particular. In the recent literature, only the O-RAN Alliance has described the possible application of ASMs for O-RAN. This topic is addressed in the next subsection.

TABLE I: Characteristics of 3GPP Advanced Sleep Modes.

Sleep level	**Deactivation duration**	**Minimum sleep duration**	**Activation duration**
SM$_1$	35.5 μs	71 μs	35.5 μs
SM$_2$	0.5 ms	1 ms	0.5 ms
SM$_3$	5 ms	10 ms	5 ms
SM$_4$	0.5 s	1 s	0.5 s

C. O-RAN energy saving techniques

Research applying ES techniques within the O-RAN architecture is still in its infancy. However, there have been some recent efforts that have sought to implement and evaluate the impact of ES techniques on various parts of the O-RAN architecture. Importantly, to the best of the author's knowledge, no work to date has considered the application of ASMs in O-RAN, aside from a reference in an O-RAN Alliance document [16] which will be considered later in this section. In [17], the author proposed three xApps: for TS, beamforming, and cell sleeping. The network has small and macro BS' connecting users with multiple radio access technologies (RATs). The cell sleeping xApp is designed to shut down the less busy small BS' based on traffic load ratio and queue length. Simulation results showed an increase of up to 37.9% in the average energy efficiency compared to a baseline approach. In [18], LTE BS' are deployed for better coverage, and 5G small cells are deployed to achieve high throughput. A traffic prediction xApp is used to indicate if demand will drop below a threshold in which case a cell sleeping rApp is initiated to optimize the cell's energy. The evaluation shows a 39.7% increase in energy efficiency when compared to an 'always on' cell-sleeping rApp.

The O-RAN Alliance has described an ES mechanism [16] considering different use cases including carrier and cell switch off/on, RF channel reconfiguration off/on, and ASM selection. ES is achieved by turning off a certain number of carriers or entire cells during low traffic load. A load-dependent ES rApp is designed to manage the traffic of the network. The O-RAN Alliance proposes that the ES rApp is defined in the Non-RT RIC, and it receives carrier/cell characteristics that include traffic load per cell and carrier, user mobility information, and energy consumption measurement reports through the O1 interface. When the traffic load decreases, the ES-rApp creates a TS policy to move all users from the capacity cell to the coverage cell and sends the TS policy to the TS xApp through the A1 interface. The TS xApp performs a handover procedure based on the policy received from the rApp and cell load, RSRP information through the E2 node. An estimated potential ES of up to 340W per O-RU can be achieved by carrier shutdown during low-load traffic. A channel switching technique also runs an AI/ML model training host on the non-RT RIC and an AI/ML model inferencing host in the near-RT RIC, with the model trained on data received from the O1 interface. An estimated ES gain of up to 80W in each O-RU is achieved in the case 4T4R by turning off 2TX/RX arrays. The O-RAN Alliance also described an ES mechanism utilising ASMs. Cell-specific parameters are executed at the Near-RT RIC using data, including cell traffic load, certain energy KPI measurements, and user service type. The Non-RT and Near-RT RIC include a unique identifier for each SM and additional parameters that include ASM activation/deactivation times as well as duration times between SMs.

As such, given the potential energy savings to be gained if ASMs are applied in O-RAN, the next section describes an analytical O-RU power model to measure, at a component level, the energy savings that could be gained. Subsequent sections detail the projected energy savings, plus open research challenges relating to their deployment.

