Toward Energy Efficiency in O-RAN: A Digital Twin for Carbon-Aware 6G Networks

Elif Ak, *Member, IEEE,* Berk Canberk, *Senior Member, IEEE,* Hyundong Shin, *Fellow, IEEE,* Trung Q. Duong, *Fellow, IEEE*

***Abstract*—This paper provides a comprehensive analysis of energy efficiency and carbon awareness in emerging sixth generation of cellular networks (6G) open radio access network (O-RAN) systems. Guided by the proposed 6S lifecycle model (sense, shrink, sleep, shift, steer, and show), the paper identifies key energy efficiency drivers such as artificial intelligence (AI)-enabled control, computational optimization, digital twin, and carbon-aware integration. A central contribution is the in-depth examination of metering methodologies and the standardization gaps stemming from O-RAN's disaggregated, virtualized design, along with a reassessment of its main energy consumption sources. This paper also consolidates industrial and standardization efforts by presenting essential key performance indicators and measurement approaches for transparent, accurate, and trustworthy energy and carbon accounting.**

***Index Terms*—Carbon awareness, digital twin, energy efficiency, open ran, o-ran, net-zero, sustainability, 5G, 6G.**

I. INTRODUCTION

ENERGY consumption in communication networks has become a major sustainability concern, with the information and communication technology (ICT) sector accounting for roughly 2–2.8% of global greenhouse gas emissions, a share expected to rise significantly by 2030 [1]. In mobile networks, the radio access network (RAN) remains the most energy-intensive component, responsible for about 80% of total network consumption [2]. As a result, energy efficiency (EE) has emerged as a core design priority in the evolution toward 6G, driven by the growing computational demands of next-generation services and the corresponding increase in operational costs and environmental impact.

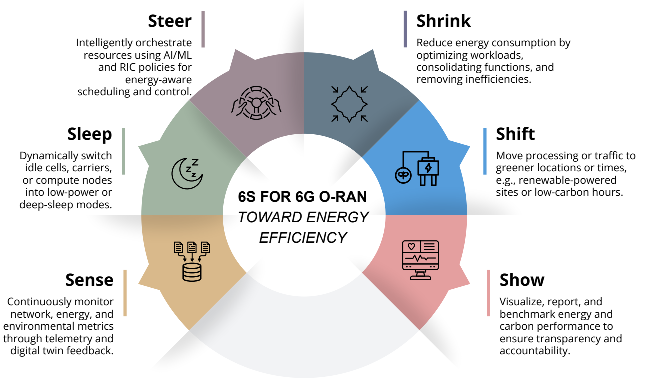

Fig. 1. The 6S model for life cycle of carbon-aware and energy-efficient 6G O-RAN

Encouragingly, progress toward decarbonization is visible: between 2019 and 2023, mobile operators reduced operational emissions by 8%, even as mobile connections grew 9% and data traffic nearly quadrupled [3]. Over the same period, global CO_2 emissions rose 3%, indicating that the mobile communications sector is outperforming most industries in decoupling growth from emissions, an achievement largely attributed to continuous improvements in EE and the increasing use of renewable energy sources [3].

However, sustaining this momentum toward net-zero and carbon-aware network operations demands a fundamental rethinking of how communication infrastructures are designed, deployed, and managed. Within this transformation, the open RAN (O-RAN) architecture plays a pivotal role. Positioned as a cornerstone of the next-generation RAN, O-RAN fosters openness, continuous innovation, softwarization, and AI/ML integration to advance sustainability objectives [4]. Yet, unlike traditional monolithic base stations, whose energy metering has been clearly defined by European Telecommunications Standards Institute (ETSI) and 3rd Generation Partnership Project (3GPP) standards [1], [5], O-RAN introduces disaggregated and virtualized components deployed over commercial off-the-shelf (COTS) hardware. This architectural shift creates a metering standards gap, as energy consumption must now be measured, attributed, and optimized across distributed software and hardware layers. Additionally, integration with external

E. Ak is with the Faculty of Engineering and Applied Science, Memorial University, St. John's, NL A1C 5S7, Canada. (e-mail: elif.ak@mun.ca).

B. Canberk is with the School of Engineering and Built Environment, Edinburgh Napier University, Edinburgh EH10 5DT, UK, (e-mail: b.canberk@napier.ac.uk).

H. Shin is with the Department of Electronics and Information Convergence Engineering, Kyung Hee University, 1732 Deogyeong-daero, Giheung-gu, Yongin-si, Gyeonggi-do 17104, Republic of Korea (e-mail: hshin@khu.ac.kr).

T. Q. Duong is with the Faculty of Engineering and Applied Science, Memorial University, St. John's, NL A1C 5S7, Canada, and with the School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Belfast, U.K, and also with the Department of Electronic Engineering, Kyung Hee University, Yongin-si, Gyeonggi-do 17104, South Korea (e-mail: tduong@mun.ca).

This work was supported in part by the Canada Excellence Research Chair (CERC) Program CERC-2022-00109 and in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant Program RGPIN-2025-04941. The work of B. Canberk is partially supported by The Scientific and Technological Research Council of Türkiye (TÜBİTAK) 1515 Frontier R&D Laboratories Support Program for Türk Telekom 6G R&D Lab under project number 5249902. The work of H. Shin was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (RS-2025-00556064), and by the Ministry of Science and ICT (MSIT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2025-RS-2021-II212046), supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation). The work of E. Ak is partially supported by TÜBİTAK 1515 Frontier R&D Laboratories Support Program for BTS Advanced AI Hub: BTS Autonomous Networks and Data Innovation Lab project number 5239903.

