On Energy Efficient Resource Allocation in Shared RANs: Survey and Qualitative Analysis

Fatma Marzouk, Student Member, IEEE, João Paulo Barraca, and Ayman Radwan, Senior Member, IEEE

Abstract—An expansion of services and unprecedented traffic growth is anticipated in future networks, aligned with the adoption of the long-awaited Fifth Generation (5G) of mobile communications. To support this demand, without exposing mobile operators to the pressure of CAPEX and OPEX, 5G uses new frequency bands, and adopts promising trends, including: densification, softwarization, and autonomous management. While the first technology is proposed to handle the traffic growth requirements, the softwarization and autonomous management are expected to play, in synergy, to ensure the desired trade-off between reducing the CAPEX and OPEX, while guaranteeing the quality of service (QoS). Softwarization is expected to transform the network design, from one size fits all, to more demand oriented adaptive resource allocation. In this work, we focus on this point, by discussing how these technologies act in synergy towards enabling RAN sharing. Particularly, we focus on how they fit into the issue of energy efficient Multi-Operator Resource Allocation (MO-RA). After a survey and classification of schemes leveraging this synergy for distinct resource allocation (RA) objectives, we present a detailed survey and qualitative classification of RA schemes with respect to energy efficiency. This work presents an innovative survey, since it concentrates on multiple operators, and the enabling of Mobile Virtual Network Operators (MVNOs), which will come into play with the complete virtualization of mobile networks. Based on the deep literature analysis of the different operations that can bring energy savings to MO-RA, we conclude the work with listing open challenges and future research directions.

Index Terms—V-RAN sharing, resource allocation, energy efficiency, SON, ML, NFV, SDN, SDR, QoS.

I. INTRODUCTION

With the vast emergence of new services in the recent years, design requirements of wireless and mobile networking are becoming more stringent, placing greater pressure on Mobile Network Operators (MNOs) to provide capacity and high-quality service, while still being able to sustain profitability. The Fifth Generation (5G) of mobile networks is expected to provide unprecedented high data rates, ultra-low latency and high reliability services, catering for a massive number of connected devices, and enabling the wide adoption of Internet of Things (IoT). A vital solution, for accommodating the capacity requirements, is built on multi-tier highly dense Heterogeneous Network (HetNet), encompassing multiple cell types and multiple Radio Access Technologies (RATs) into a unified network. The upcoming set of solutions presents the opportunity to jointly meet the two requirements thanks to the adoption of two prominent paradigms: Softwarization and Autonomous Management. The first paradigm is represented by technologies, such as Network Function Virtualization (NFV), Software Defined Network (SDN) and Software Defined Radio (SDR). The second is characterized by the exploitation of more enhanced use cases of SON and machine learning capabilities. The two paradigms are expected to play in synergy within the 5G RAN landscape, paving the way towards new business opportunities, where the RAN infrastructure and the spectrum can be shared between multiple third party service providers, called Mobile Virtual Network Operators (MVNOs). It can also afford a shift from the concept of static and rigid sharing to an adaptively shared RAN. In such shared RAN, also called Virtual RAN (V-RAN), NFV leads to the virtualization of all RAN resources, a first step towards real active sharing. On the other hand, SON will take full advantage of the NFV flexibility and the Software Defined Network Control (SDNC) and programmability. The influence and coverage map of SON would, hence, include enhanced SON use cases, supporting Multi-RAT optimization and Multi-Tenant V-RAN provisioning. A challenge arises, when it comes to achieving efficient RAN resource allocation that reflects the MVNOs’ desire to optimize a metric or to attain the desired trade-off between reducing CAPEX/OPEX and maximizing the offered Quality of Service (QoS). Ensuring energy efficient allocation is directly associated with reduced OPEX. Several research works [1]–[5] focused on energy efficient resource allocation processes, leveraging on the application of the aforementioned paradigm. Resources, in that context, can be antennas or Baseband Units (BBU).
There have been valuable research efforts on the application of SDN/NFV and autonomous management of RAN, in 5G RAN sharing [3] and leveraging these for 5G RAN Slicing [4]–[5]. On the other hand, others have proposed EE RA for C-RANs [1]–[2]. Further works [6] address EE RA with multi-operator sharing, by proposing a roaming based sharing scheme that switches-off low-load Mobile Network Operator (MNO) Base Stations (BSs), and roam their traffic to active BSs, operated by other MNOs. However, C-RAN flexibility is not fully exploited, by shutting off Remote Radio Heads (RRHs) and the corresponding BBUs Virtual Machines (VMs). Hence, such scheme could potentially lead to coverage holes. To the best of our knowledge, the research works in [7]–[8] are the only ones that suggest C-RAN sharing based on coalition game theory, which could present advantages for heterogeneous scenarios. The scheme is consistent with the C-RANs, where MNOs share their infrastructure (RRH, BBU Pools), aiming to set the rules for profitable collaboration. We believe that, with respect to this research topic, a looming question is: Going beyond the multi-operator RAN sharing scope to the on-demand Multi-Tenant Heterogeneous RAN sharing, how can MO-RA leverage from softwarization and autonomous management, to bring the desirable energy efficiency for both players: infrastructure provider and MVNOs.

The novelty of this survey is that it elaborates on this specific use case and discusses, from a design perspective, how these paradigms can fit into the jigsaw of resource allocation in shared multi-tenant environments, and how they bring energy savings. After discussing the impact of the aforementioned paradigms in the general 5G landscape, we provide a survey of RAN resource allocation schemes with respect to different technologies. Then, we further present a more detailed survey of the state-of-art resource allocation research efforts, with respect to Energy Efficiency.

The remainder of this paper is organized as follows. In Section II, we present an overview of the state-of-the-art on solutions toward meeting the capacity requirement. In Sections III and IV, we describe the main softwarization and autonomous management technologies. In Section V, we explain the differences between those technologies, while highlighting the opportunities and synergies arising from the combination of these paradigms on the RAN. In Section VI, we present relevant standardization activities and efforts that are being developed in this line of research. Section VII is devoted to the survey and classification of the RAN resource allocation schemes, with respect to the type of the resource, the used technology and the targeted metric. Section VIII includes a qualitative review and classification of energy efficient resource allocation schemes and highlights the current gap in the literature. Section IX discusses the open challenges, still not addressed by the existing approaches, as well as key future research directions. Finally, Section X concludes. At the end of the paper, in the Appendix, a table listing all acronyms, used in the paper, is presented.

II. RAN DENSIFICATION IN 5G

With mobile operators running out of spectrum, network densification is considered crucial for enhancing spectral efficiency. Densification can be achieved over space, using smaller cells, and over frequency, by making use of larger portions of the radio spectrum, and is considered a key mechanism toward meeting the 1000x challenge [9]. In this section, we elaborate on the different forms and mechanisms of densification.

A. Densification Using Small Cells in 5G

Small cells are low-power Access Points (APs) that can be easily installed anywhere, enabling telecoms to avoid the deployment costs of macro-BSs, while improving the macro cell’s edge capacity and overall network efficiency. They traditionally range from micro/metro cells (the largest) to femtocells (the smallest), also called home BSs [10], [11]. Furthermore, today’s wireless cellular networks feature an extensive form of distributed densification, where the baseband processing occurs at a base band unit, separated from the remote radio heads (RRHs). Several efforts have been devoted to the study of densification, in particular using static small cells. For instance, some authors [12] investigated the applicability of small cells that use the same unlicensed spectrum as WiFi, along with a Time Division Duplex (TDD) based coordinated interference avoidance scheme. The work showed that this approach can increase the overall network capacity. Other authors provided a comprehensive tutorial on the use of densification, using mmWave small cell [13], while applying different metrics to study the optimal location of small cells. Authors in [14] targeted the maximization of the network spectral efficiency, while authors in [15]. Focused on maximizing the throughput, through local optimization methods. These efforts assumed that traffic distribution is invariant, and did not consider the high costs of static deployment of small cells.

B. From Static to Mobile Small Cells

In contrast to the aforementioned works, more recent work [16]–[22] started to explore the concept of having a more dynamic and opportunistic small cell deployment, to suit the dynamic nature of mobile users. Through an analytical benchmark and simulations, authors demonstrated the gain induced by having small cells deployed in vehicles, with focus on metrics such as outage and error probabilities [17], [21]. In [19], authors targeted the maximization of service time, provided by deployment of mobile small cells. Also, with the aim of offloading traffic of congested macro cells in urban scenarios, the authors evaluated the efficiency of deploying small cells on buses (which follow a more predictable route), showing this solution to be beneficial, in particular when the small cells are near a hotspot. In [22], the authors used probabilistic graph algorithms in a network with mobile small cells, in order to solve resource allocation problem. Despite the existence of several algorithms for hotspot localization [25], [26], using mobile small cells leads to either sporadic gain or to a performance degradation, mainly due to the resulting added interference, when the small cell is moving away from the hotspot location. To tackle this limitation, some recent works [23], [24], proposed the idea of using on demand
mobile small cells, that can be implemented using user devices located in the vicinity. In both works, the advantages of such deployment have been strongly demonstrated. While the proposed virtual small cells formation scheme in [24] focused on the minimization of the number of communication links to the macrocell for connecting all users directly or through small cells, authors in [23] showed the gain in throughput, spectral efficiency, and Signal to Interference Noise Ratio (SINR). Authors of [25] extended this vision into an SDN controlled RAN paradigm, by proposing an SDN southbound protocol API running on top of smartphones, so that they can act as SDN-enabled routers.

C. Multi-RAT Licensed and Unlicensed-Band Capability

5G heterogeneity based on small cells is expected to add new dimensions with respect to coverage and mobility, as well as with respect to the integration of a wider range of RATs and spectrum bands. Indeed, the 5G air interface will comprise the evolved existing RATs (2G, 3G, WLAN, 4G) that enable wide area coverage, as well as millimeter wave (mmWave) and centimeter wave (cmWave) RATs, addressing an area for capacity boosting and support of Ultra-Reliable Low-Latency Communications (URLLC). The key here will be resource sharing and works [26], [27] have demonstrated that a cost-saving and opportunistic form of multi-RAT can be achieved by the reusability of RRHs, along with the agile mapping between RRH and BBU resources. The licensed/unlicensed band capability is embodied by coexistence schemes that aim to exploit unlicensed band for Long Term Evolution (LTE) communications. Several efforts already defined in relation to coexistence with WiFi includes: Qualcomm LTE-Unlicensed (LTE-U) [28], 3GPP release 13 - Licensed Assisted Access (LAA) [29], and LTE-WiFi Aggregation (LWA) [30].

III. SOFTWAREIZATION TECHNOLOGIES FOR THE 5G RAN

The increase in capacity achieved with densification comes at the cost of an increased complexity for the management of the RAN. Softwareization technologies on the RAN are considered among the main 5G drivers, presenting the opportunity to address such complexity. In this section, we give a brief overview of some technologies for a virtualized RAN.

A. Network Function Virtualization (NFV)/Wireless Network Virtualization (WNV)

Using NFV, network services (or functions) are relocated from dedicated devices to generic servers, and abstracted as virtual network functions (VNF), which provide savings in CAPEX and OPEX [31], while enabling a great amount of flexibility. Indeed, hardware overprovisioning induced costs can be saved, due to the on-demand capability of scaling up/down server capacity via software, bringing elasticity concepts to telecom networks. Baseband processing pooling [32], [33] is a form of enabling NFV [34] on the RAN, denoted as Cloud RAN or V-RAN. Figure 1 illustrates the baseband processing virtualization in C-RAN.

