i14y Energy Efficiency Testing Framework

White Paper

Marcin Hoffmann, Marcin Dryjanski, Adrian Kliks (Rimedo Labs)

Ajesh Pulyaar Keerthi, Mohammadreza Razmi (Deutsche Telekom / i14y Lab)

Sabine Demel, Norbert Entstrasser (Deutsche Telekom)

This content has been produced by

Table of Contents

1 Introduction and Motivation

Saving energy consumption has been one of the key challenges investigated in wireless communications for many years. Continuously increasing prices for energy and environmental awareness have even further emphasized the need to improve the energy efficiency (EE) of communication networks significantly. So, of course, the minimization of energy consumption became one of the most important topics within O-RAN. It is widely expected that by deployment of an appropriate set of rApps and xApps, intelligently controlled by an AI-driven RAN Intelligent Controller (RIC), significant savings can be achieved in this area. Indeed, advanced applications deployed in Non- and Near-RT RICs, also ingested with enrichment information, have great potential to adjust the network operation mode to the actual situational context. Such expectations have also been confirmed by numerous demonstrations and experiments showcased during the last O-RAN Global PlugFests, where various interesting and inspiring solutions have been implemented and – to some extent – tested.

However, although the experiments are very stimulating and very often highly advanced, they have revealed the strong need for unification of the EE testing framework. The following challenges have been identified:

- The experiments conducted have focused on improving EE in some specific situations and aspects; yet they did not consider the broader potential impact of the proposed solution on other sectors of the network or the entire network.
- The experiments are typically highly dependent on the selected testing scenario, as well as on the hardware-software stack used for running tests. The obtained gains in EE may be much different if the scenario is modified (e.g., to one less or more challenging, with more or fewer users, with higher or lower mobility, etc.) or if the applied hardware is changed (e.g., the computation performance of the devices is different, the delays between the particular network modules is changed, etc.).
- Contemporary networks benefit from virtualization; the realization of various functionalities in a virtual way must be carefully tackled while drawing conclusions on the energy efficiency of the network.

The above discussion led to the observation that there is a need for a unified, and complete **end-to-end (E2E) energy efficiency (EE) testing framework**, which can be used as a benchmarking environment for EE testing of O-RAN applications and networks. There should be widely accepted procedures for verifying the true energy consumption of particular network elements and modules, testing the real energy improvement achieved by any rApp or xApp, and measuring the overall energy usage in the entire network. Until this time, multiple standardization bodies contributing to this topic. However, these activities are only slightly related to each other, to wit:

- **ETSI** proposes some methods to evaluate the energy consumption of the base station but lacks an O-RAN context.
- **3GPP** does not define the EE evaluation procedures but has a lot of reference scenarios and mathematical models, which could be useful for large-scale simulations.
- **O-RAN ALLIANCE** has a Test and Integration Focus Group (TIFG) aimed at providing a test framework for EE use cases. However, they lack, for example, large-scale scenarios for evaluating xApps and rApps under unified conditions.
- **O-RAN Global PlugFest** activities exploit various setups from single and multiple vendors, where some power consumption measurements are demonstrated and ES xApps and rApps are evaluated under RICs and emulated RAN networks. However, the measurements and evaluations used lack unified procedures and scenarios.

This white paper addresses the above aspects in detail, trying to build foundations for the creation of the first, complete E2E EE testing framework; it highlights the importance of a robust EE testing framework, which is essential to serve as a platform that allows comparison of various solutions in an end-to-end environment. In detail, the white paper provides a summary of the standardization efforts by the ETSI, 3GPP, O-RAN ALLIANCE, NGNM, and TIP, as well as of industry solutions from Keysight, VIAVI, Rohde & Schwarz, Kepler (Red Hat Emerging Technologies), and the Aether project by ONF. Based on this analysis, the **E2E EE testing framework** is proposed. The framework provides 4 views on the E2E EE testing:

- **Component-level**, intended for the evaluation of the EE of a single stack (O-RU+O-CU+O-DU) of the network components.
- **Global/feature-based**, intended for the evaluation of the EE algorithms deployed in the form of the xApps and rApps under a large, emulated network.
- **Deployment view**, intended for covering the EE related to the cloud deployment options.
- **Multi-link/wide network,** as a large-scale EE evaluation environment utilizing multiple cells together with RIC, and various cloud deployment options.

The paper is organized as follows:

- It starts by providing a review of energy efficiency-related aspects, focusing on its scope.
- It then presents a review of the documents relevant to EE measurement and evaluation in mobile networks from standardization bodies like ETSI, 3GPP, and O-RAN ALLIANCE.
- This is followed by reviews of the existing power consumption measurements and network emulation solutions.
- Based on the above reviews, the paper identifies the gaps and proposes a high-level E2E EE testing framework that synthesizes different approaches proposed by various standardization bodies.
- In its conclusion, this document provides recommendations for potential future directions.

2 Energy Efficiency Aspects

To fully understand the necessity of developing the E2E EE testing framework, the basic concepts and challenges behind EE in 5G and future 6G networks are introduced in this section. While EE is identified as one of the key KPIs for both 5G and 6G networks, one of the related questions is how to define and measure it in the context of mobile networks. The widely used definitions come from the 3GPP technical specifications (based also on the ETSI documents) [TS28.310]:

$$EE_{DV} = \frac{data\ volume}{energy\ consumption}\ \left[\frac{bit}{J}\right]$$

$$EE_{CoA} = \frac{coverage\ area}{energy\ consumption}\ \left[\frac{m^2}{J}\right]$$

The first definition is related to the amount of energy, which the mobile network must spend to serve a particular data volume, while the second is related to the energy cost of providing a certain coverage. However, there is an ongoing discussion that apart from the above-mentioned definitions of EE, the **network quality metrics** should also be considered, such as user throughput, coverage quality, and connection quality [TR28.880]. In addition, the EE can be measured at the different levels of mobile network scope, which are:

- **whole networks** (i.e., end-to-end),
- **subnetworks** (e.g., the radio access network),
- **single network elements**,
- **telecommunication sites**, which contain network elements and site equipment.

Moreover, some recent studies within the 3GPP (see [TR38.913]) suggest that the fair evaluation of EE within the above-mentioned scopes should be based on multiple evaluation scenarios, such as urban and rural, or Mobile Broad Band (MBB) and Ultra Reliable Low Latency Communications (URLLC).

From the perspective of the EE evaluation, there are two key viewpoints:

- **EE of the network equipment** related to the power consumed by the RAN and core network components provided by different vendors.

- **EE of the network algorithms** related to the intelligent automation algorithms, which allow reconfiguring of network equipment to the network conditions using e.g., cell sleeping modes, or MIMO antenna array reconfigurations.

Most importantly, these two viewpoints are difficult to test jointly. While the network components stack can be evaluated within the single base station (e.g., based on the ETSI procedures [ETSI_2027061_TS] [ETSI_1027062_TS]), a large (most probably simulated) network is necessary to evaluate cell sleeping modes. Moreover, due to the progressing virtualization of the mobile networks, different measurement methods should be selected for the Physical Network Functions (PNF) and Virtual Network Functions (VNF) [TR32.927]. The latter would require new dedicated tools to extract power consumption of the single VNF container deployed on the server along with other ones. On a high level, the NF energy consumption can now be estimated as follows [TS28.310]:

- The energy consumed by the NF is the sum of the energy consumed by all its constituent VNF/VNFC instances.
- For each VNF/VNFC instance, its estimated EC is a proportion of the NFVI node EC on which it runs.
- This proportion is equal to the vCPU mean usage of the VNF/VNFC instance relative to the sum of the vCPU mean usage of all VNF/VNFC instances running on the same NFVI node.