Corresponding authors: Trung Q. Duong and Hyundong Shin.

resources, such as renewable energy sources, policies, and other RANs, requires an overlay system that coordinates these elements while allowing testing and validation of EE and carbon-aware actions without disrupting live networks. Digital twins (DTs) could address this need if designed carefully. DTs have matured in recent years, and their application to RAN systems shows promise. Herein, a DT creates a virtual replica of the physical RAN. For energy-efficient O-RAN, DTs could enable operators to test energy-saving strategies under different scenarios, evaluate trade-offs between energy savings and quality of service (QoS), and generate insights without affecting production networks.

Accordingly, this paper explores the intersection of EE and carbon awareness in the evolving 6G O-RAN ecosystem. We discuss recent approaches by the O-RAN Sustainability Focus Group (SuFG) [6] and connect these approaches to drivers of EE and carbon awareness, highlighting the role of the DT. To date, efforts toward energy-efficient and sustainable RANs have remained fragmented. Standardization bodies such as ETSI and 3GPP define energy metrics and measurement procedures [1], [5], the O-RAN Alliance specifies component-level energy saving mechanisms [6], [7], and academic contributions tend to optimize individual dimensions, e.g., AI-driven sleep modes, digital twin simulation, or renewable energy integration [8]–[10]. Although these works are highly valuable, what is missing is a *unifying operational lifecycle* that systematically links capabilities from monitoring through action to accountability, while being grounded in O-RAN's disaggregated design. Here, we refer to it as the 6S for 6G O-RAN, as shown in Fig. 1, which provides a structured roadmap for EE and long-term sustainability. The 6S model presents the lifecycle of energy- and carbon-aware management in O-RAN, from continuous telemetry and DT-RAN feedback (*sense*) through active resource optimization (*shrink, sleep, steer*), to carbon-aware workload migration (*shift*), and finally to transparent reporting and benchmarking (*show*). Since it is particularly grounded in O-RAN, each phase maps directly onto O-RAN's disaggregated entities and interfaces. For example, *steer* leverages the hierarchical RIC framework through xApps and rApps, *shift* exploits the DT-RAN's integration with external carbon-intensity APIs and the SMO's orchestration capabilities, and *show* builds on standardized O1/O2 telemetry. Specifically, this paper addresses these three questions: (i) What are the principal drivers of EE in 6G O-RAN, and how do they interact synergistically? (ii) What metering standards and methodologies are missing due to O-RAN's disaggregated, virtualized architecture, and how can hardware and software metering be reconciled? (iii) How can a DT-RAN serve as the integrating layer that connects O-RAN components, green energy sources, carbon-aware services, and other network operators throughout and what KPIs for transparent and trustworthy accounting? Building on this, Section II outlines the main EE drivers. Section III details the digital twin RAN (DT-RAN) and its integration with O-RAN entities, drivers and external services through interfaces, and then discusses energy-aware service delivery, validation and trustworthiness. Finally, Section IV summarizes key insights and envisions a path toward carbon awareness.

II. ENERGY EFFICIENCY DRIVERS FOR 6G O-RAN

Current standards and technical reports on EE [3], [5], [6], combined with recent literature and promising experimental results, have highlighted several interrelated drivers that collectively aim to minimize network energy consumption in 6G O-RANs. This section will discuss these drivers individually, though their full synergistic potential will be explored in subsequent sections.

a) Artificial Intelligence and Machine Learning: A primary driver for O-RAN EE is the integration of artificial intelligence (AI) and machine learning (ML) capabilities, realized through the RAN Intelligent Controllers (RICs). O-RAN uses a hierarchical RIC framework as non-real-time (non-RT) RIC and near-real-time (near-RT) RIC to embed intelligence in network management. AI/ML algorithms deployed as xApps (via near-RT RICs) and rApps (via non-RT RICs) that perform real-time and long-term tasks, respectively. These algorithms enable predictive analytics to forecast network energy consumption and traffic loads, enabling proactive, adaptive resource management rather than the reactive, rule-based configuration typically found in conventional systems [2]. However, in energy-efficient 6G systems, AI/ML functionalities must themselves be designed to minimize energy consumption, ensuring that the vision of "AI-native 6G" does not come at the cost of higher power consumption [4]. This highlights the need for sustainable, green AI [11] as a core aspect of future network design, supported by approaches such as federated learning, incremental learning, synthetic data generation and edge AI [4], [12].

b) Computing and Architectural Efficiency: EE in O-RAN depends not only on how network functions are processed but also on where and how resources are deployed. Virtualization and containerization on general-purpose processors (GPPs) tend to increase energy consumption due to non-optimized processing pipelines [1]. To address this, the use of hardware acceleration through specialized components that offload computationally intensive tasks has been increasing [1], [11]. For example, empirical studies show that field-programmable gate arrays (FPGA) implementations for network functions achieve substantially lower power consumption than central processing unit (CPU)-based designs, giving a balanced trade-off between energy, latency, and cost [1]. Likewise, "AI-native 6G" should also leverage energy-efficient hardware and select the appropriate processing units based on AI/ML requirements, such as application-specific integrated circuits (ASICs), tensor processing units (TPUs), graphics processing unit (GPU) and FPGAs, as these purpose-built devices, when used appropriately, can deliver significantly higher performance per watt than GPPs [11].