B. Software Defined Networking (SDN)

SDN implies decoupling the network control plane from the forwarding plane and implementing it in a logically centralized controller (SDN controller) to transform the network into: i) directly programmable and programmatically configurable, through the aggregation of control functions; ii) centrally manageable via an SDN controller with a global view; iii) agile in its dynamic adjustment of traffic flow to meet different QoS requirements; iii) open standard-based and vendor-neutral facilitating the development of new SDN architectures. Figure 2 depicts a general schematic of SDN architecture, along with its different layers.

When applied to the RAN, SDN presents a key enabler for addressing the heterogeneity induced complexity. Indeed,
the SD-RAN control plane would have the merit to be programmable and configurable, via an innovative and advanced SD-RAN application plane and be in charge of translating the SDN application requirements down to a set of configurations, to be adjusted by the data plane via the SDN southbound interface. The data plane would be composed of the RAN networking nodes. On the other hand, the application plane would host a panoply of evolved decision maker modules for the management and optimization of the complex RAN operations.

C. Software Defined Radio (SDR)

SDR is an RF communication system, involving a significant amount of processing components, traditionally implemented in hardware, but now realized by means of an easily reconfigurable software pipeline. SDR helps in tackling the hardware induced shortcomes: cost; manufacturer compatibility, and standards restrictions; scalability issues; major modifications for minor upgrades. Softwarization of these systems leads to high flexibility, as modifications can be easily integrated into existing software. More specifically, multi-standard operations are rendered possible, using one single SDR enabled device, allowing different technologies (Wi-Fi, GSM, LTE or 5G) to operate according to the transceiver pipelines, which are embedded in software. Several works agreed that the major benefit of SDR relies in the flexibility offered, which can be exploited to develop cognitive radios, by exploiting self-perception, and dynamic spectrum coexistence schemes [37], [38]. Indeed, cognitive radio SDRs are self-configured and can dynamically adjust their parameters, such as power, band, or modulation, to operate in the existence of a multitude of limitations. Hence, applying SDR on RAN offers an extra degree of freedom, to support the programmability and the control of radio resources.

IV. AUTONOMOUS MANAGEMENT TECHNOLOGIES IN RAN

Research efforts, focusing on autonomous management technologies, namely legacy SON and SON empowered with Machine Learning (ML) and big data techniques, aligned with 4G/5G C-RAN systems, are discussed here.

A. Self-Organizing Networks (SON)

The application of Self-Organizing (SO) to the field of mobile networking systems leads to Self-Organizing Networks (SONs), where traditional network management procedures are improved thanks to the automation capability, brought by SO. SON driven automation is enabled, by adding more intelligence to the network to support autonomous installations, optimization, and configuration and healing operations. Therefore, human involvement would be decreased or eventually completely eliminated [39]. For MNO, SON technology leads to: i) simplification of operational and management tasks, in HetNets, ii) improvement of network performance, iii) reduction of time to market of new services, and iv) OPEX savings. Benefits include more than 50% decrease in dropped calls, OPEX savings of more than 30%, and an increase in service revenue by 5-10% [40]. A set of self-x functions is defined within the SON scope, covering self-configuration, self-optimization and self-healing. Self-configuration builds on the idea that ideally a BS added to a network should configure itself, following a “plug and play” principle. Self-optimization aims to ensure that network elements are almost always operating at the most efficient point, by defining processes that analyze network performance and auto-tune the mode and parameters of network elements, to meet demands [41], [42]. Some examples of self-optimization include Mobility Robustness Optimization (MRO), Mobility Load Optimization (MLO), and Capacity and Coverage Optimization (CCO). For instance, CCO adapts RRM parameters, such as antenna tilt, or transmitter power to maximize coverage, while optimizing capacity.

Research in the area of SON, especially in HetNets, has attracted the attention of a large segment of research community. The contributions of the research efforts have spanned different challenges, including identifying SON challenges [43]–[45], proposing algorithms in support of a given self-function [46]–[49]. In [46], authors discussed the notion of cloud coverage, provided by densely deployed self-organized small cells under-laying the macro-cell. The proposed technique relies on a fuzzy-logic-based framework, allowing the macro-cell to control when to activate the cloud. A dynamic access class barring and Traffic Adaptive Radio Resource Management (TARRM) for Machine Type Communications (MTC) over LTE-A has been proposed [47]. The work suggested an ECACB scheme, which dynamically controls the barring factor, in order to avoid collision induced from numerous MTC devices. The authors based the solution on the prediction of access intensity, measured at a standard eNodeB. Once the MTC devices have been granted access, the TARRM scheme schedules their radio resources. The scheduling scheme is based on the use of cognitive radio networks, to detect unused Physical Resource Blocks (PRBs). Research work [48] falls under the self-healing category, and proposes a framework for the detection and diagnosis of faults. Other authors presented a fast approach for the self-planning of uplink Fractional Power Control (FPC) settings [49], contributing to self-configuration. The approach starts by the division of the large-scale multivariable optimization problem into multiple simple optimization problems. The approach, then, calculates optimal FPC by an exhaustive search method. These selected research efforts are examples of algorithms for control or optimization, which are not covered by the 3GPP-SON standardization.

Several works agreed that SON in future networks will not be a nice-to-have cost-saving mechanism (as it is in LTE), but will be considered an integral part of the RAN [50]–[52]. Authors, in [50], argue that the evolved small cell-based landscape will require faster operation of SON algorithms, for automatic interference and load-balancing control in terms of Network Sensing, Network Health Checking, and Network Adjustment. The authors also emphasize the requirement of an approach for higher decentralization. The exploitation of the 5G small-cells and of smart city data for the provision of the network with
user-related information, has been discussed in [50] and [53], with the aim of making them, pre-emptive in response to device needs. Other innovative approaches include: self-protection [54]–[57], SON for mmWave [58]–[60], SON for MIMO [61], [62], SON for NFV-based networking [63], [64], and SON for Multi-RAT optimization and spectrum sharing [65]. Self-Protection consists of adding a new family of use cases related to automated-security, including attack detection and countermeasures [54]–[57]. SON for mmWave is required to adapt to the different propagation characteristics of the mmWave links. Preliminary insights, about mmWave-oriented proactive SON paradigm, are presented [66]. Directional cell search for mmWave 5G has also been investigated [58], considering a network graph based approach for self-organized beam assignment. Simulation results show improved user SINR ratios to potential handover beams and better directional cell discovery. The improvement of the random-access technique for mmWave with directional beams, by proposing two different types of preambles, reducing the overall processing, and the suggestion of SO-power allocation algorithm, based on Q-learning, was the focus in [59], [60]. SON for adaptive resource allocation in mmWave MIMO 5G, has been designed in [61], where adaptability is defined with respect to power and resource block assignment. Another research work [62] focused on the design of 5G MIMO precoding schemes in the context of an HC-RAN, where the baseband pool is used for integrating the large amount of spatial domain information, and performing null-space calculation concerning massive MIMO coordination. The use-case of SON for NFV is also a highly attractive research topic. While NFV placement and traffic steering problem is formulated as a mixed integer linear problem and solved via heuristics in [63], it has been derived as an algorithmic resource management framework solved in a heuristic, inspired by the principles of market competition in [64]. Towards SO-schemes that ensure multi-RAT optimization, authors [65] propose a SO-beam scheduling in multi-RAT deployment, where BSs and APs, deployed in unlicensed bands co-exist. Finally, SON for EE radio management is another use-case of SON application, for 5G networks that is currently attracting lots of interest.

B. Learning Based Management

The first instance of SONs can be described as adaptive and autonomous systems, based on control loops and threshold comparison. In order to handle more complex scenarios, current state of the art is investigating the application of advanced techniques, such as Machine Learning (ML), data mining to SON [51], [67]–[70]. ML algorithms can be categorized in multiple ways, with a stronger focus on supervised learning, unsupervised learning, and reinforcement learning. In supervised learning, the system is trained using the expected output from a given input. In unsupervised learning, the system has to learn on its own and construct the mapping between inputs and outputs, according to the characteristics of the data. Finally, reinforced learning combines the features of the two previous categories. Indeed, the system needs to learn by itself what the expected output should be, same as in unsupervised learning. Once the output is determined, a reward mechanism (a heuristic) is used to enable the system to evaluate the input-output pairing. Applying ML to SON paves the way to proactive SONs that can infer the environmental network context, predict future network behavior, and take in advance better decisions, leading to considerable gains in network performance. Several efforts of applying ML to SON are available in the literature to enhance MRO, MLB, CCO, self-healing, resource allocation, and energy saving. Some authors proposed the use of ML for context aware mobility management in HetNets [71]. BSs jointly learn their long-term loads and optimal cell’s range and schedule their UE, based on their velocities and historical data, leading to an improvement in throughput and fairness. Mobility management, particularly handover, has also been addressed [72], by using ML and big data to forecast HO and detect abnormalities. An example of the application of reinforcement learning is presented in [73], which proposes two algorithms to balance the traffic between macro and femto-cells, and to decrease the call block rate of highly loaded cells; Deep Reinforcement Learning has been applied for CCO [74], to align a group of user signal strength to efficiently support the user scheduling for massive MIMO system, ensuring better tradeoff between system capacity and service coverage than traditional Fixed Optimization and Proportional Fair Optimization. A similar learning based approach addressing massive MIMO beamforming is proposed [75]. A neural network is trained to generate realistic user mobility patterns, which are used by a second neural network to produce relevant antenna diagrams. Similar deep learning techniques have also been applied to the cell coverage compensation in ultra-dense networks (UDN) [76]. Reinforcement Learning for self-healing SON has been studied in [77] and [78], and further enhancements of RRM, in SON with ML, have been also been studied [79]. The work gave an overview of the challenges and opportunities of the application of ML to RRM in 5G. Additionally, authors proposed learning algorithm for resource allocation in 5G [80], [81]. A practical framework based on SON, SDN, NFV, and Big Data is proposed [70], aiming to cluster, forecast and manage traffic from a huge number of BSs with different statistical traffic and multiple RATs. Authors of [82] leverage the recent trend in ML, to ensure coordination and avoid conflicts among different SON functions, with different time scales and inconsistent objectives, which would be the case in the Ultra-Dense Small-Cell Networks. Finally, applying ML for energy saving is another recent use case. Research works [81], [83] are examples of applying ML for self-powered UDN. While authors in [81] apply deep Q-Learning based dynamic resource allocation, authors in [83] use a layered learning solution, where local and network-wide layers interact for providing EE RRM policies for energy harvesting small cells.

V. 5G RAN Adaptive Sharing Enabled by Slicing

In the two previous sections, several efforts have been devoted to draw a 5G RAN landscape, where softwarization or autonomous management are the main building blocks. In
this section, we highlight how these building blocks play in synergy for enabling RAN sharing/network slicing, as well as energy efficiency.

A. Enabling Adaptive RAN Sharing / Application Specific NS

The building blocks for enabling RAN sharing are depicted in Figure 3. In this architecture, the RAN is fully virtualized, thanks to the application of NFV for the virtualization of the RAN functions, the SDR to slice the radio antenna elements (RRHs), and SDN to slice the networking device. The RAN is controlled, by a software defined unified control plane (SD-UCP). NFV/SDR enables RAN sharing by different tenants (MVNOs), while the SD-UCP translates the decisions of the enhanced SO-VRM algorithms, back to the radio physical nodes, to enable the dynamic allocation and flexible management of resources, according to SLA and load from each MVNO.

