Optimizing Flying Base Station Connectivity by RAN Slicing and Reinforcement Learning

DICK CARRILLO MELGAREJO[1,2,3], (Member, IEEE), JIRI POKORNY[4,5], PAVEL SEDA[4],
ARUN NARAYANAN[1], (Member, IEEE), PEDRO H. J. NARDELLI[1], (Senior Member, IEEE),
MEHDI RASTI[1], (Senior Member, IEEE), JIRI HOSEK[4], (Senior Member, IEEE),
MILOS SEDA[6], DEMÓSTENES Z. RODRÍGUEZ[7], (Senior Member, IEEE),
YEVGENI KOUCHERYAVY[5], AND GUSTAVO FRAIDENRAICH[2]

[1]Department of Electrical Engineering, School of Energy Systems, Lappeenranta-Lahti University of Technology (LUT), 53850 Lappeenranta, Finland
[2]School of Electrical and Computer Engineering, State University of Campinas (UNICAMP), Campinas 13083, Brazil
[3]Nokia Bell Labs, 02610 Espoo, Finland
[4]Department of Telecommunications, Faculty of Electrical Engineering and Communication, Brno University of Technology, 601 90 Brno, Czech Republic
[5]Unit of Electrical Engineering, Tampere University, 337 20 Tampere, Finland
[6]Institute of Automation and Computer Science, Faculty of Mechanical Engineering, Brno University of Technology, 601 90 Brno, Czech Republic
[7]Department of Computer Science, Federal University of Lavras, Lavras, 37200 Minas Gerais, Brazil

Corresponding author: Dick Carrillo Melgarejo (dick.carrillo.melgarejo@lut.fi)

ABSTRACT The application of flying base stations (FBS) in wireless communication is becoming a key enabler to improve cellular wireless connectivity. Following this tendency, this research work aims to enhance the spectral efficiency of FBSs using the radio access network (RAN) slicing framework; this optimization considers that FBSs' location was already defined previously. This framework splits the physical radio resources into three RAN slices. These RAN slices schedule resources by optimizing individual slice spectral efficiency by using a deep reinforcement learning approach. The simulation indicates that the proposed framework generally outperforms the spectral efficiency of the network that only considers the heuristic predefined FBS location, although the gains are not always significant in some specific cases. Finally, spectral efficiency is analyzed for each RAN slice resource and evaluated in terms of service-level agreement (SLA) to indicate the performance of the framework.

INDEX TERMS Flying base stations, UAVs, location optimization, wireless communication, deep-reinforcement learning.

I. INTRODUCTION

Extensive developments in the field of unmanned aerial vehicles (UAVs) have opened many opportunities for new applications in both private and public domains, such as surveillance, transportation, environmental monitoring, industrial monitoring, agriculture services, and disaster relief [1], [2]. Recently, the increasing number of use cases employ UAVs as wireless hotspots or relays to extend network coverage in areas where it is required. More-over, nowadays, there are UAV applications used as a tool for communications at the application level, for example,

The associate editor coordinating the review of this manuscript and approving it for publication was Jiankang Zhang.

information sharing in social media or searching for missing persons. Another example is the recent floods in Germany [3], which showed that the infrastructure is still quite vulnerable. Therefore, it is worth pursuing solutions to overcome problems when the regular communication infrastructure stops working. Thus, the use of UAVs provides an essential resource for allowing the continuity of communications and supporting human operators to continue to communicating in search and rescue operations, thereby guaranteeing efficient operation [4]. In such scenarios, the option of rapidly and efficiently deploying a fleet of drones is crucial in quickly establishing a communication network capable of saving lives, especially as it might be difficult to use terrestrial means comprising temporary networking equipment, such as

a cell on wheels in natural disasters. This feature makes UAVs unique and crucial for deployment in such use cases [5].

In addition, deploying UAVs as flying base stations (FBS) has also recently emerged as a feasible response to highly localized traffic demands in next-generation cellular networks [6], [7]. Using UAVs in such a way provides an opportunity to exploit their agility of motion to improve the air-to-ground link capacity by optimal air placement [8], [9]. Typically, the above-mentioned use cases consider significantly large areas where multiple UAVs must be used. However, this leads to two major problems. First, the UAVs must be positioned to optimally cover as many users as possible [10], [11]. Second, the intracell and intercell interference must be mitigated [12]. The first problem can be effectively approached by using heuristic algorithms. These algorithms can provide a solution with a low computational time and good results, as shown in [13], for example. In the case of intracell interference, the system performance can be improved with a variety of multiple access techniques, such as orthogonal frequency-division multiple access (OFDMA). When intercell interference is taken into account, some popular schemes, such as frequency reuse, graph theory, and cooperative multi-point (CoMP) [9], [14]–[18], can be employed.

In our previous work [13], we addressed the UAVs position optimality using an heuristic methodology. However, the radio channel interference problem was not addressed in detail. The present work extends [13] by employing a representative wireless channel model. In addition, we proposed a radio access network (RAN) slicing framework that enables the allocation of radio resources (slices) carrying specific data services. Our proposed framework aims to accommodate a diversity of services over a single shared fifth generation (5G) infrastructure and lays the foundation for fine-grained service management in FBS networks. We have considered that an agile RAN slicing framework is an appropriate solution to achieve the performance requirements introduced by verticals on 5G communication networks. The RAN slicing framework comprises several interworking functional components, aiming at a flexible instantiation of radio services, that can cope with the increasing complexity of supporting FBS services. In our work, we consider three slices: enhanced mobile broadband (eMBB), ultra-reliable low-latency communication (URLLC), and massive machine-type communications (mMTC).

The allocation of these slices is achieved by optimizing a cost function that is directly related to the spectral efficiency (SE) of the downlink data transmission, which is constrained by the maximum power transmission and the number of RAN slices. A cellular network based on subchannels usually has a high probability of intercell interferences in the edge cell. To solve this intricate allocation problem, we introduce an intelligent component in the framework—i.e., a deep reinforcement learning (DRL) model—that improves the system performance and manages the radio resource allocation minimizing the interference. By using our proposed interference management methodology to optimize the SE

on each RAN slice, specific service-level agreement (SLA)[1] can be achieved between the network service provider and the customer.

To facilitate readers comprehension of this paper, the main contributions of this paper are summarized as follows:

- enhancement of the UAV location distribution algorithm proposed in [13], using a proper air-to-ground channel model to enable aa appropriate interference analysis.
- a novel RAN slicing framework is proposed to enable the use of advanced machine learning techniques, such as DRL.
- we propose a distributed DRL approach to mitigate the downlink interference, in which each FBS operates as an independent learning agent.
- three representative scenarios of FBSs are described and analyzed in detail to compare the SLA performance between the DRL and the benchmark.
- a multiagent learning technique is proposed to optimize a nonconvex problem in the FBS system model.

The rest of the paper is organized as follows. The model finding optimal placement of UAVs in a given area, used as benchmark in this research work, is presented in Section II. In Section III, the RAN slicing framework is defined, including the system model and the DRL methodology to allocate the radio resources. A detailed description of the optimization sequence is given in Section IV. The simulation setup is presented in Section V. Numerical results together with a thorough comparative performance analysis are discussed in Section VI. Our concluding remarks and future work are presented in Section VII.