IV. O-RAN O-RU POWER MODELLING

Accurately estimating and measuring the effectiveness of ES techniques such as ASMs for O-RAN is vital to prove their impact and guide their use. Abubakar et. al. [19] proposed a power consumption model to represent this at an overall system level, based on the components of the radio network and the transport network. The model can be expressed as follows.

$$P_{\text{Total O-RAN}} = P_{\text{Cloud}} + \sum_{l=1}^{L} P_{\text{O-NU}i} + \sum_{i=1}^{L} P_{\text{O-RU}i} \tag{1}$$

P_{Cloud} represents the power consumption by O-Cloud in O-RAN. The primary components of P_{Cloud} include the O-CU, O-DU, and RIC, and are dependent on different deployment scenarios. The $P_{\text{O-NU}}$ represents the power consumption in an optical network unit (ONU). It includes the energy consumption of optical transceivers, transponders, switches, and multiplexers. $P_{\text{O-RU}}$ is the power consumption of O-RUs in the network. The O-RUs are usually deployed as a PNF with proprietary hardware and are the primary source of power consumption within an O-RAN architecture. In the next subsection, an analytical power model for O-RU is presented, quantifying the per-component power contribution.

A. Proposed O-RU power consumption model

This section presents a novel O-RU analytical power model to evaluate the power consumption of the O-RU at a component level. This evaluation can enable the identification of the power profile of individual components. Firstly, the various analog and digital sub-components of the O-RU are discussed assuming a functional split 7.2x. Subsequently, the analytical power model for O-RU is presented. This granular model is essential for identifying the most energy-intensive parts of the O-RU and for evaluating ASM performance for the O-RU.

1) O-RU components: Figure 2 illustrates the O-RU components assuming a 7.2x functional split. The primary power consumers in an O-RU are the RF and digital processing units. The RF processing unit is responsible for transmitting and receiving electromagnetic signals and comprises a PA and an analog FE. The analog FE includes digital to analog/analog to digital converters, low noise amplifiers, up/down mixers, frequency synthesis, buffers, and other components that perform basic operations that consume much less power. The digital processing unit performs low layer PHY functions, including digital/analog beam forming, base-band filtering, FFT/IFFT, and some functions in the digital front-end that include digital pre-distortion (DPD), up/ down sampling, and digital up/down conversion. The digital control unit performs some additional functions such as activating different parts of the system, managing data flow, and back-hauling to the core network. All of these components consume power based on the processing load, task complexity, and different input parameters that include bandwidth, number of antennas, spectral efficiency, system load, number of spatial streams, and quantization.

TABLE II: Scaling Exponents for Digital Sub-components

Subcomponent	BW	Ant.	Load	Streams	Q	Mod.
FFT/IFFT	1	1	0	0	1.2	0
Beamforming	1	2	0	1	1.2	1
Filtering	1	0	0	0	1.2	0
Up/Down Sampling	1	0	1	0	1	1
DDC/DUC	1	0	0	0	1.2	1
DPD	1	0	1	0	1.2	0
Control	0	0.5	0	0.2	0.2	0
Network	1	1	0	1	0	0

To derive an analytical model for an O-RU on a per component level, we leverage the work of Desset et. al [20], who presented a power model to compute the scaled power consumption, P_{total}, of an LTE BS on a component level. This is expressed as follows:

$$\begin{aligned} P_{\text{total}} = & \underbrace{\sum_{i \in I_{\text{Digital}}} P_{i,\text{ref}} \prod_{x \in X} \left(\frac{x_{\text{act},i}}{x_{\text{ref}}} \right)^{s_{i,x}}}_{P_{\text{digital}}} \\ & + \underbrace{\left(\sum_{i \in I_{\text{Analog}}} P_{i,\text{ref}} \prod_{x \in X} \left(\frac{x_{\text{act},i}}{x_{\text{ref}}} \right)^{s_{i,x}} \right) + P_{\text{amp}}}_{P_{\text{analog}}} \end{aligned} \tag{2}$$

This model combines the power consumption of digital components (P_{digital}), and analog components (P_{analog}). Each sub-component's power consumption is influence by 6 different input parameters, denoted as X. These include bandwidth (BW), number of antennas (N), system load (L), modulation scheme (M), number of spatial streams (K), and quantization (Q). x_{act} and x_{ref} are actual and references parameter values. Actual values represent the specific operations based on real-world deployments and scenarios. Reference values represent baseline parameters for establishing power consumption under the standard and default operational reference conditions. Scaling exponents describe how the power consumption of each sub-component scales with changes in operational parameters. These exponents are derived from empirical data and theoretical models from a diversity of sources including data-sheet measurements, scientific publication manufacturer