Another key driver is architecture-level efficiency, where O-RAN can flexibly manage the level of functional centralization across the disaggregated RAN. In other words, the separation of the open radio unit (O-RU), open distributed unit (O-DU), and open central unit (O-CU) introduces diverse deployment options. Centralized deployments in regional data centers can exploit workload consolidation and virtualization gains, whereas distributed configurations reduce transport overhead by processing closer to the radio edge. The energy-optimal point typically depends on traffic demand, latency constraints, use case and the available compute infrastructure [8].

Finally, emerging physical-layer technologies such as massive multiple-input and multiple-output (mMIMO) and reconfigurable intelligent surfaces (RIS) are expected to further influence the energy profile of 6G O-RAN [4]. RIS, in particular, has shown strong potential for enhancing energy efficiency, with studies indicating network-wide energy savings of up to 3.5 times compared to baseline configurations [13].

c) Sustainability Integration: 6G mandates aligning network operation with environmental goals, making sustainability considerations an intrinsic driver for O-RAN EE [4]. A forward-thinking driver is the integration and enhanced coordination between O-RAN elements (O-DU, O-CU, RICs) and external power sources, specifically recognizing the type (renewable or non-renewable) and available capacity [6], [13]. O-RAN management systems, such as service management and orchestration (SMO) or RICs, are envisioned to be *energy-aware*, coordinating network operations based on the availability and cost of green energy [13]. This *energy-awareness* enables the network to dynamically prioritize energy savings when operating on limited battery power or, conversely, focus on optimizing user experience when renewable sources are abundant. Due to its disaggregated nature, O-RAN architectures natively support the use of renewable power sources (e.g., wind, solar) integrated with smart grids [1]. O-RAN also promotes the use of standardized, general-purpose hardware, facilitating the reuse and upgrading of equipment with software updates, which directly reduces the generation of electronic waste (i.e., *e-waste*) compared to proprietary monolithic solutions [4].

III. SYNERGISTIC INTEGRATION OF ENERGY EFFICIENCY DRIVERS IN 6G O-RAN

While the O-RAN architecture already embeds intelligence through its hierarchical RICs and orchestrates services through the SMO, these components mainly operate in the live network under strict service level agreements, latency limits, and reliability requirements. As a result, their primary role is to maintain stable and continuous network operation. In contrast, as also seen in Fig. 2, the DT-RAN introduces an orthogonal layer of intelligence, that connects with renewable energy sources and external application programming interfaces (APIs) to gather carbon intensity data from regional authorities and grid operators, such as the *UK's National Grid ESO*, *Electricity Maps*, and *WattTime* [9]. It interfaces with O-RAN through standardized protocols and using DT-specific interfaces for testing, validation, and trust management. In the rest of this section, we will first clarify energy consumption sources in O-RAN to better contextualize the DT-RAN's role across its components and its interaction with other EE drivers.

A. Energy Consumption Modeling and Metering

Energy consumption modeling is the process of mathematically or computationally representing how different parts of a system use energy under varying operating conditions [13]. In the context of O-RAN, this, together with metering, means building models that describe how much power each network function (such as O-RU, O-DU, and O-CU) consumes, depending on factors such as traffic load, functional split, hardware configuration, and environmental conditions. These models can be analytical (based on equations linking utilization to *watts*), empirical (fitted from measured data), or AI-based (using ML to predict consumption patterns). All activities related to energy consumption modeling and metering fall under the sensing step of the 6S model.

1) Energy Consumption Sources in O-RAN: The main energy consumption sources in O-RAN are outlined here to clarify what the DT-RAN must model, monitor, and optimize within its framework. For a more detailed analysis of these sources, readers may refer to prior studies [1], [8], [10].

O-RU: The O-RU is among the most energy-intensive elements of the O-RAN architecture, located at the cell site and responsible for radio-frequency (RF) and lower physical-layer (PHY) functions. It connects to the O-DU through the open fronthaul (O-FH) interface and handles tasks such as beamforming, cyclic prefix operations, and RF amplification under strict latency constraints. The O-RU is implemented in hardware, with no support for virtualization or dynamic containerized scaling. Empirical studies show that its power draw scales weakly with traffic load or physical resource block (PRB) utilization [8], as static hardware elements, particularly the power amplifier (PA), RF front-end, and converters, remain active continuously. Industry reports also indicate that the O-RU and its RF chain account for 40–60% of total base-station power, with the PA alone exceeding half of this share [2], making the O-RU a central target for EE and sustainability improvements.

O-DU: The O-DU is a crucial element in the O-RAN functional split, handling computationally intensive tasks of the physical layer (PHY-high), media access control (MAC), and radio link control (RLC) layer functions. In particular, the PHY-high is one of the dominant power consumers, as it performs computationally intensive operations such as forward error correction (FEC), encoding and decoding, modulation and demodulation, and resource mapping. The O-DU's total power draw scales with both the complexity of these sub-functions (e.i., FEC decoder, MAC scheduler), represented by their reference complexity (i.e., how much computational work it performs per unit time or data) and technology-dependent parameters (i.e., how efficiently the chosen hardware platform can carry out that work). Its overall energy consumption also depends on the number of connected O-RUs, which determines the aggregated processing demand. Consequently, the O-DU's energy profile can be modeled as a function of component-level computational complexity and hardware EE parameters [8].