The use of a unified control plane and one SO-RA entity would allow the adaptability in multi-RA, such that all established slices can share the available bandwidth on the different existing RATs, depending on the MNO policy/MVNO SLA. Moreover, it allows greater efficiency, enabling rational use of resources. As the QoS of users of a given MVNO, should not be affected by activity of users in other MVNOs, dynamic resource allocation should also include a minimum throughput for each isolated MVNO’s application specific slice. Allocation of resources reflects the MVNO’s policy and slice QoS, as well as the MNO’s preference to maximize a certain metric, such as spectral efficiency, energy efficiency, or a trade-off between both. The multi-operator resource allocation aims to ensure appropriate end to end logical slice for each MVNO slice, establishing a service over shared resources, being: RAT setting/bandwidth allocation; a set of NFVs. Table I provides differences, opportunities and arising synergies between softwarization and autonomous management, for enabling adaptive RAN sharing.

B. Enabling Energy Efficient MO-RA Operations

In wireless communications, energy efficiency is defined as the ratio of reliably transmitted data to the total energy consumed (i.e., the number of delivered information bits per energy unit) [84]. Energy aware multi-operators RAN resource management operations rely on powerful SO algorithms that leverage big data and ML and the fully virtualized RAN to allow multiple operators to coexist, and adaptively share RAN resources, while reducing energy consumption. These functions include:

a) EE radio resource allocation operations of the RRH antenna resources, including UE pairing, opportunistic bandwidth allocation on a licensed or unlicensed band, Transmit Power Control, for each MVNO. In that context, radio resource optimization problems should converge to a trade-off between


**TABLE I**

**DIFFERENCES, OPPORTUNITIES AND SYNERGIES OF NFV/WVN SDN, SDR, SON EMPOWERED ML/BIG DATA**

|                               | NFV/WVN | SDN       | SDR       | SON (Empowered with ML/Big Data)                                                                 |
|-------------------------------|---------|-----------|-----------|---------------------------------------------------------------------------------------------------|
| **Scope of Control / Programmability** | Orchestration and management of the different NFVs via MANO framework. | Software Control of networking resources. | Software Control of Radio resources. | Control for the management and optimization of networks elements / RAN-Core. |
| **Virtualization level from network point of view** | Applications/network Function | Packet Flow level | Radio level |                                                                                                   |
| **Possible Deployment in the C-RAN** | Different functions: BBU processing Pool, RAN NFV. | Deployment defining SD-RAN controller; SD-RAN data plane defined by SDN enabled BBU pool and the internetworking SDN switches; SD-RAN application Plane would host the and resource and traffic management operations. | RRH | - SON algorithms can be hybrid combining algorithms that are centralized running at the SDN application plane and requiring global context information, and ones that are decentralized and driven from the network elements. |
| **Major Independent Benefit** | Enable infrastructure partitioning into multiple logical infrastructure considered as the essence of RAN Sharing. | Way to flexible innovation and efficient RAN operations. | Improved spectrum utilization when combined with dynamic coexistence schemes. | Simplification of operational and management tasks in the most complex and heterogeneous networking environments, the improvement of network performance, the reduction of time to market of new services and to OPEX savings. |
| **Synergy Aspects for Energy Efficient Allocations** | RA Controller can be implemented as NFV | Efficiency in the allocation of resources up to Mac layer with respect to different Metrics reflecting VNO/MVNO policy. | Offers extra degree of freedom for the management and allocations of the physical layer resource (Radio). | - Multi-tenant resource allocation that can also leverage spectrum-sharing opportunity supported by SDR as way to improve spectrum utilization. |

meeting the intra-MVNO slice QoS and reducing the consumed energy of the overall RAN. Schemes can include the optimization of the radio allocation operations, not only on intra-operator scale, but also inter-operators. This includes operations such as: optimization of the inter-operator beam-forming vectors, reduction of global power consumption, and
optimization of global bandwidth allocation; all performed with the aim of reducing energy consumption, while meeting the MVNO slice QoS. Another dimension is the selection of mobile UE belonging to a given MVNO to act as a small cell, upon detection of coverage hole. This would save the cost incurred by the addition of a legacy small cell. The challenge would be the development of business models for small cells sharing.

b) Performing efficient VNFs placement at the various physical network locations (for NFV enabled resources), efficient migration from under-utilized and high energy cost to low energy physical locations, and ensuring efficient traffic routing, for minimizing energy consumption. This calls for the use of proactive ML frameworks that can predict future demands and consequently pre-organize optimal NFV placement/migration and optimal routing. An example of NFV optimal placement is the placement of an optimal number of required VMs, on the available BBU pool of physical servers, with the aim of mapping the MVNOs load from virtual RRH antennas to a total minimal required number of BBU VMs. Load variation can then be exploited to reduce BBU’s required cooling and power consumption. Nevertheless, the challenge of this EE multi-operators computational sharing would be to ensure the use of resources from shared location in a trustworthy manner.

Figure 4 shows the high-level description of the framework required to achieve the aforementioned EE operations. SON/ML algorithms use the various context information collected from all network resources as input. Using ML, the collected data will be analyzed, and specific operation optimizations will be developed. Ultimately, the output is a set of optimal parameters to be adapted, by the different physical and virtual RAN resources.

VI. STANDARDIZATION AND RESEARCH ACTIVITIES

In this section, we present the recent standardization efforts related to SDN, NFV, SDR, SON and virtualized 5G RAN.

A. Standardization Efforts

SDN, NFV, and SDR have been the subject of wide standardization activities. Table II includes a summary of the relevant standardization efforts for each technology. Several standardizations bodies have also carried out efforts, toward the combination of SDN and NFV. In [85], the ETSI GS NFV established a report on SDN usage in NFV architectural framework. It considered that SDN brings to NFV a connectivity and service configuration assistance and flexibility, while NFV can provide SDN with service resources. In other perspective, ONF also undertook a step toward investigating the differences and perspectives of a common ground of applications of SDN/NFV [86].

SON, network sharing, SDN and NFV are all technologies that have been either adopted or recognized to be relevant by 3GPP RAN specifications. SON has been considered as the main enabler of the minimization of the lifecycle cost of running mobile networks, through self-configuration, self-optimization and self-healing. SON has been introduced for the first time by 3GPP in Rel. 8 [87], as a key component of LTE network. Additional functionality and enhancements to SON have been added to subsequent releases.

In terms of network sharing, the first form of network sharing appeared in 3GPP Rel. 99, as a passive network sharing. Later, new requirements were needed. As a result, active network sharing appeared and sharing requirements
were inferred into five main scenarios [88]. The 3GPP standardization does not specify how capacity is shared among the several core network operators, competing for radio access. However, 3GPP is progressively adopting the virtualization mechanism, as an emphasis on their evident impact in the move from the network sharing to multi-tenancy concept. Indeed, on the aspect of NFV adoption, starting from the 3GPP Rel. 13, the potential impacts of the inclusion of virtualized core network functions following the ETSI NFV MANO on the existing 3GPP Service and System 5 (SA5) network management architecture has been identified. 3GPP Rel. 14 has introduced three specifications [89], [90], [91] regarding the architecture requirements for virtualized network management. The New Generation Core (NGC) [92], part of 3GPP Rel. 15, has considered the complete separation of control plane from user plane, aligning with the SDN paradigm and a modularization of the control plane, aligning with VNF paradigm, towards increased flexibility. It is clear from the survey of the standardization efforts, networking in general, and mobile networking specifically, is rapidly moving closer towards a fully virtualized/softwarized networking concept. This was also clear from the number of research efforts, listed in previous sections.

### B. Research Activities

Several 5G related research actions have been launched in Europe. The most visible one is Horizon 2020 (H2020). Additionally, within H2020, 5G-PPP (5G Infrastructure Public Private Partnership) [108] was created, as a joint initiative between the European Commission and European ICT industry (ICT manufacturers, Telecommunications operators, service providers, SMEs, and research institutions), mainly to deliver solutions, architectures, technologies and standards for the coming decade (i.e., 5G and beyond). Many research projects have been created in the context of these initiatives.

### Table II: Standardization Efforts

| Technology | YEAR | MAIN CONTRIBUTING/STANDARDIZATION BODIES | EFFORT DESCRIPTION |
|------------|------|----------------------------------------|-------------------|
| SDN        | 2011 | Open Networking Foundation (ONF) [93]  | Standardization of OpenFlow protocol for programming the data plane using the control plane [94]. |
|            |      | Internet Research Task Force (IRTF) [95]| Investigation of scalability perspectives, abstractions and programming language for the SDN. |
|            |      | Software Defined Networking Research work group (SDN-RG) [96] | |
|            |      | Internet Engineering Task Force (IETF) [97] | Investigation of interfaces to routing system. |
| NFV        | 2012 | ETSI NFV ISG (founded by Global telecom and industry players) | Developed the NFV MANO framework [98] for the management and orchestration of the virtual network functions. |
| SDR        | 1987 | Air Force Rome Labs (AFRL). | First SDR design consisting of a programmable modem. |
|            | 1992 | Joe Mitola | Presentation of critical analysis and future directions of software defined radio at the IEEE National Telesystems Conference. |
|            | 1996 | SDR forum/ (WINNF), first industry association [99] | Industrial aspect of SDR and definition of standardization interfaces to facilitate the use of software across different hardware vendors. WINNF is currently working collaboratively with standardization and regulatory bodies. |
|            | 1997 | Joint Tactile Radio System (JTRS) [100] | Definition of a Software Communication Architecture (SCA) to assist in the development of software defined radio communication system. |
| 2001       | GNU RADIO | Foundation of an open source SDR development toolset. |
| SOM        | 3GPP Rel. 8 [87] | Introduction of SON as a key component of LTE network with functionality regarding initial equipment installation and integration. |
|            | 3GPP Rel. 9 [42] | Addition of Self-Optimization. |
|            | 3GPP Rel. 11 [101] | Enhancement of energy saving and functions related to management of heterogeneous networks. |
|            | 3GPP Rel. 12 [102] | Optimization and enhancement of small cells, including deployment in dense area. |
|            | 3GPP Rel. 13 [103] | Improvement of Operations, Administration, and Maintenance (OAM) activities and the LTE operation in unlicensed band. |
|            | 3GPP Rel. 14 [104] | Addressed the use of unlicensed spectrum in a fair manner, support for carrier aggregation, energy efficiency at OAM level. |
|            | 3GPP Rel. 16 [105] | Planned to include a study on SON for 5G elaborating on the requirements of enhancing the SON with the latest advances in AI and machine learning. |
| Network Sharing | 3GPP Rel. 99 [103] | First form of network sharing: Passive Sharing. |
|            | 3GPP Rel. 11 [107] | 3GPP Architecture Working Group SA2 specified two distinct active RAN sharing architectures. |
specifically to bring breakthroughs to the field of network management and self-organization, from which we highlight the most relevant.

One of the first projects is SPEED-5G [109], which is among the projects under the 5G PPP-Phase 1 [108] that focused on the autonomous RRM and self-optimization of spectrum utilization in heterogeneous small cells based networks. The H2020 SELFNET [110] and COGNET [111], as well as FP7 SEMAFOUR [112], are projects that were devoted to self-organized network management. 5G SONNET [113] and SELFNET [110] are H2020 projects that evolved the paradigm of SO RRM, with the proposal of an SDN/NFV architecture. Sesame [114], which is a Phase 1 5G-PPP project, aimed to leverage the NFV/SDN for the implementation of the cloud enabled small cells (CESCs) concept, as a major enabler for both radio access and edge computational capability. Building on the concept of Small Cell-as-a-service (SCaaS), these small cells are envisioned to be shared between multi-tenant and supporting multi-service. The management and orchestration of the platform is ensured through the traditional 3GPP network management elements, as well as the novel recommended functional blocks for the management of the NFVs. FlexRan [115] is a project, co-funded by H2020 and FP7, which proposes a flexible and programmable SD-RAN platform, where the control plane is separated from the data plane through customized southbound API. The platform is envisioned to touch on the SDN application plane, through the support of real-time RAN control and management applications. Aligning with the direction of Elastic 5G RAN, several other projects investigated the SDN/NFV based resource management, such as 5G-PPP 5G-MONARCH [116], COHERENT [117], SELFNET [110], Sesame [118], ESSENCE [119] (a project that builds on the inherited implementation of small cells as NFV from Sesame). The focus on 5G network control for the support of multi-tenancy and network slicing continued with 5G PPP Phases 2 projects, such as 5GCity [120] and SliceNet [121], as well as Marie Curie Innovative Training Network (ITN) projects, such as 5G-AURA [122], and SECRET [123]. Other projects, from the earlier program FP7, were also relevant. For instance, iJOIN focused on the design, network operation and management algorithms required for RAN-as-a-Service (RANaaS) concept [124], while CROWD suggested a hierarchical SDN for the efficient operation of the very dense heterogeneous wireless access networks [125]. Additionally, the 5G Innovation Centre (5GIC) in U.K., hosted by the University of Surrey, represents the world’s largest academic research center devoted to 5G and wireless connectivity.