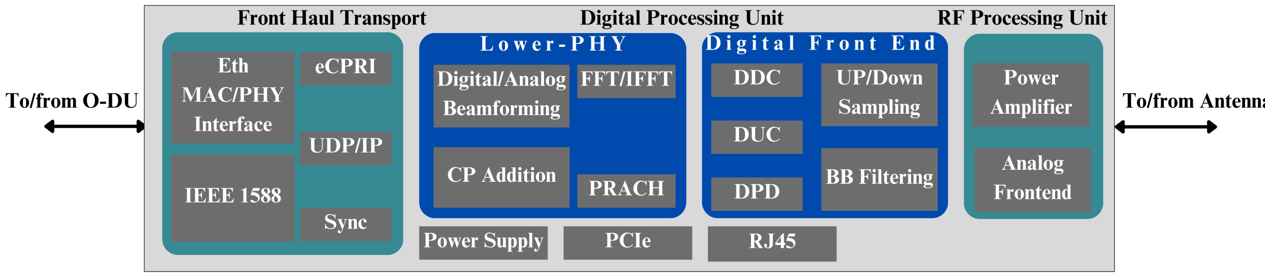

Fig. 2: O-RAN O-RU assuming functional split 7.2x.

information, and technology trends. $s_{i,x}$ is the scaling coefficient for the analog and digital sub-components. If the relationship between any sub-components and the system parameter is linear, the scaling coefficient is set to 1 and for nonlinear, it is configured greater than 1. If the relationship is independent, it is configured as 0. The scaling exponents of different digital and analog components for the different input parameters are given in Tables II and III. $P_{i,\text{ref}}$ is the reference power of sub-component i, where i is part of a set of I_{Analog} or I_{Digital} sub-components forming the O-RU under

TABLE III: Scaling Exponents for Analog Sub-components.

Subcomponent	BW	Ant.	Load	Streams	Q	Mod.
Predriver	1	0	1	0	0	1
Modulator	1	0	1	0	0	1.2
Buffer	1	0	1	0	0	0
Freq. synthesis	0	0	1	0	0	0
Feedback VCO	1	0	1	0	0	0
Feedback mixer	1	0	1	0	0	0
Clock generation	1	0	0.5	0	1	0
DAC	1	0	1	0	0	1
Feedback ACD	1	0	1	0	0	0
LNA	1	0	1	0	0	1
Second LNA	1	0	1	0	0	1
Mixer	1	0	1	0	0	0
VGA	1	0	1	0	0	0
ADC	1	0	1	0	1	1

standard reference conditions that include specific technology node size and different input parameters. In the next subsections, we describe the methods to compute $P_{i,\text{ref}}$ for the digital and analog components.

2) Power consumption of digital components: The reference power of the digital components illustrated in Figure 2 depends on the computational complexity of digital components and the intrinsic energy efficiency. Complexity refers to the computational workload required to perform various signal-processing tasks. Complexity is measured in Giga Operations Per Second (GOPS) and encompasses all the computational tasks required for signal processing, including arithmetic operations, data transfers, and memory accesses. It is a critical parameter when evaluating a digital component's power efficiency and performance in communication systems. Each digital component has a specific complexity value, dependent on the operational task and reference input parameters. Complexity numbers of digital components are

TABLE IV: Technology Node Sizes by Year from [21].