O-CU: The O-CU, a logical node, responsible for higher-layer processing functions such as radio resource control (RRC), packet data convergence protocol (PDCP), and service data adaptation protocol (SDAP) [9], consumes energy mainly through processing and networking operations. Its processing load arises from packet-based computations, virtualization overhead, and platform control tasks such as orchestration,

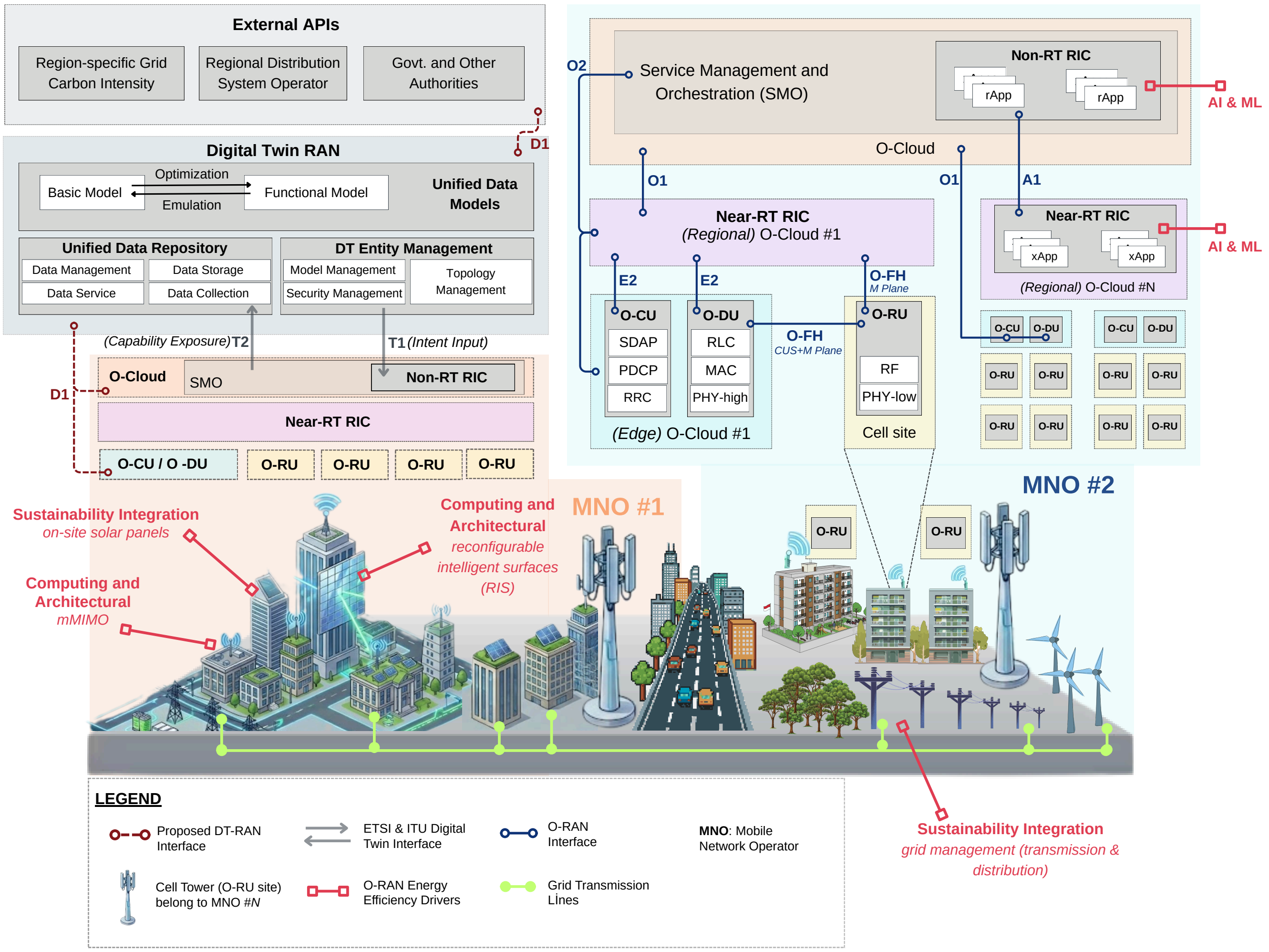

Fig. 2. The Digital Twin for carbon-aware and energy-efficient 6G O-RAN, integrating green energy sources with O-RAN components.

load balancing and resource scheduling.

RICs, SMO, O-Cloud Infrastructure: Beyond the core RAN nodes, energy consumption also arises from O-RAN's control, orchestration, and cloud layers, namely non-RT RIC, near-RT RIC, and SMO. While the RICs are central to O-RAN's intelligence and automation (also supporting EE through *xApps* and *rApps*), they remain computationally intensive due to continuous data collection, analytics, and AI model training and inference for managing O-RAN components [2]. The SMO also adds constant energy overhead since it continuously analyzes metrics and sends control messages across multiple network domains. This includes managing faults, configuration, accounting, performance, and security (FCAPS); collecting telemetry through O1 and O2 interfaces; and orchestrating the lifecycle of virtualized or containerized network functions. The SMO's energy profile is therefore correlated with the number of managed nodes and the intensity of monitoring activities [8].

All these control entities above are hosted within the O-Cloud, potentially with virtualized instances of the O-DU and O-CU, compounding its overall energy footprint. Empirical studies show that virtualized O-RAN functions such as the O-CU and O-DU exhibit limited power scalability since virtual machines (VMs) and containers stay reserved for latency-critical tasks [5], [14]. What DT can provide for EE is strategies such as workload consolidation, selective node deactivation, and adaptive resource scaling through the SMO and non-RT RIC [6], [7]. This coordinated control provided by DT can eventually reduce the persistent energy overheads of O-RAN's cloud and management layers, especially when guided by energy-aware policies and services.

