Energy Efficiency of Open Radio Access Network: A Survey

Attai Ibrahim Abubakar[1], Oluwakayode Onireti[1], Yusuf Sambo[1], Lei Zhang[1],
G. K. Ragesh[2], Muhammad Ali Imran[1]
[1]James Watt School of Engineering, Unversity of Glasgow, Glasgow, UK
[2] Indian Institute of Information Technology, Kottayam, India
Corresponding Email: Attai.Abubakar@glasgow.ac.uk

***Abstract*—The Open Radio Access Network (O-RAN) architecture has been identified as a promising technology for enhanced network deployment, innovation, improved competition, and reduction of capital and operating expenses (CAPEX/OPEX) of 5G and beyond networks because of its open interfaces, disaggregated network entities and functions, virtualization of network hardware and software, and intelligent control. However, the effect of this improved technology on the energy consumption of the RAN needs to be carefully investigated, so that the many advantages that can be obtained from the O-RAN are not overwhelmed by increased energy consumption. Hence, in this paper, we investigate the O-RAN from an Energy efficiency (EE) perspective by reviewing the state-of-the-art power consumption models, and EE techniques that have been proposed to minimize the energy consumption of O-RAN. In addition, the challenges associated with the optimization of the EE of O-RAN and opportunities for further research are highlighted.**

I. Introduction

The radio access network (RAN), which is the part of the network that connects users to the core network, has undergone a series of evolution and improvements since the advent of second-generation networks [1]. The first deployment strategy of the RAN followed a distributed architecture where each base station site hosts its baseband unit (BBU) and remote radio head (RRH) until the centralized/cloud-RAN (C-RAN) architecture was introduced which pooled the BBU in a central location and enable connections of many RRHs to a single BBU pool. The C-RAN architecture was a major improvement on the distributed RAN as it enabled efficient hardware utilization, enhanced EE, reduced capital and operating expenditure (CAPEX/OPEX), more efficient resource allocation, and coordination of network operations. However, the C-RAN architecture still utilizes propriety hardware and software, employs closed interfaces, and the centralization of BBU processing results in increased capacity overhead and stringent latency requirements for the fronthaul links [2], [3]. In order to overcome the challenges associated with C-RAN, there was a need for further improvement to made to the RAN architecture, which led to the introduction of the open-RAN (O-RAN).

The O-RAN architecture comprises the disaggregation of RAN in three network entities [4]: Centralised Unit (CU), Distributed Unit (DU), and Radio Unit (RU). It also involves the use of open radio interfaces whereby both hardware and software from different equipment vendors can communicate together. Another feature of O-RAN is the virtualization of network functions such that network functions can be abstracted from dedicated hardware and installed on commercial-off-the-shelf (COTS) hardware comprising general purpose processors (GPPs) using virtual machines (VM). In addition, there is the introduction of RAN automation using intelligent controllers so that native support is provided for the application of artificial intelligence and machine learning for network management and optimization. On one hand, the move from proprietary to general and open hardware and software results in cost savings and broadens the supply chain of equipment vendors. On the other hand, virtualization makes it possible for network functions and resources to be performed and allocated to different parts of the networks in a dynamic matter thus making the RAN-as-a-service rather than as dedicated hardware as obtained in the previous generations of RANs [2].

In general, the open RAN architecture facilitates network scalability, intelligent network management and orchestration, efficient resource allocation, energy efficiency (EE), and reduced CAPEX/OPEX. It would also lead to enhanced competition among various equipment vendors as well as foster innovation. The research into the standardization of O-RAN and optimization of various aspects of the O-RAN are the subject of attention not only for the industry [5] but the academia and even the government of different countries. For example, the UK government has invested heavily in the development of O-RAN architecture as it is envisioned to be a major driver for the massive deployment of beyond 5G networks across the country [6].

Even though O-RAN has many potential advantages, the effect of this architecture on the energy consumption of the network is still an area of open research. Hence, this paper investigates the O-RAN from the perspective of EE by reviewing the power consumption models that have been proposed for quantifying energy consumption, and the techniques that have been proposed in the literature for improving the EE of O-RAN. In addition, it also presents open challenges and future research directions for EE enhancement in O-RAN. As this is the first survey paper on the EE of O-RAN, we believe that this work would keep researchers abreast of the state-of-the-art in O-RAN EE and power consumption modelling and point them to areas that need more research attention.

The remaining parts of the paper are organized as follows: Section II presents RAN evolution from distributed to O-RAN while Section III highlights the power consumption components and discusses the various power models proposed for O-RAN. The different techniques for optimizing the EE of O-RAN are discussed in Section IV, open challenges are highlighted in Section V, while Section VI concludes the paper.

II. RAN Evolution: From D-RAN to Open RAN

In this section, we discuss the evolution of the RAN from distributed to the current O-RAN architecture alongside the various factors that were responsible for these advancements.

