Guerra-Gómez, Rolando; Ruiz-Boqué, Silvia; Garcia-Lozano, Mario; Saeed, Umar

Energy and Cost Footprint Reduction for 5G and Beyond with Flexible Radio Access Network

Published in:
IEEE Access

DOI:
10.1109/ACCESS.2021.3120765

Published: 15/10/2021

Document Version
Publisher's PDF, also known as Version of record

Published under the following license:
CC BY

Please cite the original version:
Energy and Cost Footprint Reduction for 5G and Beyond With Flexible Radio Access Network

ROLANDO GUERRA-GÓMEZ, SILVIA RUIZ-BOQUÉ, (Member, IEEE), MARIO GARCÍA-LOZANO, AND UMAR SAEED
Department of Signal Theory and Communications, Universitat Politècnica de Catalunya, 08034 Barcelona, Spain
Department of Communications and Networking, Aalto University, 02150 Espoo, Finland
Corresponding author: Rolando Guerra-Gómez (rolando.guerra@upc.edu)
This work was supported in part by the COST2020-INTERACT (Intelligence-Enabling Radio Communications for Seamless Inclusive Interactions) EU Project, in part by MCIN/AEI/10.13039/501100011033 and by “ERDFA way of making Europe” under Grant RTI2018-099880-B-C32, and in part by Grant FPI-UPC.

ABSTRACT This paper focuses on Beyond fifth generation (B5G) non-linear data modeling and decision-making tools to optimize cost reduction versus coverage-QoS trade-off, in other words, the number of active Remote Radio Heads or Units (RRHs) needed according to traffic demands. The cost and energy optimization are analytically expressed by modeling the complex relationships between input and output system parameters using realistic scenarios and traffic profiles for low, medium, and high traffic environments. The optimization tool is based on a multi-objective integer linear programming model, designed to reduce the network cost while maintaining a good coverage-QoS and accounting for capacity constraints, User Equipments (UEs), and different slices. Results at 3.6 and 28 GHz are presented by analyzing and comparing several Cloud Radio Access Network (C-RAN) split options in a heterogeneous deployment with Macro-RRHs (MRRHs) and Small-RRHs (SRRHs). Cost reductions ranging from 30 % to 70 % have been obtained depending on the scenario. This proposal allows mobile network operators (MNOs) to achieve further optimization, while providing better network diagnostics.

INDEX TERMS 5G, C-RAN, radio network optimization.

I. INTRODUCTION

The Fifth-Generation (5G) of cellular networks has a service-oriented architecture that significantly increases both performance and flexibility of the offered services to users and service providers. This is done by introducing new radio modes such as ultra-Reliable Low Latency Communication (uRLLC), Massive Machine Type Communications (mMTC) and enhanced Mobile Broadband (eMBB). Robust and dynamic radio network management solutions should be implemented by operators to tackle such complexity and flexibility.

The future Sixth-Generation (6G) network ecosystem is a step further that should implement a fully cloud-native architecture capable of dealing with a network of sub-networks, offering Tbps of data throughput, sub-ms latency and extremely low packet error rate, increased device density, ultra-low-energy consumption, very high security and cm-level accuracy localization among others [1].

The network design of Beyond Fifth-Generation (B5G) must benefit the whole society by being a human-centric reliable infrastructure. Moreover, future mobile networks should support immersive communication, cognition and twinning, deterministic end-to-end applications, and high-resolution sensing services. Sustainability is crucial to support these services and network requirements. It is fundamental to reduce: energy consumption, resources usage, and emissions footprints to avoid an excessive growth in power consumption. The enormous increase in the number of devices, data amounts, and data rates implies an increase in the overall data traffic and required capacity, while the energy reduction is not automatically guaranteed.

The 3rd Generation Partnership Project (3GPP) included Cloud Radio Access Network (C-RAN) in New Generation RAN (NG-RAN) to facilitate the implementation of ultra-dense mobile networks while reducing the Total Cost of Ownership (TCO). The centralization of the baseband
The optimum deployment of 5G and B5G networks is a challenging problem. It should account for different frequency bands, channel propagation models, the massive amount of User Equipments (UEs), and devices with realistic traffic profiles. Additionally, it could consider cooperation among BSs to reduce interference and increase throughput in a heterogeneous and ultra-dense network.

This work proposes an optimization framework with two purposes. Firstly, the efficient deployment of 5G and B5G radio networks on C-RAN ecosystems. Secondly, the activation/deactivation of the RRHs to maintain the coverage and Quality of Service (QoS) while minimizing the network cost. The proposed algorithm selects the optimal distribution of RRHs from a set of realistic candidate locations. It is based on the data traffic of the UEs, which generate Guaranteed Bit Rate (GBR) or Best Effort (BE) services. The algorithm includes the possibility of implementing cooperation strategies between cells, automatically selecting the cells that should cooperate to satisfy the traffic demand of a specific zone in the map.

The main contributions are summarized as follows:

- The complex relationships between input and output parameters of the system in a dense B5G environment is modeled. This is done analytically, by obtaining a complete set of nonlinear equations.
- A formulation and software implementation of decision-making rules to optimize the number of required active RRHs under different traffic patterns in a heterogeneous environment is presented. The software could identify the optimum constellation of RRHs to provide the best performance in terms of coverage, user equipment density, QoS, and overall cost deployment.
- On the other hand, the impact of different split options in the deployment cost and coverage-QoS of the radio network is detailed. The selected split option introduces a cost ratio between types of BSs, especially when heterogeneous network deployments with Macro and Small-BSs are considered. The cost ratio between macro-RRHs (MRRHs) and small-RRHs (SRRHs) increases when they contain more baseband functionalities. Thus, this cost ratio achieves the maximum value in distributed RAN (option 1) scenarios, while it is minimum in full C-RAN (option 8).
- Furthermore, a comprehensive analysis and comparison of the performance at 2.6 (MRRHs), 3.6, and 28 GHz carrier frequencies for SRRHs in urban environments is presented, which is of high interest for researchers and Mobile Network Operators (MNOs).
- Through a careful analysis of the performance on a realistic scenario, it is demonstrated that the proposed optimization could help MNOs to a) improve network planning by detecting problems in advance while offering solutions, b) provide a network diagnostic tool, and c) optimize and control the network cost-effectiveness and energy efficiency.
As far as our knowledge, the optimization of the RRH deployments of B5Gs in terms of energy and cost-saving considering: flexible radio access network, both frequency ranges (sub 6GHz and mm-wave), coverage and capacity constraints, realistic propagation models, and heterogeneous mobile networks with MRRHs and SRRHs is not available in the literature.

The rest of this paper is organized as follows. Section II presents a literature review on radio network deployment algorithms. Section III models the radio network and the proposed optimization problem. Section IV describes the considered radio network platform and its different planes or layers: RRH plane, demand plane, UE plane and propagation model. Section V evaluates the results. Finally, section VI summarizes the most relevant results.

II. RELATED WORKS

In parallel with the standardization process, some research works have been focused on the definition of different strategies to deploy and optimize 5G and B5G networks [5]–[11].

A mixed-integer linear programming (MILP) algorithm is proposed in [5] to minimize the network deployment cost and latency of a C-RAN with Mobile Edge Computing (MEC) nodes. Moreover, they propose a heuristic algorithm because of the complexity of the MILP approach. The main goal of this paper is to optimize the MEC nodes placement and the C-RAN deployment. Although the proposed strategies are novel and could be of interest because they implement a joint optimization considering MEC and C-RAN, the analysis is limited to the transport network and the placement of BBU pools and MEC nodes, without considering the RRH deployment and the mobile network demand plane.

In [6], the authors propose an energy-effective radio network deployment where the system could select a subset of RRHs according to the traffic demand simulated using Traffic Demand Points (TDPs), which concentrate the data rate of a specific zone to satisfy the QoS requirements of the potential UEs. However, the problem is divided into two sub-optimal problems: RRH-TDP association and RRH selection, which could reduce the possibility of finding the optimal solution for the network deployment. On the other hand, the authors consider two synthetic scenarios to validate the results. The first scenario depicts a dense square region of 250 m × 250 m with two micro-RRHs and seven pico-RRHs, while the second represents an area of 500 m × 500 m with 3 micro-RRHs and 13 pico-RRHs.