Common Site Infrastructure (CSI): In addition to the compute and virtualization domains, another share of energy usage at O-RAN sites originates from the CSI. This category includes power supply systems, rectifiers, cooling and ventilation units, environmental controls, and backup batteries that support the operation of O-RAN nodes and the O-Cloud. These non-computational elements can account for approximately 20–40% of the total site power in both conventional and O-RAN deployments [2], [13]. And DT-RAN extends these efforts by enabling dynamic fan speed control, and renewable power integration to mitigate CSI energy inefficiencies.

2) O-RAN Energy Metering: Since O-RAN functions are distributed and disaggregated across diverse processing hardware, accurately measuring energy consumption requires multiple complementary approaches. This process relies on

TABLE I
ENERGY CONSUMPTION, DIGITAL TWIN MITIGATION, AND MEASUREMENT APPROACHES ACROSS O-RAN COMPONENTS

O-RAN Component	**Primary Energy Consumption Sources**	**Data Path for EE Telemetry**	**DT-RAN Mitigation Mechanisms**	**Measurement & Tools**
O-RU	Power Amplifier (PA) bias and RF front-end losses; static hardware load (clocking, converters) largely independent of traffic	O1 (via M-plane), E2 (via CUS-Plane)	DT models predict PA utilization and optimize bias voltage, DTX/DRX cycles, and deep-sleep scheduling. Mitigation actions enforced by near-RT RIC (xApps) for short-term control and non-RT RIC (rApps) for policy optimization.	Metered PDUs, GW-Instek GPM-8213 or DC-capable power analyzers like the Keysight N6705C and environmental measurement tools.
O-DU	High-PHY computation (FEC, modulation/demodulation), MAC scheduling, virtualization overhead on COTS servers	E2, O2, D1	DT simulates traffic load, functional split selection, and DU consolidation to forecast energy savings. Insights consumed by near-RT RIC xApps for real-time DU load balancing and SMO for orchestration policies.	RAPL (CPU/DRAM), *powerstat* and *turbostat*, Kepler (per-container energy), Redfish, turbostat, Prometheus metrics via O2, *nvidia-smi* or NVIDIA management library (*NVML*), or *cAdvisor* or *PowerTOP*, if applicable.
O-CU	Packet processing, virtualization baseline, NIC and backhaul I/O power	O1, O2, D1	DT-driven orchestration recommends VM/container consolidation and adaptive CPU/GPU frequency scaling. Policies applied through SMO orchestration and monitored via non-RT RIC for service optimization.	Redfish (server-level), NVML (GPU), RAPL, NIC telemetry (DPDK), Kepler metrics via O2, *cAdvisor* or *PowerTOP*
RICs (Near-RT / Non-RT)	Continuous AI/ML inference and control signaling overhead	A1, E2, D1, O2	DT validates AI/ML-driven xApp/rApp policies offline, reducing retraining frequency and optimizing energy-aware decision-making across layers.	Kepler + NVML for power tracking; Redfish for host monitoring; RIC logs via O2.
Common Site Infrastructure	cooling, PSUs, rectifiers; persistent orchestration load	O2, D1	DT aligns SMO orchestration with renewable availability and carbon-cost signals, recommending carbon-aware workload placement and O-Cloud consolidation.	Redfish/IPMI thermal sensors, PDU logs, smart grid API data, Prometheus exporters.

telemetry and is addressed in this section through *hardware metering* and *software metering*.

Hardware Metering: Hardware metering refers to the physical measurement of energy consumption and related performance indicators directly from the underlying infrastructure, which includes servers, storage, and network components. For O-RAN metering, there are three measurement methods: out-of-band, direct power, and in-band measurements.

- **Out-of-band measurement:** The industry convergence for managing and measuring COTS hardware centers heavily on the Redfish data model [15]. Redfish can collect real-time and cumulative power consumption metrics (in Watts or kWh) at multiple levels: i) server, ii) power supply unit (PSU), or iii) individual components such as CPUs and memory. It also provided detailed thermal data, including temperatures, fan speeds, and cooling status, along with sensor health information like PSU and voltage readings. Redfish runs on the baseboard management controller (BMC), a specialized microcontroller embedded on the motherboard of a server, such as those in an O-Cloud environment that host network functions like the O-CU, O-DU, and RIC. Similarly, another standard, intelligent platform management interface (IPMI) [15], the predecessor technology to Redfish, is used for the management of physical infrastructure, for collecting the physical status of the hardware components, such as temperature, voltage, and fan status [5].
- **Direct power measurement:** Various physical tools and software-based estimation methods can also be used for O-RAN metering, similar to conventional RAN deployments in bare-metal environments, to directly measure power consumption. For example, metered power distribution units (PDUs) can also provide ground-truth measurements of total O-RU power consumption, including RF, PA, and converter losses. Other relevant measurement approaches are listed in Table I for reference.
- **In-band measurement:** For the underlying COTS servers hosting the O-CU, O-DU or RICs, in-built software tools are also used to estimate power draw. The outputs of these tools (listed in Table I) should be integrated into the DT data ingestion pipeline to enable closed-loop analysis.