In addition to European efforts, other global initiatives and institutions devoted efforts towards 5G promotion and development. The 5G initiative across Asia features two main institutions, among others. In China, The IMT-2020 (5G) [126] Promotion group was established with the support from three ministries of China representing a main platform for 5G research and development promotion and including a variety of mobile operators, universities, research institutions and vendors. In Korea, the 5G Forum [127] was founded by the Ministry of Science and ICT, with the aim of contributing to the development of standard and its globalization. Similar vision led to the development of the Fifth Generation Mobile Communications Promotion Forum (5GMF) [128] in Japan. In Brazil, 5G Brazil project was established by several Brazilian mobile operators, as well as vendors, with the aim of driving 5G forward in Brazil. From the USA side, 5G Americas [129] is composed of leading telecommunication service providers and manufacturers, and stands as the industry trade organization. 5G Americas focuses on fostering the advancement of LTE and its evolution to 5G. Joint research activities and efforts towards standardization are conducted through several cooperation projects, such as 5G DRIVE [130], and global events, such as the IEEE world forum [131] and the 5G World Summit [132]. Figure 5 shows a schematic representation of
VII. CLASSIFICATION OF RAN RESOURCE ALLOCATION SCHEMES

Several research works have been devoted to the improvement and innovation of the RAN allocation schemes, using virtualization paradigms, software control, or autonomous management. In this paper, we highlight the state of the art in most of these valuable works and summarize it in Figure 6. We first classify the different efforts, according to the resource to be allocated, as whether such resource: i) belongs to one single mobile network operator (i.e., legacy MNO); ii) is an abstracted resource, meaning a virtualized resource, belonging to one single operator; or iii) is a resource to be allocated following a multi-tenant paradigm. More distinctively, we emphasize on resource allocation and perform a second classification, depending on the context of the resource or the particularity of the operation. The final taxonomy for all the RAN operation schemes is presented, with respect to the technology/approach on which the operation relies, as well as the metric. The considered technologies can either be one of the aforementioned softwarization techniques, autonomous management approaches or hybrid synergies.

Resource allocation aims to ensure the accordance between the provided service to users and the required QoS, by the different global efforts, divided between the European Union (EU) and global initiatives.

![Classification of RAN Resource Allocation works.](image_url)
dynamically and efficiently scheduling the available transmis-
sion resources and assigning resources to the user to satisfy
the required QoS.

Resource definition/scope has evolved with:

a) The Heterogeneity in terms of Radio Access technol-
ogy (RAT), leading to the inclusion of the additional question:
what RAT is better to be selected as a resource, where the
different RAT coexists in a RAN owned by a single oper-
ator. Addressing this question was the objective of research
work falling into subcategory: 1) resource allocation for single
operator network.

b) The application of WNV, SDN and V-RAN to differ-
ent RAN aspects implying the possibility of: i) virtualizing
the multi-RAT; ii) virtualizing hardware/network function
resources; iii) virtualizing the spectrum; and iv) chaining
the virtualized network functions for the establishment of
service-oriented network slicing. Research works, applying
the different virtualization forms to the scope of a RAN owned
by the legacy mobile operator network, are included in subcate-
gory: 2) Abstracted resource allocation for a single operator’s
network.

c) The shift to multi-tenant RAN is mostly enabled via the
synergy between these virtualization paradigms and the au-
nomic management. The same RAN virtualization granularity,
reported in subcategory 2, can be applied in that context of
sub-category 3) Virtual resource allocation for multi-operator
network.

Based on the Previous Points, We Further Detail Our
Classification, as Follows:

A. Resource Allocation for Single Operator Network

Classical resource allocation techniques in LTE are spec-
ified to be handled by the eNodeB and usually consist of
assigning RBs to users. Assignment algorithms are implemen-
tation and vendor dependent. Several research works aimed
to give refinement of the legacy 3GPP RRM specifications
in LTE, to cater for resource requirements of specific LTE
release, specific service, given use-case, or a specific degree
of heterogeneity. Considering heterogeneity as a criterion, we
perform a second classification. Different works focused on
physical resource allocation, in the legacy 4G. For instance,
the legacy packet scheduling procedure is revisited, to ensure
good use of the multiple carriers stipulated by the LTE-
A carrier aggregation deployment [161]. Following another
direction in [162], the scheduling procedure has been formu-
lated as a cross-layer scheme that considers dynamic adaptive
streaming over related information, among scheduling cri-
teria, to improve the transmission of videos over cellular
wireless systems. With the advent of densification in 4G,
several research works suggested using more advanced mech-
anisms, catering for the complexity of the scenario, such
as cluster-based resource allocation [135], and Q-learning for
self-organized resource allocation [136]. Similar efforts were
performed by others teams to leverage the coexistence of dif-
ferent RATs in the scenario of HetNets and design schemes
for optimal RAT selection, towards optimizing user Quality
of Experience (QoE), or increasing revenue. End user driven
RAT selection presents the advantage of requiring less com-
unication and less change in the network. Game theory
framework [163], [164] is one of the most used approaches
for this category. Toward meeting operator’s objective, while
individually maximizing mobile devices’ utility, other works
evolved the user decision delegation to be network assisted,
using Markov decision process [138] (a Markov chain with the
addition of an action model and a performance criterion) and
learning method, where the network would derive the network
information for the mobile device to make decision [140].

B. Abstracted / Virtual Resource Allocation for Single /
Multiple Operator Networks

Different research works shed the light on resource alloca-
tion problems, where a resource can be part of a virtualized
RAT, spectrum, infrastructure or slice. In this subsection, we
give an overview of resource allocation issues, with respect to
this type of virtualized resource. It is important to note that, at
least in the near future, both single as well as multiple operator
type of networks will exist simultaneously.

Virtualized RAT Selection: Virtualized RAT selection will
be different from the concept of multi-RAT selection used
in today’s network. Although in both cases a multitude of
RATs would coexist, a major difference is how RAT con-
trol is ensured. In today’s network, each RAT is controlled
by a specific separate entity. An illustrative example is the
case of LTE control by the MME and PCF and the WLAN
control via the WLAN controller, with an inter-operation
ensured by the Access Network Discovery and Selection
Function (ANDSF), which is defined as a stand-alone server
that assists user devices in the selection and discovery of differ-
ent RATs. On the contrary, RAT virtualization implies shifting
to a unified control plane, where heterogeneity of coexistent
orthogonal and non-orthogonal RAT resources can jointly be
measured and managed, ultimately to a fully driven network
approach that completely shields the physical details of the
user. Building on the existence of a common multi-RAT con-
trol plane for joint management is the widely argued concept
that literature approaches share for RAT virtualization. Yet,
the definition/technology, used for the common control plane,
is a particularity of each work. As a matter of fact, while
authors in [148] define the common multi-RAT control plane
as a virtualized medium access control that converges RAT
protocols and performs inter-RAT resource allocation, it is
realized through a normalization of the heterogeneous radio
resource, based on a deep learning method that derives RATs
consumption model from the accumulated historical data for
efficient radio resources RAT selection, in [147]. In [149],
a virtual BS pool, supporting different RATs, was implemented
thanks to SDR.

RA in Virtualized Infrastructure: Virtualization of the RAN
infrastructure refers to decoupling the functionality from the
hardware of some or all RAN physical entities. This form of
virtualization appeared with the advent of the H-CRAN,
allowing to realize a form of NFV, by decoupling the dis-
tributed baseband, and substituting them by lower power radio
units (RRHs), while mapping their baseband processing to the
cloud to be processed by software-based baseband processing unit, executed by virtual machine, part of a pool, running on general purpose server.

When combined with softwarized control paradigms, such as SDN and SDR, this form of virtualization leads to the already mentioned SD C-RAN. It allows intelligent resource allocation by a software defined virtual resource management entity, located after the BBU pool. This model is different from the traditional concept of distributed resource allocation, handled by the eNodeB. In this context, virtualized resource allocation involves the allocation of virtual radio related resource, e.g. frequency band, power, or resource block, virtual processing resource (vBBU), and virtual computation resource (vCPU). Resource allocation in SD C-RAN would also involve low latency coordination cooperation between cells [151], stemming from the fact that data between BBUs, associated to different cells, can be shared locally, without data transfer between distant cell sites, which eliminates the need for handover [152]. This idea paves the way towards more efficient spectrum, interference, and throughput aware allocation schemes. Distinctively, the exploitation of SDN/NFV/SON in the C-RAN, also, brings about energy efficient allocation of the radio, processing and computational resources. While NFV/SDN processing and computational resources can be scaled up and down and steered to where the demand is, SON enables automatic activation/deactivation of small cells, as suggested in [165]. This way, operator does no longer need to provision the network for maximum peak capacity, which is crucial in the minimization of operational costs. Several works have been investigated on energy efficient allocation that leverages this C-RAN flexibility. For instance, one research work targeted energy and throughput aware radio resource allocations, by the proposal of a SO strategy programmable on SDN that allocates power adaptively to traffic demand in a soft fractional frequency reuse (SFFR) framework [153]. On the other hand, dynamic and energy efficient processing resource allocation in the C-RAN was the focus of [166] that experimentally illustrated their concept of the efficiency of a modeling approach for the joint allocation problem for both radio and processing resources in V-RAN [167].

Beside enabling more efficient resource allocation particularly in terms of energy consumption, another major benefit of that type of RAN infrastructure virtualization, when coupled with the programmable and software control (SDN/SDR), is that it represents a key factor for network sharing and the corner stone of RAN customization on per-tenant basis. Several frameworks have been proposed in the literature to support RRM for multi-operator virtualized environment, focusing on metrics, such as flexibility, energy efficiency and revenue maximization, established on the multi-tenant business model base [118], [155]. Authors, in [155], propose OpenRAN, an SD-RAN architecture for multi-operators, where computing and storage resources are created and allocated via SDN, and radio resources via SDR. Maximizing the revenue for each of the MVNO/MNO, in perspective, along with energy saving through the proposal of a model for a two-level Auction for resource allocation in Multi-tenant C-RAN was investigated [168]. Ensuring an efficient use of resources, while meeting the QoS requirements of the services provided, was the target of the work in [118]. The authors proposed a framework for implementing virtualized RRM/SON functions in multi-tenant small cell networks, as VNFs. In the efforts toward network slicing, the research efforts conceive that some RRM/SON functions would be instantiated per tenant to carry tenant-related RA and other common ones, instantiated to carry common RRM/SON operations.

Network Slice Provisioning: A network slice refers to a managed group of a subset of network functions/VNFs, establishing a logical link customized to meet the need of an industry service or vertical, or operator. Network slice provisioning is a multi-resource allocation operation, where end-to-end resources including a RAT, a bandwidth (spectrum allocation), and a set of NFVs (assignment of virtualized infrastructure), optimized for the slice use-case. Although establishing a network slice can involve virtualized (shared) resources, a network slice is an end-to-end independent logical network. Given that, network slice provisioning is expected to fully leverage the capacity of SDN and NFV.