However, a further standardization effort is required to obtain a unified method of measuring and exposing the power consumption of the VNF to the network management layer [TR28.880]. Possibly tools like Kepler by Red Hat or Redfish can be used.

Things are getting even more complicated when introducing the O-RAN concept of disaggregating the base station into the O-RAN Centralized Unit (O-CU), O-RAN Distributed Unit (O-DU), and O-RAN Radio Unit (O-RU), and their potential multiple deployment configurations along with cloud implementations. Moreover, the O-RAN networks come with the RIC, which allows deployment of the third-party algorithms, i.e. xApps for Near-RT RIC and rApps for the Non-RT RIC. From this perspective, the O-RAN ALLIANCE identifies four ES use cases to be potentially address by xApps and rApps [ORAN_NES_TR]:

- **Carrier and Cell Switch Off/On (COOS)** is aimed at dynamic switching on/off either of the full cells or individual carriers to save energy without compromising the QoS, e.g., by switching off cells during low-load hours.

- **RF Channel Reconfiguration Off/On (RCR)** is a use case dedicated to Massive MIMO. While large antenna arrays utilize a significant number of hardware components, e.g., power amplifiers, this use case aims to scale the number of active RF channels to the user throughput demands.

- **Advanced Sleep Mode Selection (ASM)** aims at providing energy savings on a micro-time scale, e.g., to switch off some hardware components at the level of single frames or even symbols depending on the short-term characteristics of network traffic.

- **O-Cloud Resource Energy Saving Mode (ORES)** is oriented toward providing energy savings in the O-Cloud, by scaling the cloud resources to the traffic demands, e.g., switching off O-Cloud Nodes during idle times, or adjusting the CPU frequency.

The disaggregated and virtualized nature of the O-RAN networks in connection with third-party xApps and rApp creates even more complexity for the evaluation of the network EE. Challenges can arise in:

- EE of the network equipment:
    - Multiple deployment options of the gNB, including different endpoints for the measurements of power consumption, e.g., O-RU, O-DU, O-CU all separated, or a joint O-CU and O-DU, with a separate O-RU.
    - Different cloud deployment options, to wit: local cloud, global cloud, O-CU and O-DU on the same server or on different servers.
    - Comparison between the O-RAN hardware stack and the Single RAN (SRAN) deployments, to wit: demand for unified setups, and measurement methodologies.
    - Reliable tools to evaluate power consumption of the hardware and software components, e.g., how much energy can be saved by using a certain energy-saving xApp compared to the power consumption of the RIC software.
- EE of the network algorithms:
    - Unified evaluation scenarios to compare energy saving between different xApps and rApps vendors, optimally based on real-world data.

- The operation of the energy-saving algorithms is highly affected by the coexisting mechanisms, or other xApps/rApps, e.g., traffic steering, or radio resource scheduler.
- Cloud deployment of the RIC, and related latency issues. For example, to see whether it is possible to run a Near-RT control loop with a global cloud deployment.

These challenges provide a strong reason to explore the existing standardization effort from different bodies like O-RAN ALLIANCE, 3GPP, NGNM, etc. to see how their knowledge can be synthesized to move towards the unified E2E EE testing framework.

3 Standard and Industry Review

Following up on the energy efficiency aspects covered in Chapter 2, the standardization efforts from ETSI, 3GPP, NGMN, O-RAN ALLIANCE, TIP (Sec. 3.1), and vendor solutions (Sec. 3.2) were analyzed. This chapter aims to summarize these to identify useful and missing elements in the context of building an E2E EE testing framework.

3.1 Energy Efficiency Aspects Covered by Standardization Bodies

This section summarizes the standardization efforts by different organizations. The scope of the documents defined by each of them along with their gaps concerning an O-RAN-related EE framework are summarized in Table 1.

ETSI specifications provide very well-defined procedures for the measurement of BS EC, including temperature, voltage requirements, UE traffic models, and distribution. However, they lack an O-RAN architecture context, procedures for cloud deployment of network functions, and scenarios for the evaluation of large-scale networks necessary, such as for the evaluation of xApps and rApps controlling tens of cells.

3GPP technical reports provide models that can be used for large-scale, system-level simulations to test xApps and rApps under unified conditions, e.g., PC models, and reference deployments (like urban, rural, V2X). However, there are no procedures for E2E EE/EC testing.

NGNM is focused mostly on the cloud/virtual deployment of the mobile networks, and measurement of power consumption of VNFs (per Kubernetes cluster, Pod, container). However, it is missing the O-RAN context, and NGNM documents do not provide technical specifications.

O-RAN ALLIANCE specifications for E2E testing are not fully developed yet. They contain test procedures (functional and performance) only for a COOS use case (and only for one type of control, i.e., using an rApp within Non-RT RIC). However, the test cases are still missing large-scale scenarios – they rely on the ETSI dynamic load measurements that are good for evaluating a single O-RU, O-DU, or O-CU stack. Also, some KPIs proposed by the O-RAN ALLIANCE might not be available in practice. Finally, VNF and cloud aspects are present in O-RAN documents, but they are at an early stage of definition.

Finally, TIP defines EE-oriented use cases together with requirements, configuration, measurements, goals, and interface requirements. However, they are not related to the testing procedures themselves.

The main observation is that each entity provides documents containing important notes and (in some cases) normative specifications related directly to the EE framework. However, to produce a comprehensive E2E EE testing framework, the information should be combined, as each of them touches separate, sometimes disjoint aspects.

Table 1 Summary of standardization efforts

Documents	Scope	What is missing?
ETSI	• Defines metrics and measurements for EC/EE of a base station. • Provides models for BSs and requirements for measurement equipment. • Describes standardized test setups and procedures for power measurements (static and dynamic) for a single base station. • Standardized reports from the measurements. • Good starting point for O-RAN extension.	• Lacks O-RAN context and architecture, e.g. 7.2 split. • Lacks scenarios / configurations for large-scale network power measurements. • Missing procedures for testing mobile networks with VNF, e.g., virtual CU/DU. • Missing QoS in EE definitions.
3GPP	• Frequently refers to ETSI regarding energy efficiency / power consumption. • Defines reference scenarios, and models for system-level simulations. • Demand on measurement of VNFs' EE/PC.	• Lack of O-RAN context. • Lack of procedures for E2E EE testing. • EE/Power consumption KPIs are not available per HW unit or VNF. • Most of the material referring to EE/EC is from TRs, not TS, i.e., not normative.