II. UAV LOCATION OPTIMIZATION

The effective deployment of UAVs across a selected area is a difficult task that falls into the category of $\mathcal{NP}-$complete class of problems [19]. To address this task, we enhance the model presented in [13] for location covering of the UAV deployment in on-demand connectivity scenarios. The enhancement focuses in the elimination of interference for all pairs of newly added centres and for new and existing centres. The proposal in [13] did an extended explanation of the UAV location methodology. The main idea of this deployment was to select a feasible locations where UAVs can be located by this heuristic methodology. Based on that, the optimization algorithm selected the suitable UAV location to compose a list of UAVs and their respective locations.

A. DEPLOYMENT MODEL

To facilitate the understanding of the model, we provide the terminology used in the rest of this paper adapted to the terms used in the literature in Table 1. To localize the suitable positions for the FBS deployment, location optimization

---

[1]SLAs establish customer expectations regarding the service provider's performance and overall quality. It is a contract between the network service (NS) provider and the customer.

TABLE 1. Mapping mathematical terminology to communication networks terminology.

Mathematical Terminology	Wireless Networks Terminology
Facility	UAV or FBS node.
Demand	A user demanding certain level of network connectivity in a given area.
Capacity	Throughput that is requested by the sum of user requirements in a given area to be covered.
Multiple service	A user requires to be potentially covered by the x UAV or FBSs.
Existing service	Usually FBS nodes that already exist in the area to be covered and should remain after the reconfiguration or deployment phase.

problems is taken as inspiration. Currently, there exist several facility location problems dealing with many real-world use cases. Simply, they can be divided into location set covering problems (LSCPs) [20], and maximal covering location problems (MCLPs) [21]. The LSCP targets the minimization question in which the number of facilities that satisfy the network requirements and the need to be located is minimized. On the other hand, in the MCLP, a predefined number of resources tries to maximize its coverage. The main division is based on the available resources. Because these models have been used for a wide range of applications, they are not tied to telecommunication network deployment only. Hence, to the best of our knowledge, there is a gap in the literature that [13] aims to bridge for these models and the use case of UAV deployment. As a gap, we see the following factors (or their combination in one model):

(i) Separating the capacity of facilities or locations covering both downlink and uplink. This may differ for each location or facility.

(ii) Consideration of the existing services; for the use case of UAVs, it is essential to consider the existing infrastructure that can serve at least some demand from the locations to be covered ad hoc.

(iii) Splitting capacity requirements from one location to only one facility at a given moment.

(iv) Covering some locations with zero or a higher number of facilities. This is crucial for the must-have locations where it is not acceptable to lose the connectivity.

(v) The oversimplified wireless interference is based on the overlaps between cellular cells. We eliminate the interference by Eq. (8) and (9).

In [13], it is assumed that coverage availability is guaranteed. Capacity considerations are critical in the 5G-and-beyond deployments that expect a significant increase in network traffic. This is due to the growth of services that have considerably higher network throughput requirements, such as the growth of high-definition videos, augmented reality (AR) / virtual reality (VR), machine-to-machine communication, and other very intensive or demanding services in terms of network requirements. In particular, we have to deal with a high density of users that are simultaneously connected. For the existing facilities E_f and their corresponding decision

variables x_i, where $i \in E_f$, we set this parameter to 1, which means that all the existing facilities are taken into account. The allocation of capacity requirements between uploads and downloads represents a split of 100 Mbps to 80 Mbps for download (more extensive) and 20 Mbps for upload. In addition, we still need to satisfy the requirement that the demand j both for download and upload must be assigned to the same facility i.

To derive a mathematical model, let us set the following notation:

- I = a set of facility sites (UAV or FBS) $1, 2, \ldots, m$;
- J = a set of demand areas (customers) $1, 2, \ldots, n$;
- d_{ij} = the shortest distance between facility i and demand j;
- D_{max} = maximum distance which will be accepted for operation between the facilities and demands;
- l_j = number of facilities required for servicing demand j;
- $x_i \in \{0, 1\}$, where $x_i = 1$ means that facility i is selected, while $x_i = 0$ means that it is not selected.
- $N_j = \{i | d_{ij} \leq D_{max}\}$ = the set of facilities i that can cover the demand location j;
- C_i^u = upload capacity of facility i;
- C_i^d = download capacity of facility i;
- a_j^u = upload amount of demand at j;
- a_j^d = download amount of demand at j;
- $y_{ij} \in \{0, 1\}$ = nonfragmented demand from location j is assigned (1) or is not assigned (0) to facility i.

Now, we set out the following model extracted from [13] to minimize the number of required FBSs and maximize the cellular coverage area.

$$\min \sum_{i \in I} x_i, \tag{1}$$

$$\text{subject to } \forall j \in J : \sum_{i \in N_j} x_i \geq l_j \tag{2}$$

$$\forall j \in J : \sum_{i \in N_j} y_{ij} = 1 \tag{3}$$

$$\forall i \in N_j : C_i^u x_i \geq \sum_{j \in J} y_{ij} a_j^u \tag{4}$$

$$\forall i \in N_j : C_i^d x_i \geq \sum_{j \in J} y_{ij} a_j^d \tag{5}$$

$$(\forall i \in I)(\forall j \in J) : y_{ij} \leq x_i \tag{6}$$

$$\forall i \in E_f : x_i = 1 \tag{7}$$

$$(\forall i \in I - E_f)(\forall j \in I - E_f)(i \neq j) :$$
$$d_{ij} \geq (x_i + x_j - 1)d_{\min} \tag{8}$$

$$(\forall i \in I - E_f)(\forall j \in E_f) : d_{ij} > d_{\min} x_i \tag{9}$$

$$\forall i \in I : x_i \in \{0, 1\} \tag{10}$$

$$(\forall i \in I)(\forall j \in J) : y_{ij} \in \{0, 1\} \tag{11}$$

Constraint (3) guarantees that the demand j is assigned to only one facility at a given moment. All selected facilities must have a sufficient sum of their capacities for uploads and downloads to cover all upload and download demands (in

practice, this is an ideal case that network operators are trying to reach with the available resources), this is guaranteed by constraints (4) and (5). If a facility is selected to be removed from the network infrastructure, none of the demand should be assigned to it; this constraint is given by (6). In [13], the interference is simplified to minimize the coverage overlaps defined by the cells. Finally, consider the following: if d_{ij}, $i \in I$, $j \in I$ is the distance between the facilities i and j, then we can set that for all pairs of selected facilities, the facilities will have a distance greater or equal than a certain threshold, which is guaranteed by constraint (8).

Further, as it is typical in the state of the art dealing with location coverage with model enhancements, authors in [13] provided an alternative maximization model that considers a predefined number of new facilities (not yet optimized) to be located and covered as much area as possible:

$$\max \sum_{i \in N_j} \sum_{j \in J} y_{ij}(a_j^u + a_j^d), \qquad (12)$$

subject to the same constraints as in the minimization model, but with the addition of the following constraint for a predefined number of new facilities.

$$\sum_{i \notin E_f} x_i = p \qquad (13)$$

Note that this model targets the localization of FBS nodes, which defines the benchmark. This information is used as input in the optimization of FBS based on (RAN) slicing framework, which is detailed in Section III-E.

The brief explanation of the algorithm implementing the model is detailed in Algorithm 1. The *GetInputData* part represents the list existing facilities, expected/existing demands with coordinates, and additional important metadata. For the *CoveringModel* computation part the heuristics needs to implement repair operator satisfying all the model constraints (e.g., adding new UAVs or FBSs to the list of solution to satisfy the capacity requirements).