Research works, addressing network slice provisioning in the context of one single mobile operator, focused on multi-resource allocation for the establishment of service-oriented network slice (i.e. a network slice to serve one type of the 5G services, e.g. ultra-reliable services, ultra-high-bandwidth communication or extremely low latency, to be delivered to a given group of end users). The work in [137] falls in that category, as it proposes an SDN-based network slicing solution, for enabling efficient coexistence of enhanced mobile broadband (eMBB) and IoT services, sharing same RAN, while considering the QoS requirements of both services. The authors validated the approach through emulation techniques, using OAI software and FlexRAN SDN controller. Motivated by similar incentive, research work [142] proposes a service-oriented network resource slicing framework that evolves Time Division Duplex (TDD) network towards 5G, in an SDN based architecture, aiming to improve system performance. The performance was evaluated using system-level simulation.

Other efforts have been devoted for network slice provisioning in 5G multi-tenant. Authors, in [143], propose a framework and a proof of concept of a flexible and dynamic RAN slicing that builds on the FlexRAN concept and adapts the shared resources, based on the tenant SLA and traffic QoS. Towards the same targeted metrics, research works [144] used a different SDN based resource allocation approach that weighs the significance of the resources per given network slice, by measuring the correlation structure of the different QoS requirements of that slice. The slice allocation metric has been shifted toward combining the profit optimization of slice provider/customer and resource efficiency in [145], through the establishment of an optimization framework for slice dimensioning that builds on the tradeoff that exists between these two considered metrics. Focusing particularly on the bandwidth efficiency and processing resources, an optimization framework has been proposed to enable real time slice auto-scaling and tuning [146].
RA in Virtualized Spectrum: Radio spectrum acquisition in cellular networks is costly and governed by licenses and agreements. Spectrum virtualization enables efficient use of the available spectrum, by allowing different radio infrastructure, services, or operators to share the spectrum. Hence, it is considered as the main enabler of spectrum sharing. This form of virtualization implies the isolation and dynamic allocation of spectrum bands to each of these entities. Different technologies and methods have been proposed to ensure this isolation, while sharing resources, including time based, frequency based, space and code based. Other solutions are derived from the concept of spectrum virtualization layer [169], which calls for RF front-end that are entirely decoupled from the protocol and can be programmable, thanks to the SDR paradigm and to dynamic spectrum access schemes, which enable opportunistic spectrum sharing.

Some efforts presented a dynamic spectrum-level slicing algorithm, to share the physical radio resources across the HetNet and proved that it leads to the reduction of dropped packets, along with spectrum efficiency advantage [157]. In contrast to the previous approach, an architectural framework was proposed that combines SDN-SDR and highlights the advantages in terms of bandwidth utilization and latency [160].

Other efforts have focused on spectrum resource allocation, performed by the infrastructure provider/Mobile operator that acts like the facilitator of spectrum sharing between MVNOs [156], [158], [159]. The virtualization technology, on which the allocation of radio resources to the different MVNO relies, differs between different efforts, including spatial multiplexing scheme for the allocation of Massive MIMO cell [158], SDN-enabled spectrum management application [156], and heuristic algorithm [159]. The focus of [158] and [159] was revenue maximization and fairness, through an auction-based resource allocation.

VIII. CLASSIFICATION OF EE RESOURCE ALLOCATION EFFORTS

Optimizing the energy efficiency (EE) aspect of the RAN, by capitalizing on the flexibility offered by virtualization or self-organization, has been widely investigated by various research works, driven by the need for bringing more CAPEX/OPEX savings. We, here, classify those efforts into three separate categories, based on the resource type to be optimized: antenna RRH resource elements, BBU resource elements, or hybrid resource element focusing jointly on RRH and BBU elements.

For each category, we consider different resource allocation tasks, as performance metrics for the qualitative classification. When examining antenna resources allocation works, the different allocation tasks can be a radio resource allocation including PRBs and/or a power allocation to users, depending on their channels’ estimation and required data rate, with the aim of optimizing EE. Other radio related tasks include user pairing to an optimal RRH or to a set of RRH antenna resources, in a coordinated multipoint CoMP and/or beamforming design. Traditionally, the aim of CoMP was the mitigation of interference; hence, the improvement of spectral efficiency and overall system throughput [151]. Indeed, they rely on cooperation principle among RRH antenna elements, in the transmission of precoded data. The signal, observed at each UE, represents a superposition of signals from these multiple active RRHs, leading to an interference that is out-of-phase. A major challenge used to be the considerable power consumed, by the several-coordinated RRHs. Recently, several efforts have proved that CoMP can be used, while also bringing EE to the C-RAN [170], [171], [172], [173], [174], [175]. This can be achieved by using the coordinated beamforming CoMP, where RRH antennas can steer the signals in narrow beams, leading to improved signal to noise ratio; hence, lower transmission power [173]. Also, automatically powering on/off RRHs, according to load variation, is another way that leverages self-optimization capability for achieving EE. This applies for CoMP scheduling, as well as all the other antenna resource allocation efforts; thus, we consider it among other performance metrics.

For BBU computation resource allocation, performance metrics include, among others: RRH-BBU mapping or per task BBU scheduling, the consideration of load dependent baseband pool server (BPS), virtualization support, as well as the automatic switching on/off of BB servers/VMs. In RRH-BBU mapping, the workload assignment is done according to one-to-one or many-to-one mapping relationship, with the aim of minimizing the number of active BBU instances, reducing energy consumption. In BBU scheduling, the optimal BB computation resources are dimensioned to meet the requirement of RRH/UE dynamic flows. Requirement consists of ensuring a trade-off between energy efficiency and latency/QoS.

Regarding hybrid resource allocation studies, we examine if they include all the above performance tasks. Additionally, for all categories, we shed the light on another key performance metric, namely the contribution of the work to multi-operator RA and simulation/implementation based evaluation.

We start with EE Antenna resource allocation schemes, which represent most of the EE allocation contributions. As explained, in such category, resource management schemes perform the optimization of antenna elements serving UEs, targeting the improvement of energy-efficient power allocations, user paring to RRH or to a set among different RRHs (CoMP/beamforming optimization), allocation of physical resource blocks, or optimization of the number of active RRH antenna elements. Joint allocation has been formulated into a mixed combinatorial or non-convex optimization problem, in many of these efforts, and solved after decomposition to elementary steps. RRH selection and RRH on/off problems have been solved via heuristic algorithms, such as greedy activation ([73], [176], [178]). Power/bandwidth allocation has been solved, by relaxation and decomposition techniques [179] or game theory [180]. Table III summarises the qualitative classification of EE antenna resources allocation works.

Other EE RM schemes have focused on the BBU resources optimization solely, driven by the need to reduce the significant power consumption of the BBU pool. These efforts targeted the minimization of the number of BBUs,
when mapping RRHs workloads to BBUs, mostly by formulating the problem as a bin packing problem and solving it via heuristic algorithms [193], [194], or meta-heuristic [2]. Most of these works considered the many to one mapping [2], [195], [196], [197], by considering RRH that can be grouped into a cluster and mapped to one BBU server. The BPM formulation, in these works, considers fixed number of UEs, or resource blocks per BBU solely. Although these efforts proved to be efficient for minimizing the active number of BBUs and considered the BBUs processing virtualization-induced gain, they do not account for the user QoS requirement and the level of interference in the network, when forming RRH clusters. Recent works included the QoS constraint by formulating the problem to modified BPM [198] or to set partitioning problem (SPP) [199]. Other works considered the optimization of the number of virtual machines or virtual BBUs, solely [200]. In [201], joint service time and power minimization have been targeted in a per BBU Joint load scheduling scheme. Table IV summarizes the qualitative classification of EE BBU resources allocation works.

Unlike the majority of works that approached the energy efficiency optimization on the C-RAN and overlooked computational or radio optimization aspects, recent research works [1], [108], [207], [209], [210] undertook steps towards designing EE multi-resource allocation. This category of research caters for the nature of resources in cloud-based RAN environment and provides a more overall optimization, by considering both resource aspects; hence, very relevant to our research work. Therefore, besides including their qualitative classification in Table V, we provide (in Table VI) a summary of their main contributions in terms of allocation tasks, overall energy efficient aspect, and RA problem formulation/solution. As depicted in Table V, all research works, included in this category, focused on at least two allocation tasks, touching on the optimization of radio and BBU computation resources and devised a 2-step resolution method for the optimization problems. For instance, a hybrid allocation was proposed [211], targeting the optimization of the virtual BBU (vBBU) computation capacity, the set of selected RRHs, and the beamforming vectors at the active RRHs, in order to minimize the overall system power consumption for C-RAN. The optimization problem has been formulated as a mixed-integer nonlinear programming (MINLP). In the first stage, the power consumed, by the BBU, is minimized, by adjusting the VMs’ capacity according to the load. To solve the problem, the authors considered two approximation approaches. After solving the power minimization problem and getting the optimal achievable data rate, an RRH selection is performed, based on the Shaping-and-Pruning algorithm. In the same line, scheduling of RRH in CoMP is formulated [208], to a network-wide beamforming vectors optimization problem, and it is solved by weighted minimum mean square error (WMMSE) approach, while the BBU scheduling problem is formulated as a bin packing problem and solved by heuristic algorithm based on Best-Fit Decreasing (BFD). Different from the former approach which focused on BBU capacity scheduling, the later considered per RRH-BBU mapping task, where BBUs are VMs, each serving one workload from a UE. Only few hybrid resource allocations [207] considered mapping load from many RRHs to BBUs. None of the hybrid research works considered leveraging the baseband servers’ virtualization induced gain and including in the cross-layer optimization problem dimensioning the optimal number of virtual machines, ensuring a trade-off between: reduced power consumption, and the latency with respect to the different QoS. We elaborate that some work considered VMs dimensioning [200],

| Ref | [170], 2014 | [171], 2017 | [181], 2015 | [182], 2015 | [177], 2015 | [183], 2016 | [175], 2016 | [181], 2016 | [186], 2016 | [178], 2016 | [187], 2016 | [181], 2016 | [189], 2017 | [191], 2017 | [191], 2018 | [179], 2018 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Subchannel/PRB allocation | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Power allocation | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| CoMP | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Beamforming | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| User-RRH pairing/ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| RRH ON/OFF | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| InP/MVNO | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Simulation Based Evaluation | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Evaluation Based Implementation | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
without catering for scheduling or mapping tasks of the RRHs’ workloads to the vBBU, neither for the optimal allocation of antenna resources elements. Table V lists the qualitative classification of EE hybrid resources allocation works. As mentioned before, due to the high relevance of these hybrid EE RA works, Table VI provides further brief summary of the different research efforts towards providing energy efficient hybrid resource allocations schemes.

### IX. OPEN ISSUES AND CHALLENGES

Based on our thorough analysis of state of the art and existing literature, in the current section, we highlight some open issues in this field, which represent some of the main topics of future research in the general area of wireless networking.