Year (Product)	$\text{Node}_{\text{size}}$ (nm)
2010	65
2012	40
2014	28
2016	20
2018	14
2020	10
2022	7
2024	5
2026	3
2028	2

converted into reference power by dividing them with intrinsic energy efficiency η_{int}, which is expressed in (GOPS/W) as defined in [20].

$$P_{i,\text{ref}} = \frac{C_i(\text{GOPS})}{\eta_{\text{int}}(\text{GOPS/W})} \tag{3}$$

To derive the reference power of a given digital sub-component i, it is firstly necessary to calculate the computational complexity of said digital component. This can be derived through different expressions defined in [22], [21], and [23]. Once the computational complexity of digital components is derived, it is next necessary to derive the energy efficiency. This represents the total number of GOPS performed per watt of power consumed. It depends on several factors, including semiconductor technology nodes (usually measured in nanometres (nm)), different input parameters, and circuit design. Due to the advancement of technology, newer nodes typically enhance efficiency because of smaller transistor sizes and lower power consumption. In [21], the authors present a scaling table that scales down the CMOS technology node every two years with different power savings shown in both digital and analog components based on this CMOS reduction. This ranges from 65nm in 2010 to an anticipated value of 2nm in 2028 and is shown in Table IV. In [21], digital components estimate a consistent power saving of 50% up until 2016, after which there is a reduction in power savings of 35%. Based on this, we can derive Equation 4 to compute the intrinsic energy efficiency for any technology year, $\eta_{\text{int}}^{\text{tech\_cur}}$.

$$\eta_{\text{int}}^{\text{tech\_cur}} = \frac{\eta_{\text{int}}^{\text{tech\_ref}} \times \prod_{i=1}^{n} \left(\frac{1}{1-f_i}\right)}{\left(\frac{Q}{Q_0}\right)^{s_{i,x}} \times \text{overhead}} \tag{4}$$

The reference intrinsic energy efficiency of any technology node, $\eta_{\text{int}}^{\text{tech\_ref}}$, can be computed by taking 75 GOPS/W as a reference value for 65nm. f_i is the percentage of power saving per year, depending on the reduction in technology node size. Intrinsic energy efficiency is also scaled to the selected quantization level based on the scaling exponent, and in Equation 4, Q and Q_0 represent the targeted and reference technology node quantization level values. The overhead factor is also incorporated to include the power consumption of data transfer and some memory operations.

3) Power consumption of the analog components: Analog components consume a specific amount of power depending on the CMOS technology node size. To compute the reference power $P_{,\text{ref}}$, technology scaling is used for analog components also. The $P_{,\text{ref}}$ for any selected CMOS technology node can be computed through the following expression [20].

$$P_{i,\text{ref}}^{\text{cur}} = P_{i,\text{ref}}^{\text{tech\_ref}} \left(1 + \frac{\left(\frac{N_s^{\text{ref}}}{N_s^{\text{cur}}} - 1 \right)}{2} \right) \quad i \in I_{\text{Analog}} \tag{5}$$

$P_{\text{tech\_ref},i,\text{ref}}$ is the reference power of different analog components for reference CMOS technology. $\frac{N_s^{\text{ref}}}{N_s^{\text{cur}}}$ represents the ratio for the reference node size to current technology node size.

4) Power consumption of power amplifier: PAs are the primary source of energy consumption within an O-RU. The power consumption of a PA does not depend on the single reference power number or scaling rules. Rather, it depends on the actual measurements and P_{amp}, can be expressed as follows, as presented by Al-Karawi et. al in [24].

$$P_{\text{amp}} = \frac{P_T}{\eta_{\text{amp}}(P_T)} \tag{6}$$

It depends on the transmitted power, P_{T}, and the total efficiency of the amplifiers η_{amp}, while observing that efficiency also varies with respect to transmitted power $\eta_{\text{amp}}(P_T)$. Thus, P_{amp} can be minimised by increasing the efficiency of $\eta_{\text{amp}}(P_T)$.

V. Application of O-RU model for power modelling in O-RAN

By using the proposed analytical power model derived in the previous sections, it is now possible to estimate O-RU power consumption on a per-component level, assuming the network parameters defined in Table V. The power consumption of O-RU components is computed at full load and also at different SM levels. The O-RU components that can be shut down for each ASM sleep level in an O-RAN architecture are shown in Table VI. O-RU components for a functional split 7.2x with the same transition time can be turned off at each SM level. The depth of SM can be selected based on network conditions without degrading the QoS of the network. TableVII represents the detailed power profile for a given operational scenario. At full load, the total power consumption is is 393W, which drastically drops to 3W in SM4.