B. CONSIDERATIONS OF THE COMPUTATIONAL COMPLEXITY MODEL

The size of the search space is determined by the number of all possible selections of facilities. For m facilities, according to the binomial theorem, it is equal to

$$\binom{m}{1} + \binom{m}{2} + \cdots + \binom{m}{m} = (1+1)^m - 1 = \mathcal{O}(2^m). \qquad (14)$$

Furthermore, we need to find the most complex condition in extended models for $m < n$ (where n is the number of demand areas) to find the resulting computational complexity. In the minimization model, these are (6) and (11) in the corresponding equations of the maximization model, which require $m \cdot n$ operations. This is based on the fact that the resulting time complexity of these models is $\mathcal{O}(2^m mn)$ [13].

Algorithm 1 Algorithm That Optimizes the UAVs or FBSs Location

▷ Main part
1: **function** findLocations
2:    *GetInputData*
3:    Generate theoretical possible UAVs locations
4:    Apply CoveringModel() with these data
   ▷ Covering model
5: **function** CoveringModel()
6:    Generate possible solutions
7:    Apply repair operator providing feasible solutions
8:    Apply selected heuristics to find optimal solution
   Suitable locations for the UAV deployment

III. RAN SLICING FRAMEWORK

To complement the benchmark described in Section II, in this section we describe the proposed RAN slicing framework. We employ the standardized definitions of RAN slicing in 5G, the system channel model, the radio optimization problem formulation, and our proposed approach using DRL.

A. FRAMEWORK DESCRIPTION

The diverse performance requirements introduced by 5G communication networks are fertile ground for the application of an agile RAN slicing framework. Our proposed framework aims to accommodate a diversity of services over a single shared 5G infrastructure and lays the foundation for fine-grained service management in FBS networks. This RAN slicing framework primarily comprises several interworking functional components, aiming at a flexible instantiation of radio services. The proposed architecture is devised to cope with the rising complexity of supporting FBS services, achieving not only more manageable RAN slices but also conforming the business propositions sought by network operators and service provider stakeholders.

This framework comprises orthogonal physical resources that split the available bandwidth to support a specific number of network slices. In this specific work, we consider three slices: eMBB, URLLC, and mMTC. In a cellular network based on subchannels, the RAN slicing framework is exposed to a high probability of intercell interference, specially in the edge cell. To address this issue, we incorporate an intelligent component in the framework to manage the radio resource allocation using DRL. This interference management aims to achieve specific SLA policies between the network service provider and the customer by optimizing the SE on each RAN slice. A graphical description of the concept of the proposed framework is shared in Fig. 1. The system model and the DRL methodology are detailed in the following section.

B. SYSTEM MODEL

Consider a set $\mathcal{I}$ of I FBS providing downlink wireless service to a group of user equipments (UEs) in a geographical area $\mathcal{A}$. Each FBS $i \in \mathcal{I}$ serves an area $\mathcal{A}_i$, such that
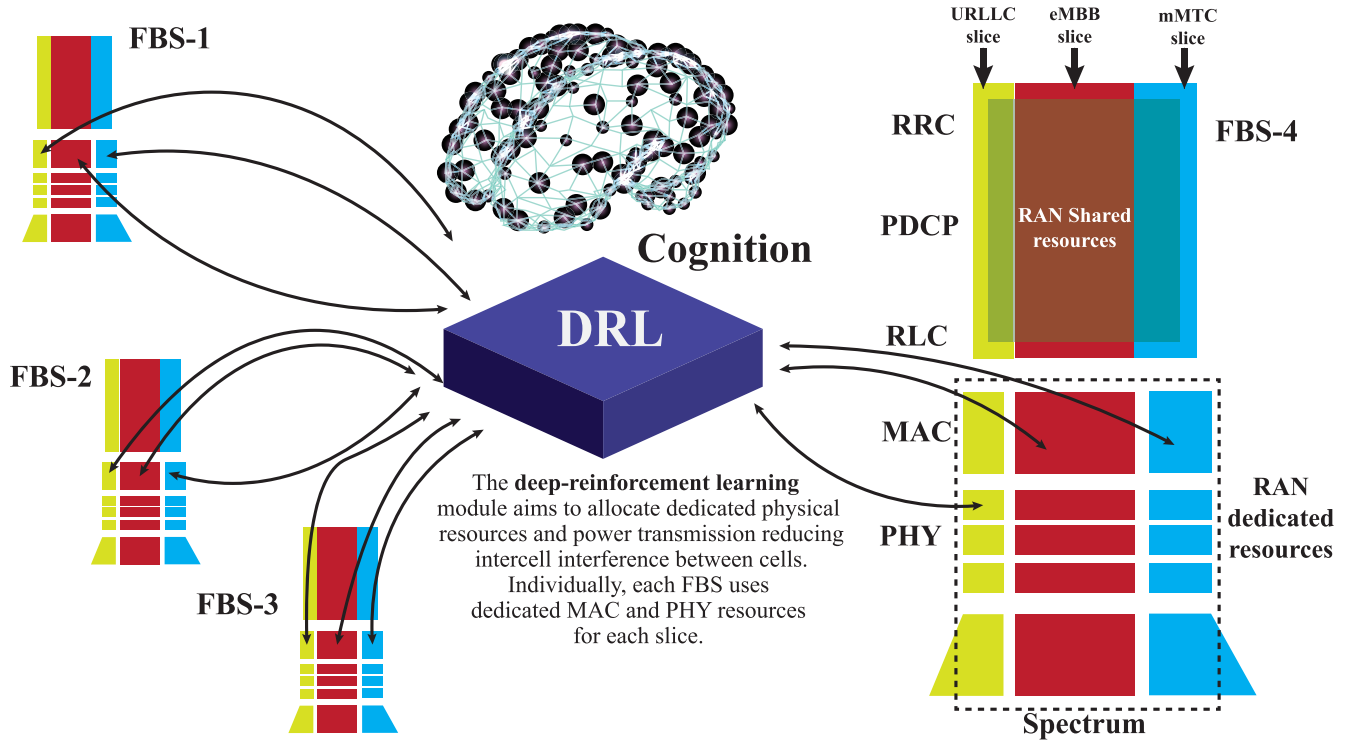
FIGURE 1. The RAN Slicing framework emphasizing the role of the DRL module. In this particular case we consider that each slice uses dedicated physical resources in each FBS. Thus, the cognition element optimizes the interference reduction using the deep-reinforcement learning approach.

$\cup_{\forall i \in \mathcal{I}} \mathcal{A}_i = \mathcal{A}$ and $\mathcal{A}_i \cap \mathcal{A}_k \neq \emptyset$ for any $i \neq k \in \mathcal{I}$. In other words, we consider that when UAVs are allocated by the optimization algorithm presented in Section II, it is possible that some cells have a significant intersection between them.