- **Traditional RAN or Distributed RAN (D-RAN):** In this architecture, the base station is segmented into two parts, the BBU, which is used for signal processing, resource management, etc., and may be located at the base of the cell site, and the RRH, which is responsible for RF functions and is located close to the antenna. In addition, each RRH is linked to its own dedicated BBU with the fronthaul [3]. The D-RAN employs dedicated hardware and software that is proprietary and the interface between the BBU and RRH is closed, thus preventing the possibility of inter-operation among different vendor equipment. In addition, since each BBU serves a single RRH, as the number of users increases, more base stations need to be deployed to accommodate the surge in network demand, thereby resulting in increased CAPEX/OPEX. The need to implement network densification more efficiently necessitated the shift in RAN architecture from a distributed to a centralized approach [7].

- **Centralized-RAN (C-RAN):** In this architecture, rather than connecting each RRH to a dedicated BBU as in D-RAN, BBUs are pooled together in a central location to host network resources and several RRHs are connected to the BBU pool. By pooling BBUs, the C-RAN architecture enhances resource utilization, improves network scalability, improves network coordination and load balancing, and enhances EE. In terms of EE, three energy-saving gains have been identified in C-RAN [8]: i) Stacking gains, which is the energy saving obtained due to enhanced utilization of RAN hardware; ii) Pooling gain, which results from the use of more powerful and energy-efficient BBUs to minimize total power consumption or the ability to dynamically map the BBUs to RRHs based on traffic demand thus adapting its energy consumption to the variations in the network load; iii) cooling benefits which are derived from effective cooling and efficient power usage due to the implementation of the centralized cooling system compared to distributed cooling in D-RAN. In addition, cooperation among RRHs is possible, thus reducing the distance between RRHs and user equipment, and minimizing interference among RRHs which leads to a further reduction in the energy consumption of C-RAN [3]. However, the C-RAN requires a high-capacity and low-latency transport network to support the fronthaul traffic which is capital intensive, has security concerns, and uses vendor-specific hardware and software with closed interfaces. Further, there is the problem of a single point of failure due to BBU pooling.

- **Virtual-RAN (V-RAN):** This architecture builds on the C-RAN architecture by introducing network virtualization. Virtualization enables the softwarization of various components and functions of the RAN and their deployment on generic hardware (servers). The introduction of V-RAN enabled the BBU pool of the C-RAN to be deployed as software on general-purpose servers rather than being attached to dedicated hardware as in the conventional C-RAN. The introduction of network virtualization laid the foundation for the segmentation of network protocol stack and the ability to deploy them on different nodes in the network. The advantages of V-RAN architecture include improvement in flexibility and scalability of network deployment and reduced OPEX and CAPEX. However, it also introduced new challenges in terms of complexity in network management and orchestration. In addition, its interfaces are still closed, thereby preventing the interoperability of different vendor equipment [9].

- **O-RAN:** The O-RAN architecture separates the RAN into three logical nodes: O-CU, O-DU, and O-RU. It involves the disaggregation of network functions into eight (8) split options such that different segments of the protocol stack can be executed either in the CU, DU, or RU [4]. It also embraces the concept of virtualization of the network functions which enables network functions to be separated from dedicated hardware and installed on general-purpose servers using virtual machines. It employs open interfaces which enable different vendor hardware and software to work together irrespective of the manufacturer. In addition, intelligent controllers are also included thus providing native support for artificial intelligence and machine learning for the orchestration and optimization of the O-RAN [2]. The O-RAN architecture enhances the flexibility of network deployment by enabling the dynamic allocation of network resources. In addition, the inclusion and support for intelligence and virtualization also make the RAN configuration adjustable to suit the user demand in different locations of the network [9].

III. Power consumption of O-RAN

Since the implementation of O-RAN comprises the use of COTS hardware comprising GPP servers and the virtualization of networks functions using software applications, a major part of the power consumption of O-RAN is related to data processing or computing in the cloud or edge serves where the CUs and DUs are located and that due to RF functionalities and power amplification in the RUs. The other aspect of power consumption that cannot be ignored is due to data transportation through the backhaul, midhaul, and fronthaul links. Therefore, in this section, we first highlight the power consumption components of both the radio and transport network of the O-RAN, then, we discuss the various power consumption models that have been proposed in the literature for quantifying the energy consumption of O-RAN. The power consumption components of O-RAN are illustrated in Fig. 1.