	• EE evaluation based on multiple scenarios: MBB, URLLC, mMTC/Rural, urban.	
NGMN	• Cloud vs. Physical deployment for a mobile network. • Presents aspects relevant to cloud-type/environment measurements. • Metering servers, storage, and network components. • Solutions for VMN/CNF power measurement. • Optimization of RAN deployment. • Points to gaps / missing elements between the O-RAN and cloud aspects.	• Missing O-RAN aspects. • Gathers different aspects but does not fill the gap between the cloud itself and the network. • Not a specification.
O-RAN	• Related directly to O-RAN aspects. • Defines NES methods, metrics, and requirements associated with the O-RU (COOS and RCR) (quite mature). • Concepts, requirements, and use cases optimizing for O-Cloud resources (ORES) along with metrics (early stage). • Test procedure, test equipment definition, test setup and configuration,	• Missing details and tech spec for O-Cloud aspects (early stage on this). • Different maturity stages for different use cases regarding ES and E2E testing. • Different documents for different parts of the system (treating O-Cloud separately from COOS, RCR, ASM). • Test specs only for COOS rApp (early stage). • Several options for measurements and KPIs (not

	test criteria for E2E testing for O-RAN ES (COOS).	all are implemented in practice). • Missing large-scale testing scenarios for xApp/rApp evaluation.
TIP	• A use case, being a framework for a method of reducing EC. • Requirements, configuration, measurements, goal, interface requirements.	• Not related to testing. • Only for a single use case and very specific. • Only generic parameters.

3.2 Energy Efficiency Related Solutions

In this section, we summarize the solutions that could be used to develop an E2E EE testing framework. These are gathered in Table 2. The first 3 rows contain the companies that provide commercial-grade measurement equipment for the evaluation of mobile networks.

Keysight provides full-stack equipment for component-level power measurements, e.g., O-RU, O-DU/CU, UE emulators, and power sensors, as well as dedicated software for simulation studies mainly oriented towards RIC xApps and rApps under large-scale networks.

Similarly, VIAVI provides TeraVM RIC Test software to emulate the O-RAN protocol stack, allowing for the large-scale testing of EE provided by the xApps and rApps. For the HW components evaluation, VIAVI provides an O-RU tester that emulates the DU, synchronizes and configures the O-RU, and offers several test scripts to verify O-RU EE under different load conditions. VIAVI products are presented together with Rohde & Schwarz high-quality equipment for network signal measurements: oscilloscopes, signal analyzers, and a smart platform that uses real smartphones (UEs) to benchmark end-to-end service performance with standardized QoE and QoS scoring methods.

A very important solution is Kepler, founded by Red Hat. It allows monitoring of Kubernetes clusters EC at the POD and container level. This enables PC measurements of a VNF, as well as for RIC, xApps, and rApps.

During recent O-RAN PlugFests, EE/EC and its measurements were among the key use cases tested. Examples such as “Setting up O-RAN-based end-to-end system with Verification of O-RAN Specified ES use case (O-RAN PlugFest hosted by Korea Telecom)” or “Validation of an E2E ES use case with Rakuten's Near-RT RIC, Non-RT RIC, xApp, rApp, and AI/ML Platform and E2 Node emulated by VIAVI RIC Test (O-RAN PlugFest hosted by Nakao Lab – University of Tokyo)”, can provide an opportunity to discover potential partners for building the E2E EE testing platform.

Another interesting recent activity is conducted by Aether (previously ONF). They proposed a Platform for O-RAN Energy Efficiency Testing (POET). It is built on top of an open-source FlexRIC and Open Air Interface, while utilizing Kepler to measure VNF’s PC, such as O-CU/O-DU RIC xApps. However, the platform does not come with standardized procedures and scenarios. At this point, it measures the EC.

The above can serve as building blocks / partners for the development of the E2E EE testing framework.

Table 2 Summary of solutions

Vendor	Products Overview

Keysight	• Full-stack approach to the PC measurements of HW components, based on the ETSI traffic profile, including:
	  ◦ O-RU emulator
	  ◦ O-DU emulator
	  ◦ Power sensor
	  ◦ Power supply / analyzer
	• Provides a RIC Tester software to emulate the O-RAN protocol stack, allowing for the large-scale testing of EE provided by the xApps and rApps

Viavi	• Provides TeraVM RIC Test software to emulate the O-RAN protocol stack, allowing for the large-scale testing of EE provided by xApps and rApps

	• Provides an O-RU tester that emulates the DU, synchronizes and configures the O-RU, and offers several test scripts to verify O-RU EE under different load conditions.
Rohde & Schwarz	• The equipment can monitor dynamic device activities versus PC. • The oscilloscope monitors energy dynamics under various traffic conditions by tracking power changes over time. • The power supply – besides powering the O-RU – also provides high measurement resolution and accuracy over a long period. • The R&S Smart platform uses real smartphones (UEs) to benchmark end-to-end service performance with standardized QoE and QoS scoring methods.

Red Hat	• Kepler is a Red Hat Emerging Technologies project
	• Kepler utilizes software counters and power to measure power consumption by hardware resources within a Kubernetes cluster:
	  ◦ Per cluster
	  ◦ Per POD
	  ◦ Per container
	• Kepler can be used to evaluate virtual deployments of CU/DU as well as RIC xApps and rApps

O-RAN Global Plugfests 2024	• The ES techniques and PC measurements are emerging topics: • Setting up O-RAN-based E2E system with verification of O-RAN specified ES use case (O-RAN PlugFest hosted by Korea Telecom) • Validation of an E2E ES use case with Rakuten's Near-RT RIC, Non-RT RIC, xApp, rApp, and AI/ML platform as well as E2 Node emulated by VIAVI RIC Test (O-RAN PlugFest hosted by Nakao Lab – University of Tokyo)

AETHER	• POET: A Platform for O-RAN Energy Efficiency Testing • Utilizes open-source solutions: flexRIC and Open Air Interface (OAI) • Physical O-RU and O-cloud deployment of CU/DU RIC with Kepler for PC measurements • At the date of writing this report, basic evaluation supported – measurement of power and throughput without defined procedures

4 E2E EE Testing Framework

After defining the scope and analyzing the current status of development from both the standards and solutions perspectives, this chapter aims to provide a proposal for a high-level E2E EE testing framework (see Sec 4.1). However, it should be noted that it is a starting point that indicates research directions for the complicated topic of EE testing, rather than a complete solution to this advanced problem. The proposition of the high-level E2E EE testing framework is followed by the discussions of the different views on E2E EE testing (Sec. 4.2), such as EE measurement at component-level in small lab-networks, or large-scale network scenarios to evaluate xApps and rApps performance that can be realized using network simulators (i.e., RIC testers). Finally, in Sec. 4.3, we present the building blocks of the E2E EE testing framework, together with an indication of potential vendor partners for the next phases of the work.

4.1 High-Level E2E EE Testing Framework

Based on the analysis of the documents and vendor solutions in Sec. 3, we have identified what is already defined and what is missing. Based on this analysis, we are proposing a high-level E2E EE testing framework in Fig. 1.