The path loss of the air-to-ground communication link from a typical FBS located at $x_i \in \mathbf{R}^3$ to a typical ground UE that is located at $y \in \mathbf{R}^3$ is given as follows [22]:

$$h_{v \to n}^{(t)}[dB](x_i, y) = 20 \log_{10} \left(\frac{4\pi f_c \|x_i - y\|}{c} \right) + \xi(x_i, y),$$

$$(15)$$

where f_c is the carrier frequency of FBS downlink communications, $\|x_i - y\|$ is the FBS–UE distance, c is the speed of light, and $\xi(x_i, y)$ is the additional path loss of the air-to-ground channel, compared with the free space propagation. The value of $\xi(x_i, y)$ can be modeled as a Gaussian distribution with different parameters $(\mu_{\text{LOS}}, \sigma_{\text{LOS}}^2)$ and $(\mu_{\text{NLOS}}, \sigma_{\text{NLOS}}^2)$ for line-of-sight (LOS) and non-line-of-sight (NLOS) links, respectively. Then, the downlink spectral efficiency achieved by the RAN slice m, the user n, at the time slot t from the FBS located at x_i to a UE located at $y \in \mathcal{A}_i$ is

$$\mathcal{C}_{n,m}^{(t)}(x_i, y) = \log_2 \left(1 + \gamma_{n,m}^{(t)}(x_i, y) \right), \quad (16)$$

where $\gamma_{n,m}^{(t)}(x_i, y)$ is the signal-to-interference-plus-noise (SINR) at the user n, on the RAN slice m, at the time slot t,

which is defined by (17)

$$\gamma_{n,m}^{(t)}(x_i, y) = \frac{\left[\beta_{l,m}^{(t)} \, g_{l \to n,m}^{(t)}(x_l, y) \, p_l^{(t)} \right]_{l=n}}{\sum_{v \neq l} \beta_{v,m}^{(t)} \, g_{v \to n,m}^{(t)}(x_v, y) \, p_v^{(t)} + \sigma_n^2}, \quad (17)$$

where $\beta_{v,m}^{(t)}$ is the binary variable that indicates the RAN slicing selection m transmitted from the UAV v at time t, $g_{v \to n,m}^{(t)}(x_v, y)$ indicates the downlink channel gain from the FBS v to the user n on the RAN slice m in the time slot t when the UE is located in the position y and the FBS in the position $x_v \in \mathbf{R}^3$, $p_v^{(t)}$ is the transmit power of the UAV v in the time slot t, and σ^2 is the additive white Gaussian noise power spectral density at the user receiver n.

$$g_{v \to n,m}^{(t)}(x_v, y_n) = h_{v \to n}^{(t)}(x_v, y_n) \left| \alpha_{n \to l,m}^{(t)} \right|^2 \quad t = 1, 2, \cdots, \quad (18)$$

where $h_{v \to n}(x_v, y_n)$ is the path loss in a linear scale, which is calculated in (15), and $\alpha_{n \to l,m}^{(t)}$ is the small-scale Rayleigh fading.

The probability of having an LOS link between the FBS j located at x_j and the UE located at y is given by [22]:

$$P_{\text{LOS}}(x_j, y_i) = \frac{1}{1 + a \exp \left(-b \left[\frac{180}{\pi} \sigma(x_j, y) - a \right] \right)}, \quad (19)$$

where a and b are constant values that depend on the communication environment, $\sigma(x_j, y) = \sin^{-1}\left(\frac{H_j}{\|x_j - y\|}\right)$ is the elevation angle, and H_j is the altitude of the FBS j. Then, the average downlink SE between an FBS i and the UE at n located in $y_n \in \mathcal{A}_i$ will be:

$$\bar{\mathcal{C}}_{n,m}^{(t)}(x_i, y_n) = P_{\text{LOS}}(x_i, y_n)\mathcal{C}_{n,m}^{(t)\text{LOS}}(x_i, y_n) \\ + (1 - P_{\text{LOS}}(x_i, y_n))\mathcal{C}_{n,m}^{(t)\text{NLOS}}(x_i, y_n). \quad (20)$$

C. RAN SLICING IN 5G

A simplified 5G logical architecture is composed of a core cloud, an edge cloud, and an RAN. The core cloud provides generic control plane signalization, slice management, mobility management, and authentication. The edge cloud performs some user plane functions as a packet/service gateway (P/S-GW) to improve latency communication on critical applications. It also enables data forwarding, control plane functions, and mobile edge computing platforms, such as content storage servers. In the radio access plane, the 3rd Generation Partnership Project (3GPP) defines the next-generation RAN (NG-RAN), which is comprised of next-generation NodeBs (gNBs) connected to the core network. This architecture is used to support the network slicing approach proposed in 5G. In this aspect, there are two types of subnets in the 5G slicing architecture: core network slice subnets and RAN slice subnets.

In the core network slice subnets, the network slicing operation used in the core network is controlled by the network slicing management. It is composed of the virtualized network function management (VNFM), the software-defined network (SDN) controller, the management and orchestration unit, and the virtualized infrastructure management (VIM). The VNFM maps the physical network functions to virtual machines (VMs); the SDN controller manages and operates the entire virtual network; the VIM allocates virtualized resources to VMs; and the management and orchestration unit creates, activates, and deletes network slices based on the service requirements.

In the RAN slice subnets, the gNB is a crucial enabler of network slices. It provides RAN slice subnets that are composed of a centralized unit (CU), multiple distributed units (DUs), and multiple radio units (RUs). The gNB functionalities are distributed in a flexible manner between the CU, DUs, and RUs. To manage their life cycles, the standard specifies the RAN network slice subnet template (NSST) and two management entities, such as the RAN network slice subnet management function (NSSMF) and the network function management functions (NFMFs) [23].

The core network slice subnets have been studied and developed in the current 5G with outstanding results. However, the RAN slicing is still an open topic, and it is not yet standardized. The RAN slicing aims to improve the efficient usage of available physical radio resources and simultaneously guarantees the SLA policies imposed in each slice.

D. REINFORCEMENT LEARNING AIDED UAVs

Machine learning is an approach that has become increasingly popular for sequential decision-making on wireless communication networks with applications in many diverse areas, such as smart grids, self-driving cars, and robotics. There are three machine learning categories, depending on the nature of the information or feedback available to the learning system: (i) supervised learning; (ii) unsupervised learning; and (iii) reinforcement learning (RL). In this paper, we use RL as the primary approach for optimizing the cost function based on spectral efficiency. RL is a technique that is concerned with how agents should determine the sequences of actions in an environment that will maximize cumulative rewards [24], [25]. It is a trial-and-error process where an agent interacts with an unknown environment in a sequence of discrete time steps to achieve a task. At time t, the agent first observes the current state of the environment, which is a tuple of relevant environment features and is denoted as $S^{(t)} \in \mathcal{S}$, where S is a set of possible states. It then takes an action $a^{(t)} \in \mathcal{A}$ from an allowed set of actions $\mathcal{A}$ according to a policy that can be either stochastic, i.e., π with $a^{(t)} \sim \pi(.|\mathcal{S}^{(t)})$ or deterministic, i.e., μ with $a(t) = \mu(\mathcal{S}^{(t)})$. Because the interactions are often modeled as a Markov decision process, the environment moves to a next state $S^{(t+1)}$ following an unknown transition matrix that maps state–action pairs onto a distribution of successive states, and the agent receives a reward $S^{(t+1)}$. Overall, the above process is described as an experience at $t + 1$ denoted as $e^{(t+1)} = (S^{(t)}, a^{(t)}, r^{(t+1)}, s^{(t+1)})$.