Besides, the authors in [7] recently proposed a hybrid fronthaul solution based on fibers and Free-Space Optics (FSO) to minimize the deployment costs in dense urban scenarios. They formulated and compared two Integer Linear Programming (ILPs): joint and disjoint approaches. The disjoint approach splits the whole problem into two sub-problems: the RRH placement and the fronthaul deployment, while the joint strategy only solves one optimization problem to deploy the whole C-RAN. The authors conclude that the joint approach is better than the disjoint strategy in terms of deployment cost. Although they propose an interesting solution to reduce the deployment cost of a C-RAN with hybrid fronthaul, there is room for improvement by introducing realistic UEs to model the traffic demand, validating the results under realistic RRH possible locations.

In [8], the authors propose a Multi-Objective Optimization (MOO) problem for small cell planning which considers fiber and wireless backhaul technologies and two types of BSs. The MOO aims to determine the optimum type and location of the deployed BSs. The authors also propose a joint cell and fiber backhaul planning algorithm employing heuristic techniques. This work is also of interest because it focuses on the last standard of Passive Optical Networks (PONs), called Next-Generation Passive Optical Network 2 (NG-PON2), for the fiber deployment of the backhaul; however, it does not consider a C-RAN environment.

In [9], the authors design a joint optimization framework considering the costs of the mobile network and its fronthaul in a C-RAN ecosystem. Deployment cost is analyzed under different scenarios; they also extend the work to consider three optical fronthaul technologies: Common Public Radio Interface (CPRI), Physical Layer Split (PLS), and Analogue Radio-over-Fiber (ARoF). Although a detailed analysis of the fronthaul network and the C-RAN deployment is provided, traffic profiles are simplified due to the complexity of the model. Hotspots are considered to generate traffic without accounting for different services and UEs. On the other hand, the radio network deployment is simplified by introducing a fixed coverage radio per RRH, instead of modeling the Signal-to-Noise-plus-Interference-Ratio (SINR) using a suitable propagation channel model.

On the other hand, the authors in [10] propose an optimization problem that minimizes the number of RRHs in a C-RAN context. Following the same approach of [6], they use the concept of TDPs to simulate the traffic demand, where the TDPs are allocated at the center of the demand zones. The algorithm starts with all the possible RRHs on, and connects each TDP to the nearest RRH. Then, the proposed algorithm turns off some of the RRHs at each iteration until the percentage of not satisfied TDPs is greater than 0.1 %. However, for the sake of simplicity, the traffic demand of each TDP is established without UE and service modeling, considering only a capacity constraint, and the RRH-TDP association is based on a minimum distance approach.

In [11], the authors propose a framework to improve resource efficiency at the BS level. They employ a joint optimization problem to efficiently allocate the resources of the network slices, the cell-slice association, and the UE-BS connections. They include SINR requirements and different slice services in the network optimization problem. This work demonstrates that realistic scenarios with UEs and services can be modeled.

The works mentioned above propose promising radio network deployment strategies. However, there is room for improvement because they skip traffic generation by
considering only demanding points without accounting for slices, services, and UEs. Moreover, the RRH coverage and the RRH-TDP association are simplified, which results in synthetic scenarios, usually with a small number of cells, that do not reflect the complexity of mobile networks. To the best of our knowledge, there are no published papers that include cell cooperation in radio network deployment algorithms, saving energy, and reducing costs at different frequency ranges and split options in a realistic scenario. For these reasons, it is impossible to compare the presented results with the available literature. However, the initial deployment, where all the cells are active, is considered as a benchmark. This baseline is extracted from the realistic cell deployment provided by the European COST Action IRACON in [12], which has been widely used by the research community.

III. RADIO NETWORK: MATHEMATICAL MODEL

This section presents the mathematical model developed to optimize the number and distribution of the active RRHs required to minimize the deployment cost while simultaneously maximizing the coverage and satisfying QoS requirements for any BBU–RRH split option. The model also allows for cooperation among RRHs.

A. MODEL DEFINITION

Let $\mathcal{R}$ be the set of candidate RRHs and their locations. Information provided by the MNO about the already deployed cellular networks is used (Fourth-Generation (4G) and 5G) and complemented with additional locations at lamp poles and street corners. In general, locations with feasible access to the power grid and line of sight propagation have been considered. Two types of RRHs can be used, MRRHs and SRRHs, which have a deployment cost of $C_{\text{MRRH}}$ and $C_{\text{SRRH}}$, respectively. A binary vector $\eta = \{\eta_1, \eta_2, \ldots, \eta_{|\mathcal{R}|}\}$ indicates what kind of cell could be deployed at each possible location, where the notation $|\mathcal{R}|$ denotes the number of elements of the set $\mathcal{R}$. Equation (1) shows the definition of the elements of $\eta$. The set $\mathcal{R}$ is subdivided into the sets $\mathcal{M}$ and $\mathcal{S}$ that represent MRRHs and SRRHs respectively, such that $\mathcal{M} \cup \mathcal{S} = \mathcal{R}$.

\[
\eta_r = \begin{cases} 
1 & \text{if } r \in \mathcal{M} \\
0 & \text{if } r \in \mathcal{S}
\end{cases} \tag{1}
\]

On the other hand, UEs have been modeled to generate the traffic demand and also have been represented mathematically by the set $\mathcal{U}$. Each UE is subscribed to a unique Service Provider (SP) and, for the sake of simplicity, each SP is associated to one slice. A slice can support multiple services and the data flow related to a given service is represented by a Service Function Chain (SFC). Thus, each UE generates only one service and is associated to one SFC of its SP. Those services could be GBR or non-GBR (Best Effort) services such as High-Definition (HD) video streaming and File Transfer Protocol (FTP), respectively.

Especially, the GBR services must guarantee a minimum bit rate to each UE, which is denoted as $D_u^{\text{min}}$, where $u \in \mathcal{U}$.

This parameter is selected by the SP according to the minimum QoS that should be assured for each service. In order to provide the minimum bit rate to the GBR-UEs, a minimum SINR should be maintained (denoted as $\gamma_u^{\text{min}}$), which can be estimated employing the Shannon’s equation (2).

\[
\gamma_u^{\text{min}} = 2^{\frac{B_u}{P_u}} - 1 \quad \forall u \in \mathcal{U} \tag{2}
\]

where $B_u$ depicts the bandwidth assigned to the UE $u$.

The geographical area under analysis is divided into TDPs, where each of them aggregates and concentrates the data rates of the UEs inside it. Let $\mathcal{Z}$ be the set of demand zones or TDPs in the region and $D_z$ the demand bit rate of the zone $z \in \mathcal{Z}$.

The algorithm should select the optimum RRH distribution that reduces the deployment cost while increases the coverage. For this reason, the binary decision vector $\mathbf{\rho} = \{\rho_1, \rho_2, \ldots, \rho_{|\mathcal{R}|}\}$ has been defined to indicate the RRH distribution. $\rho_r$ is a binary variable that indicates if the candidate RRH $r \in \mathcal{R}$ is activated or not, see (3).

\[
\rho_r = \begin{cases} 
1 & \text{if } r \in \mathcal{R} \text{ is selected as RRH} \\
0 & \text{otherwise}
\end{cases} \tag{3}
\]

Besides, a binary decision matrix $\mathbf{X}$ of dimension $|\mathcal{R}| \times |\mathcal{Z}|$ is employed to represent the association between RRHs and TDPs. The elements of $\mathbf{X}$ are represented by the binary variable $x_{r,z}$, which is defined as (4):

\[
x_{r,z} = \begin{cases} 
1 & \text{if } z \in \mathcal{Z} \text{ is served by } r \in \mathcal{R} \\
0 & \text{otherwise}
\end{cases} \tag{4}
\]

Next, constraint (5) takes into account cooperation among RRHs to improve the network capacity (e.g. Joint Transmission (JT) or any other Coordinated Multipoint (CoMP) technique) by the introduction of the integer $\mu$, that limits the number of RRHs that can cooperate to increase the bit rate while mitigating the interference of the zones. It guarantees that each zone is served by a maximum of $\mu$ RRHs. Its minimum value is $\mu = 1$, when the cooperation techniques are not allowed. MNOs should select $\mu$ before the optimization process and according to the available cooperation technology of the considered network.