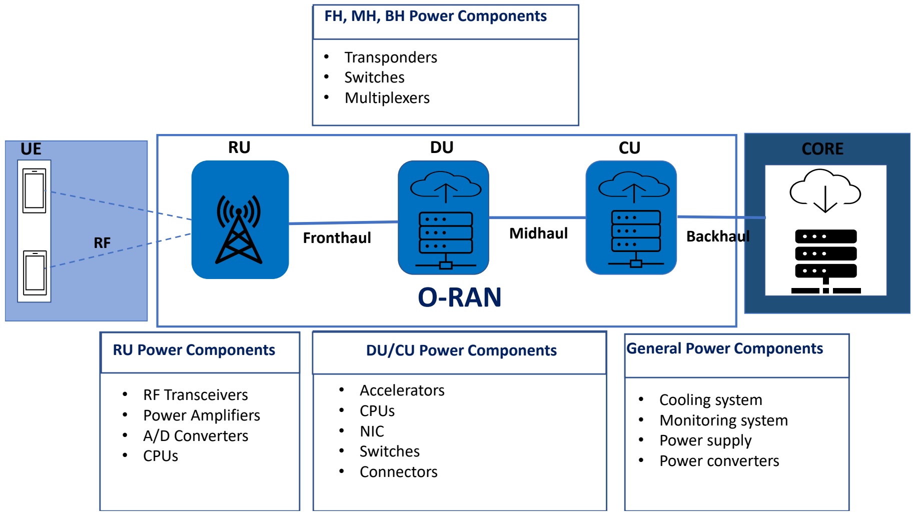

Fig. 1. The power consumption components of O-RAN (where FH, MH and BH denote fronthaul, midhaul, and backhaul, respectively).

A. Power consumption components of O-RAN

- **Radio Network:** The power consumption of the radio network is the aggregation of the power consumed by the hardware and software that are used to implement CUs, DUs, and RUs. In addition, the location of the CUs, DUs and RUs on the network nodes also affects the power consumption of the radio network. Since O-RAN employs network virtualization such that the CUs, DUs and some parts of the RUs are implemented using virtual machines on COTS servers, a bulk of the energy consumption in the radio unit would be due to the computation overhead incurred by the GPPs [10]. Typical components of the radio network are the central processing units (CPUs), accelerators and network interface cards (NIC) which are parts of the server that hosts the CUs and DUs. In addition to the aforementioned components, the RU also comprises RF transceivers and power amplifiers. There are also common site infrastructures including cooling, monitoring, alarm, power supply, and conversion systems [11]. The energy consumption of the radio network also depends on the kind of functional split implemented as the more centralized the network functions are at the CU the lesser the energy consumption of the RU, even though it has its associated latency and transport network challenges. Hence, a trade-off between EE and other QoS metrics such as latency, throughput, etc., is often the focus of most optimization problems.
- **Transport Network:** The transport network is made up of fronthaul, midhaul, and backhaul and the associated switches. The power consumption of the transport network depends on the type of technology, network configuration or topology, capacity requirement, and the number of connections between CUs, DUs and RUs. The power consumption components of the transport network are mainly switches, transponders, and multiplexers. Examples of transport technologies are point-to-point (P2P) fibre, passive optical network (PON), microwave radio, coarse wavelength division multiplexing (CWDM), Ethernet, etc [12]. The power consumption of the transport network also varies with the type of split option adopted. The work in [13] investigated the effect of the transport network on the total power consumption of a C-RAN with three split options: 6, 7, and 8. Their findings reveal that the transport network contributes about 2%, 30%, and 60% respectively, with split options 6, 7, and 8. Hence, although lower functional splits have the advantage of centralization gain, this can also lead to higher energy consumption due to increased capacity that has to be accommodated by the transport network [8].

B. Power consumption model of O-RAN radio network

The shift in O-RAN implementation from dedicated hardware to V-RAN using software implemented on GPPs means that the power consumed by the processors or CPUs would become the predominant contributor to the energy consumption of the O-RAN. Hence, most of the power consumption models that have been proposed for O-RAN are mainly related to computational power consumption and function or data migration from one CPU to another within the CU, DU or RU. These models can be broadly categorized into analytical models which are based on mathematical derivations, and statistical models which are based on experimentation. Hence, in the following paragraphs, we discuss the various analytical and statistical power consumption models that have been proposed in the literature.

- **Analytical Models:** The authors in [14] proposed power consumption models for a 5G RAN which utilizes the V-RAN architecture and considers dynamic functional splits. The first power consumption model is related to processing operations that take place in the CUs and DUs while the second power model is related to the migration of network functions between CU and DU. For the power consumption due to processing, they took a cue from EARTH's power model [15] to develop a power consumption model for the CUs and DUs that is a function of the average CPU load and the CPU processing