The goal is to learn a policy that maximizes the cumulative discounted reward at time t, defined as follows:

$$R^{(t)} = \sum_{\tau=0}^{\infty} \gamma^\tau r^{(t+\tau+1)}, \quad (21)$$

where $\gamma \in (0; 1]$ is the discount factor.

RL has been growing in popularity because it does not require an extensive network model. Instead, its learning process is based on the interactions with the environment that produces its optimal strategies. Owing to the possibility of combining RL with deep learning [26], DRL is a highly suitable method for solving problems with a high number of states and low prior knowledge, which is the case of the resource allocation scenario in RAN slicing.

DRL has recently been used in problems related to UAVs [27]–[31]. Moreover, the DRL approach has been exploited and applied to the problem of UAV position and resource allocation. In [30], the authors proposed a DRL algorithm based on echo state network (ESN) cells for optimizing the UAV path, cell association to minimize the intercell interference level, transmission delay, and transmit power level. In [32], the ESN algorithm was used based on a multiagent Q-learning approach, which was employed to predict the future positions of UEs and determine the positions of UAVs. However, this work did not consider UAV cooperation and capacity limitations of fronthaul links between UAVs and

regular base stations. Further, for the optimization of FBS placement, the studies in [27], [28], [33] used an RL algorithm. In this work, we present an RAN slicing framework based on a DRL methodology that complements the location optimization model obtained in [13] by adding a radio channel model and optimizing the RAN slice resources between multiple FBSs covering an arbitrary area.

E. RADIO OPTIMIZATION PROBLEM FORMULATION

To apply the DRL methodology explained in the previous subsection, we define the radio optimization problem that this paper aims to optimize. Thus, details of the cost function and its constraints are defined in the following.

Denoting RAN slices and power vectors in the time slot t as $\boldsymbol{\beta}^{(t)} = \left[\beta_{1,1}^{(t)}, \beta_{1,2}^{(t)}, \cdots, \beta_{N,M}^{(t)} \right]^T$ and $\boldsymbol{p}^{(t)} = \left[p_1^{(t)}, \cdots, p_N^{(t)} \right]^T$ respectively, we define the sum-rate maximization problem as

$$
\max_{\boldsymbol{p}^{(t)}, \boldsymbol{\alpha}^{(t)}} \sum_{n=1}^{N} \bar{\mathcal{C}}_n^{(t)}(x_i, y_n)
$$
$$
\text{s.t. } 0 \leq p_n^{(t)} \leq P_{\max}, \quad \forall n \in \mathcal{N},
$$
$$
\beta_{n,m}^{(t)} \in \{0, 1\}, \quad \forall n \in \mathcal{N}, \quad \forall m \in \mathcal{M},
$$
$$
\sum_{m \in \mathcal{M}} \beta_{n,m}^{(t)} = 1, \quad \forall n \in \mathcal{N}, \qquad (22)
$$

where $\bar{\mathcal{C}}_n^{(t)} = \sum_{m=1}^{M} \bar{\mathcal{C}}_{n,m}^{(t)}(x_i, y_n)$.

The nonconvex problem in (22) requires a highly complex approach that could also increase the computational complexity. To handle this nonconvex problem, we consider a multiagent learning scheme, where each transmitter, mounted in each FBS, operates as an independent learning agent. Each agent successfully executes two policies to determine its associated RAN slice and transmission power level. The proposed multiagent approach is easily scalable to more extensive networks and can operate with local information after training.

The components of the DRL methodology considered based on the system model described before is composed by:

- **Agents:** in the multiple learning approach, the FBSs represent the agents.
- **Policies:** two well defined policies are considered. π_1 to choose an specific RAN slice, and the π_2 to select a proper power level for each user.
- **Actions:** we consider two well defined group of actions. The discrete actions related to the selection of RAN slices, and the continuous action to choose the power transmission for each individual user.
- **States:** It is composed by a tuple of information related to the RAN slice allocation, the SE, interference in each individual user, gain and interference in each user.
- **Rewards:** a proportional value of the SE in each receiver (UE) is used as reward. It considers the following criteria: the SE is evaluated in every user with the condition of one neighbourhood base station (BS) or agent is not transmitting. Thus, if the SE value is significant, then

the BS being evaluated is penalized. In contrast, if the SE remains, then the BS is rewarded.

At the beginning of each time slot, each agent successively executes two policies to determine its associated transmission power level and RAN slice selection. For this purpose, the DRL considers two optimization approaches. The first considers a Deep Q-network to optimize a stochastic policy that aims to improve the RAN slice selection. A second Deep Q-network optimizes a deterministic policy to select a suitable power transmission value. The agent of the second Deep Q-network requires the RAN slice decision of the first approach to determine its state input before setting the transmit power of the agent. A brief explanation of this approach is done in Algorithm 2.

Algorithm 2 Algorithm of the DRL Approach
> Main Loop
1: **while** Stop Criteria not met **do**
2:     RAN slice Selection()
3:     Power Control()

> RAN slice selection
4: **function** RAN slice Selection()
    > Action Selection
5:     action $\leftarrow a_n^{(t)} \in \mathcal{A}_{\text{RAN-Slice}} = \{1, \cdots, M\} = \mathcal{M}$
6:     State set design $\leftarrow s_{n,m}^{(t)}$
7:     Training by a deep-Q Network
8:     Reward function design
9:     Update Policies: π_1

> Power Control
10: **function** Power Control()
    > Action Selection
11:     action $\leftarrow a_{n,a_n^{(t)}}^{(t)} \in \mathcal{A}_{\text{power}} = [0, 1]$
12:     State set design $\leftarrow s_{n,a_n^{(t)}}^{(t)}$
13:     Training by a deep-Q Network
14:     Reward function design
15:     Update Policies: π_2

IV. DESCRIPTION OF THE PROPOSED SOLUTION

This paper aims to complement and enhance the output obtained in [13], which is not optimal if evaluated in a real scenario. Thus, the resource allocation that we propose is conditioned to the prelocation of each FBS obtained using the methodology presented in Section II. As the benchmark does not consider any channel model to evaluate the intracell and intercell interference, it can get suboptimal results in practical wireless scenarios. In particular, we aim to address the following research questions that arise during the FBS network deployment in a real scenario:

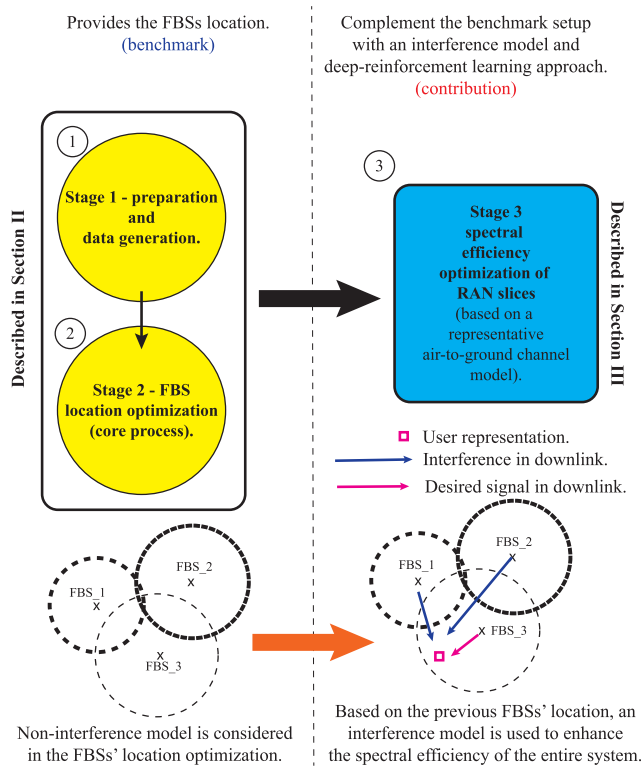- What can be the potential improvements based on the FBSs prelocation defined in [13]?,

FIGURE 2. Optimization flow diagram showing major steps and stages. Here, we emphasis the optimization done in the benchmark (in yellow) and highlight the DRL approach (blue).