\[
\sum_{r \in \mathcal{R}} x_{r,z} \leq \mu \quad \forall z \in \mathcal{Z} \tag{5}
\]

Additionally, a key point is the establishment of a relationship between the decision variables $\rho_r$ and $x_{r,z}$ to ensure that, if a possible RRH $r \in \mathcal{R}$ is selected, it must serve at least one zone $z \in \mathcal{Z}$; and if a RRH is associated to a zone $z$, it must be active. Equations (6) and (7) account for these conditions respectively.

\[
\sum_{z \in \mathcal{Z}} x_{r,z} \geq \rho_r \quad \forall r \in \mathcal{R} \tag{6}
\]

Moreover, equation (7) is also a capacity constraint. It ensures that a selected RRH has enough capacity to satisfy the demand of its associated TDPs or zones.

\[
\xi_r \leq \xi_{\text{max}} \rho_r \quad \forall r \in \mathcal{R} \tag{7}
\]
$$\xi_r = \sum_{z \in Z} \frac{x_{r,z} D_r}{D_r \sum_{r' \in R} x_{r',z}} \quad (7)$$

In (7), $D_r$ represents the achievable throughput at the RRH $r$. This parameter depends on the RRH configuration, e.g., Multiple-Input Multiple-Output (MIMO) order, modulation order and bandwidth. Besides, the real variable $\xi$ depicts the traffic load of the GBR services through the RRH $r$. The MNOs establish a partition of the resources between the GBR and best effort services by controlling the parameter $0 \leq \xi_r \leq \xi^{\text{max}} \leq 1$. For instance, if $\xi^{\text{max}} = 0.8$ it means that 80% of the radio resources of the RRHs could be dedicated to satisfy the traffic of the slices with GBR services. The remainder 20% of the RRH capacity is reserved for the non-GBR traffic. The parameter $\xi^{\text{max}}$ should be selected according to the demand of the different types of services.

As it has been mentioned, to guarantee the QoS, it is important that the UEs experience a SINR greater than a minimum ($\gamma_u^{\text{min}}$). However, due to the enormous amounts of UEs that are expected in 5G networks, it is not scalable to define a constraint that maintains an independent SINR requirement for each UE in the optimization algorithm. For this reason, equation (8) guarantees that the SINR constraint is accomplished while relaxing the specifications by moving them to an upper level (zone plane).

$$\gamma_z \geq \gamma_z^{\text{min}} \kappa^{\text{adj}} \quad \forall z \in Z \quad (8)$$

In (8), $\gamma_z$ represents the perceived SINR at the TDP or zone $z \in Z$, while $\gamma_z^{\text{min}}$ depicts the minimum SINR that must be kept at TDP of the zone $z$, which is taken equal to the required SINR of the UE with highest demand in the zone $z$. This approach is highly restrictive and it does not take into account techniques such as enhanced Inter-Cell Interference Coordination (eICIC). For this reason the factor $\kappa^{\text{adj}}$ is introduced, which should be adjusted by the MNO to consider the mitigation of the interference by dynamic resource allocation techniques.

$$\sum_{r \in R} x_{r,z} P^{\text{Rx}}_{r,z} \geq \gamma_z^{\text{min}} \kappa^{\text{adj}} \left( \sum_{r \in R} \rho_r (1 - x_{r,z}) P^{\text{Rx}}_{r,z} + N \right) \quad \forall z \in Z \quad (9)$$

Equation (8) is transformed into (9) by introducing an estimated value for $\gamma_z$. The left side of (9) represents the useful received power at TDP $z$, while the expression in brackets of the right term models the interference plus noise power. The parameter $N$ represents the average UE thermal noise power. It is important to mention that according to equation (8), all the active RRHs that are not serving the considered TDP are a source of interference, which means that the mobile network is designed with a frequency reuse factor equal to unity. However, operating at the same frequency band in the entire network will produce a high level of interference. For this reason, the proposed algorithm considers that MRRHs and SRRHs operate at different frequency bands. In this work, the MRRHs operate at 2.6 GHz while the SRRHs are able to operate at multiple frequency bands (for instance, 3.6 GHz and 28 GHz). As a result, equation (9) is split into (10) and (11)

$$\sum_{r \in M} x_{r,z} P^{\text{Rx}}_{r,z} \geq \gamma_z^{\text{min}} \kappa^{\text{adj}} \left( \sum_{r \in M} \rho_r (1 - x_{r,z}) P^{\text{Rx}}_{r,z} + N \right) \quad \forall z \in Z \quad (10)$$

$$\sum_{r \in S} x_{r,z} P^{\text{Rx}}_{r,z} \geq \gamma_z^{\text{min}} \kappa^{\text{adj}} \left( \sum_{r \in S} \rho_r (1 - x_{r,z}) P^{\text{Rx}}_{r,z} + N \right) \quad \forall z \in Z \quad (11)$$

where, as mentioned above, $M$ and $S$ are subsets of $R$ that contain the sets of MRRH and SRRH respectively, such that $R = M \cup S$. The parameter $P^{\text{Rx}}_{r,z}$, which represents the received power at TDP $z$ from the RRH $r$, is calculated by using the link budget equation (12)

$$P^{\text{Rx}}_{r,z} = \frac{P^{\text{TX}}_r G_r G^{\text{UE}}_r}{L_{r,z}} \quad (12)$$

where $P^{\text{TX}}_r$ and $G_r$ denote the transmission power and the antenna gain of the RRH $r$, respectively. $G^{\text{UE}}_r$ represents the UE antenna gain, while $L_{r,z}$ takes into account the radio link losses. Namely, $L^{\text{RRH}}_r$ and $L^{\text{UE}}_z$ account for the losses due to connectors, transmission lines and other mismatches at the RRH and the UE respectively. Besides, $L^{\text{FD}}_{r,z}$ represents the path-loss from the RRH $r$ to the TDP $z$. Finally, $L^{\text{FD}}$ is a random variable modelling the slow fading. It is important to notice that this parameter should be eliminated if the considered channel model already takes into account the shadowing effects.

Additionally, equation (13) ensures that if a RRH $r$ is serving the zone $z$, the received power ($P^{\text{Rx}}_{r,z}$) is greater or equal than the sensitivity of the UE-receivers ($P^{\text{Rx}}_{\text{min}}$).

$$x_{r,z} P^{\text{Rx}}_{r,z} \geq P^{\text{Rx}}_{\text{min}} \quad (13)$$

As it has been stated, the proposed algorithm allows for RRH cooperation with the introduction of the parameter $\mu$. On the other hand, constraint (5) does not limit the cooperation between MRRH and SRRH, which introduces a high complexity to the UEs because they would have to operate simultaneously at different frequency bands, for instance in a dual connectivity operation mode. For this reason, constraint (14) guarantees that the cooperation is limited to a specific frequency band (known as inter-site aggregation). In other words, the cooperation in a specific zone is carried out by only one type of RRH.

$$\sum_{r \in M} x_{r,z} \leq 0 \lor \sum_{r \in S} x_{r,z} \leq 0 \quad \forall z \in Z \quad (14)$$

where $\lor$ stands for the logical disjunction (logical OR operation), guaranteeing that the zone $z \in Z$ is not served by RRHs of different classes.

The goal of the proposed algorithm is to select the optimum distribution of the RRHs, minimizing the radio network
deployment cost while simultaneously maximizing the coverage and satisfying QoS requirements. The deployment cost is computed as in equation (15).

\[
F_1 = \sum_{r \in R} \rho_r (\eta_r C_M + (1 - \eta_r) C_S)
\]

\[
F_1 = C_S \sum_{r \in R} \rho_r (\eta_r \sigma + (1 - \eta_r))
\]

\[
\sigma = \frac{C_M}{C_S}
\]

(15)

where \( \sigma \) represents the ratio between the MRRH and SRRH costs, \( C_M \) and \( C_S \), respectively. This parameter is useful to consider different kind of scenarios; for instance, to represent heterogeneous mobile network deployments (\( \sigma > 1 \)). Furthermore, it is used in this work to consider different BBU–RRH split options because it modifies the cost ratio between MRRHs and SRRHs. On the other hand, it allows the normalization of the RRH costs by considering \( C_S = 1 \). The MNOs should carefully select the cost ratio (\( \sigma \)) according to the cost of the considered network devices.