time. The model comprises the fixed power consumption component which covers the cost of running the server such as cooling, power amplifiers, network switches, etc., and the dynamic power consumption component which varies with the load on the machine and is due to data processing operations. The power consumption model for function migration maps the volume of data transferred during function migration between CU and DU to the power utilized in performing the operation. The work in [16] considered the peculiarities of virtual base stations such as dynamic computational resource allocation, and proposed a computational aware power consumption model which takes into consideration the active CPU cores, clock speed, and CPU load. The authors in [17] introduced a power consumption model for a GPP which comprises the static power consumption (e.g., cooling and monitoring system), and dynamic power consumption component (which depends on the complexity of CPU operation and is measured in Giga operations per second (GOPs), and computational requirements at the RU). A power consumption model for DUs was proposed in [18]. The model considered the power consumed by all the hardware components where the DU is hosted, including the processing cards, switches, cooling, power conversion, and power supply.
- **Statistical models:** The work in [19] investigated the power consumption of virtual base stations based on experimentation while considering certain parameters such as signal-to-noise-ratio (SNR), modulation and coding scheme (MCS), traffic load, channel quality, and airtime. Since the CPU is the most dormant aspect of the virtual base station, they proposed two statistical models based on experimental measurements to quantify the power consumption of the CPU. The first model considers the SNR and duty cycle and showed a linear relationship between the CPU power consumption and these two parameters. In developing the second model, they first investigated the effect of the choice of MCS on CPU power, then, based on their findings, they proposed a holistic CPU power consumption model that considers MCS, SNR, and duty cycle. In [20], the authors developed an experimental model for virtual C-RAN that employs GPPs which is a function of MCS, physical resource block, CPU clock frequency, and traffic load on the network.

C. Power consumption model of O-RAN transport network

As stated earlier, the transport network also known as the crosshaul network comprises the backhaul, midhaul, and fronthaul and has been implemented using different technologies. The power consumption of the transport networks depends on the type of technology, and topology of the transport network employed. Therefore, in the following paragraphs, we review the few power models for transport networks that are compatible with the O-RAN that have been proposed in the literature. The authors in [21] proposed various power models based on simulations for five (5) different kinds of technologies including P2P fibre, PON, microwave radio, CWDM, and Ethernet that can be used for mobile crosshaul in O-RAN. The authors in [8] considered various optical transport network configurations and proposed power models for quantifying the power consumption of each type of network topology.

IV. Energy efficiency techniques in O-RAN

In this section, we first classify the EE techniques proposed in the literature into three (3) categories. The first category deals with the dynamic allocation of resources and network function placement between the CUs and DUs. The second approach considers the dynamic location of DUs and CUs on physical network nodes and user association with the DUs. Thirdly, there are indirect EE optimization approaches that do not directly consider EE as the performance metrics but other metrics whose optimization directly impacts the EE of the O-RAN such as computational cost, routing cost, etc. Then, we present and discuss the various EE optimization techniques that have been proposed in the literature under these three (3) categories in the following subsections. Table I summarizes the proposed EE techniques in O-RAN.

A. Dynamic resource allocation and network function placement (DRA&NFP)

Research works in this area focus on the segmentation of processing operations between the CUs and DUs via the process of dynamic network function placement. The goal is to ensure that more network functions are performed in a few DUs and CUs, so that dormant virtual machines hosting the CUs and DUs can be turned off. Since the power consumption of CU and DU is proportional to their active processing time, hence minimizing the number of active CUs and DUs while meeting satisfying QoS requirements such as latency and packet delivery ratio would gratefully improve the energy efficiency of O-RAN. In the following paragraphs, the research works that have focused on this approach of energy optimization in O-RAN are discussed.

The authors in [22] proposed the dynamic relocation of network functions, the selection of the optimal number of DUs, and the switching off of redundant DUs in order to minimize the total energy consumption of O-RAN while considering as constraints the latency of the transport network, and the computation capacity of the DUs. To achieve their goal, a RL algorithm was applied to determine the optimal location of the network functions among the DUs that will result in improved EE while meeting QoS requirements. The authors in [23] investigated the impact of the level of V-RAN centralization in next generation (NG)-RAN on the energy consumption of the network and observed that some scenarios exist where increased centralization does not impact the energy consumption of the network. The work in [24] investigated the problem of CU-DU mapping in NG-RAN in order to minimize energy costs without affecting the quality of experience (QoE) of the users. To achieve their objective, they proposed a heuristic algorithm that will minimize the number of inter-CU user handovers by ensuring that neighboring DUs within user proximity are mapped to the same CU and also ensuring

TABLE I
VARIOUS APPROACHES TO ENERGY EFFICIENCY IN O-RAN

Paper	**Year**	**EE Technique**	**Proposed Solution**
[22]	2022	DRA&NFP	Actor-critic RL
[23]	2022	✓	e-constraint method
[24]	2020	✓	Heuristic algorithm
[14]	2021	✓	Lagragian decomposition simulated annealing
[25]	2021	✓	MILP
[26]	2022	✓	Adversarial bandit learning
[27]	2021	✓	Q-learning, SARSA
[28]	2020	✓	Deep RL
[11]	2022	DCDP&UA	MILP, heuristic algorithm
[17]	2021	✓	MILP, graph-based heuristic
[29]	2020	✓	Heuristic algorithm
[30]	2018	✓	analytical solution based on Constraint programming
[31]	2022	✓	MILP, heuristic algorithm
[32]	2018	IOEE	analytical solution based on Bender's decomposition
[33]	2022	✓	Deep RL
[34]	2022	✓	Heuristic algorithm
[35]	2022	✓	Deep Q-networks
[36]	2022	✓	Deep deterministic policy gradient (DDPG)
[37]	2022	✓	BILP

that the number of active CUs in the CU pool is minimized while considering three functional splits options.