- What is the performance of the simulation setup when different services are supported by the FBS network?,
- Is it possible to develop a practical optimization method that is capable of improving the performance of FBSs?.

To address these questions, we use a standard simulation model defined by 3GPP and a RAN slicing framework proposed in Section III, which is proposed to analyze the network performance using a DRL methodology to optimize the radio resources.

The whole optimization process is divided into three phases— data matrix generation, FBS location minimization, and RAN slicing optimization—as shown in Fig. 2. Here, Stages 1 and 2 have already been applied and verified in the original publication [13]. Details of each stage are described in the following paragraphs.

A. STAGE 1–DATA MATRIX GENERATION

In this stage, users are generated randomly and uniformly in one specific and fixed area. The FBSs are generated according to the desired radius. The larger the radius is, the lower the number of FBSs is required. A larger radius causes a higher percentage of overlaps. Note that Stages 1 and 2 do not take interference into consideration. Thus, interference has to be avoided by limiting the radio overlapping. The users are given random data rate requirements according to the traffic mix. The output of this stage is a matrix whose rows represent all FBSs and columns represent all users. The matrix is filled

with zeros if the user is inside the specific FBS cell radius, and otherwise, it is filled with ones. This matrix is completed and used as the input for Stage 2.

B. STAGE 2–FBS LOCALIZATION

This stage is composed of a software service that uses the output of Stage 1. This stage aims to optimize the rows of the generated matrix so that the customer requirements are met. In this stage, we apply the optimization model from Section II-A. Because the complexity of selecting the optimal rows (FBSs) from the matrix is $\mathcal{O}(2^m)$, heuristic algorithms are applied. As a consequence, we use the differential evolution and cuckoo search algorithms with repair operators presented in [13]. The output of the computation is a set of FBSs (with their locations from the previous stage) that should be used in the network deployment. Further, the location of each FBS is used as an input to Stage 3, considering a representative interference channel model.

C. STAGE 3–OPTIMIZATION OF RAN SLICES

We add an interference model based on the 3GPP standard to complement the previous stages to analyze and mitigate downlink interference. Then, we calculate the SINR, defined in (17), for each user considering the signal strength coming from the serving UAV and from the FBSs that are interfering with that specific user n. Using (22), our simulation model calculates an optimal spectral efficiency for each user n following the proposed DRL approach on the RAN slicing framework defined in Section III.

V. SIMULATION SETUP

Two scenarios were designed to emphasize the gain of the optimization models that were previously explained. In both cases, 1000 users were deployed following a uniform distribution in a specific area. Subsequently, in accordance with the constraints in each scenario, the FBSs location optimizer determined the coordinates of each FBS. The constraints are mostly related to the cell radius, that depends on the radio frequency operation, and throughput demand of each user or group of users.

Using the predetermined FBSs location, we consider a specific wireless system model. This system model considers an infinity backhaul capacity on each FBS. However, the offered bandwidth is limited in each FBS. The wireless system simulation considers that only a specific number of users are capable of getting the access stratum (AS)[2] interface. Thus, most of the users are in the RRC_IDLE and RRC_INACTIVE states, as defined in the 5G new radio (NR) standard. The remaining users that passed the random access procedure at the medium access control (MAC) layer are in the RRC_CONNECTED state.

To facilitate the analysis, we consider that each FBS is capable of supporting a fixed number of users N in the

---

[2]AS is a functional interface that is responsible for transporting data over the wireless connection and managing radio resources.
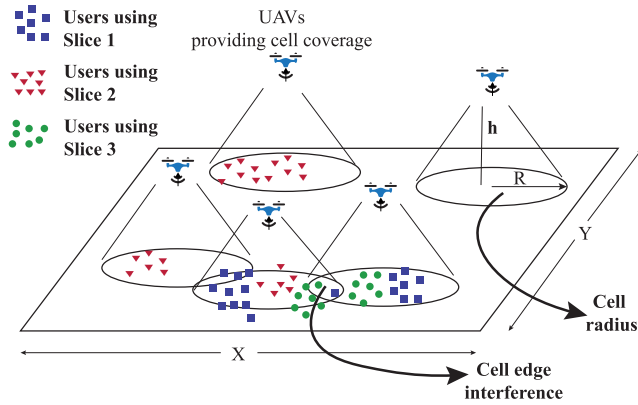
FIGURE 3. Generic simulation system model, considering the key constraints to be solved by the deep reinforcement learning approach.

RRC_CONNECTED state. Based on this assumption, the cellular network is simulated with L FBSs, and each FBS supports K RAN slices. As we consider only downlink transmission, the interference that each user suffers from nonserving FBSs is calculated following the assumptions defined in the 3GPP TR 36.931 Ver. 16 specifications. The height altitude for all FBSs is 20 m.

The DRL and the wireless system model are deployed in Python. The machine learning implementation is done using TensorFlow libraries to implement the deep Q-network setup. The deep neural network used in the simulation has 3 hidden layers with 200, 100, 50 fully connected neurons. The batch size is 128. The epsilon-Greedy Algorithm, used in this work, considers a maximum ϵ equal to 0.1 and an ϵ_{decay} of 0.9995. The implementation and further hyper-parameters are available in the following Github URL *http://www.github.com/TBD*.

Three different scenarios were defined, considering aspects such as a fixed number of FBSs (Scenario 1), and coverage maximization with a fixed number of users in the desired coverage area and in each cell (Scenarios 2 and 3, respectively).

A. SCENARIO 1–FIXED NUMBER OF AVAILABLE FBSs
This scenario considers $L = 20$ FBSs covering a specific area and supporting $N = 100$ users. To facilitate the analysis, it is considered that each FBS has $N/L = 5$ attached users. This scenario aims to identify an optimal cell radius supporting N users in the RRC_CONNECTED state. A graphical description of this scenario is shown in Fig. 4.

B. SCENARIO 2–MAXIMIZING THE NETWORK COVERAGE I
In this case, the number of FBSs (L) is defined by the optimization outputs of Stages 1 and 2, which were described in Section IV. Using Fig. 3 as a reference, the values of X and Y are 400 m. Thus, the number of FBSs is a function of cell radius; e.g., the larger the cell radius is, the smaller the number of FBSs needs to be deployed. In this scenario, the number of users N in the RRC_CONNECTED state is

TABLE 2. Comparison of parameters used in Scenarios 2 and 3.

Number of Users (N) – Scenario 2	Number of Users (N) – Scenario 3	Number of FBS (L)	Cell Radius in meters
114	152	38	50
114	76	19	75
114	44	11	100
114	32	8	120
114	24	6	140
114	16	4	160
114	16	4	180
114	12	3	200
114	8	2	300

always the same in the area defined by X and Y. This scenario is illustrated in Fig. 5.