The coverage–QoS is estimated by the number of served zones, computed as in equation (16).

\[
F_2 = \sum_{z \in Z} u(\sum_{r \in R} x_{r,z} - 1)
\]

\[
F_2 \geq F_2^{\min} |Z|
\]

\[
u[n] = \begin{cases} 
1 & \text{if } n \geq 0 \\
0 & \text{if } n < 0
\end{cases}
\]

(17)

where \( u[\cdot] \) denotes the Heaviside sequence and the operator \(| \cdot |\) stands for the cardinal of the considered set. Moreover, the MNO has the flexibility to establish the minimum coverage-QoS that must be provided by controlling the parameter \( 0 \leq F_2^{\min} \leq 1 \) in constraint (17). Finally, the optimum radio network deployment algorithm is formulated as an integer optimization problem in equation (18).

\[
\text{minimize } \quad F_1, -F_2
\]

subject to: \( (1) - (7), (10) - (14), (17) \)

(18)

Table 1 summarizes the sets, variables and parameters of the proposed algorithm.

**B. REFORMULATION AS INTEGER LINEAR OPTIMIZATION PROBLEM**

The algorithm formulated in section III-A is a non-linear integer programming model. The non-linearity is introduced by the constraints (7), (10), (11), (14) and the coverage function (16). In order to solve the proposed algorithm employing a linear optimization solver, a reformulation of these expressions to linear equations is needed. In this section, the problem is transformed into an ILP model.

The linearization of the previously mentioned expressions uses theorem 1:

**Theorem 1:** Let assume \( D \subseteq \mathbb{R}^n, f: D \rightarrow \mathbb{R}, M \in \mathbb{R} : M \neq 0; M \geq \max\{f(\phi) | \phi \in D \} \) and \( \delta \) a binary variable such that \( \delta \in \{0, 1\} \). Then, the following expressions are equivalent,

\[
i: \quad \delta = 0 \implies f(\phi) \leq 0
\]

\[
ii: \quad f(\phi) - M\delta \leq 0
\]

(19)

The linearization of the constraint (7) implies the modification of the term \( \sum_{r \in R} \sum_{z \in Z} x_{r,z} \). Following the constraint (5), the denominator of this term is an integer (\( k \in \mathbb{N} | 0 \leq k \leq \mu \)), bounded by the maximum number of RRHs that could cooperate to serve a zone (\( \mu \)). So, it is possible to reformulate the expression as it is described in (19).

\[
\sum_{r \in R} x_{r,z} = \begin{cases}
\sum_{k=1}^{\mu} k^{-1} x_{r,z} \delta_{k,z} & \text{if } k \neq 0 \\
0 & \text{if } k = 0
\end{cases}
\]

(19)

where \( \delta_{k,z} \) are binary variables indicating if the TDP \( z \) is served by \( k \) RRHs, as described in (20).

\[
\delta_{k,z} = 1 \implies \sum_{r \in R} x_{r,z} = k \quad \forall z \in Z, k \in [0, \mu]
\]

(20)

However, the equations (19) and (20) are also non-linear expressions that should be linearized. The product of binary variables in (19) is substituted by another binary variables such that \( x_{r,z} \delta_{k,z} = \psi_{r,z,k} \), which is equivalent to (21).

\[
\psi_{r,z,k} = 1 \iff x_{r,z} + \delta_{k,z} = 2
\]

(21)

After this mathematical procedure, the equations (20) and (21) could be converted to linear expressions employing the **Theorem 1**. Equation (22) shows the linear equivalent expressions of the constraint (7).

\[
\sum_{z \in Z} D_z \sum_{k=1}^{\mu} k^{-1} \psi_{r,z,k} \leq \xi_{\max} \rho_r D_r
\]

(22a)

\[
\psi_{r,z,k} \leq x_{r,z}
\]

(22b)

\[
\psi_{r,z,k} \leq \delta_{k,z}
\]

(22c)

\[
\psi_{r,z,k} \geq \delta_{k,z} + x_{r,z} - 1
\]

(22d)

\[
\sum_{r \in R} x_{r,z} \leq \mu - \delta_{k,z} (\mu - k)
\]

(22e)

\[
\sum_{r \in R} x_{r,z} \geq k \delta_{k,z}
\]

(22f)

\[
\sum_{k=0}^{\mu} \delta_{k,z} = 1
\]

(22g)

For the sake of simplicity, the domain of the indexing subscripts \( r, z, k \) has been made explicit only when it is different from the defined domain.

The other non-linear constraints are (10), (11) and (14). As it has been explained, each TDP could be served by \( k \) RRHs, where \( k = 0 \) means there is no RRH that can serve TDP \( z \) satisfying the constraints. This situation has not been considered by constraints (10) and (11), which do not hold for this special case. For this reason, the binary variables \( \rho_{z,M}^r \) and \( \beta_z^S \), that are used to indicate if the zone \( z \) is served by MRRHs or SRRHs, are introduced in (23) and (24).

\[
\rho_{z,M}^r = 1 \iff \sum_{r \in M} x_{r,z} \leq 0
\]

(23)
### TABLE 1. Glossary of terms of the RRH selection algorithm.

<table>
<thead>
<tr>
<th>Terms</th>
<th>Sets</th>
<th>Input Parameters</th>
<th>Binary Variables</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{R}$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>set of possible RRHs and their locations</td>
</tr>
<tr>
<td>$\mathcal{M}$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>set of possible MRRHs and their locations</td>
</tr>
<tr>
<td>$\mathcal{S}$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>set of possible SRRHs and their locations</td>
</tr>
<tr>
<td>$Z$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>set of zones or TDPS</td>
</tr>
<tr>
<td>$\mathcal{U}$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>set of UEs</td>
</tr>
<tr>
<td>$\rho_r$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>RRH distribution, 1 if $r \in \mathcal{R}$ is selected as RRH, 0 otherwise</td>
</tr>
<tr>
<td>$x_{r,z}$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>1 if RRH $r \in \mathcal{R}$ manages the traffic demand of zone $z \in Z$, 0 otherwise</td>
</tr>
<tr>
<td>$C_{\text{RRH}}$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>deployment cost of an RRH</td>
</tr>
<tr>
<td>$C_{\text{MRRH}}$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>deployment cost of a MRRH</td>
</tr>
<tr>
<td>$\eta_r$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Identifier of RRH type, 1 if $r$ is a MRRH, 0 otherwise</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>ratio between MRRH and SRRH costs</td>
</tr>
<tr>
<td>$D_{\text{min}}$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>required bit rate of the UE $u \in \mathcal{U}$</td>
</tr>
<tr>
<td>$\gamma_{\text{min}}$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>SINR required by the UE $u \in \mathcal{U}$ to satisfy $D_{\text{min}}$</td>
</tr>
<tr>
<td>$\gamma_{\text{adj}}$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>minimum SINR to satisfy at zone $z \in Z$</td>
</tr>
<tr>
<td>$\gamma_z$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>received SINR at zone $z \in Z$</td>
</tr>
<tr>
<td>$P_{r,z}$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>transmission power of the RRH $r \in \mathcal{R}$</td>
</tr>
<tr>
<td>$G_r$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>antenna gain of the RRH $r \in \mathcal{R}$</td>
</tr>
<tr>
<td>$G_{\text{UE}}$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>antenna gain of the UEs</td>
</tr>
<tr>
<td>$L_{\text{z}}$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>transmission loss from the RRH $r \in \mathcal{R}$ to the TDP at zone $z \in Z$</td>
</tr>
<tr>
<td>$\mathcal{L}_{\text{RRH}}$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>loss introduced at the RRHs (e.g., transmission lines and connectors losses)</td>
</tr>
<tr>
<td>$L_{\text{UE}}$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>loss introduced at the UEs (e.g., transmission lines and coupling losses)</td>
</tr>
<tr>
<td>$L_{\text{PD}}$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>loss introduced by the fading effects</td>
</tr>
<tr>
<td>$L_{\text{PL}}$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>path loss from the RRH $r \in \mathcal{R}$ to the TDP at zone $z \in Z$</td>
</tr>
<tr>
<td>$\beta_0^c$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>minimum normalized coverage–QoS</td>
</tr>
<tr>
<td>$\beta_0^d$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>bandwidth of the UE $u \in \mathcal{U}$</td>
</tr>
<tr>
<td>$D_s$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>traffic demand (bit rate) of the TDP at zone $z \in Z$</td>
</tr>
<tr>
<td>$D_{\text{t}}$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>achievable throughput at RRH $r \in \mathcal{R}$</td>
</tr>
<tr>
<td>$\mu$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>number of allowed simultaneous connections of each UE to the RRHs</td>
</tr>
<tr>
<td>$\xi_r$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>resource usage ratio of the RRH $r \in \mathcal{R}$</td>
</tr>
<tr>
<td>$\xi_{\text{max}}$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>maximum resource usage ratio of the RRHs</td>
</tr>
<tr>
<td>$N$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>noise power</td>
</tr>
</tbody>
</table>