Software Metering: In virtualized O-RAN systems, software metering estimates function-level energy consumption by linking hardware measurements with resource usage of virtualized workloads. The fundamental challenge of software metering arises because the energy consumption of specific O-Cloud components cannot be measured directly, as it is a subset of the compute resources on a physical node. It must be derived by collecting infrastructure-level data, aggregating it, and aligning it with workloads' resource usage. This requires coordinated data provisioning and utilizing advanced tools that ingest hardware-metered data to build models for energy attribution. O-RAN's cloud-native architecture, often deployed atop Kubernetes clusters, naturally opens opportunities to leverage existing Kubernetes sustainability mechanisms to improve network EE. *Kepler*, which is an open-source tool, integrates kernel-level telemetry and processor performance counters to estimate per-pod and per-container power consumption, and can expose energy metrics for O-DUs, O-CUs, and apps (e.g., xApps, rApps) running in containers across the O-Cloud. When real-time data are unavailable, Kepler also utilizes regression-based models. Similarly, another ready-

TABLE II
O-RAN Energy and Carbon Accounting KPIs

Physical/Hardware KPIs (related to hardware metering)		**Sustainability & Carbon Accounting KPIs**	
Power Consumption	Avg./min./max. power consumed over the period (Watt, W)	**Operational Carbon (direct)**	Direct emissions from owned/controlled sources (fuel, vehicles, generators).
Energy Consumption	Total energy consumed (kWh); e.g., O-RU per carrier or per antenna array	**Operational Carbon (indirect)**	Indirect emissions from purchased electricity, steam, or cooling.
Environmental Temperature	Avg./min./max. temperature during period (°C)	**Renewable Electricity Share/Purchased**	Proportion of electricity sourced from renewables (TWh).
Environmental Voltage	Voltage (V)	**Electricity Use per Connection**	Total electricity use divided by total connections (kWh/connection).
Environmental Current	Current (A)	**Carbon Intensity per Connection**	Avg. GHG emissions per mobile connection (kg CO_2e/connection).
Environmental Humidity	Humidity percentage (0–100)	**Carbon Savings**	Reduction from refurbished/repaired network devices (e.g., UE, servers, GPPs).

Virtualized/Containerized (related to software metering)	
Total Energy Consumption of Container/Pod	Total energy consumption within a container (e.g., using Kepler/Prometheus technologies) (Kilowatt-hours, kWh)
Energy Consumption of Specific Components	Energy consumption attributed to specific components (CPU cores, DRAM, uncore components) used by a container (Kilowatt-hours, kWh)
Energy Efficiency (general)	The ratio between useful output (performance) and energy/power consumption calculated by the SMO. General energy efficiency metric for the O-Cloud nodes (O-CU, O-DU, etc.) (N/A)
Energy Efficiency (data volume)	EE based on PDCP SDU data volume (bit/J)
Energy Efficiency (coverage)	EE based on coverage (network extent, topology-wise) (m^2/J)
Energy Efficiency (latency)	EE based on latency, for example Ultra-Reliable Low Latency Communication (URLLC) slices (1/ms·J)
Energy Efficiency (active UE)	EE of registered subscribers (connected UE) (user/J)

to-use tool, Kube-Green, automatically shuts down idle pods and worker nodes based on policy or time scheduling. When applied to O-RAN, this capability resembles the cell and carrier shutdown or sleep mode features standardized by the O-RAN SuFG [6]. Prometheus and Grafana form the backbone of O-RAN observability, as also noted in multiple O-RAN experimentation platforms [7], [14]. It is also worth noting that Kubernetes-based orchestration is not always applicable to certain O-RAN functions, particularly O-DU and O-RU components, which are sometimes deployed on bare metal, real-time Linux, or lightweight Docker runtimes. In such cases, full Kubernetes orchestration may add unnecessary overhead or complexity; alternative tools are presented in Table I.

B. Energy-Aware Service Delivery

Energy-aware service delivery refers to techniques that improve the EE of individual O-RAN components while also considering carbon-aware strategies, such as integrating renewable energy sources into network operations [13], without degrading the QoS. Guided by the 6S vision, and continuing with *sleeping* and *steering* steps, the non-RT RIC can be designed to predict traffic and mobility patterns, enabling proactive resource adaptation by selectively deactivating cells, carriers, or antenna chains to reduce idle power consumption, while neighboring cells absorb the load and maintain service continuity [2]. Complementing this, near-RT RIC xApps can be developed to dynamically steer transceiver activity and transmission power, leveraging CSI feedback and adaptive power amplifier biasing for energy efficiency gains [6], [13]. Continue with *sleeping* action in near-RT, xApps dynamically adjust the number of active transmit and receive arrays within the O-RU, particularly essential for massive MIMO deployments [6]. This process, also known as RF channel switch off/on, is contingent upon traffic volume and aims to minimize energy consumption while maintaining the requisite QoS. At the infrastructure level, the non-RT RIC and SMO consolidate workloads by migrating O-DU and O-CU functions onto fewer O-Cloud nodes, scaling CPU frequency and leveraging deep low-power states to *shrink* computational demand and suppress idle energy draw. More importantly, DT-RAN enhances this optimization by introducing a *shifting* capability, integrating external factors such as energy cost, grid carbon intensity, and renewable availability to guide workload migration and reduce both operational expenses and emissions. By incorporating awareness of energy sources like solar and wind, along with their temporal variability, DT-RAN enables O-RAN management entities (SMO, RICs, O-CU, O-DU) to dynamically optimize RAN operations, including O-RU power usage and function placement, in alignment with sustainability objectives. Moreover, DT-RAN enables energy-aware service delivery models that adapt to user and operator preferences. For example DT-RAN can prioritize renewable-powered or low-consumption service paths when possible, and scheduling tasks during periods of high renewable availability to reduce grid dependency and operational costs.