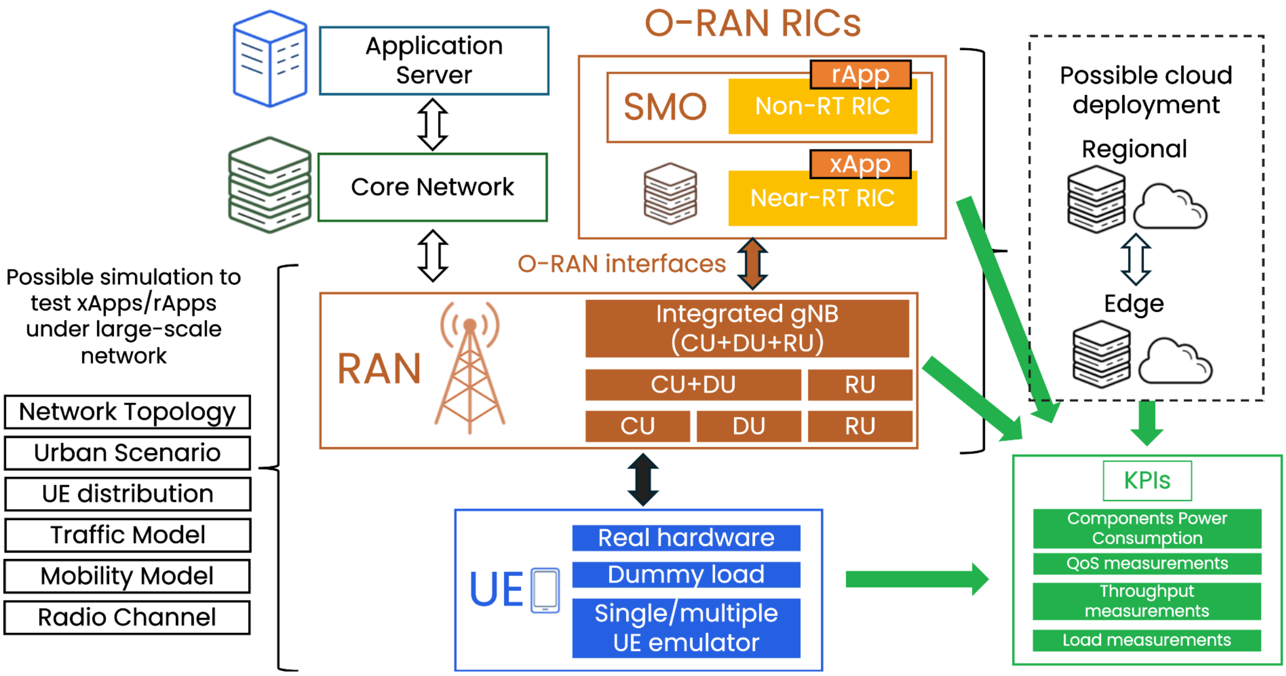

Fig. 1 High-Level E2E EE Testing Framework

The framework aims to evaluate RAN components (both deployed as PNF and VNF), as well as RICs, xApps and rApps, under unified conditions. However, it should be seen as a first attempt to capture the broad topic of E2E EE testing in O-RAN. The proposition indicates framework components and fields of further research and refinement in the future phases of the project.

The main components of the framework are:

- **Application server** to generate the traffic of UEs. Possibly different traffic types should be considered, from full-buffer to specific traffic classes like low latency, low bitrate, or high allowed latency, and high bitrate. Note, that the application server should be used in an approach, where a real UE or full UE emulator is used.

- **Core Network** that feeds the RAN with data demand created by the application server. As the core network is not evaluated, the deployment of the framework can be based on one reliable core network component.

- **RAN components** are a crucial part of the E2E EE testing framework, as their EC and KPIs will be under evaluation. They can follow various configurations: integrated gNB (CU+DU+RU), integrated CU+DU and separated RU, or full O-RAN split with separated CU, DU, and RU. Moreover, the gNB components can be from different vendors and can be realized either as PNF or VNF. VNF realization of gNB components opens up different deployment options: local server, edge cloud, or regional cloud. One should notice that PC measurement of PNFs is done with dedicated equipment, while for VNFs software solutions like Kepler must be used.

- **UE** transmits/receives the data from the gNBs, their placement and movement are related to the specific radio channel variations, creating more or fewer opportunities to save energy. The following options for the UEs can be considered for the E2E EE testing framework (note that the use of a particular one can imply the use of different elements in the framework. For example, when using a simulated UE or dummy traffic, the application server may not be needed at all)**:**

    - **Real hardware:** real network users being served by the network build of components under test, or a few UEs in a lab environment. This is the most realistic approach. However, the hardest to control and reproduce.

    - **Dummy load** can be used for static power measurements (e.g., based on an extension of ETSI static power measurement in Sec. 3.1.1) of RAN components, supported or unsupported by xApps/rApps.

    - **Single/multiple UE emulators** can be used to generate more advanced traffic patterns and profiles for dynamic power measurements (e.g., based on an extension of ETSI dynamic power measurements from Sec.

3.1.2), that could be easily repeated for different set of RAN components under the test, as well as xApps/rApps. Here proper vendor solutions should be considered.

- **O-RAN RICs** are the second crucial part of the E2E EE testing framework, as they host an xApps/rApps analyzing the RAN data and controlling actions to increase EE. The O-RAN RICs in Fig. 1 represent the**:**
    - **O-RAN interfaces** connecting RICs with O-RAN components. The most important are: E2 interface to connect E2 Nodes with Near-RT RIC, A1 interface to connect Non-RT RIC and Near-RT RIC, O1 interface to connect SMO with E2 Nodes, O2 that connects SMO with O-Cloud and O-FH interface to connect O-RUs with E2 Nodes, or SMO.
    - **SMO** which hosts the Non-RT RIC and provides termination of O1 interface.
    - **Non-RT RIC/Near-RT RIC** hosts the rApps/xApps. It can also be the unit under test to evaluate its latency, and PC related to the processing of RIC messages, e.g., indications, subscriptions.
    - **xApps/rApps** receive data from Non-RT/Near-RT RICs, process them and perform control actions aimed at improvement of EE/ES. They do this either directly through for example cell shutdown, or RF channel reconfiguration, or indirectly by for example performing load balancing. The power consumption of xApps/rApps should be measured along with provided gains, e.g., extensive utilization of CPU/GPU must be taken into account to assess E2E EE.

As SMO, RICs and xApps/rApps are VNF for monitoring of their PC, Kepler should be used.

- **Simulation scenarios** are a necessary part of the E2E EE testing framework. In many cases, especially while evaluating RICs, xApps, and rApps, it is required to test measures like cell on/off switching or load balancing in multiple traffic scenarios, user distributions and motion patterns in a large-scale network. This is difficult to achieve in a lab environment; thus an EE testing framework must support options of simulating the RAN and UEs. Here, to create a unified scenario, 3GPP models (see Sec. 3.2.3) can be useful, as well as the cooperation with MNO to obtain traffic profiles based on the real-network data.
- **KPIs** are crucial for the evaluation of RAN components possibly controlled by the RICs in terms of EE/EC. The 3GPP specification that follows the ETSI

document defines EE as data volume divided by the EC (see Sec. 3.2.1). Such metrics are also mentioned in O-RAN ALLIANCE documents (see Sec. 3.4.2). They will be captured in the framework. However, following the recent 3GPP reports, we suggest that new KPIs should be considered in an E2E EE testing framework that take the QoS of individual UEs into account. Also, PC should be measured for both PNFs and VNFs as suggested by NGNM, 3GPP and O-RAN ALLIANCE. We propose that the E2E EE testing framework should capture the load and state of RAN components, so that it is able to produce reliable power consumption models of O-RU, O-CU, O-DU, e.g., based on the 3GPP model from Sec. 3.2.3.