$$\beta_0^c = 1 \iff \sum_{r \in \mathcal{S}} x_{r,z} \leq 0 \quad (24)$$

$$\sum_{r \in \mathcal{T}} x_{r,z} P_{r,z}^{\text{t}} + L_{\text{z}}^{\text{t}} \beta_z^{\text{min}} \geq \gamma_z^{\text{adj}} (\sum_{r \in \mathcal{T}} P_{r,z}^{\text{t}} - \sum_{r \in \mathcal{T}} x_{r,z} P_{r,z}^{\text{t}} + N) \quad \forall z \in Z \quad (25)$$

where $\mathcal{T}$ is equal to $\mathcal{M}$ or $\mathcal{S}$ to represent the constraints (10) and (11) respectively. An alternative is to write both constraints in the same expression. It is important to notice that the non-linear expression $\rho_{r,x}$ has been substituted by $\rho_{r,z}$, because the constraint (7) ensures that if $x_{r,z} = 1$ then $\rho_{r,z} = x_{r,z}$. The parameter $L \in \mathbb{R}$ is a large number to hold the constraint (25) when the zone $z$ is not served by this type of RRH.

On the other hand, the constraint (14) is an inclusive disjunction that could be rewritten combining (23) and (24) with the expression $\beta_0^c = 1$. Theorem 1 has been employed to obtain the linear expressions of (23) and (24), which are shown in (26).

$$\sum_{r \in \mathcal{M}} x_{r,z} \leq \mu (1 - \beta_0^c) \quad (26a)$$

$$\sum_{r \in \mathcal{S}} x_{r,z} \geq \epsilon (1 - \beta_0^c) \quad (26b)$$

$$\sum_{r \in \mathcal{S}} x_{r,z} \leq \mu (1 - \beta_0^c) \quad (26c)$$

$$\sum_{r \in \mathcal{S}} x_{r,z} \geq \epsilon (1 - \beta_0^c) \quad (26d)$$

$$\beta_0^c + \beta_0^s \geq 1 \quad (26e)$$

Besides, the coverage is estimated in equation (16), as the number of served zones. Following this definition, the multi-objective optimization problem (18) should minimize the deployment cost while maximizing the coverage. However, maximizing the number of served zones is equivalent to minimizing the zones without service. This approach allows for the reuse of the binary variable $\delta_{0,z}$ that indicates if the zone $z$ is not served. Then, constraint (17) is expressed as (28) and the underlying linear coverage–QoS function is shown in equation (27).

$$\sum_{z \in Z} \delta_{0,z} = F_3 = |Z| - F_2 \quad (27)$$

$$\sum_{z \in Z} \delta_{0,z} \leq F_2^\text{min} - |Z| \quad (28)$$

This strategy reduces the complexity of the proposed algorithm by eliminating the additional variables in the linearization of the Heaviside sequence $u[.]$ in (16).

Finally, the ILP model of the proposed algorithm is summarized in equation (29). The multi-objective optimization
problem is solved employing the weighted–sum method, where the weights \( \omega_1 \) and \( \omega_3 \) should be carefully selected by the MNO in order to obtain an optimal point on the Pareto Front. Additionally, the subscript \( n \) in equation (29a) means that the objective functions have been normalized to guarantee that their values are in the same range. In this case, each function has been divided by its maximum value. The maximum of \( F_1 \) and \( F_3 \) are \( C_S(\sigma |M| + |S|) \) and \( |Z| \), respectively.

\[
\begin{align*}
\text{Min } & \omega_1 F_{1n} + \omega_3 F_{3n} \\
\text{subject to : } & \sum_{r \in R} x_{r,z} \leq \mu \\
& \sum_{r \in R} x_{r,z} \geq \rho_r \\
& x_{r,z} p_{r,z}^R \geq p_{\max}^R \\
& \sum_{z \in Z} D_z \sum_{k=1}^{k=m} k^{-1} \psi_{r,z,k} \leq \xi \max \rho_r D_r \\
& \psi_{r,z,k} \leq x_{r,z} \\
& \psi_{r,z,k} \leq \delta_{k,z} \\
& \psi_{r,z,k} \leq \delta_{k,z} + x_{r,z} - 1 \\
& \sum_{r \in R} x_{r,z} \leq \mu - \delta_{k,z}(\mu - k) \\
& \sum_{r \in R} x_{r,z} \geq k \delta_{k,z} \\
& \mu \sum_{k=0} \delta_{k,z} = 1 \\
& \sum_{r \in R} x_{r,z} p_{r,z}^R + L \beta_z T \geq \\
& \gamma_z \min_{k \in \{P,M,S\}} (\sum_{r \in T} \rho_r p_{r,z}^R - \sum_{r \in T} x_{r,z} p_{r,z}^R + N) \\
& \sum_{r \in M} x_{r,z} \leq \mu (1 - \beta_z^M) \\
& \sum_{r \in M} x_{r,z} \geq \epsilon (1 - \beta_z^M) \\
& \sum_{r \in S} x_{r,z} \leq \mu (1 - \beta_z^S) \\
& \sum_{r \in S} x_{r,z} \geq \epsilon (1 - \beta_z^S) \\
& \beta_z^M + \beta_z^S \geq 1 \\
& \sum_{z \in Z} \delta_{0,z} \leq \gamma_z \min |Z| \\
\end{align*}
\]

\( \rho_r, x_{r,z}, \delta_{k,z}, \psi_{r,z,k}, \beta_z^M, \beta_z^S \) binary variables.

### IV. DESCRIPTION OF THE SCENARIO

This section describes the scenario used to validate the proposed algorithm. A realistic radio network deployment over the Vienna City is considered (see Fig. 2), which has been detailed in [13]; it has been widely used to validate different optimization algorithms in mobile communications [12], [14]. Furthermore, this same scenario has been used in our previous works to validate the performance of adaptive resource allocation strategies on C-RAN and 5G networks [4], [15].

The proposed scenario has a map resolution of 5 m with 205 \( \times \) 291 points, which is equivalent to an area of 1025 \( \times \) 1455 m\(^2\).

To complete the scenario, three layers in a hierarchical structure have been defined: RRH plane, Demand plane and UE plane as it is represented in Fig. 3. Details of each layer are given in the following subsections as well as the interrelations between them.

#### A. RRH PLANE

Fig. 2 shows the candidate locations of the RRHs in the urban region of Vienna. Red stars and green dots represent the MRRH and SRRH possible locations, respectively. Sectorization can be used at MRRHs locations. As shown on Fig. 2, the SRRHs are located at street corners to facilitate line-of-sight propagation. The parameters of the RRHs have been chosen to describe a realistic 5G radio network deployment. Transmitted powers of MRRHs and SRRHs are \( P_r \leq 43 \) dBm and \( P_r \leq 24 \) dBm respectively. When both, MRRH and SRRH, operate at FR1 (sub 6 GHz) the antenna gains are 18 dBi and 2 dBi respectively. When FR2 is considered (mmWave, 28 GHz) the antenna gain of the SRRHs is increased to 12 dBi, because more elements can be added at the antenna array, according to [16].