C. SCENARIO 3–MAXIMIZING THE NETWORK COVERAGE II
In this case, the number of UAVs (L) is defined by the optimization outputs of Stages 1 and 2, which were described in Section IV. Based on Fig. 3, the value of X and Y is fixed at 400 m. In this scenario, the number of N users in the RRC_CONNECTED state varies when the number of FBSs changes, as it is stated in Table 2. Each cell radius has a fixed number of N users in the scenario. Fig. 6 depicts this scenario.

VI. ANALYSIS AND RESULTS
A. ANALYSIS OF SCENARIO 1
The spectral efficiency obtained in Scenario 1 is presented in Fig. 7. Here, different cell radii between 50 m and 300 m are evaluated. In the same figure, the first 500 iterations of the first episode[3] (between 1 and 10000 iterations) are compared with the first 500 iterations of the second episode (between 10001 and 20000 iterations). The policy, defined in Subsection III-D, is almost null in the first iterations; the spectral efficiency is very low, similar to a random allocation of RAN slicing resources. At the beginning of episode 2, the spectral efficiency improves quickly despite that a new user deployment is considered.

To facilitate the interpretation of the simulation results, the average spectral efficiency in each scenario is calculated. For instance, Fig. 8 presents the results of Scenario 1 before and after applying the RAN slicing resource allocation. Before applying the resource allocation, the maximum spectral efficiency is almost 1.5 bps/Hz for the cell radius of 50 m. The spectral efficiency decreases for the other cell radii when the cell radius is increased. This output represents the network performance when only Stages 1 and 2, described in Section IV, are considered. However, the performance is improved in all cell radii after applying the DRL approach. In all cases, the spectral efficiency is improved. For instance, the cell radius of 75 m achieves a spectral efficiency of more than 3.5 bps/Hz, representing a gain of 2 bps/Hz compared with the setup without optimization.

---

[3]With *episode* we refer to a complete sequence of interactions, from start to finish. In our simulations, one episode is completed after 10 000 iterations.
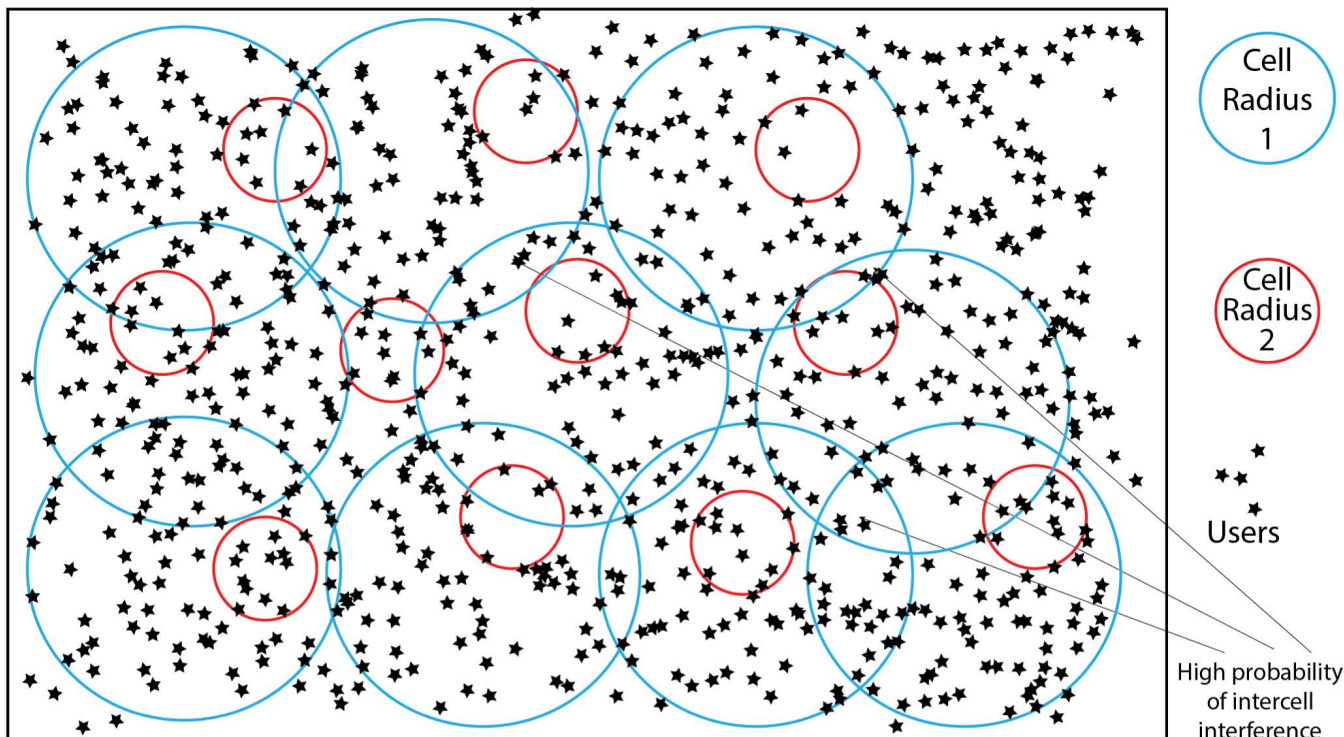
FIGURE 4. Visual description of Scenario 1. Here, the number of FBSs is a fixed, e.g., in this figure there are 11 UAVs. When the cell radius has a relatively high value with respect to the area of interest, the chances to have intercell interference is increased (cell radius 1). However, when the cell radius is small (cell radius 2), the chances of cell interference are reduced.
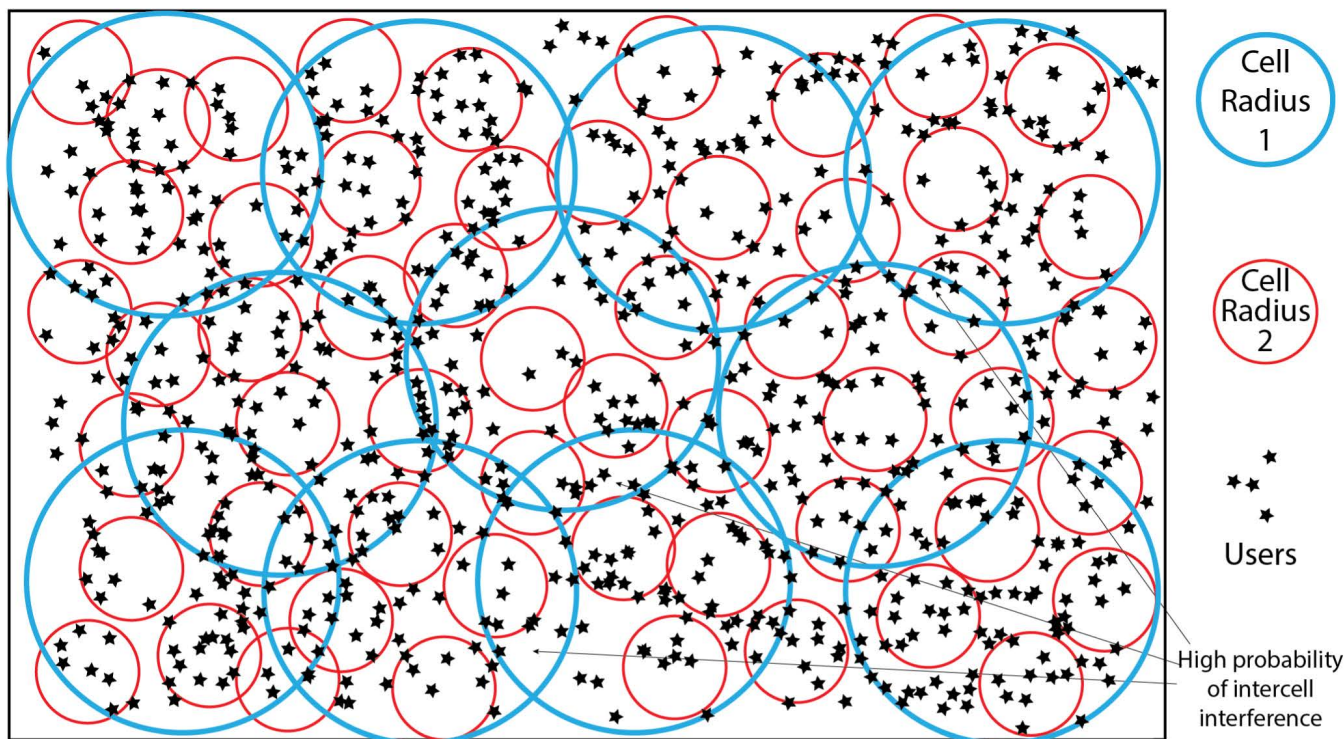