C. Testing, Validation and Trustworthiness

Following the 6S model, the *show* step emphasizes transparency, verification, and continuous improvement of EE and carbon-aware strategies across the O-RAN.

In this context, DT-RAN functions as a high-fidelity sandbox for testing and validating energy-saving mechanisms (such as carrier or cell deactivation, RF port sleep, and workload consolidation) without risking service degradation in live networks. By executing what-if analysis, as detailed in [7], DT-RAN evaluates the trade-offs between energy savings, QoS, and coverage, guiding RIC-based decision loops with empirically validated thresholds. The outcomes of these analyses feed into the non-RT RIC, as illustrated in Fig. 2 via T1 and T2 interfaces, enabling it to refine its policy models and provide optimized guidance to the Near-RT RIC via the A1 interface for enforcement on E2 nodes (O-CU/O-DU). A key part of this validation process is accurate metering, calibration and metric compatibility. In O-RAN, energy data is collected from various tools and sources, each with its own KPIs, units, timestamps, and sampling intervals, requiring careful alignment to get meaningful comparisons and complete EE observation.

Standardized KPIs: Various standardization bodies and research efforts [5], [6], [10] have defined metrics to evaluate energy-saving mechanisms, energy efficiency and carbon footprint, as summarized in Table II. The core principle of EE KPIs is quantifying output per unit of energy (e.g., bit/Joule) [2]. ETSI also defines the core metric as mobile network data energy efficiency (EEDV), which is the ratio of delivered data volume to network energy consumption [5]. 3GPP complements this approach with the network energy efficiency metric (EEglobal), summing efficiencies across multiple deployment scenarios and traffic load levels. Moreover, since DT-RAN employs different data models such as YANG, JSON, and XML, these differences must be considered during calibration to maintain consistency and interoperability across measurements [5], [6].

Beyond data consistency, the operational trustworthiness of the DT-RAN depends on fidelity, that is, the degree to which the digital twin accurately and realistically replicates the physical, behavioral, and time-varying characteristics of the real representation. In O-RANs EE operations, this fidelity can be evaluated by comparing the twin's reconstructed values (states and KPIs) against the corresponding measured telemetry for each variable group (e.g., traffic/load, power/energy, thermal metrics, or carbon-intensity inputs). If certain data cannot be properly synchronized with the twin, the fidelity score reflects this mismatch. This enables a graduated trust model for DT-RAN recommendations. When the fidelity remains within operator-defined tolerance limits, the DT-RAN's optimization decisions (such as sleep mode activation, workload migration, or RF deactivations) can be automatically executed through the RICs and SMO. However, if the fidelity falls outside the acceptable range, the DT-RAN switches to an advisory mode. In this case, its recommendations are flagged for human review before being applied. The acceptable tolerance levels depend on the network function type and timing of the decision.

Another challenge related to KPIs, validation and trustworthiness is that the KPIs in Table II can conflict with each other. For example, expanding coverage by activating more cells or increasing transmit power may reduce data energy efficiency if the additional area serves little traffic. The 6S model with DT manages these trade-offs through the steer and show phases. In the steer phase, the operator defines priorities (e.g., carbon reduction, coverage guarantees, or latency limits) as policy inputs to the non-RT RIC. The DT-RAN then evaluates different operating configurations against these multiple objectives and identifies a point that is similar to a Pareto-efficient point, rather than a single optimum.

IV. CONCLUSION

We presented an integrated perspective on energy-efficient and carbon-aware O-RAN drivers, revisiting the main sources of energy consumption in O-RAN and critically examining current mitigation strategies proposed by DT-RAN, along with existing energy measurement tools and KPIs. Guided by the 6S principles (sense, shrink, sleep, shift, steer, and show), we illustrate how DT-RAN can enable a carbon-aware and energy-efficient 6G lifecycle. Our key finding is that reliable hardware metering, complemented by software-based metering, provides the foundation for data to be used in EE operations that DT can process and use to make accurate decisions. Future work should prioritize developing telecom-specific energy measurement tools, as current solutions like Kepler and Kube-Green are designed for general cloud workloads and fail to account for the timing constraints and real-time processing demands of O-RAN. Building on improved measurement capabilities, fine-grained energy attribution methods are needed to accurately map power consumption to individual virtualized network functions across distributed O-Cloud resources, particularly for O-DU and O-RU deployed on bare metal. These attribution methods are then needed to enable carbon-aware workload migration algorithms that can dynamically balance conflicting KPIs. Finally, exploring inter-DT relationships will be critical, as multiple digital twins operating across different network domains or operators must coordinate to optimize energy efficiency at a broader scale without creating conflicting policies or redundant resource allocation.

ACKNOWLEDGEMENTS

This work was supported in part by the Canada Excellence Research Chair (CERC) Program CERC-2022-00109 and in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant Program RGPIN-2025-04941. The work of B. Canberk is partially supported by The Scientific and Technological Research Council of Türkiye (TÜBİTAK) 1515 Frontier R&D Laboratories Support Program for Türk Telekom 6G R&D Lab under project number 5249902. The work of H. Shin was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (RS-2025-00556064), and by the Ministry of Science and ICT (MSIT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2025-RS-2021-II212046), supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation. The work of E. Ak is partially supported by TÜBİTAK 1515 Frontier R&D Laboratories Support Program for BTS Advanced AI Hub: BTS Autonomous Networks and Data Innovation Lab project number 5239903.