- **A test/measurement system** is necessary to put all the components together. It must be developed to enable automated configuration of the network components and simulation setups, depending on the purpose of tests and measured components. It is also responsible for running the baseline algorithms under a configured test-case. Finally, it captures the KPIs and produces the test results.

- **Procedures** are an important part of the E2E EE testing framework. They must define how the measurements are to be prepared, conducted, and reported. Good starting points are the ETSI and O-RAN ALLIANCE documents. However, the definition of procedures for the proposed framework is a separate field of future research. Different procedures are required for:

    - **EE evaluation of single RAN components –** This inlcudes questions such as: how to perform EE/EC measurement of O-RU, O-DU, O-CU, what traffic profiles to use, how to define baseline configurations, and replace individual components in a unified way, and - if xApps/rApps are used - what are the baseline xApps/rApps? Testing of RAN components can be done in a lab environment with only one gNB.

    - **EE evaluation of xApps/rApps components –** one approach could be to create a unified procedure independent of the use cases, while the use cases should be used as examples and extensions to the procedure for the xApps/rApps being evaluated in a per use-case manner. There should be a common set of scenarios, for example urban, rural, eMBB, URLLC etc, to compare xApps. This requires baseline xApps/rApps to compare results against. Testing of most of xApps requires a large network.

    - **Providing measurement reports –** can be based on the ETSI and O-RAN ALLIANCE testing reports. Must identify the unit under the test, scenario, baseline algorithms, set of utilized RAN components

- **Extraction of models –** procedures must be defined on how to process the results in order to formulate specific models, e.g., what kind of measurements should be conducted to obtain power consumption model of O-RU, and how the model is parametrized

4.2 Different Views on E2E EE Testing

The proposed high-level E2E EE testing framework captures the network components that can be used for EE evaluation. However, there are multiple levels at which EE can be measured (e.g., RAN component, network-wide) and multiple units can be under test, e.g., RIC, xApps/rApps, O-RU, O-CU, or O-DU. Depending on the unit under test and the measurement scope, different components of the high-level E2E EE testing framework can be selected and put together to fit the purpose, e.g., one can evaluate the PC of an individual RAN component like an O-RU, while another might aim at the comparison between the performance of the RIC and the xApps/rApps deployed therein. In this section, we present four views on E2E EE testing to cover different angles. Again, each of the views presented is a starting point for further definitions in the next phases of the project.

4.2.1 Component-level Measurements

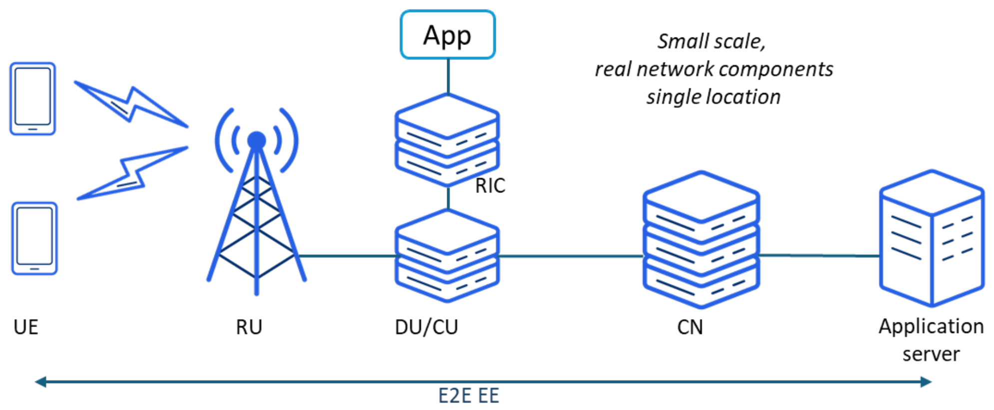

Fig. 2 Component-level E2E EE testing

The first view is the Component-level E2E EE testing as depicted in Fig. 2. This kind of setup aims to evaluate an EE/EC of a single stack of O-RAN components, i.e., O-RU, O-DU, and O-CU, inspired by the ETSI setup for BS Power Consumption measurements. We expect that this deployment utilizes separate entities: O-RU, O-CU, O-DU connected to the core network (possibly emulated). The UEs might be real or emulated. Such a setup can be used for:

- **Evaluation of the EE/EC of a single O-RAN component in a default configuration** (without xApps/rApps), i.e., different vendors' equipment can be mixed together and compared against each other under unified traffic conditions. In addition, such an evaluation can produce a PC model of a certain O-RU/O-CU/O-DU as a function of load utilized, bandwidth, number of active TRX chains, and transmit power.

- **Evaluation of xApps/rApps against baseline approach** for NES features applicable to the single O-RU+O-DU+O-CU stack, things like ASM or RF Channel Reconfiguration can be tested under various vendors' hardwares and under unified traffic profiles and scenarios, and compared against the state-of-the-art baseline algorithms. Here power consumption should also be measured for RIC and xApps/rApps to see if their internal computations are too energy consuming.

- **Evaluation of RICs against baseline approach** for NES features applicable to the single O-RU+O-DU+O-CU stack, e.g., for the fixed set of baseline O-RU+O-DU+O-CU and xApps/rApps providing NES features one can replace the RICs to test their internal power consumption associated with O-RAN interfaces and xApps/rApps deployments.

4.2.2 Global/feature-based testing of EE

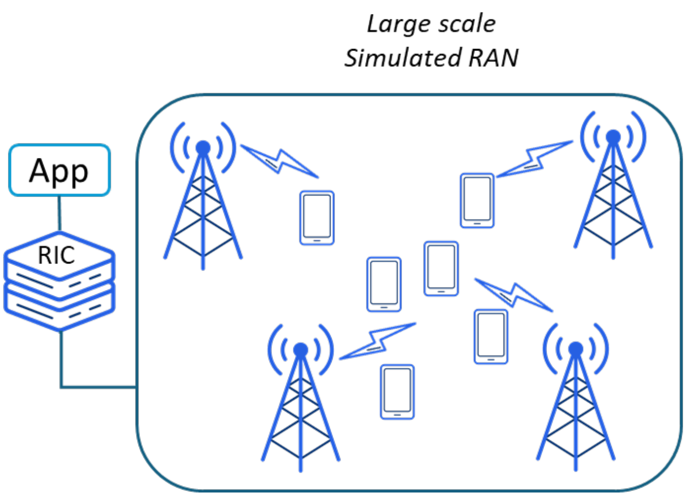

Fig. 3 Global/feature-based E2E EE testing

While the component-level E2E EE testing allows for the measurement of EE/EC of single O-RAN components, its application for the evaluation of the RIC, xApps, and rApps is limited to only a few use cases focused on the NES features, like ASM, RF Channel reconfiguration in a single gNB. Other use cases like Traffic Steering, or Cell On/Off Switching require large-scale networks to be properly evaluated. As building a large enough network in the lab environment is hard for this purpose, we propose a global/feature-based view on the E2E EE testing, as depicted in Fig. 3. This view on E2E EE testing assumes a deployed RIC, together with xApps/rApps and a large-scale emulated network, e.g., hundreds of cells, under different scenarios and traffic profiles (rural, urban, dense urban). For this kind of E2E EE evaluation, the unified models and definitions of scenarios are crucial. Moreover, the results of component-level measurements can be used here (PC models of O-RAN components provided by particular vendors), as well as the traffic models obtained based on the MNOs data. This view on E2E EE testing can be used especially for:

- **Evaluation of xApps/rApps against baseline approach algorithms** using the emulated realistic large-scale network under unified scenarios, cell deployments, and traffic distributions. The EE/EC of the emulated cells should be monitored as well as the EE/EC of the RIC and xApps/rApps themselves, e.g., some algorithms might require significant computational resources, e.g., to run large ML models or optimization tools. This approach allows xApp/rApp vendors to compare their solutions under common conditions to make results more reliable.