#### A. Multi-Operator Hybrid RA in C-RAN Environments

Throughout this survey, we can see that although the numerous research effort brought about valuable contributions in terms of hybrid EE allocation, there are still multiple unanswered challenges that need to be addressed, including EE multi-operator hybrid RA in heterogeneous RAN environment. Indeed, among all selected research works, the proposal in [191] was the only one to introduce an EE multi-operator RA scheme. The scheme leverages the sleep mode criteria in BSs of cooperating MVNOs and builds on inter-band non-contiguous carrier aggregation, by segmenting the licensed spectrum of each into private and shared bands, in which UE access is mutually exclusive. While spectrum is improved as a result of the multi-operator spectrum scheme, energy efficiency is improved thanks to the optimization of intra and inter-operators beamforming vectors. However, the work does consider the RRH antenna resources solely and does not address how the BBU computational resources can be shared among operators in an EE way. As already mentioned, the design of EE-MO-RA schemes calls for the investigation of new energy aware inter-operator joint optimization. While radio optimization would approach, among others, inter-operator beamforming vector optimization, power control, and bandwidth allocation operations, computational optimization would encompass VNF/VM optimal placement. The unified target of these sub-operations would be the reduction of the overall power consumption of the RAN, while meeting the per-slice QoS.

#### B. Heterogeneity in Terms of RAT

Heterogeneity is another dimension that would allow coexisting operators to enhance their performance, with respect to energy, interference, and throughput. Nevertheless, from the literature survey, we could only identify one proposal that investigates, in the context of non-shared RANs, the inclusion of heterogeneity to the hybrid allocation problem [210]. The proposal considers the joint assignment of radio aspects allocation, considering RRH and low-power femto access point (FAP), as well as computation aspect allocations, by performing VNF migration for EE along with RRH to FAP offloading. Aligning with this logic, future contributions can address this joint EE allocation problem catering for the coexistence of multiple operators and different RATs. The question to answer is how these different available RATs can be exploited by the coexisting MVNOs, to meet the growing demand and the expansion of applications, while ensuring optimal energy efficiency. This can be achieved by the design of new schemes for sharing, catering for the radio, as well as the computational resources of the heterogeneous V-RAN. When formulating this problem, a foreseen challenge would be transposing RAT virtualization and cooperation, in terms of not only cooperative radios, but also cooperative VMs.

#### C. On-Demand/Mobile Small Cells

Few recent research efforts propose the idea of using mobile small cells that can be set up on-demand [23], [216] using existing UEs, which can provide service using an array of RATs to choose from. Although this concept offers many advantages, it also raises several challenges. One challenge is related to determining when a small cell is required in a certain location by relying on the number of connected devices, their activities (i.e. current services), mobility, etc. The processing
and communication capabilities of the current smart devices facilitate this, but the methods still need to be researched and developed. A second challenge facing such promising concept is determining what communication technologies to use for the provisioning of the fronthaul, in addition to the one used by the small cell’s RAT. In particular, the synchronizing of the technologies used at both ends is required in order to optimize the overall resource usage, by minimizing interference, while maximizing the QoS offered.

In addition to interference management among different small cells, there are technological challenges that require particular attention, so that the concept can achieve its high potential gains. Mobility management, in such highly random and variable environment, is also one of the biggest research challenges, as the network attachment points are typically considered to be static, facilitating frequency assignment and interference management. An obvious solution would be limiting the use of small cells to low-mobility devices. Ultimately, a future research direction is foreseen, when dealing with adding mobile small cells belonging to a particular MVNO, among several coexisting MVNOs, in the scope of a shared RAN infrastructure. New business models and new sharing schemes, catering for the mobile small V-RAN complexity and the wireless fronthaul nature and constraints, need to be established in support of this use case. At the same time, the addition of user owned devices, which are heterogeneous in nature, brings challenges from the point of view of availability, trust and accountability. All these challenges require the development of solutions that orchestrate and account the effectively provided capacity, as well as the effective resources used by each user.

D. Fronthaul Provisioning

In mobile networks up till recently, the challenge of meeting the required high data demands was concentrated on the RAN side, since the backhaul was provided, using wired optical connectivity that usually provides much higher capacity than the RAN. Recently, mobile networking has moved towards wireless fronthaul concept, to meet the envisioned flexibility of data demands, in addition to avoid the high costs of CAPEX required for the planning and provisioning of wired fronthaul.

In this paper, we have focused on the challenges arising when allocating RAN resources in shared multi-tenant environments with a particular focus on BBUs and RRHs resources; however, it is very important to elaborate that the construction of an efficient wireless fronthaul is a major challenge, in the current and future mobile networks. In order to meet the network KPIs, fast and reliable fronthaul connections should be established between the RRHs and BBUs, which should also meet strict performance requirements, such as capacity, delay, and jitter, etc. Different research works [217], [218] agreed that the fronthaul design is a major challenge for the practical implementation of the next generation mobile networks. Wired fronthaul designs present the advantage of providing high data rate connection between fixed stations; however, it incurs high costs and does not meet the required flexibility of future networks. To overcome the cost drawback, wireless transport technologies have been proposed as a more flexible and cost effective solutions. Converged optical and wireless design has been the subject of multiple recent contributions [218], [219], [220]. These efforts bring the vision of a fronthaul connectivity strategy that varies based on the targeted use-case (i.e., application/scenario). However, this vision still presents a major challenge to be considered for the design of Multi-Operator and Multi-RAT resource sharing schemes, especially when considering the type of connection to be used on the fronthaul links as one dimension of the optimization, while taking into account the required data rates and delays. Fortunately,

| TABLE V | QUALITATIVE CLASSIFICATION OF EE HYBRID RESOURCE ALLOCATION WORKS |
|---------|--------------------------------------------------|
| EE Hybrid Resource Allocation | Referenced References |
| Subchannel/PRP allocation | ✓ | ✓ | ✓ | ✓ | ✓ |
| Power allocation | ✓ | ✓ | ✓ | ✓ | ✓ |
| CoMP | ✓ | ✓ | ✓ | ✓ | ✓ |
| Beamforming | ✓ | ✓ | ✓ | ✓ | ✓ |
| User -RRH pairing | ✓ | ✓ | ✓ | ✓ | ✓ |
| BBU Scheduling | ✓ | ✓ | ✓ |
| RRH-BBU Mapping | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Virtualized BBUs | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Load dependent BPS Power Consumption | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| RRH ON/OFF | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| BBU ON/OFF | ✓ | ✓ | ✓ | ✓ | ✓ |
| InP/MVNO | ✓ | ✓ | ✓ | ✓ |
| Simulation Based Evaluation | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Implementation Evaluation Based | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
the fronthaul design can be more optimized with the addition of existing technologies, such as distributed caching, data compression, and mobile edge computing for the reduction of traffic amount on the fronthaul. Additionally, with the emergence of mmWave technology integrated with multi-hop communications, wireless fronthaul can take advantage of such new area, exploring its potential to provide very high speed and low delay communications.

| Ref       | Allocation Tasks/Overall EE Aspect                                                                 | Resource Allocation Problem Formulation / Solution                                                                 | Observation                                                                                                                                 |
|-----------|--------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------|
| [212], 2014 | Allocation of frequency resource among cells in H-2CRAN, while mitigating co-channel interference and reducing energy consumption, by decreasing the number of BBUo   | The frequency reuse and allocation has been performed using a traffic aware graph coloring method.               |                                                                                                                                            |
| [211], 2015 | Cross-layer resource allocation including the optimization of VMs computation capacities in the BBU pool, the set of selected RRHs, and the beamforming strategies at the active RRHs in order to minimize the overall system power consumption for C-RAN. | The cross-layer allocation is formulated as an MINLP, by minimizing the system power consumption induced from the BBU VM, fiber fronthaul link and the transmission power on the RRHs. To solve the problem, the paper considers 2 approximation approaches. After solving the power minimization problem and getting the optimal achievable data rate, an RRH selection is performed based on the Shaping-and-Pruning algorithm. | Includes in the allocation problem the dynamic adjustment of the BBU VMs’ computational capacity. The VM power consumption is a function of the former. |
| [207], 2016 | Downlink OFDMA power and resource allocation as well as BBU-RRH assignment in C-RAN. Minimal power consumption and significant BBU’s savings are achieved, as a result of power on/off RRH, power allocation and the possibility to map several RRH to one single BBU. | Resource and power allocations are formulated into a mixed integer linear problem for real-time fluctuating traffic of UEs. The problem is solved using a dynamic approach, based on the branch and cut algorithm. The optimal number of BBUs is computed. BBU-RRH assignment problem is formulated to Knapsack model and solved with the linear solver IBM CPLEX. | RA in this work includes user resource allocation and its pairing to UE. Mapping many low load RRH to one BBU is considered. |
| [11], 2018 | Cloud assisted management for power saving through BBU scheduling including the possibility to switch off low traffic BBU’s beamforming and LTE-WiFi spectrum sharing. | A graph-coloring scheme is developed to label new formulated fronthaul clusters of femtocells using power as the performance metric. Additional power savings are obtained through efficient allocations of the virtualized baseband units (BBUs) subject to the arrival rate of active fronthaul interfacing requests. Moreover, the proposed solutions are used to reduce power consumption for virtualized LTE networks operating in the Wi-Fi spectrum band. | The work includes the induced power saving reduced from LTE operating on unlicensed band.                                                   |
| [213], 2016 | Joint resource scheduling scheme considering both the computation resources in BBUs and the antenna resources provided by RRHs as the beamforming vector, in time varying Cloud-RAN, with delay sensitive traffic, targeting energy efficiency and traffic delay trade-off. | Using a Lyapunov optimization, the problem is partitioned to two sub-problems. The first sub-problem is derived as convex optimization problem and solved with a convex solver. The second sub-problem uses WMMSE approach to obtain network-wide energy efficient beamforming vectors, taking the limited fronthaul capacity and UE QoS requirements into consideration. It is solved by the convex solver CVX. | The scheme is proposed for the downlink.                                                                                                 |
| [108], 2017 | Scheduling of RRH antenna resources in a CoMP transmission as well as BBU computation resources. Targeting the minimization of number of BBUs. | The scheduling of RRH in CoMP is formulated to a network-wide beamforming vectors optimization problem, and it is solved by weighted minimum mean square error (WMMSE) approach. An algorithm to get RRH-Ue clusters is proposed as a solution, based on the optimized beamforming vector. | A BPS is defined to have a number of BBUs each hosting a number of VM. A VM can serve one UE.                                               |
| [214], 2017 | Dynamic coordination scheme based on witch eNodeB’s dynamically agree on the use of resource blocks. Energy efficiency achieved via ON/OFF activation and also reduced from the use of fewer RBs to serve the same traffic. | Coordinated scheduling has been formulated to a Quadratic Assignment Problem (QAP) and linearized by introducing overlap vectors and solved by a general-purpose solver CPLEX. | RRH and BBU on/off is enabled by the prototype. However, the work does not include algorithm that conclude the minimum number of nodes. |
| [215], 2018 | QoS-aware joint BBU-RRH mapping and user association in C-RAN, aiming to minimize the system cost of both RRHs and the BBU pool. | The problem has been formulated into an integer linear programming (ILP) problem. Suboptimal algorithms have been proposed as a solution to avoid computational complexity. |                                                                                                                                            |
| [209], 2018 | Joint optimization of the remote radio head (RRH) selection and computing resource provisioning. | The energy-efficient C-RAN virtualization problem is formulated and solved in two steps. The first step is the RRH grouping algorithm to cluster RRHs into groups. Then a mapping between each RRH cluster and a vBBU is performed. In the second step, the RRH selection is performed through the optimization of cooperative beamforming. | The paper considers the H-CRAN system and possibility of offloading RRH traffic to low power femto access points (FAPs) tier. |
| [210], 2019 | Jointly assigning the (RRHs/APs), sub-carrier, transmit power, RRH, front-haul link and BBU. BBU-RRH mapping includes the migration of VNf's from under-utilized and high energy cost BBUs to low energy BBUs. Also associating UEs to the under-utilized BBUs and their related RRHs can be offloaded to (FAPs) tier. Underutilized BBU can be switched off. These two schemes improve network utility and EE. | The energy efficient allocation tasks are formulated to an energy efficient optimization problem with is non-convex and NP-hard, suffering from high computational complexity. To solve it, authors followed an efficient two-step iterative algorithm based on the successive convex approximation (SCA) and complementary geometric programming (CPG). |                                                                                                                                            |
E. Security Trustworthy Use of the V-RAN Shared Resources

As in all cloud environments, security using the virtualized shared mobile network raises challenges when it comes to providing privacy to MVNOs and their users. Recent efforts [221], [222] revealed that ensuring secure isolation of network slicing in multi-tenant networks is still open research challenge. The noisy neighbor issue, present in most multi-tenancy (or even multi-process) infrastructures, can introduce side effects into each tenant, due to contention at the virtualized resources. Explicit attacks, in the form of a Denial of Service (DoS), or simply an explicit degradation of services from a competitor, is envisioned. In some cases, shared caches or shared structures can provide the opportunity to create side channels funnelling information between tenants, which can include sensitive information from clients, or operations. This situation will be particularly more critical in the light of the application of energy efficient multi-operator schemes, calling for more cooperation and sharing of the virtualized infrastructure, with the aim of decreasing energy consumption. This challenge would be more accentuated, when the sharing of mobile small cells is added to the V-RAN scenario.