REFERENCES

[1] W. B. Abbas, Q. Z. Ahmed, F. A. Khan, N. S. Mian, P. I. Lazaridis, and P. Sureephong, "Designing future wireless networks (fwn)s with net zero (nz) and zero touch (zt) perspective," *IEEE Access*, vol. 11, pp. 83 301–83 321, 2023.

[2] L. Kundu, X. Lin, and R. Gadiyar, "Toward energy efficient ran: From industry standards to trending practice," *IEEE Wireless Commun.*, vol. 32, no. 1, pp. 36–43, 2025.

[3] GSMA, "Mobile Net Zero 2025: State of the Industry on Climate Action," Research Report, Jun. 2025. [Online]. Available: https://view.gsma.com/mobile-net-zero-2025

[4] H. Ahmadi, M. Rahmani, S. B. Chetty, E. E. Tsiropoulou, H. Arslan, M. Debbah, and T. Q. S. Quek, "Toward sustainability in 6G and beyond: Challenges and opportunities of open RAN," *IEEE Commun. Stand. Mag.*, vol. 9, no. 3, pp. 126–135, Sep. 2025.

[5] NGMN Alliance, "Green Future Networks: Metering in Virtualised RAN Infrastructure," Final Deliverable Version 1.0, Apr. 2024, lead Editor: Sabine Demel; Project: Green Future Networks. [Online]. Available: https://www.ngmn.org/wp-content/uploads/GFN_Metering-Virt.-RAN-Infrastructure_V1.0.pdf

[6] O-RAN Alliance Sustainability Focus Group (SuFG), "O-RAN Energy Measurements Analysis Report," Tech. Rep. O-RAN.SuFG.TR.NES-Analysis-R004-v01.01, Apr. 2025. [Online]. Available: https://specifications.o-ran.org/

[7] O-RAN next Generation Research Group (nGRG), "Digital Twin RAN: Key Enablers," O-RAN Alliance, Contributed Research Report RR-2024-09, Oct. 2024, release version (v1). [Online]. Available: https://www.o-ran.org/research-reports/digital-twin-ran-key-enablers

[8] G. Baldini, R. Bolla, R. Bruschi, A. Carrega, F. Davoli, C. Lombardo, and R. Rabbani, "Toward sustainable o-ran deployment: An in-depth analysis of power consumption," *IEEE Trans. Green Commun. Netw.*, vol. 9, no. 2, pp. 429–444, Jun. 2025.

[9] A. Kharbouch, F. H. Aghdam, N. Gholipoor, and M. Rasti, "Digital-twin-6G empowered future smart grid applications," *IEEE Wireless Commun.*, vol. 32, no. 3, pp. 90–97, Jun. 2025.

[10] K. Alam, M. A. Habibi, M. Tammen, D. Krummacker, W. Saad, M. D. Renzo, T. Melodia, X. Costa-Pérez, M. Debbah, A. Dutta, and H. D. Schotten, "A Comprehensive Tutorial and Survey of O-RAN: Exploring Slicing-Aware Architecture, Deployment Options, Use Cases, and Challenges," *IEEE Commun. Surveys Tuts.*, vol. 28, pp. 1637–1678, 2026.

[11] R. Schwartz, J. Dodge, N. A. Smith, and O. Etzioni, "Green ai," *Commun. ACM*, vol. 63, no. 12, p. 54–63, Dec. 2020.

[12] G. M. van de Ven, T. Tuytelaars, and A. S. Tolias, "Three types of incremental learning," *Nature Machine Intelligence*, vol. 4, no. 12, pp. 1185–1197, Dec. 2022.

[13] O-RAN Alliance Sustainability Focus Group (SuFG), "Enhancing Hardware Energy Efficiency of O-RAN Ecosystem," White Paper, Feb. 2025, release version 1. [Online]. Available: https://mediastorage.o-ran.org/ecosystem-resources/O-RAN_WP_HW_Energy_Efficiency_in_O-RAN_Ecosystem_v1.pdf

[14] J. X. Salvat, J. A. Ayala-Romero, L. Zanzi, A. Garcia-Saavedra, and X. Costa-Perez, "Open Radio Access Networks (O-RAN) Experimentation Platform: Design and Datasets," *IEEE Commun. Mag.*, vol. 61, no. 9, pp. 138–144, Sept. 2023.

[15] G. Goncalves, D. Rosendo, L. Ferreira, G. L. Santos, D. Gomes, A. Moreira, J. Kelner, D. Sadok, M. Wildeman, and P. T. Endo, "A standard to rule them all: Redfish," *IEEE Commun. Stand. Mag.*, vol. 3, no. 2, pp. 36–43, Jun. 2019.

Elif Ak is with Memorial University, Canada. She was recognized by the IEEE ComSoc Communications Software Technical Committee and Istanbul Technical University for her Ph.D. thesis, "Network Digital Twins: Tackling Challenges and Enhancing Wireless Network Management."

Berk Canberk is a Professor in the School of Computing, Engineering and The Built Environment at Edinburgh Napier University, United Kingdom.

Hyundong Shin is a Professor at the Department of Electronic Engineering, Kyung Hee University, South Korea.

Trung Q. Duong is a Canada Excellence Research Chair (CERC) and Full Professor at Memorial University, Canada. He is also an adjunct professor at Queen's University Belfast, UK, a visiting professor at Kyung Hee University, South Korea and Edinburgh Napier University, UK.