In [14], the authors studied the problem of energy-efficient orchestration in 5G V-RAN while considering different functional splits. They aimed to determine the optimal splitting of network functions between CUs and DUs, the location of the CUs and DUs, that is whether in the cloud and or at the edge site that would lead to minimum energy consumption in the network. An energy optimization problem that jointly considers the energy consumption due to processing and function migration was formulated as an integer quadratic programming (IQP) problem and solved using lagrangian decomposition and a simulated annealing algorithm. The work in [25] considered the joint problem of resource allocation and DU selection in O-RAN to maximize EE while guaranteeing the delay requirements of low-latency traffic. The problem was first linearized and solved using mixed integer linear programming (MILP). The authors in [26] investigated the problem of energy-aware scheduling of virtual base stations in O-RAN in order to adapt the performance of the RAN to its energy consumption. To achieve this, an online learning algorithm based on adversarial bandit was proposed to determine the optimal virtual base station configuration that will utilize minimal computation resources and memory while ensuring that the overall energy consumption of the RAN is optimized.

The problem of dynamic function splitting in disaggregated and virtualized green O-RAN was investigated in [27] while considering varying renewable energy supplies (RES) and traffic conditions. The problem was formulated as an OPEX minimization problem and two reinforcement learning solutions based on Q-learning and SARSA were proposed to determine the optimal functional split that best utilizes the RES and reduces the cost of network operation. The work in [28] studied the problem of virtual network formation and forward graph embedding (FGE) placement problem in order to determine the optimal placement policy in a virtual network infrastructure that would lead to the minimization of the overall energy consumption of the network. The problem was modelled as a constrained combinatorial optimization problem and solved using a neural combinatorial optimization (NCO)-based reinforcement learning framework.

B. Dynamic CU and DU placement and user association (DCDP&UA)

The goal of this approach is to examine how the CUs and DUs can be implemented in a few network nodes or physical machines such that dormant nodes or physical machines (servers) can be shut down and energy savings can be achieved due to more efficient use of computational resources via more centralized processing. The following paragraphs present the few works that have adopted this approach of energy optimization in O-RAN. In this regard, the work in [11] considered the problem of power consumption in 5G RAN by optimizing the placement of CUs and DUs in the network in an energy-efficient manner while considering different functional splits, capacity, and latency as constraints. Two optimization strategies; MILP and a reduced complexity heuristic algorithm were proposed for small and large network topologies, respectively. The authors in [17] considered the problem of energy-efficient DUs and CUs placement in an optical 5G metro access network. Their goal was to aggregate the CUs and DUs into fewer nodes as well as select the optimal transmission part that would lead to minimal energy consumption in the network. To achieve this, they first modeled the problem using MILP and proposed a graph-based heuristic algorithm to optimize the DUs and CUs placement while considering the power consumption due to processing and network components.

The authors in [29] proposed a heuristic algorithm for the dynamic placement of DU and CU virtual machines over network nodes in an optical metro access network. Their work aimed to consolidate the baseband functions in a few nodes based on traffic flows in order to minimize the power consumption of the nodes while ensuring that the constraints including link capacity and fronthaul latency are satisfied. In [30], the authors considered a virtual RAN model where the RAN was disaggregated into CU, DU and RU. Their focus was on the joint optimization of the bandwidth of the midhaul and total energy consumption of the system while considering the constraints of processing and midhaul bandwidth capacity. The joint optimization problem was modelled as a constrained programming problem and solved analytically. The numerical results obtained revealed that the extent of energy savings that can be obtained depends on the availability of capacity in the transport network. The work in [31] investigated the problem of DU, CU, and mobile edge computing (MEC) deployment in next-generation cellular networks in order to minimize the power consumption of the network. The DU, CU, and MEC deployment problem was modelled as MILP after which a heuristic algorithm was proposed to find the optimal deployment policy that will maximize the amount of energy savings that can be achieved in the network.