Besides the sensitivity and SINR constraints, the capacity constraint should also be satisfied. To this end, it is fundamental to introduce an estimation of the maximum bit rate capacity of the RRHs. To do so, it is necessary to know additional RRHs configuration parameters such as modulation, MIMO order, operating frequency band, bandwidth and 5G numerology.

As previously discussed, to reduce the inter-cell interference MRRHs and SRRHs operate at different
frequencies. In this work, 2.6 GHz (FR1) is selected for the MRRH, while SRRHs could operate at 3.6 GHz (FR1) or 28 GHz (FR2). In particular, n41 (2496-2690 MHz) and n77 (3300-4200 MHz) frequency bands are considered when MRRHs and SRRHs operate at FR1, with a bandwidth of 100 MHz for each. When the SRRHs operate at FR2, the selected frequency band is n257 (26.50-29.50 GHz) with 300 MHz of bandwidth.

While the maximum modulation order considered for all the frequency ranges is 256QAM, lower values would be dynamically assigned according to the interference and propagation conditions. Regarding the MIMO, 8 × 8 and 16 × 16 are considered in FR1 and FR2 respectively. Finally, the theoretical RRHs maximum capacity is estimated according to [17], and using the values summarized in Table 2.

In the scenario used to test, the RRH plane is composed of 41 possible RRHs locations, with 8 of them for possible MRRHs while the remainder are for possible SRRHs.

B. DEMAND PLANE

The demand plane contains the TDPs that are allocated in the set \( Z \) and their UEs. The whole map is divided into demand regions.

The proposed optimization algorithm does not depend on the strategy used to split the demand plane. This work considers two options for splitting. The first one is based on a regular map division: the whole map is divided into \( \sqrt{|Z|} \times \sqrt{|Z|} \) homogeneous zones. This simple approach has the advantage that by increasing the number of divisions (so, by decreasing the area of one zone), a finer tuning is obtained at the expense of increasing the computational complexity. The second is an irregular map division: the number of zones is equal to the number of candidate RRHs, being the center of the zone the point where the RRHs is located. In other words, the zones are defined using a Voronoi diagram, which is a better approach for those cases where traffic is not uniformly distributed and there are hotspots or areas with a high demand. In this diagram each map point is associated with the zone of the nearest RRH. As result, the dimension of the zones decreases in areas with higher RRH spatial density and higher concentration of UEs. Once the demand plane has been split, the traffic demand of each zone should be estimated. It depends on the demand of the UEs that fall inside the zone.

C. UE PLANE

As it has been mentioned above, each UE is associated to a SFC provided by a slice of a specific SP. Different kind of services have been modeled to generate the traffic demand. The voice and video services on 5G networks will be delivered based on the IP Multimedia Subsystem (IMS). In general, this kind of services are enclosed in the standardization of Voice/Video over New Radio (VoNR) [18]. In particular, three examples of these services, which have been specified by [19], are considered in this work: conversational voice, HD video and Standard-Definition (SD) video.

The GBR of each service must be selected by the SP to guarantee a specific QoS. In this case, the conversational audio service uses Enhanced Voice Services (EVS) codec (EVS), which has different bit rates configurations with a maximum value of 128 kbps; this is the value considered as the GBR to provide the maximum QoS to the end user. On the other hand, the video services use H.265 and EVS codecs. In the proposed scenario, two different video qualities are considered: HD video and SD video with 10 Mbps and 2 Mbps of GBR respectively, which is consistent with [20].

Table 3 summarizes the service parameters, where the parameter SP mix represents the percentage of UEs subscribed to each SP/service.

A random user distribution can be used when the demand plane is regular. When the map is split using a Voronoi diagram, more users are allocated in the SRRH regions. In a general case, UEs should be allocated according to the traffic measures and reports that the MNO has about the region.

Fig. 4 shows the demand plane with Low traffic (LT) profile that contains 30000 randomly distributed UEs in a regular divided map of 49 zones (7 × 7) or TDPs. Fig. 4a shows the traffic demand of each TDP in Mbps \( D_z \). On the other hand, Fig. 4b shows the number of UEs that belong

![FIGURE 3. Hierarchical structure of the scenario.](image-url)
TABLE 2. RRH plane: configuration and maximum capacity.

<table>
<thead>
<tr>
<th>Frequency Range</th>
<th>Bandwidth (MHz)</th>
<th>Modulation</th>
<th>MIMO</th>
<th>( D_r ) (Gbps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FR1</td>
<td>100</td>
<td>256QAM</td>
<td>8 \times 8</td>
<td>4</td>
</tr>
<tr>
<td>FR2</td>
<td>300</td>
<td>256QAM</td>
<td>16 \times 16</td>
<td>28</td>
</tr>
</tbody>
</table>

TABLE 3. Service parameters.

<table>
<thead>
<tr>
<th>Service</th>
<th>GBR/Best-effort</th>
<th>( D_{\text{min}} ) (Mbps)</th>
<th>SP mix (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>voice</td>
<td>GBR</td>
<td>128 kbps</td>
<td>25</td>
</tr>
<tr>
<td>HD video</td>
<td>GBR</td>
<td>10 Mbps</td>
<td>15</td>
</tr>
<tr>
<td>SD video</td>
<td>GBR</td>
<td>2 Mbps</td>
<td>30</td>
</tr>
<tr>
<td>FTP</td>
<td>Best-effort</td>
<td>-</td>
<td>15</td>
</tr>
<tr>
<td>Web</td>
<td>Best-effort</td>
<td>-</td>
<td>15</td>
</tr>
</tbody>
</table>

FIGURE 4. Traffic distribution of the demand plane at each zone or TDP with a total of 30000 UEs: a) traffic demand in Mbps, b) number of UEs.

to each zone. In order to analyze the performance of our proposal, medium and high traffic profiles have also been considered, with 60000 and 300000 UEs, respectively.

D. PROPAGATION MODEL

The use of an adequate propagation model is fundamental to analyze a realistic system; especially in a millimeter-wave 5G environment. In this paper, a 3D ray-tracing map-based propagation model is employed [21]. It is similar to the METIS model for urban macro and micro cells [22]. This channel model provides correlation and spatial consistency because it employs deterministic and physical principles accounting for scattering mechanisms, such as diffraction, scattering and blocking.

It is important to mention that the proposed analysis is limited to the downlink and outdoor communications. In particular, it is unfeasible to provide indoor communications at 28 GHz. Table 4 summarizes the parameters of the scenario.

The air interface establishes not only the considered propagation model but also the RRH–TDP association, which has been described by the binary variables \( x_r, z \), including cooperation among RRH to serve the same TDP or demand zone, as represented in Fig. 3.

V. RESULTS: ANALYSIS AND DISCUSSION

This section presents the results of testing a radio network deployment using the proposed optimization algorithm. These results demonstrate how the algorithm reduces the number of required active RRH. Besides, it offers acceptable coverage and satisfies the UE requirements in terms of QoS. This is crucial because it entails a considerable reduction in the network cost, with the consequent improvements in energy-saving, necessary when a large amount of cells are deployed as it is the case of 5G and beyond.

The algorithm could adapt to variations in traffic patterns and load, recalculating the set of RRHs that should be active for each case. Without the optimization procedure, the MNO would activate the 41 available RRHs, not benefiting from the resource, cost, and energy-saving improvements. The optimization gains will be compared with this baseline situation [12] through the different figures and tables analyzed in this section.

As 3GPP includes different split options of the protocol stack between BBU and RRH to reduce bandwidth and latency requirements, it is fundamental to consider them in the optimization process. For this reason, the following splits are analyzed: split option 8 that corresponds to a fully centralized C-RAN; split 6, where MIMO precoding and Orthogonal Frequency-Division Multiplexing (OFDM) modulation are maintained at the RRH side; and split 1 that represents a traditional architecture where all the baseband functions are allocated at the RRH (see Fig. 1).

Besides, a comparison of results when SRRHs operate at different frequency ranges (FR1 and FR2) is included. In particular, MRRHs operate at 2.6 GHz, while SRRHs could work at 3.6, 5, and 28 GHz. For simplicity, only the 3.6 and 28 GHz cases are analyzed since there are no significant differences between the results obtained at 3.6 and 5 GHz.
Finally, it is interesting to stress the algorithm considering different traffic loads. For this purpose, three data traffic options have been considered: Low, Medium, and High Traffic patterns, with 30000, 60000, and 300000 UEs, respectively. The split of service parameters with the different percentages of GBR and Best-Effort users is given in Table 3.