- **Evaluation of RICs against a baseline set of xApps/rApps** in the emulated realistic large-scale network under unified scenarios, cell deployments, and traffic distributions. The EE/EC of various vendor RICs can be compared under various signaling overhead on O-RAN interfaces.

- **Evaluation of the possible impact of conflicts** generated by the baseline set of xApps/rApps on the EE/EC achieved by the xApp/rApp under the test.

4.2.3 Deployment view for EE testing

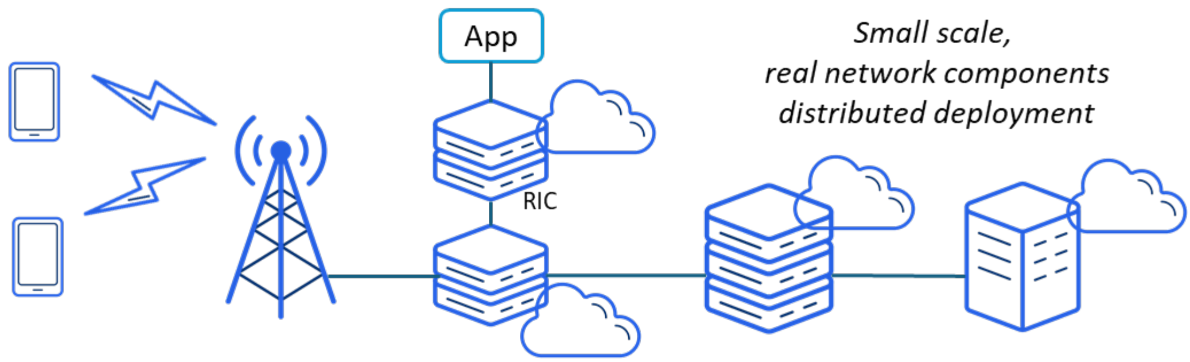

Fig. 4 Deployment view for E2E EE testing

The component-level E2E EE testing assumed that all components except RIC are in the same location. Another point of view on E2E EE testing is the possibility of moving some of the network functions to the cloud or even various cloud locations. This is the deployment view on EE testing depicted in Fig. 4. It is an extension to the component-level E2E EE testing where some, or most of the components are put into the cloud in various PoPs, e.g., virtual O-CU, O-DU, RIC as defined by the O-RAN WG6 [O-RAN.WG6.CADS-v07.00]. The evaluation cases are the same as for the component-level view:

- Evaluation of the EE/EC of a single O-RAN component in a default configuration.
- Evaluation of xApps/rApps against the baseline approach.
- Evaluation of RICs against the baseline approach.

However, measurement techniques would be different, probably mostly relying on the Kepler software. The variety of possible variations of PNF+VNF together with putting VNF in an edge-cloud or global cloud makes the number of possible evaluation scenarios large.

4.2.4 Multi-link/wide network EE testing

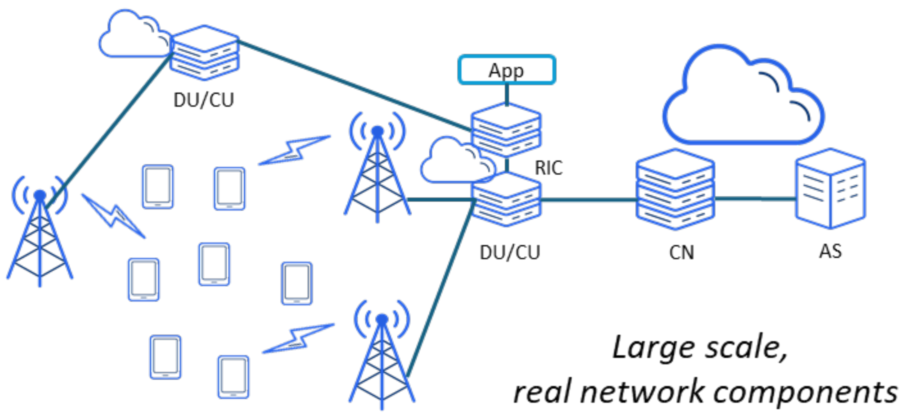

Fig. 5 Multi-link/wide network E2E EE testing

The last view on E2E EE testing is the combination of all previous views and constitutes a target evaluation environment together with connected RIC and xApps/rApps. This is the multi-link/wide network E2E EE testing depicted in Fig. 5. It is a large-scale network build of the real components, e.g., tens of hundreds of gNBs following potential cloud deployments (edge cloud, centralized cloud), and different PNF/VNF deployments, i.e. O-RU and integrated CU+DU, single gNB, separate O-RU + O-DU + O-CU. The network is equipped with a RIC (Non-RT RIC and Near-RT RIC) placed in the cloud, where the xApps and rApps are deployed. This could be a large test network or a part of the commercial network. This view allows mainly for conducting large-scale tests of xApps/rApps in a continuation of the previous tests in a real environment:

- Evaluation of xApps/rApps against baseline algorithms.
- Evaluation of RICs against a baseline set of xApps/rApps.
- Evaluation of the possible impact of conflicts.

Under such conditions, new challenges might impact the results, e.g., the delay between data reports and their processing by xApps/rApps deployed in the RIC which is placed in the centralized cloud. Some decisions might be outdated, e.g. ASM. However, new opportunities occur such as data capture to improve global/feature-based view on E2E EE testing.

4.2.5 Relations between component level and global E2E EE testing

While there are multiple views on the E2E EE testing, they might be within one framework, and they can be combined to constantly improve each other. Fig. 6 presents the information exchange between the component level (see section 4.2.1) and global view (see section 4.2.2) on the E2E testing of EE. Here, while performing the component-level measurements of EE/EC, the measurement data can be collected to formulate a parametrized PC model of a particular RAN component, e.g., RU/CU/DU. This model can be then transferred to the simulation environment utilized for the global/feature-based testing of EE to model this component. Moreover, the obtained model can also be used again for component-level measurements but in a different context, e.g., one can measure O-RU with an emulated O-DU and obtain a PC model. Then the O-DU can be measured to obtain its PC model under emulated O-RU and use both models to combine the results to obtain E2E EE/EE.