TABLE V: Network Parameters.

Parameter	Value	Parameter	Value
BW	100 MHz	N	4
Q	12 bits	P_T	40 W
$\text{Node}_{\text{size}}$	5 nm	K	4
f_c	3.5 GHz	M	64-QAM
f_s	30.72 MHz	T_s	71.4 µs

Figure 3 also demonstrates the effectiveness of ASMs when applied to both digital and analog processing. Figure 4 presents the power consumption of the O-RU for different 3GPP functional splits. ASMs are applied to functional splits 6, 7.2x, and 8 with power consumption for different operational conditions computed. Split 6 consumes the highest power, followed by 7.2x, with split 8 consuming the least power. However, split 8 also requires low latency and a high capacity front haul connection, which is expensive to implement. Split 7.2x is preferred in high-capacity urban scenarios because of its balanced approach to network efficiency and flexibility with significant savings in terms of power consumption.

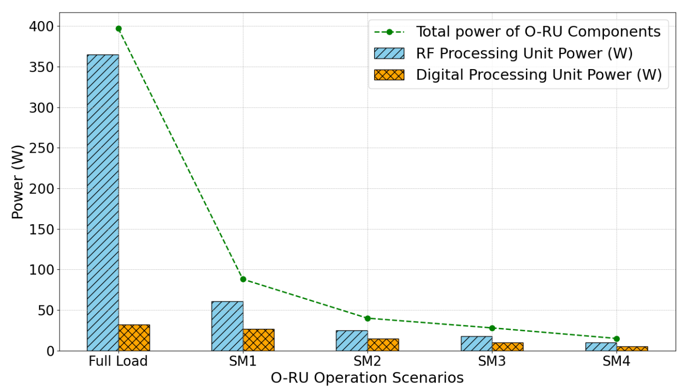

Fig. 3: O-RU power savings for different ASM levels assuming functional split 7.2x.

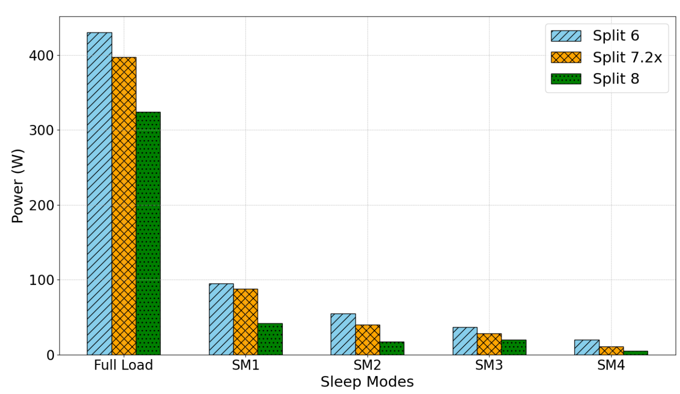

Fig. 4: O-RU power savings for different functional splits.

VI. Open research challenges and opportunities

Certain challenges exist in implementing ASM strategies in next-generation O-RAN architectures. However, some emerg-

TABLE VI: Deactivation of O-RU sub-components as a function of ASM level.

Components	Sub-Components	SM1	SM2	SM3	SM4
RF Processing Unit	Power Amplifier	✓	✓	✓	✓
	Analog Front End				
	LNA		✓	✓	✓
	Frequency Synthesis	✓	✓	✓	✓
	ADC		✓	✓	✓
	VGA	✓	✓	✓	✓
	CLK Generation		✓	✓	✓
	Mixer	✓	✓	✓	✓
	Feedback Mixer	✓	✓	✓	✓
	Modulator	✓	✓	✓	✓
	DAC		✓	✓	✓
	Feedback DAC		✓	✓	✓
	Pre-driver			✓	✓
	Buffer			✓	✓
Digital Processing Unit	**Digital Front-end**				
	Digital Pre-distortion (DPD)			✓	✓
	Digital Up/Down Conversion			✓	✓
	Up/Down Sampling	✓	✓	✓	✓
	Low-PHY Layer Functions				
	FFT/IFFT		✓	✓	✓
	Digital Beamforming	✓	✓	✓	✓
	BB Filtering		✓	✓	✓

TABLE VII: Power profile for 4T4R O-RU.