The simulations and modeling have been carried out in MatLab. Specifically, the convex programming software CVX [23] with the Mosek solver [24] have been used to solve the optimization problem. An MSI Prestige 15 A10SC computer with a Core i7 10th gen. CPU and 32 GB of RAM has been used to carry out the simulations.

### A. COST REDUCTION

As it has been explained in subsection III-B, the cost is normalized using the maximum cost of the considered split option, which is $C_S(\sigma \ |M| + |S|)$, where $\sigma$ represents the ratio between the cost of MRRHs and SRRHs. The value of $\sigma$ changes depending on the split option. For split 8 (fully centralized C-RAN), it corresponds to a value of $\sigma = 1$, whereas for split options 6 and 1 the assigned values are $\sigma = 10$ and $\sigma = 50$, respectively. However, this parameter should be adjusted according to the cost of the available devices. Cost differences associated with power amplifiers and antennas have not been considered in the value of $\sigma$, nor the additional cost of the hardware equipment when working at higher frequencies, because the purpose is to measure the impact of different split options. However, they could be easily included by changing $\sigma$ values.

It is fundamental to fix the weights $\omega_1$ and $\omega_3 = 1 - \omega_1$ to solve the multi-objective optimization problem. Equation (29a) shows that these weights are associated with the cost and coverage-QoS optimization, respectively. The considered values for $\omega_1$ cover the range from 0 to 1 with a step of 0.2. For instance, if $\omega_1 = 0.2$ and $\omega_3 = 0.8$, the algorithm provides more importance to coverage than cost reduction optimization.

Table 5 shows the comparison before and after running the optimization algorithm for the combinations of the considered parameters. The first three columns indicate the traffic pattern (Low, Medium, or High), the frequency range for the SRRH, which can be 3.6 GHz or 28 GHz, and the considered splits (8, 6, and 1). The fourth column displays the cost of the deployment before the optimization, that is, assuming that all the RRHs in the scenario are active. Values under the Normalized Cost columns give the normalized cost factors after the optimization for different $\omega_1$: from 0, meaning that the optimization is focusing on cost reduction, to 1, meaning that the optimization is focusing on cost reduction. Absolute cost values could be obtained by multiplying the normalized factor by the cost value before optimization. The final columns under the Coverage-QoS label provide the percentage of covered zones after the optimization.

The data from Table 5 can be used to extract multiple conclusions:

- Assuming that only the solutions with a final Coverage-QoS higher than 95 % are acceptable, it is possible to see that some combinations of parameters should not be used. It is the case of Medium Traffic (60000 users) at 3.6 GHz where, regardless of the split option and the considered weights, the requirements are never achieved. Even when Coverage-QoS is prioritized ($\omega_1 = 0$), results show that all the 41 RRHs need to be active, but the maximum achieved Coverage-QoS is only 84 %. For this reason, the combination High Traffic (300000 users) at 3.6 GHz is not analyzed, as it is known in advance that this combination will never accomplish the coverage-QoS requirement.

- On the other hand, there is always a solution that guarantees a Coverage-QoS higher than 95 % in the remaining combinations, and in most cases, there is more than one solution. If this is the case, the best one in terms of cost reduction is the solution associated with the higher $\omega_1$ value, because it is the solution that maximizes the cost reduction of the scenario, allowing for a higher number of inactive RRHs with the consequent energy-saving.

- Cost reduction is indirectly given in Table 5 as the complementary of the Normalized Cost value. For example, in the first row, the value is 0.68 for LT, FR1, split 1, and $\omega_1 = 0$. In this case, the algorithm provides a cost reduction of 32 %.
B. COST VS COVERAGE-QoS

Table 5 shows the trade-off between the achieved cost reduction and the Coverage-QoS of the UEs in the scenario. Fig. 5 illustrates this trade-off. Each circumference represents a different Coverage-QoS percentage, starting with 50% for the most internal, meaning that only 50% of the TDPs of the scenario have been covered with the required QoS, to 100% for the most external, meaning that all the coverage-QoS requirements have been fulfilled. On the other hand, each radial shown in Fig. 5 represents a different cost reduction value, ranging from 0 to 100%, written at the edge of the radial. Remember that cost reduction is calculated with respect to the maximum cost, obtained when all the RRHs remain active and shown in Table 5. The colored region in green is the area where the coverage-QoS is higher than 90%. The light blue area represents the region where the cost reduction is higher than 20%. Each point (circles, triangles, or rhomboids) is obtained after running the optimization and represents the result for different $\omega_1$ values. The points closest to the most internal circumference are associated with the maximum value $\omega_1 = 1$. Furthermore, the distance to the center of the circumference increases as the value of $\omega_1$ decreases, indicating that coverage-QoS is gaining priority with respect to cost reduction. The blue, red, and green lines in Fig. 5 connect the points with the same input simulation parameters (FR, split, and traffic level).

The blue-continuous lines represent split 8, while red-discontinuous and green-punctured lines represent split 6 and 1, respectively. The cases shown in Table 5 that do not achieve a good coverage-QoS after the optimization have not been represented in Fig. 5, as they are not considered valid solutions.

The best solution for each case (above 95% of coverage-QoS) is represented by the symbol located at the outermost end of the line. There are other symbols of the same type showing a better Coverage-QoS, even in some cases close to 100%, at the expense of an increasing cost. They are represented in Fig. 5 by the corresponding symbols, but they appear isolated (not connected to the line) to distinguish them. Despite this analysis, the MNO could select the solution point that best reflects the network requirements, addressing the trade-off between coverage-QoS and cost reduction.

Firstly, it should be appreciated that split 1 provides the highest cost reduction (around 95%) while offering a good coverage-QoS. This extreme cost reduction is due to the higher cost of the RRHs, as they contain all the baseband functionalities. Additionally, split 6 shows cost reductions of around 80-90%, while split 8 exhibits cost reductions of 70-50-40% for the different combinations of carriers and traffic. The cost is also reduced when moving to higher frequency bands because, as wider bandwidths are assigned to the RRHs, they are able to serve more UEs. However, if some system parameters change (as for example antenna gains or transmitted power), the number of RRHs needed to satisfy receiver sensitivity requirements, could increase when working at 28 GHz.
On the other hand, analyzing LT and Medium traffic (MT) cases at 3.6 GHz, it is shown that when the traffic profile is close to the maximum capacity of the whole network, the cost reduction decreases since most of the RRHs should be active to satisfy the demand. This behavior is similar at 28 GHz. However, in this case, the algorithm saves at least 20% of the network cost because of the higher capacity of the network at FR2. The cost reduction reaches approximately 75% when LT demand is considered, while the coverage-QoS reaches 100%.

Fig. 6 complements the previous results by representing the minimum cost that guarantees a Coverage-QoS higher than 95% after solving the multi-objective problem where Fig. 6(a), 6(b) and 6(c) stand for split 8, 6, and 1 respectively. Fig. 6 also shows the cost values before optimizing, to facilitate the comparison. The blue bars represent valid solutions, while the red bars represent solutions that do not satisfy the 95% of Coverage-QoS and neither reduce the cost. In terms of absolute cost deployment, optimal resource management, and computational capacity efficiency, split 8 is the best option for C-RAN networks, as has been widely shown. As the cost to deploy a new RRH is lower than with splits 6 and 1, the cost reduction when turning-off a RRH is also lower; however, it is still a significant reduction. It is fundamental to notice that the cost reduction is directly associated with an increase in the energy efficiency of the network. The lower the number of RRH required to satisfy the UEs requirements, the higher the energy-saving.

To summarize the analysis, it has been shown that the proposed algorithm is highly efficient allowing practical cost deployment reductions between 20 to 70% depending on the traffic level (Low, Medium, High), carrier frequency used and selected split option.