C. Indirect optimization of energy efficiency (IOEE)

The works considered in this subsection are those where the energy efficiency of the O-RAN was not the objective of the optimization problem, however, other factors such as computational cost, routing cost, etc. which directly impact the energy consumption of O-RAN were considered. In this regard, the work in [32] proposed a dynamic V-RAN configuration that ensures that the demand of users are satisfied while reducing the total cost relating to computing and routing. To achieve this objective, an analytical model was developed to select the optimal functional split and signal transmission paths between RUs and CUs that will minimize the RAN costs. In [33], the authors investigated the optimization of the functional split placement between the CUs and the DUs in a V-RAN in order to minimize the computation and routing cost. The optimization problem was first modelled as a constrained neural combinatorial reinforcement learning problem, then a long and short-term memory (LSTM) sequence-sequence model was applied to determine the optimal functional splitting policy. The work in [34] proposed an optimal virtual network function splitting framework for O-RAN with the goal of balancing the network load between the CUs and DUs while considering the delay requirements in the midhaul links. The proposed framework was implemented using a heuristic algorithm that is scalable.

The problem of DU-CU placement and user association in O-RAN was considered in [35]. The goal is to determine the optimal location of CU and DUs network function whether at the regional or edge cloud that would result in minimum end-to-end delays experienced by users and minimize the costs associated with network deployment. The problem was modelled as a multi-objective optimization problem, then solved using a deep Q-network. The authors in [36] considered the problem of computing and radio resource control in V-RAN to optimize CPU usage while achieving the desired performance target. A deep deterministic policy gradient-based actor-critic neural network framework was developed to determine the optimal resource control decision that would meet the QoS target of the network while minimizing CPU usage. The performance evaluation of their work using real-life data showed that huge savings in terms of CAPEX and OPEX can be achieved by the proposed method. In [37], the authors considered the problem of optimal virtual network function placement among the CUs, DUs, and RUs in V-RAN. Their goal was to maximize the aggregation level of the virtual network functions while minimizing the number of computing resources required to execute these functions. The optimal routing path was also considered in order to ameliorate the inadequacies of the crosshaul networks. The problem was modelled as a binary integer linear programming problem (BILP) and the solution was obtained using a conventional solver known as IBM CPLEX.

V. Challenges and Open Research Problems

The research on EE of O-RAN is still in its early stages, as a result, only very few works specifically focus on O-RAN architecture. Most of the research works are still aligned towards the C-RAN architecture, even though they consider the disaggregation of the RAN entities, different functional splits, and the virtualization of network functions. Hence, more research attention needs to be focused on the EE of O-RAN in order to accommodate the peculiarities of O-RAN architecture that are not found in the C-RAN architecture. This will ensure that the proposed approaches are properly situated within the O-RAN framework as this would directly impact the amount of energy savings that can be achieved.

The need for the development of a holistic power consumption model for O-RAN that captures both the hardware and software components of both the radio and transport network while considering different functional splits is an open challenge. Most of the works in the literature still directly apply the earth model or a slight modification of the EARTH model. However, the EARTH model does not capture all the features, components, and requirements of the O-RAN such as the dynamic computational resource allocation due to virtualization, the effect of functional splitting and migration, memory requirements, etc [19], [16]. Hence, research efforts need to be intensified toward developing a standardized model that can accommodate all the peculiarities of the O-RAN architecture.

Although a few works have considered the variations of energy consumption due to the dynamic placement of CUs and DUs at different locations in the network as well as the possibility of dynamic functional splitting [24], the full impact of the CU and DU placement as well as dynamic function splitting on the overall energy consumption of the network while considering various network topologies and service requirements is yet to be fully investigated. In addition, most of the work on O-RAN EE has focused on the CUs and DUs with very little attention given to the RUs. The impact of the energy consumption of RU needs to be studied alongside that of the CUs and DUs because the RU houses among other components the power amplifier, which is a major contributor to the total energy consumption of the RAN. The effect of the type of transport technology and transport network design on the overall energy consumption of the RAN while considering different split options need to be thoroughly investigated. This is because preliminary studies in [8] and [13] have already pointed out that the transport network can have a significant contribution to the overall energy consumption of the network if careful consideration is not given to it.

VI. Conclusion

In the paper, we first considered the evolution of the RAN from D-RAN to O-RAN and the major changes that took place and the factors responsible. Then we examined the power consumption components of O-RAN and the various models that have been proposed for quantifying the energy consumption in both the radio and transport network. Furthermore, various techniques that have been developed in the literature for optimizing the EE of O-RAN were presented and discussed. Finally, some open challenges are highlighted.

ACKNOWLEDGEMENT

The authors wish to acknowledge receipt of the funding we received for this research under the EPSRC UK-India Future Networks Initiative (Grant ref- EP/W016524/1) led by the University of East Anglia in UK.

REFERENCES

[1] S. K. Singh, R. Singh, and B. Kumbhani, "The Evolution of Radio Access Network Towards Open-RAN: Challenges and Opportunities," in *2020 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, 2020, pp. 1–6.

[2] P. K. Thiruvasagam, V. Venkataram, V. R. Ilangovan, M. Perapalla, R. Payyanur, V. Kumar *et al.*, "Open RAN: Evolution of Architecture, Deployment Aspects, and Future Directions," *arXiv preprint arXiv:2301.06713*, 2023.