O-RU Operation Scenarios	**RF Processing Unit Power (W)**	**Digital Processing Unit Power (W)**	**Total O-RU Components Power (W)**
Full Load	365	32	397
SM1	60∼ 61	25∼27	88
SM2	23∼25	13∼15	40
SM3	16∼18	8∼10	28
SM4	8∼10	3∼5	15

ing 6G technologies also present opportunities to deliver energy efficiencies. Open research challenges and opportunities are now discussed.

A. Network slicing

Implementing ASMs in an architecture that employs network slicing poses challenges due to differing QoS requirements, e.g. diverse latency, bandwidth, and reliability per network slice. Coordination policies are necessary across multiple slices to manage network stability to prevent unfairness to services associated with a given slice. An adaptive ASM xApp algorithm for each slice in the network should be able to handle the dynamic nature of service demands in different network slices both in the time and space domains. This is currently unaddressed in literature. Importantly, the prediction of network traffic for UEs connected to multiple slices simultaneously is complex as each slice may have different traffic patterns at peak times.

B. Energy efficient coordination amongst O-RUs

When selecting the SM level, different O-RUs may have diverse traffic patterns and policies. Inter O-RU communication managed by O-DUs is necessary to determined the dynamic SM levels for the O-RUs. Traffic mobility is also a critical challenge when implementing SM techniques. Connected & Automated Driving (CAD) presents several challenges because of the dynamic nature of mobility which requires frequent re-evaluation of system-wide energy efficiency ASM policies. This is further complicated by vehicular UEs running services that may be connected with multiple network slices to satisfy diverse QoS requirements. In contrast, the predictability of mobility of vehicular traffic may also offer an opportunity to proactively adapt ASM cell sleeping xApps. Also, when some cells are in SM, avoiding coverage gaps and ensuring continuous user connectivity is necessary. These challenges require solutions that include real-time traffic prediction, ultra-fast wake-up mechanisms, and adaptive power management strategies. Coordinated Multipoint (CoMP) transmission and reception managed at the O-DU and O-CU is required to handle the dynamically changing network load, and for real-time coordination and management of high physical layer to ensure energy-efficient handovers. A traffic steering xApp must be designed in such a way to handle these challenges so that each O-RU can select the SM level according to the overall state of the network.

C. Coordinated MAC layer scheduling in the presence of ASMs

A MAC scheduler allocates radio resources within symbols, subframes, and frames. It cannot be controlled from the Near-RT RIC perspective, which poses a significant challenge for ASMs. Another key issue for ASMs is the need for cooperation and control of ASM features with the MAC scheduler to save energy by switching off O-RU components for a symbol or subframe/slot. Coordinated and adaptive MAC layer scheduling algorithms are necessary to adapt to dynamic environments in the presence of ASMs. Such algorithms must also account for the user mobility patterns to ensure seamless transitions between different cells.

D. Mitigating conflicts amongst xApps

Different xApps must be deployed to enhance the network energy without degrading the quality of service and overcoming the above-discussed challenges. These xApps include a cell sleeping xApp for ASMs, a TS xApp for mobility and load management, and a resource management xApp for network slicing. xApps works simultaneously to optimize the network resources. Some challenges are also introduced when multiple xApps work together to optimize network resources. One of the major challenges is when there is a conflict between their objectives. Effective management of these conflicts is crucial to handle network efficiency and quality of service.