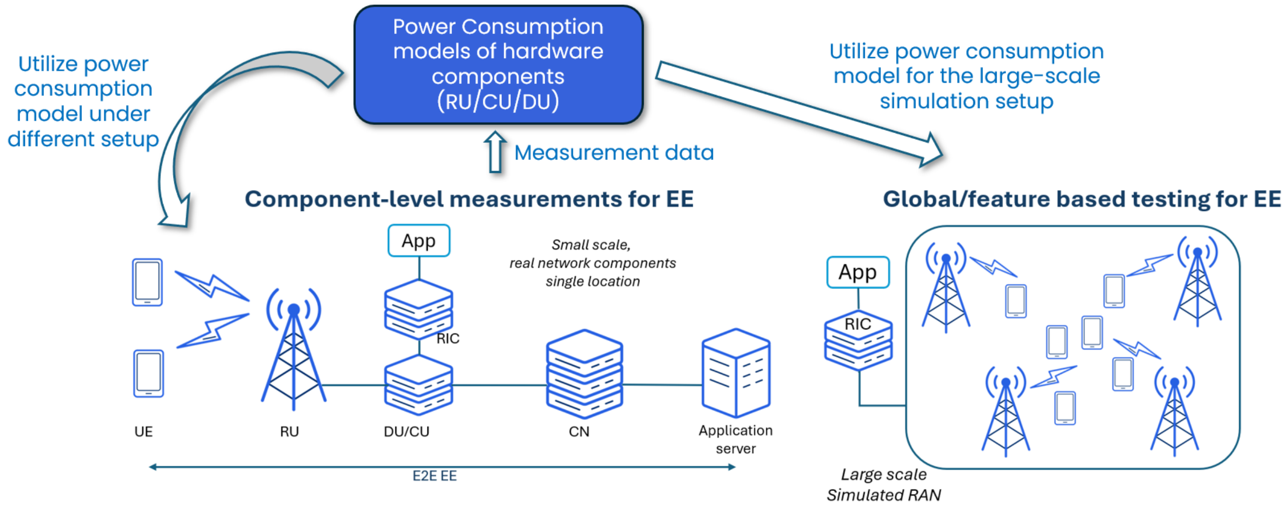

Fig. 6 Information exchange between component-level and global view on E2E testing of EE

The procedure can be summarized as follows:

Test of a network E2E full-stack, with a single gNB (left side of Fig. 6):

- Measure PC of various components, e.g., O-CU/O-DU/O-RU/RIC.
- Utilize either real, emulated UE, or dummy load.
- Produce power consumption models for simulations.
- Utilize previously measured data to create PC models of different vendors.

Simulation of a large-scale network (right side of Fig. 6):

- Test EE/ES provided by xApps/rApps.
- Test xApps/rApps signaling overhead.
- Test different RICs, e.g., their computational capability, or delay between RAN and xApps/rApps.

4.3 Possible Setups

All the views on E2E EE testing within the E2E EE testing framework must be built with the proper HW and SW components. For this purpose, it is important to identify the vendors that can potentially participate in building a test bed. The potential building blocks are depicted in Fig. 7. They are HW components, SW components, cloud components, measurement equipment, HW emulators, and SW emulators.

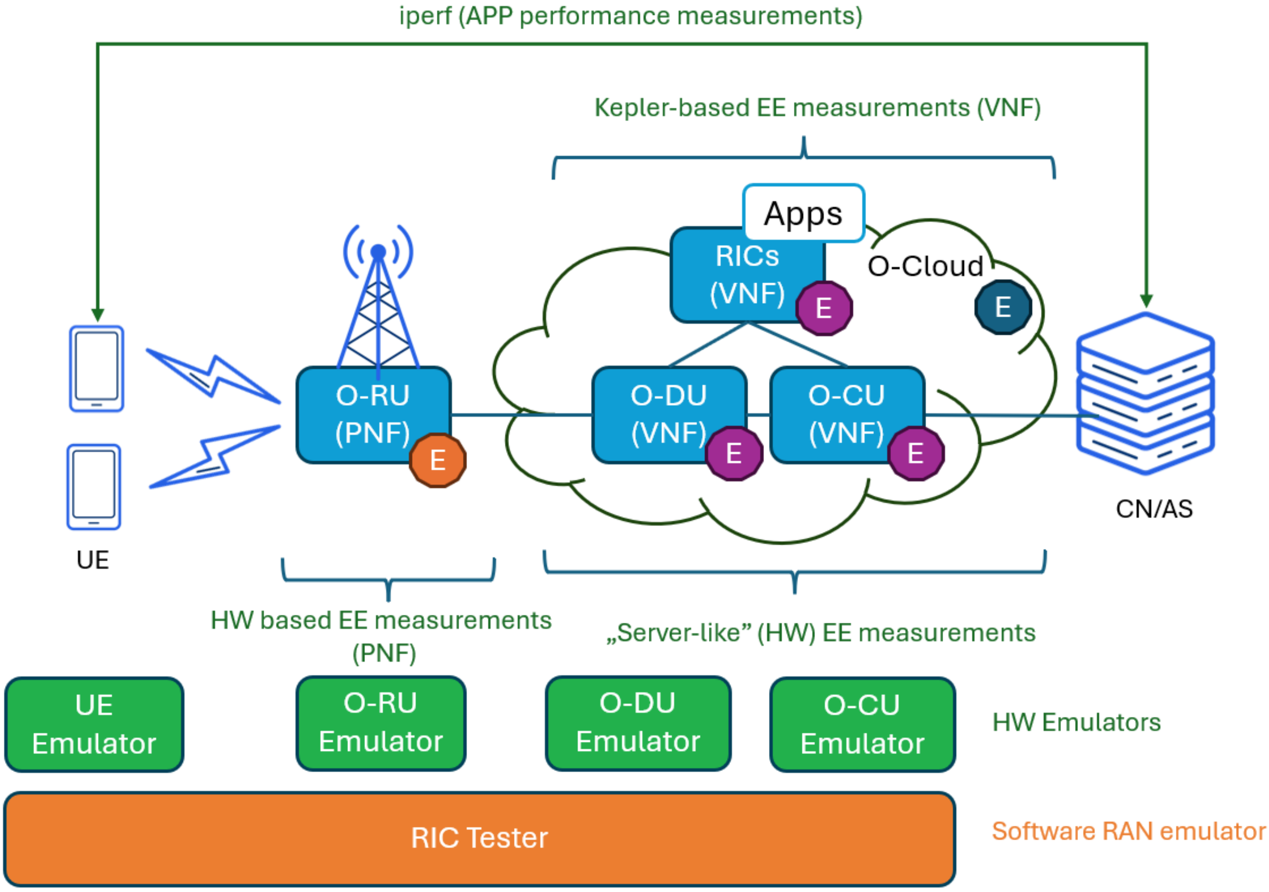

Fig. 7 Building blocks of E2E EE testing framework

The potential vendor and provider breakdown associated with each building block are summarized in Table 3. Note that there is a possibility, that some vendors can provide multiple of those blocks.

Table 3 A potential breakdown of partners associated with building blocks of the E2E EE testing framework

Building block	O-RU	O-DU/ O-CU	RIC	xApps rApps	Measurement tools	Hardware Emulators	RIC Testers
Example partner	O-RU Vendor 1	O-DU/O-CU Vendor 2	Near-RT RIC Vendor 3 Non-RT RIC Vendor 3 or 4	xApp Provider 1 rApp Provider 2	Measurement Vendor 5	O-RU, O-CU, O-DU, UE Emulator Vendor 6	Vendor 7

5 Conclusions

The document outlines a proposal for an E2E EE testing framework for telecommunication networks, highlighting key aspects and challenges.