[3] M. A. Habibi, M. Nasimi, B. Han, and H. D. Schotten, "A Comprehensive Survey of RAN Architectures Toward 5G Mobile Communication System," *IEEE Access*, vol. 7, pp. 70 371–70 421, 2019.

[4] M. Polese, L. Bonati, S. D'Oro, S. Basagni, and T. Melodia, "Understanding O-RAN: Architecture, Interfaces, Algorithms, Security, and Research Challenges," *IEEE Communications Surveys and Tutorials*, pp. 1–1, 2023.

[5] O. R. Alliance, "O-RAN: towards an open and smart RAN," *White paper*, vol. 19, 2018.

[6] M. Masoudi, O. T. Demir, J. Zander, and C. Cavdar, "Energy-Optimal End-to-End Network Slicing in Cloud-Based Architecture," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 574–592, 2022.

[7] D. Wypiór, M. Klinkowski, and I. Michalski, "Open RAN—Radio access network evolution, benefits and market trends," *Applied Sciences*, vol. 12, no. 1, p. 408, 2022.

[8] M. Fiorani, S. Tombaz, J. Martensson, B. Skubic, L. Wosinska, and P. Monti, "Modeling energy performance of C-RAN with optical transport in 5G network scenarios," *Journal of Optical Communications and Networking*, vol. 8, no. 11, pp. B21–B34, 2016.

[9] B. Brik, K. Boutiba, and A. Ksentini, "Deep Learning for B5G Open Radio Access Network: Evolution, Survey, Case Studies, and Challenges," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 228–250, 2022.

[10] W. Azariah, F. Asisi Bimo, C.-W. Lin, R.-G. Cheng, R. Jana, and N. Nikaein, "A Survey on Open Radio Access Networks: Challenges, Research Directions, and Open Source Approaches," *arXiv e-prints*, p. arXiv:2208.09125, Aug. 2022.

[11] L. M. Moreira Zorello, M. Sodano, S. Troia, and G. Maier, "Power-Efficient Baseband-Function Placement in Latency-Constrained 5G Metro Access," *IEEE Transactions on Green Communications and Networking*, vol. 6, no. 3, pp. 1683–1696, 2022.

[12] L. M. P. Larsen, A. Checko, and H. L. Christiansen, "A Survey of the Functional Splits Proposed for 5G Mobile Crosshaul Networks," *IEEE Communications Surveys and Tutorials*, vol. 21, no. 1, pp. 146–172, 2019.

[13] D. López-Pérez, A. De Domenico, N. Piovesan, G. Xinli, H. Bao, S. Qitao, and M. Debbah, "A Survey on 5G Radio Access Network Energy Efficiency: Massive MIMO, Lean Carrier Design, Sleep Modes, and Machine Learning," *IEEE Communications Surveys and Tutorials*, vol. 24, no. 1, pp. 653–697, 2022.

[14] R. Singh, C. Hasan, X. Foukas, M. Fiore, M. K. Marina, and Y. Wang, "Energy-Efficient Orchestration of Metro-Scale 5G Radio Access Networks," in *IEEE INFOCOM 2021 - IEEE Conference on Computer Communications*, 2021, pp. 1–10.

[15] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?" *IEEE Wireless Communications*, vol. 18, no. 5, pp. 40–49, 2011.

[16] T. Zhao, J. Wu, S. Zhou, and Z. Niu, "Energy-delay tradeoffs of virtual base stations with a computational-resource-aware energy consumption model," in *2014 IEEE International Conference on Communication Systems*, 2014, pp. 26–30.

[17] Y. Xiao, J. Zhang, and Y. Ji, "Energy-Efficient DU-CU Deployment and Lightpath Provisioning for Service-Oriented 5G Metro Access/Aggregation Networks," *Journal of Lightwave Technology*, vol. 39, no. 17, pp. 5347–5361, 2021.

[18] M. Shehata, A. Elbanna, F. Musumeci, and M. Tornatore, "Multiplexing Gain and Processing Savings of 5G Radio-Access-Network Functional Splits," *IEEE Transactions on Green Communications and Networking*, vol. 2, no. 4, pp. 982–991, 2018.

[19] J. A. Ayala-Romero, I. Khalid, A. Garcia-Saavedra, X. Costa-Perez, and G. Iosifidis, "Experimental Evaluation of Power Consumption in Virtualized Base Stations," in *ICC 2021 - IEEE International Conference on Communications*, 2021, pp. 1–6.

[20] U. Pawar, B. R. Tamma, and F. A. Antony, "Traffic-Aware Compute Resource Tuning for Energy Efficient Cloud RANs," in *2021 IEEE Global Communications Conference (GLOBECOM)*, 2021, pp. 01–06.