F. Development of Implementation Based Platforms

Finally, concluding from our thorough survey, we could identify that there is a lack in terms of implementation-based evaluation platforms, where potential MO-EE hybrid allocation schemes, building on SON capabilities and softwarization architecture advantages, can be emulated and evaluated. We believe that the shift to multi-tenancy can be considerably eased, if research community joins efforts towards designing emulators, combining these multi-capabilities, in addition to building solid business models.

G. Conclusion

In this article, we studied how softwarization and autonomous management schemes can be harnessed to enable adaptive sharing among MVNOs, while potentially enhancing the energy efficiency, as desired by MNOs. Initially, we presented research efforts, focusing on densification and heterogeneity, towards meeting the capacity requirement foreseen within future mobile networks. Thereafter, the survey discussed softwarization and autonomous management technologies, highlighting main and recent efforts addressing autonomous management technologies and how these paradigms can bring profitability requirements and tackle the induced complexity from densification and heterogeneity solution. The survey also highlighted standardization and research projects, emphasizing the application of softwarization technologies and autonomous management technologies to the RAN of cellular mobile networks, in Europe and worldwide. Moreover, we explain how technologies from both paradigms can fit the jigsaw of the adaptive sharing in RAN and setting application specific network slices, and how they can bring energy savings, while meeting the application demand of each of the coexisting MVNOs. Focusing on one particular use-case, we present a high-level description of an MO-EE hybrid resource allocation framework, leveraging the aforementioned paradigms, along with heterogeneous mobile services, further including the so called on-demand small cells. Although the survey concentrates on research efforts addressing energy efficient resource allocation in shared RANs among multi-operator, it is important to emphasize, here, that single operator networks will continue to exist in the future, especially in the era of 5G. To provide a deep analysis of the different resource allocation schemes, we provide a survey and classification of the RAN resource allocation schemes with respect to the context of the resource (shared or non-shared), the used technology and the targeted metric. A second survey is presented to classify and provide a qualitative analysis of EE RA schemes, existing in the literature. Finally, based on the presented survey and analysis, we concluded by identifying the open research challenges and questions, that require answers from the research community, and main players in the field. Those challenges would be steering future research in the area.

| Acronym | Definition |
|---------|------------|
| BS      | Base Station |
| BBU     | Baseband Units |
| CCO     | Capacity Coverage Optimization |
| EE      | Energy Efficiency |
| HetNets | Heterogeneous Network |
| ML      | Machine Learning |
| MNO     | Mobile Network Operators |
| MO-RA   | Multi-Operator Resource Allocation |
| MVNOs   | Mobile Virtual Network Operators |
| NFV     | Network Function Virtualization |
| QoS     | Quality of Service |
| RA      | Resource Allocation |
| RAN     | Radio Access Network |
| RANaaS  | RAN-as-a-service |
| RAT     | Radio Access Technology |
| RRH     | Remote Radio Head |
| RRM     | Radio Resource Management |
| SDN     | Software Defined Network |
| SDNC    | Software Defined Network Control |
| SD-RAN  | Software Defined-RAN |
| SD-UCP  | Software Defined Unified Control Plane |
| SDR     | Software Defined Radio |
| SLA     | Service Level Agreement |
| SON     | Self-Organizing Networks |
| SO-VRM  | SO-Virtual Resource Management |
| V-RAN   | Virtual RAN |
| VNF     | Virtual Network Functions |
| WNF     | Wireless Network Virtualization |
"Study on NM centralized coverage and capacity optimization (CCO) Description; Summary of Release 14 Work Items (Release 14) "Service aspects and requirements for network sharing (Release 10)," 3GPP Release 16: Study Items and Road Map 5G-PPP P. Neves et al. Ö. Bulakci et al. EU FP7 SEMAFOUR Project X. Foukas, N. Nikaein, M. M. Kassem, M. K. Marina, and et al. S. Mumtaz et al. "Self-organization towards reduced cost and energy per bit for future emerging radio technologies—SONNET," in Proc. IEEE Globecom Workshops, Dec. 2017. I. Giannoulakis et al., "Enabling technologies and benefits of multi-tenant multi-service 5G Small Cells," in Proc. Eur. Conf. Netw. Commun. (EuCNC), Jun. 2016, pp. 42–46. X. Foukas, N. Nikaein, M. M. Kassem, M. K. Marina, and K. P. Kontovas, FlexRAN: A flexible and programmable platform for software-defined radio access networks," in Proc. Int. Conf. Emerg. Netw. Exp. Technol., 2016, pp. 427–441. A. Balasci et al., "Overall 5G-MoNArch architecture and implications for resource elasticity," in Proc. Eur. Conf. Netw. Commun. (EuCNC), 2018, p. 5. A. Kostopoulos et al., "Scenarios for 5G networks: The COHERENT approach," in Proc. Int. Conf. Telecommun. (ICT), May 2016, pp. 1–6. J. Perez-Romero, O. Sallent, R. Ferrus, and R. Agustí, "On implementing RRM/SON in virtualized multi-tenant small cell networks," in Proc. Eur. Conf. Netw. Commun. (EuCNC), Jun. 2016, pp. 1–6. I. P. Chochliouros, et al., "Using small cells for enhancing 5G network facilities," in Proc. IEEE Conf. Netw. Function Virtual. Softw. Defined Netw. (NFV-SDN), Nov. 2016, pp. 264–269. I. Neokosmidis et al., "Are 5G networks and the neutral host model the solution to the shrinking telecom market," in Proc. Int. Conf. Artif. Intell. Appl. Inno., 2018, pp. 70–77. Q. Wang et al., "Slicenet: End-to-end cognitive network slicing and slice management framework in virtualised multi-domain, multi-tenant 5G networks," in Proc. Broadband Multimedia Syst. Broadcast. (BMSB), Jun. 2018, pp. 1–5. Application-Aware User-Centric Programmable Architectures for 5G Multi-tenant Networks, 5G-ARA Project, H2020, CORDIS, European Commission. Application-Aware User-Centric Programmable Architectures for 5G Multi-Tenant Networks, 5G-ARA Project, H2020, CORDIS, European Commission. Application-Aware User-Centric Programmable Architectures for 5G Multi-Tenant Networks, 5G-ARA Project, H2020, CORDIS, European Commission. Application-Aware User-Centric Programmable Architectures for 5G Multi-Tenant Networks, 5G-ARA Project, H2020, CORDIS, European Commission. Application-Aware User-Centric Programmable Architectures for 5G Multi-Tenant Networks, 5G-ARA Project, H2020, CORDIS, European Commission. Application-Aware User-Centric Programmable Architectures for 5G Multi-Tenant Networks, 5G-ARA Project, H2020, CORDIS, European Commission. Application-Aware User-Centric Programmable Architectures for 5G Multi-Tenant Networks, 5G-ARA Project, H2020, CORDIS, European Commission. Application-Aware User-Centric Programmable Architectures for 5G Multi-Tenant Networks, 5G-ARA Project, H2020, CORDIS, European Commission. Application-Aware User-Centric Programmable Architectures for 5G Multi-Tenant Networks, 5G-ARA Project, H2020, CORDIS, European Commission. Application-Aware User-Centric Programmable Architectures for 5G Multi-Tenant Networks, 5G-ARA Project, H2020, CORDIS, European Commission. Application-Aware User-Centric Programmable Architectures for 5G Multi-Tenant Networks, 5G-ARA Project, H2020, CORDIS, European Commission. Application-Aware User-Centric Programmable Architectures for 5G Multi-Tenant Networks, 5G-ARA Project, H2020, CORDIS, European Commission.
G. Dandachi, T. Chahed, S. E. Elayoubi, N. C. Taher, and Z. Fawal, M. Moltafet, S. Rezvani, N. Mokari, M. R. Javan, and E. A. Jorswieck, E. Aryafar, A. Keshavarz-Haddad, M. Wang, and M. Chiang, "RAT et al.

J. Li, M. Peng, J. Wu, M. Peng, and P. Zhang, "Queue-aware energy-efficient joint remote radio head activation and beamforming in cloud radio access networks," IEEE Trans. Wireless Commun., vol. 15, no. 6, pp. 3880–3894, Jun. 2016.

D. Zeng, J. Zhang, S. Guo, L. Gu, and K. Wang, "Take renewable energy into CRAN toward green wireless access networks," IEEE Netw., vol. 31, no. 4, pp. 66–68, Jul. 2017.

Y. L. Lee, L.-C. Wang, T. C. Chua, and J. Loo, "Joint resource allocation and user association for heterogeneous cloud radio access networks," in Proc. Int. Teletraffic Congr. (ITC), Sep. 2016, pp. 87–93.

G. K. Tran, H. Shimodaira, R. E. Rezagah, K. Sakaguchi, and K. Araki, "Dynamic cell activation and user association for green 5G heterogeneous cellular networks," in Proc. IEEE PIMRC, 2015, pp. 1–5.

P. Luong, L.-N. Tran, C. L. Despins, and F. Fagnoni, “Joint beamforming and remote radio head selection in limited fronthaul C-RAN,” in Proc. IEEE VTC-Fall, 2016, pp. 1–6.

T. Kim and J. M. Chang, “QoS-aware energy-efficient association and resource scheduling for HetNets,” IEEE Trans. Veh. Technol., vol. 67, no. 1, pp. 650–664, Jun. 2018.

S. Zhou, M. Dong, K. Ota, G. Wang, and L. T. Yang, “Energy-efficient resource allocation for D2D communications underlaying cloud-RAN-based LTE-A networks,” IEEE Internet Things J., vol. 3, no. 3, pp. 428–438, Jun. 2016.

M. Peng, K. Zhang, J. Jiang, J. Wang, and W. Wang, “Energy-efficient resource assignment and power allocation in heterogeneous cloud radio access networks,” IEEE Trans. Veh. Technol., vol. 64, no. 11, pp. 5275–5287, Nov. 2015.

S. Luo, R. Zhang, and T. J. Lim, “Downlink and uplink energy minimization through user association and beamforming in C-RAN,” IEEE Trans. Wireless Commun., vol. 14, no. 1, pp. 494–508, Jan. 2015.

K. Wang, K. Yang, and C. S. Magurawalage, “Joint energy minimization and resource allocation in C-RAN with mobile cloud,” IEEE Trans. Cloud Comput., vol. 6, no. 2, pp. 760–770, Jul. 2018.

Z. Yu, K. Wang, H. Ji, X. Li, and H. Zhang, “Joint user association and downlink beamforming for green cloud-RANs with limited fronthaul,” in Proc. IEEE Globecom, Dec. 2016, pp. 1–6.