Different standardization bodies address distinct aspects of energy saving, and there is a need for a unified approach to create a cohesive E2E EE testing framework. The proposal identifies four key views of EE testing:

- The first view involves a realistic small-scale setup with interchangeable components to test the EE performance of individual elements.
- The second focuses on large-scale simulation or emulation to evaluate the performance of energy-saving applications.
- The third view extends to a small-scale setup but incorporates varying cloud deployment options, such as edge or regional locations, to assess EE under diverse scenarios.
- The fourth view involves a large-scale realistic setup with different cloud deployment locations to conduct comprehensive EE testing.

Initially, the focus will be on the first two views for practicality. The reason for this is simple feasibility, i.e., a lab setup with a single/few real O-RU, O-DU, and O-CU is a typical scenario to analyze and measure which can be fed towards extended scenarios. Similarly, the verification of xApp/rApp from the perspective of scale, performance, and standards compliance is enabled with the use of RIC testers alone, before using them in a full setup with real RAN software/hardware.

There are significant challenges to address. Standardized methodologies for cloud/O-Cloud and VNFs are limited. Additionally, the various deployment configurations for xApp, CU/DU, and RIC components add complexity to the testing process. To streamline efforts, the project should target a single deployment type for logical functionality testing rather than verifying the EE of multiple configurations.

Metrics and benchmarks are essential to the framework. Core KPIs and other measures must be defined to integrate energy-saving goals with performance metrics such as outage probability, QoS fulfillment, throughput, or handover efficiency. Furthermore, the framework should include standardized benchmarks for energy savings, tailored to specific use cases like ultra-dense mobile broadband traffic, IoT deployments, and AI testing.

The framework's design and execution will involve developing a standardized core testing procedure adaptable to various scenarios. It is crucial to avoid excessive configurability of modules to prevent inconsistencies in results. Tests should operate in well-defined setups to ensure the reliability and comparability of outcomes.

As the next steps, the framework will require detailed specifications for its modules, components, and functionalities. Dependencies, such as traffic steering, must be integrated and clearly defined. Overall, the proposal emphasizes a balanced approach between standardization and adaptability, aiming to create a practical, scalable, and reliable framework for advancing energy efficiency testing in telecommunications networks. Most importantly, the aim of the white paper is not to give a direct answer on how the EE should be tested and measured in O-RAN networks but to start a wide discussion between the network components vendors, standardization bodies, and research units. We expect that this material and proposed framework can be extended and adjusted by the joint efforts to make a unified environment to evaluate EE of both O-RAN network components and various ES RIC applications.

6 References

[ETSI_103786_TS] ETSI TS 103 786, “Measurement method for energy efficiency of wireless access network equipment; Dynamic energy efficiency measurement method of 5G Base Station (BS)”, v1.2.1

[ETSI_2027061_TS] ETSI ES 202 706-1, “Metrics and measurement method for energy efficiency of wireless access network equipment; Part 1: Power consumption - static measurement method”, v1.7.1

[ETSI_1027062_TS] ETSI ES 202 706-1, “Metrics and measurement method for energy efficiency of wireless access network equipment; Part 2: Energy Efficiency - dynamic measurement method”, v1.6.1

[ETSI_138 141-1_TS] ETSI TS 138 141-1, “"5G; NR; Base Station (BS) conformance testing Part 1: Conducted conformance testing (3GPP TS 38.141-1)"

[ETSI_138 141-2_TS] ETSI TS 138 141-2, “5G; NR; Base Station (BS) conformance testing Part 2: Radiated conformance testing (3GPP TS 38.141-2)”

[intel_es_wp] Intel, Mavenir, “A Holistic Study of Power Consumption and Energy Savings Strategies for Open vRAN Systems”, White Paper, Feb. 2023

[IEC_62018] IEC 62018, “Power consumption of information technology equipment - Measurement methods”, 2003

[keysight_ep_test] Keysight, “Energy Plane (E-Plane) Test Suites for RAN / vRAN / O-RAN”, White Paper, Jun. 2023

[NGNM_GFN] NGMN Alliance, “GREEN FUTURE NETWORKS: METERING IN VIRTUALISED RAN INFRASTRUCTURE”, v1.0

[ORAN_NES_TR] O-RAN ALLIANCE, WG1, “Network Energy Saving Use Cases Technical Report”, v02.00

[ORAN_TIFG_E2E] O-RAN ALLIANCE, TIFG, “End-to-end Test Specification”, v06.00

[ORAN_TIFG_TFWK] O-RAN ALLIANCE, TIFG, “O-RAN End-to-End System Testing Framework Specification 1.0”, v01.00

[ORAN_WG6_ES]	O-RAN ALLIANCE, WG6, “Study on O-Cloud Energy Savings”, v01.00
[ORAN_WG7_NES]	O-RAN ALLIANCE, WG7, “Network Energy Savings Procedures, and Performance Metrics”, v01.00
[TR32.826]	3GPP TR 32.826, “Study on Energy Savings Management (ESM)”, v10.0.0
[TR38.864]	3GPP TR 38.864, “Study on network energy savings for NR”, v18.1.0
[TS28.310]	3GPP TS 28.310, “Management and orchestration; Energy efficiency of 5G”, v18.3.0
[TS28.552]	3GPP TS 28.552, “Management and orchestration; 5G performance measurements”, v18.4.0
[TS28.554]	3GPP TS 28.554, “Management and orchestration; 5G end to end Key Performance Indicators (KPI)”, v18.3.1
[TR23.700]	3GPP TR 23.700-66, “Study on Energy Efficiency and Energy Saving”, v1.0.0
[TR28.880]	3GPP TR 28.880, “Study on energy efficiency and energy saving aspects of 5G networks and services”, v0.3.0
[TR26.942]	3GPP TR 26.942, “Study on Media Energy Consumption Exposure and Evaluation Framework”, v0.3.0
[TR32.927]	3GPP TR 32.927, “Study on system and functional aspects of energy efficiency in 5G networks”, v18.0.0
[TR38.901]	3GPP TR 38.901, “Study on channel model for frequencies from 0.5 to 100 GHz”, v18.0.0
[TR38.840]	3GPP TR 38.840, “Study on User Equipment (UE) power saving in NR”, v16.0.0

About i14y Lab

The i14y Lab is an open lab for interoperability testing of disaggregated telco systems, such as Open RAN, led by Deutsche Telekom together with project partners BISDN, Capgemini engineering, EANTC, Fraunhofer HHI, highstreet technologies, NOKIA, Rohde & Schwarz, Telefonica, TU Berlin, Vodafone, and supported with public funding from the German Ministry of Digital and Transport (BMDV).

The i14y Lab provides infrastructure for integration tests with the aim to evaluate market readiness and accelerate production readiness of multi-vendor disaggregated telco solutions. By creating and providing a vendor-independent environment, we promote the development of an innovative, open, and interoperable telco ecosystem. For more information, go to www.i14y-lab.com.

Funded by:

Funded by: Federal Ministry for Digital and Transport on the basis of a decision by the German Bundestag

Consortium Partners

BISDN · Capgemini engineering · T · EANTC · Fraunhofer HHI · highstreet technologies Network solutions · NOKIA · ROHDE & SCHWARZ · Telefónica · TU berlin · vodafone

Associated Partners

Red Hat · TIMEDO LABS · VIAVI Solutions

Supporting Partners

ONAP · ONF Open Networking Foundation · OPEN Compute Project · O-RAN ALLIANCE · TELECOM INFRA PROJECT

Contact

info@i14y-lab.com

www.i14y-lab.com