[21] L. M. Larsen, M. S. Berger, and H. L. Christiansen, "Energy-Aware Technology Comparisons for 5G Mobile Fronthaul Networks," *Int. J. Adv. Networks Serv.*, vol. 13, no. 1-2, 2020.

[22] S. Mollahasani, T. Pamuklu, R. Wilson, and M. Erol-Kantarci, "Energy-aware dynamic DU selection and NF relocation in O-RAN using actor-critic learning," *Sensors*, vol. 22, no. 13, p. 5029, 2022.

[23] W. Pires, G. de Almeida, S. Correa, C. Both, L. Pinto, and K. Cardoso, "Bi-objective Optimization for Energy Efficiency and Centralization Level in Virtualized RAN," in *ICC 2022 - IEEE International Conference on Communications*, 2022, pp. 1034–1039.

[24] H. Gupta, M. Sharma, A. Franklin A., and B. R. Tamma, "Apt-RAN: A Flexible Split-Based 5G RAN to Minimize Energy Consumption and Handovers," *IEEE Transactions on Network and Service Management*, vol. 17, no. 1, pp. 473–487, 2020.

[25] T. Pamuklu, S. Mollahasani, and M. Erol-Kantarci, "Energy-Efficient and Delay-Guaranteed Joint Resource Allocation and DU Selection in O-RAN," in *2021 IEEE 4th 5G World Forum (5GWF)*, 2021, pp. 99–104.

[26] M. Kalntis and G. Iosifidis, "Energy-aware Scheduling of Virtualized Base Stations in O-RAN with Online Learning," *arXiv e-prints*, p. arXiv:2208.09956, Aug. 2022.

[27] T. Pamuklu, M. Erol-Kantarci, and C. Ersoy, "Reinforcement Learning Based Dynamic Function Splitting in Disaggregated Green Open RANs," in *ICC 2021 - IEEE International Conference on Communications*, 2021, pp. 1–6.

[28] R. Solozabal, J. Ceberio, A. Sanchoyerto, L. Zabala, B. Blanco, and F. Liberal, "Virtual Network Function Placement Optimization With Deep Reinforcement Learning," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 2, pp. 292–303, 2020.

[29] L. Askari, F. Musumeci, L. Salerno, O. Ayoub, and M. Tornatore, "Dynamic DU/CU Placement for 3-layer C-RANs in Optical Metro-Access Networks," in *2020 22nd International Conference on Transparent Optical Networks (ICTON)*, 2020, pp. 1–4.

[30] A. Alabbasi, X. Wang, and C. Cavdar, "Optimal Processing Allocation to Minimize Energy and Bandwidth Consumption in Hybrid CRAN," *IEEE Transactions on Green Communications and Networking*, vol. 2, no. 2, pp. 545–555, 2018.

[31] H. Li, K. Assis, A. Vafeas, S. Yan, and D. Simeonidou, "Resilient and Energy Efficient DU-CU-MEC Deployments for Service Oriented Reliable Next Generation Metro Access Network," *47th WWRF*, 2022.

[32] A. Garcia-Saavedra, X. Costa-Perez, D. J. Leith, and G. Iosifidis, "FluidRAN: Optimized vRAN/MEC Orchestration," in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, 2018, pp. 2366–2374.

[33] F. W. Murti, S. Ali, and M. Latva-Aho, "Constrained Deep Reinforcement Based Functional Split Optimization in Virtualized RANs," *IEEE Transactions on Wireless Communications*, vol. 21, no. 11, pp. 9850–9864, 2022.

[34] E. Amiri, N. Wang, M. Shojafar, and R. Tafazolli, "Optimizing Virtual Network Function Splitting in Open-RAN Environments," in *2022 IEEE 47th Conference on Local Computer Networks (LCN)*, 2022, pp. 422–429.

[35] R. Joda, T. Pamuklu, P. E. Iturria-Rivera, and M. Erol-Kantarci, "Deep Reinforcement Learning-Based Joint User Association and CU–DU Placement in O-RAN," *IEEE Transactions on Network and Service Management*, vol. 19, no. 4, pp. 4097–4110, 2022.

[36] J. A. Ayala-Romero, A. Garcia-Saavedra, M. Gramaglia, X. Costa-Pérez, A. Banchs, and J. J. Alcaraz, "vrAIn: Deep Learning Based Orchestration for Computing and Radio Resources in vRANs," *IEEE Transactions on Mobile Computing*, vol. 21, no. 7, pp. 2652–2670, 2022.

[37] F. Z. Morais, G. M. F. De Almeida, L. L. Pinto, K. Cardoso, L. M. Contreras, R. d. R. Righi, and C. B. Both, "PlaceRAN: optimal placement of virtualized network functions in Beyond 5G radio access networks," *IEEE Transactions on Mobile Computing*, pp. 1–1, 2022.