C. REDUCTION IN THE NUMBER OF ACTIVE RRH AND USAGE RATIO

The proposed optimization framework is worthy for the MNO, not only in the deployment phase but also to select the RRHs that should be active to satisfy the current traffic demand or even the predicted traffic demand. This could be achieved by combining the present algorithm, or the look-up tables that can be generated after running it, with optimized AI prediction tools allowing to analyze a dynamic scenario.

Additionally, it contributes enormously to energy-saving, a key parameter for 5G and future 6G networks. The number of required active RRHs after the optimization is shown in Fig. 7 for each split option, being Fig. 7(a) for the 3.6 GHz carrier frequency, while Fig. 7(b) shows the 28 GHz results. The first bar of each split corresponds to LT profile, while the second bars stand for MT or High traffic (HT), depending on the figure. Remember that the initial situation, without optimization, uses the 41 RRHs of the scenario, being 8 of them MRRHs. The presented solutions correspond to the weights considered in Fig. 6, which guarantee a 95% Coverage-QoS while reducing simultaneously the cost.

As expected, regardless of the frequency band, the number of active RRHs increases with the traffic demand. On the other hand, the distribution of MRRHs and SRRHs is detailed, showing that the algorithm prioritizes SRRHs when $\sigma$ increases, to reduce the cost. The MT simulation working at 3.6 GHz needs all the RRHs of the scenario to maximize the coverage. This MT solution is represented to show that the optimization algorithm could signal when the initial assigned resources are insufficient. In this case, to find a feasible solution, MNOs should increase the
number of RRHs deployed or the bandwidth allocated to them.

It is also interesting to show that in Fig. 7(a) the optimized solution ends with a similar number of required active RRHs (25-24), being the main difference that optimal split 8 requires keeping six active MRRHs while in splits 6 and 1 MRRHs are not needed at all. This also explains why in Fig. 6 there is a significant difference in cost between the three splits: in splits 1 and 6 the cost of a MRRH is very high compared to the cost of a SRRH, so the algorithm tries to avoid the activation of MRRHs when searching for the optimal solution.

From a practical point of view, it may be reasonable that the deployed MRRHs remain always active, the switch-off possibility being associated only to the SRRHs. For lack of space, these results have not been shown. However, it is implemented as a configuration option.

The operation at 28 GHz shows an enormous reduction in the number of required active RRHs, around one-third of them are needed in LT conditions while half of them are required when HT is considered. Propagation is not the limiting factor in the scenario because the cells are close enough and transmitted power and antenna gain are high enough to satisfy the UE requirements. However, only outdoor UEs have been considered, assuming that at 28 GHz, indoor users should be served by indoor Base Stations due to the large building penetration losses. The main difference with the 3.6 GHz operation is that the bandwidth associated to each RRH is higher.

The last fundamental parameter analyzed in this work is the usage ratio, \( r \), which has been previously defined as the ratio between the GBR traffic load at RRH \( r \) and its maximum

![FIGURE 7. Number of active RRHs vs split options. Left and right bars of each split option represent LT, and MT or HT cases, respectively.](image)

![FIGURE 8. Resource usage ratio \( (r) \) of the RRHs by the GBR slices in C-RAN (split option 8): (a) at 3.6 GHz with LT profile, (b) at 28 GHz with LT profile, and (c) at 28 GHz with HT profile.](image)
VI. CONCLUSION

Overall power consumption of future mobile networks should not grow beyond what it is now for 5G. A strategy that provides a sustainable optimal deployment not only for 5G but also for B5G and 6G radio networks has been provided. The main objectives are to reduce the footprint on energy efficiency, the deployment and operational costs of the network while maintaining the coverage-QoS. This complex problem has been modeled, introducing a Multi-objective ILP optimization algorithm to select the optimum distribution of the RRHs on the densest zone of the city of Vienna.

The proposed algorithm is tested by using a realistic scenario that includes 41 possible RRHs in a heterogeneous deployment with MRRHs and SRRHs, UEs modeled with different services, and an accurate 3D ray-tracing propagation model. Additionally, operation at frequency bands 3.6 and 28 GHz, as well as different C-RAN split options are studied.

As so many parameters can be compared after the optimization, it is impossible to summarize the main results in few words. A detailed explanation is already given in Section V. Only mention that the algorithm reduces the deployment cost while maintaining the coverage-QoS better than 95%. Especially, at 3.6 GHz with low traffic demand, the cost reduction is around 35%, while at 28 GHz it reaches 70% with LT profile and almost 50% under a HT condition.

The proposed algorithm also includes RRH cooperation. However, the detailed analysis and comparison of the achievements with/without RRHs cooperation has been left for a future publication, to simplify a bit the comprehension of the mathematical model exposed, which is quite complex and requires a large amount of variables. Let’s mention that the results obtained so far with cooperation are promising, showing and additional cost reduction and improving the energy efficiency of the network.

Besides, the option of reducing the cost based only on SRRHs activation/deactivation is currently being included in the algorithm. This upgrade would allow modelling a heterogeneous network where the control plane is managed by few MRRHs that always remain active, to provide better indoor communications.

Finally, this approach will be upgraded by integrating prediction tools based on AI to efficiently manage the resources centralized at the BBU pools. To do so, the AI algorithms analysed by the authors in a previous publication [4] will be used to deploy the optimal solution on a real-time or near-to-real-time basis, attending of course the limitation that an RRH cannot be activated/deactivated continuously in a real network.

This tool undoubtedly could help the MNOs to improve their network planning by detecting problems, providing a network diagnosis, optimizing and controlling the network by allowing the operator to balance between coverage-QoS and cost reduction and consequently power consumption savings.

REFERENCES


ROLANDO GUERRA-GÓMEZ received the B.S. degree in telecommunications and electronics engineering from the Technological University of Havana, Cuba, in 2015, and the M.S. degree in applied telecommunications and engineering management (MASTEAM) from the Universitat Politècnica de Catalunya (UPC, BarcelonaTech), in 2020, where he is currently pursuing the Ph.D. degree in telecommunications engineering.

From 2015 to 2018, he was a Researcher with the Wireless Communications Laboratory, Technological University of Havana, Cuba. Since 2018, he has been part of the Group of Wireless Communications and Technologies (WiComTec), Department of Signal Theory and Communications, UPC. His research interests include 5G, C-RAN, smart antennas, and resource allocation.

SILVIA RUIZ-BOQUÉ (Member, IEEE) received the Ph.D. degree in telecommunication engineering from the Universitat Politècnica de Catalunya (UPC), Spain, in 1989. In 1992, she became an Associate Professor with UPC, where she is currently the Head of the Wireless Communication and Technologies Research Group (WiComTec). She is the responsible of the NET layer WG3 of the European COST Action CA15104 Inclusive Radioaccession Networks for 5G and beyond. Her research interests include the field of radio communication systems, and more specifically in 5G, LTE-A, and NB-IoT radio network planning and optimization.

MARIO GARCÍA-LOZANO received the M.S. and Ph.D. degrees in telecommunications engineering from the Universitat Politècnica de Catalunya (UPC, BarcelonaTech), Spain, in 2001 and 2009, respectively. He has more than 15 years of experience in different radio network planning and optimization activities both at academia and industry. He is currently an Associate Professor at UPC, and his research activities are focused in the field of radio resource management and the optimization of wireless networks. He has actively participated in more than 25 competitive research projects and contracts with the industry. He currently leads the Spanish CRIN-5G Project at UPC. He was a recipient of three best paper awards and was an Advisor of the student team that won the international competition for mobile network planning organized by the company ATDI.

UMAR SAEED received the B.E. degree in electronics engineering from the National University of Sciences and Technology (NUST), Karachi, Pakistan, in 2007, the M.Sc. degree in communications engineering from Aalto University, Espoo, Finland, in 2012, and the Ph.D. degree in the area of signal processing for atmospheric remote sensing from the Department of Signal Theory and Communications, Universitat Politècnica de Catalunya, Barcelona, Spain, in 2016.

From 2017 to 2019, he was a Postdoctoral Researcher at Aalto University, working on research topics related to vehicle-to-everything (V2X) communications. Since 2019, he has been working as a System Architect for 5G mobile networks at Nokia. He is also a Visiting Researcher at Aalto University, where his current research interests include V2X communications, remote driving, and ultra-reliable low latency communications (URLLC).