B. Fan, H. Tian, and X. Yan, “Resource allocation in a generalized LTE air interface virtualization framework exploiting user behavior,” EURASIP J. Wireless Commun. Netw., vol. 2016, no. 1, p. 107, Dec. 2016.

H. M. Soliman and A. Leon-Garcia, “QoS-aware joint RRH activation and clustering in cloud-RANs,” in Proc. IEEE Wireless Commun. Netw. Conf., Apr. 2016, pp. 1–6.

M. Peng, Y. Yu, H. Xiang, and H. V. Poor, “Energy-efficient resource allocation optimization for multimedia heterogeneous cloud radio access networks,” IEEE Trans. Multimedia, vol. 18, no. 5, pp. 879–892, May 2016.

S. Wang and Y. Sun, “Enhancing performance of heterogeneous cloud radio access networks with efficient user association,” in Proc. IEEE ICC, May 2017, pp. 1–6.

B. Wang, Q. Yang, L. T. Yang, and C. Zhu, “On minimizing energy consumption cost in green heterogeneous wireless networks,” Comput. Netw., vol. 129, pp. 522–535, Dec. 2017.

C. Pan, H. Zhu, N. J. Gomes, and J. Wang, “Joint precoding and RRH selection for user-centric green MIMO C-RAN,” IEEE Trans. Wireless Commun., vol. 16, no. 5, pp. 2891–2906, May 2017.

J. Opadere, Q. Liu, and T. Han, “Energy-efficient RRH sleep mode for virtual radio access networks,” in Proc. IEEE Globecom, Dec. 2017, pp. 1–5.

N. A. Chughaita, M. Ali, S. Qaisar, M. Imran, M. Naeem, and F. Qamar, “Energy efficient resource allocation for energy harvesting aided H-CRAN,” IEEE Access, vol. 6, pp. 43990–44001, 2018.

T. Sigwele, A. S. Alam, P. Pillai, and Y. F. Hu, “Evaluating energy-efficient cloud radio access networks for 5G,” in Proc. IEEE Conf. Data Sci. Data Intensive Syst., Dec. 2015, pp. 362–367.

H. Liu, K. Wang, H. Ji, and V. C. M. Leung, “Energy saving in C-RAN based on BBU switching scheme,” in Proc. IEEE Int. Conf. Netw. Inform. Syst. Infr. Content (IC-NIDC), Sep. 2016, pp. 44–49.

Z. Kong, J. Gong, C. Xu, K. Wang, and J. Rao, “eBaser: A baseband unit cluster testbed to improve energy-efficiency for cloud radio access networks,” in Proc. IEEE ICC, Jun. 2016, pp. 4222–4227.

S. R. Aldeaebool and M. F. Abbdo, “Reducing power consumption by dynamic BBUs-RRHs allocation in C-RAN,” in Proc. Telecommun. Forum (TELFOR), 2017, pp. 1–4.
[197] W. Al-Zubaidi and H. S. Al-Raweshidy, “A parameterized and optimized BBU pool virtualization power model for C-RAN architecture,” in Proc. IEEE EUROCON, Jul. 2017, pp. 38–43.

[198] K. Boulou, M. El Helou, and S. Lahoud, “RRH clustering in cloud radio access networks,” in Proc. Conf. Appl. Res. Comput. Sci. Eng. (ICAR), Oct. 2015, pp. 1–5.

[199] J. G. Tang, W. P. Tay, and T. Q. S. Quek, “Cross-layer resource allocation in heterogeneous cloud radio access networks,” in Proc. IEEE Conf. Cloud Netw. (CloudNet), Sep. 2017, pp. 83–88.

[200] R. S. Alhumaima, R. K. Ahmed, and H. S. Al-Raweshidy, “Maximizing the energy efficiency of virtualized C-RAN via optimizing the number of virtual machines,” IEEE Trans. Green Commun. Netw., vol. 2, no. 4, pp. 902–1001, Dec. 2018.

[201] L. Ferdouse, W. Ejaz, A. Anpalagan, and A. M. Khattak, “Joint workload scheduling and BBU allocation in cloud-RAN for 5G networks,” in Proc. Symp. Appl. Comput., 2017, pp. 621–627.

[202] S. Namba, T. Warabino, and S. Kaneko, “BBU-RRH switching schemes for centralized RAN,” in Proc. Conf. Int. Conf. Commun. Netw. China, Aug. 2012, pp. 762–766.

[203] M. Khan, R. S. Alhumaima, and H. S. Al-Raweshidy, “Reducing energy consumption by dynamic resource allocation in C-RAN,” in Proc. Eur. Conf. Netw. Commun. (EaCNC), 2015, pp. 169–174.

[204] M. Y. Lyazidi, L. Giupponi, J. Mangues-Bafalluy, N. Aitsaadi, and R. Langar, “A novel optimization framework for C-RAN BBU selection based on resiliency and price,” in Proc. VTC-Spring, Sep. 2017, pp. 1–6.

[205] J. Liu and O. E. Falowo, “Traffic-aware heuristic BBU-RRH switching scheme to enhance QoS and reduce complexity,” in Proc. IEEE PIMRC, Sep. 2018, pp. 1–7.

[206] S. Guo, D. Zeng, L. Gu, and J. Luo, “When green energy meets cloud radio access network: Joint optimization towards brown energy minimization,” Mobile Netw. Appl., vol. 24, no. 1–2, Feb. 2019.

[207] M. Y. Lyazidi, N. Aitsaadi, and R. Langar, “Dynamic resource allocation for cloud-RAN in LTE with real-time BBU/RRH assignment,” in Proc. IEEE ICC, May 2016, pp. 1–6.

[208] K. Wang, W. Zhou, and S. Mao, “On joint BBU/RRH resource allocation in heterogeneous cloud-RANs,” IEEE Internet Things J., vol. 4, no. 3, pp. 749–759, Jun. 2017.

[209] Q. Liu, T. Han, and N. Ansari, “Energy-efficient on-demand cloud radio access networks virtualization,” in Proc. IEEE Globecom, Dec. 2018, pp. 1–6.

[210] N. Amani, H. Pedram, H. Taheri, and S. Parsaeefard, “Energy-efficient resource allocation in heterogeneous cloud radio access networks via BBU offloading,” IEEE Trans. Veh. Technol., vol. 68, no. 2, pp. 1365–1377, Feb. 2019.

[211] J. Tang, W. P. Tay, and T. Q. S. Quek, “Cross-layer resource allocation with elastic service scaling in cloud radio access network,” IEEE Trans. Wireless Commun., vol. 14, no. 9, pp. 5008–5019, Sep. 2015.

[212] K. Wang, M. Zhao, and W. Zhou, “Traffic-aware graph-based dynamic frequency reuse for heterogeneous cloud-RAN,” in Proc. IEEE Globecom, Dec. 2014, pp. 2308–2313.

[213] K. Wang, W. Zhou, and S. Mao, “Energy efficient joint resource scheduling for delay-aware traffic in cloud-RAN,” in Proc. IEEE GLOBECOM, Dec. 2016, pp. 1–6.

[214] N. Iardella et al., “Flexible dynamic coordinated scheduling in virtual-RAN deployments,” in Proc. IEEE ICC Workshops, May 2017, pp. 126–131.

[215] J. Yao and N. Ansari, “QoS-aware joint BBU-RRH mapping and user association in cloud-RANs,” IEEE Trans. Green Commun. Netw., vol. 3, no. 4, pp. 881–889, Dec. 2018.

[216] F. Marzouk, R. Alheiro, J. Rodriguez, and A. Radwan, “Perspectives for 5G network sharing for mobile small cells,” in Broadband Communications, Networks, and Systems (Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering), vol. 263, V. Sucasas, G. Mantas, and S. Althunibat, Eds. Cham, Switzerland: Springer, Sep. 2018, pp. 377–386.

[217] J. Bartelt, “5G transport network requirements for the next generation fronthaul interface,” EURASIP J. Wireless Commun. Netw., vol. 2017, no. 1, pp. 1, 29 Dec. 2017.

[218] G. Kalfas et al., “Next generation fiber-wireless fronthaul for 5G mmWave networks,” IEEE Commun. Mag., vol. 57, no. 3, pp. 138–144, Jan. 2019.

[219] I. A. Alimi et al., “Enabling optical wired and wireless technologies for 5G and beyond networks,” in Telecommunication Systems—Principles and Applications of Wireless-Optical Technologies. London, U.K.: IntechOpen, Apr. 2019.

[220] D. Camps-Mur et al., “5G-XHaul: A novel wireless-optical SDN transport network to support joint 5G backhaul and fronthaul services,” IEEE Commun. Mag., vol. 57, no. 7, pp. 99–105, Jul. 2019.

[221] D. Sattar and A. Mattawry, “Towards secure slicing: Using slice isolation to mitigate DDoS attacks on 5G core network slices,” in Proc. IEEE Conf. Commun. Netw. Security (CNS), Jan. 2019, pp. 82–90.

[222] Z. Konulski et al., “Ensemble constructive approximate slice isolation in 5G networks,” EURASIP J. Inf. Security, vol. 2018, no. 1, p. 2, 2018.

Fatma Marzouk received the Engineering degree in telecommunications from the Higher Institute of Computer Science, El Manar, Tunisia. She is currently pursuing the Ph.D. degree in Telecommunications with the University of Aveiro, Portugal. Since 2018, she has been working with PROEFT, Portugal, as an Early Stage MSCA Researcher of the EU-funded SECRET Project. Her research interests include mobile networks, vehicular ad hoc networks, resource management, and software defined networks.

João Paulo Barrucca undergraduate degree in computers and telematics and in electronics and telecommunications, and the Ph.D. degree in informatics engineering in 2012. He acts as an Assistant Professor with the University of Aveiro, where he has been lectures since 2008. He conducts research with the Telecommunications Institute, having led the TN-AN for in 2015 and 2016. He has close to 100 publications including both peer reviewed and technical reports. He participated in more than 20 projects (Eight EU funded), either developing novel concepts, or applying these concepts in innovative products and solutions, involving the design of IoT platforms for telecom operators and water monitoring networks. He lead the FCT/CAPES DEVNF Project in Portugal that addressed novel 5G orchestra- tion, leads the local teams in EU LIFE-PAYT, participates in EU AENEAS (European Science Cloud for Astronomy), leads the local team in the P2020 @CRUISE Project, the security team at P2020-SOCIAL, participates in the EU Interreg CISMOB focusing in smart transportation and connected buses, the Engage SRA Research Infrastructure, and the Square Kilometer Array System Team, having lead activities for the instrument Cloud Platform, among a dozen other innovation projects. His current research interests include security and protocols for the Internet of Things in cloud environments, with focus in solutions for software defined networks in future networking environments.

Ayman Radwan (Senior Member, IEEE) received the M.A.Sc. degree from Carleton University, Ottawa, ON, Canada, and the Ph.D. degree from Queen’s University, Kingston, ON. He is a Senior Researcher and an EU Project Coordinator with the Instituto de Telecomunicações, Aveiro, Portugal. Since 2010, he has been intensively active in European projects, coordinating and technically managing multiple EU projects. He is currently supervising multiple Ph.D. students at University of Aveiro, Portugal. He was the Technical Manager of the FP7 Project CPOWER and the Coordinator of the CELTIC Project Green-T. He also acted as the Coordinator of the CELTIC Plus Project MUSCLES. He is an expert in the field of 5G and future mobile communications, with specific concentration on radio resource management, and green communications. His research is also specifically e-health and intelligent transportation systems. He is an Associate Editor of IEEE COMMUNICATIONS LETTERS and IEEE NETWORK.