E. Intelligent reflecting surfaces (IRS)

IRS are passive reflective elements that can improve next-generation performance without deploying additional power-intensive infrastructure. In particular, they aim to enhance wireless network coverage and signal quality by reflecting and focusing signals toward users in low-coverage areas. The authors in [25] explored the deployment of IRS in V2X communication within the O-RAN architecture. A multi-IRS system is deployed to improve the sum rate of the network in an environment with severe signal blockage. Deploying IRS within an O-RAN architecture alongside ASM could significantly benefit network energy efficiency without additional infrastructure. This mechanism enables the BS to operate at lower power levels without degrading the network performance. However, this deployment also introduces several challenges. Real-time dynamic phase adjustment towards intended users is essential when deploying IRS with SMs. Accurate channel estimation and synchronization between the IRS and the transmitter/receiver are essential for optimal performance.

VII. Conclusion and Future Work

This paper presented a per-component analytical O-RU power model. This model can be easily scaled for different scenarios by changing parameters and values. By categorizing the various components of the O-RU into the RF and Digital Processing Units, a granular analysis of the power consumption of the O-RU across different functional splits is presented. Additionally, a calculation of the power savings up to 97% is achieved between the hardware active state and deepest sleep mode. This research contributes to the ongoing efforts to design an energy-efficient O-RAN architecture. Our future research will address the challenges associated with implementing ASMs in O-RAN, including the integration of an ASM x-App for dynamic power management.

Acknowledgements

This publication has emanated from research with the financial support of Science Foundation Ireland (SFI) under Grant number 18/CRT/6222. For the purpose of Open Access, the author has applied a CC by public copyright license to any Author Accepted Manuscript version arising from this submission.

References

[1] GSMA Intelligence. (2024) Going green: Energy efficiency in telecoms. Accessed: 2024-06-25. [Online]. Available: https://www.gsmaintelligence.com/product-news/going-green-energy-efficiency-in-telecoms/

[2] L. M. Larsen, H. L. Christiansen, S. Ruepp, and M. S. Berger, "Toward Greener 5G and Beyond Radio Access Networks—A Survey," *IEEE Open Journal of the Communications Society*, vol. 4, pp. 768–797, 2023.

[3] W. U. Khan, A. Mahmood, C. K. Sheemar, E. Lagunas, S. Chatzinotas, and B. Ottersten, "Reconfigurable intelligent surfaces for 6G non-terrestrial networks: Assisting connectivity from the sky," *IEEE Internet of Things Magazine*, vol. 7, no. 1, pp. 34–39, 2024.

[4] F. Salahdine, J. Opadere, Q. Liu, T. Han, N. Zhang, and S. Wu, "A survey on sleep mode techniques for ultra-dense networks in 5g and beyond," *Computer Networks*, vol. 201, p. 108567, 2021.

[5] O-RAN Working Group 1, "Use cases detailed specification," O-RAN Alliance, Technical Specification O-RAN.WG1.Use-Cases-Detailed-Specification-R003-v13.00, November 2023.

[6] O-RAN Working Group 6, "O-ran cloud architecture and deployment scenarios for o-ran virtualized ran," O-RAN, Alfter, Germany, Technical Specification O-RAN.WG6.CAD-v02.02, Jul. 2021.

[7] "xRAN Forum Merges with C-RAN Alliance to Form O-RAN Alliance," Online, xRAN Forum, 2018, [Online]. Available: https://www.businesswire.com/news/home/20180227005673/en/.

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

[9] A. El-Amine, H. A. H. Hassan, M. Iturralde, and L. Nuaymi, "Location-aware sleep strategy for energy-delay tradeoffs in 5g with reinforcement learning," in *2019 IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. IEEE, 2019, pp. 1–6.

[10] A. E. Amine, P. Dini, and L. Nuaymi, "Reinforcement learning for delay-constrained energy-aware small cells with exmulti-sleeping control," in *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2020, pp. 1–6.

[11] A. El Amine, J.-P. Chaiban, H. A. H. Hassan, P. Dini, L. Nuaymi, and R. Achkar, "Energy optimization with multi-sleeping control in 5g heterogeneous networks using reinforcement learning," *IEEE Transactions on Network and Service Management*, vol. 19, no. 4, pp. 4310–4322, 2022.

[12] M. Masoudi, M. G. Khafagy, E. Soroush, D. Giacomelli, S. Morosi, and C. Cavdar, "Reinforcement learning for traffic-adaptive sleep mode management in 5G networks," in *2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*. IEEE, 2020, pp. 1–6.

[13] M. Meo, D. Renga, and Z. Umar, "Advanced sleep modes to comply with delay constraints in energy efficient 5g networks," in *2021 IEEE 93rd Vehicular Technology Conference*, 2021, pp. 1–7.

[14] M. Masoudi, E. Soroush, J. Zander, and C. Cavdar, "Digital twin assisted risk-aware sleep mode management using deep q-networks," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 1, pp. 1224–1239, 2022.

[15] D. Renga, Z. Umar, and M. Meo, "Trading off delay and energy saving through advanced sleep modes in 5g rans," *IEEE Transactions on Wireless Communications*, 2023.

[16] O.-R. Alliance, "Network Energy Savings Technical Report," O-RAN Alliance Working Group 1, Tech. Rep. R003-v02-00, 2020. [Online]. Available: https://www.o-ran.org/resources/

[17] M. A. Habib, H. Zhou, P. E. Iturria-Rivera, M. Elsayed, M. Bavand, R. Gaigalas, Y. Ozcan, and M. Erol-Kantarci, "Intent-driven intelligent control and orchestration in o-ran via hierarchical reinforcement learning," in *2023 IEEE 20th International Conference on Mobile Ad Hoc and Smart Systems (MASS)*. IEEE, 2023, pp. 55–61.

[18] M. Arafat Habib, P. E. Iturria-Rivera, Y. Ozcan, M. Elsayed, M. Bavand, R. Gaigalas, and M. Erol-Kantarci, "Transformer-based wireless traffic prediction and network optimization in o-ran," *arXiv e-prints*, pp. arXiv–2403, 2024.

[19] A. I. Abubakar, O. Onireti, Y. Sambo, L. Zhang, G. K. Ragesh, and M. Ali Imran, "Energy efficiency of open radio access network: A survey," in *2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring)*, 2023, pp. 1–7.

[20] C. Desset, B. Debaillie, V. Giannini, A. Fehske, G. Auer, H. Holtkamp, W. Wajda, D. Sabella, F. Richter, M. J. Gonzalez *et al.*, "Flexible power modeling of lte base stations," in *2012 IEEE wireless communications and networking conference (WCNC)*. IEEE, 2012, pp. 2858–2862.

[21] C. Desset, P. Wambacq, Y. Zhang, M. Ingels, and A. Bourdoux, "A flexible power model for mm-wave and thz high-throughput communication systems," in *2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*. IEEE, 2020, pp. 1–6.

[22] Ö. T. Demir, M. Masoudi, E. Björnson, and C. Cavdar, "Cell-free massive mimo in o-ran: Energy-aware joint orchestration of cloud, fronthaul, and radio resources," *IEEE Journal on Selected Areas in Communications*, 2024.

[23] J. K. Chaudhary, A. Kumar, J. Bartelt, and G. Fettweis, "C-ran employing xran functional split: Complexity analysis for 5g nr remote radio unit," in *2019 European Conference on Networks and Communications (EuCNC)*. IEEE, 2019, pp. 580–585.

[24] Y. Al-Karawi, H. Al-Raweshidy, and R. Nilavalan, "Power consumption evaluation of next generation open radio access network," in *2024 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 2024, pp. 1–6.

[25] M. Abou Houran, M. Asad, G. Srivastava, J. Mirza, A. Ranjha, M. A. Javed, and X. Yang, "Intelligent reflecting surfaces assisted cellular v2x based open ran communications," *IEEE Transactions on Vehicular Technology*, 2024.