FIGURE 5. Visual description of Scenario 2. Here, the number of FBSs is not fixed. For instance, in this figure there are 11 UAVs for cell radius 1, and 75 UAVs for cell radius 2. When the cell radius has a relatively high value with respect to the area of interest, the chances to have intercell interference are increased (cell radius 1). However, when the cell radius is small (cell radius 2), the chances of cell interference are reduced. Here, the number of users (UEs) is fixed for both cell radius in the area of interest.
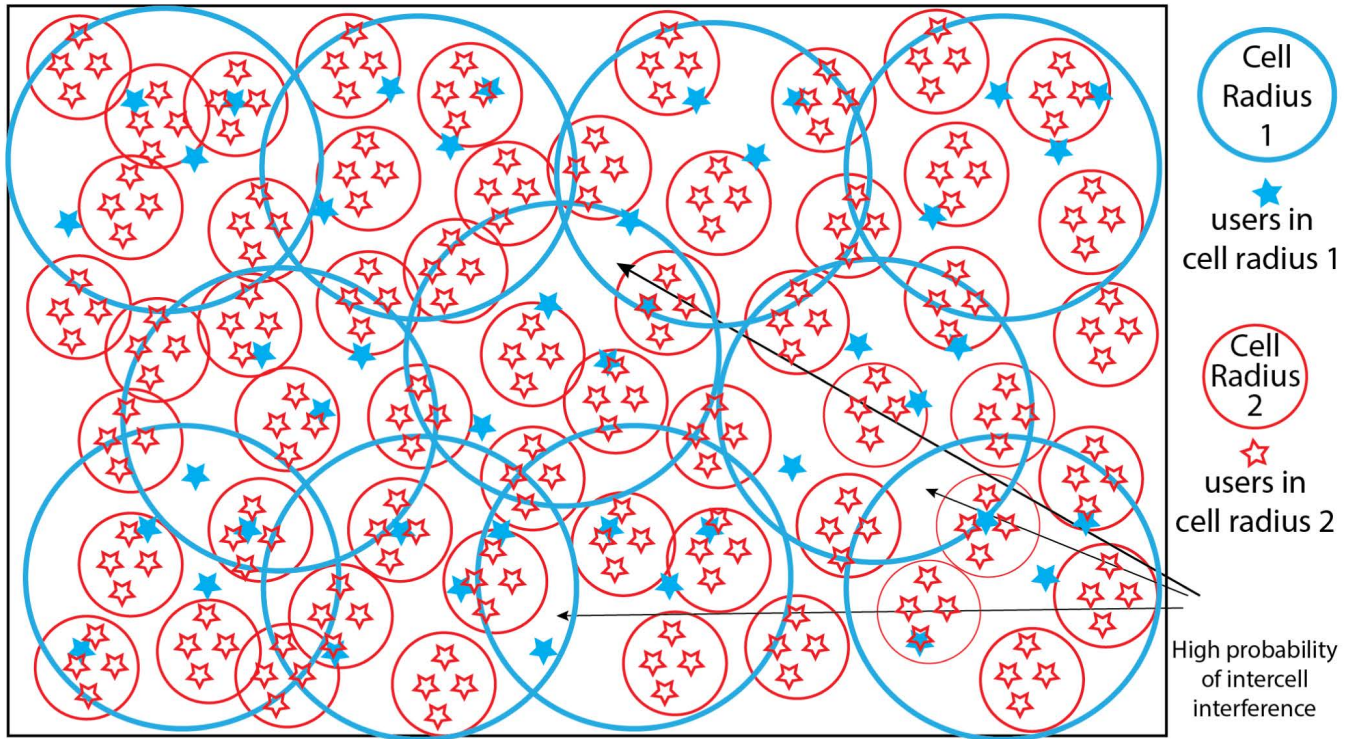
FIGURE 6. Visual description of Scenario 3. Here, the number of FBSs is not fixed. For instance, in this figure there are 11 UAVs for cell radius 1, and 75 UAVs for cell radius 2. Here, the number of users (UEs) is the same in every cell; in this illustrative representation every cell has four users. When the cell radius has a relatively high value with respect to the area of interest, the chances to have intercell interference are not significant (cell radius 1). However, when the cell radius is small (cell radius 2), the chances of cell interference are increased because the number of users in the area of interest is increased.
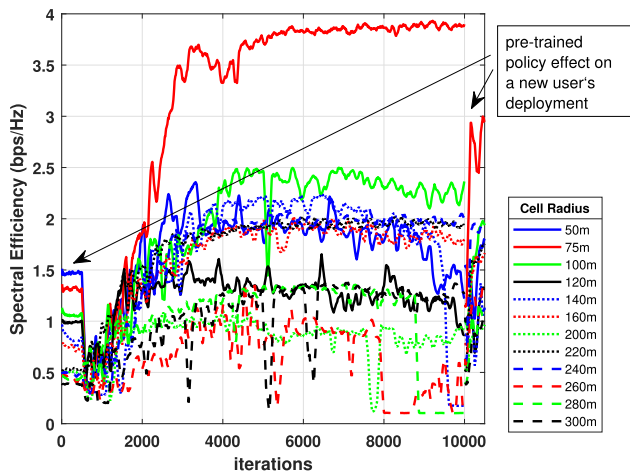

FIGURE 7. Spectral efficiency versus number of iterations obtained in Scenario 1 considering a diversity of radio cells. The first 500 iterations do not consider any pretrained DRL policy; after iteration 500 the policy gets a better understanding of the radio channel environment with a small degradation followed by a continuous improvement in most of the cells, especially in cells with small radius.

FIGURE 8. Mean of spectral efficiency of Scenario 1 comparing a diversity of cell radius before and after the RAN slicing resource allocation.

B. ANALYSIS OF SCENARIO 2

In Fig. 9, the spectral efficiencies of six different cell radii were obtained after 30 000 iterations. In contrast to Scenario 1, the reinforcement learning approach yields a moderate improvement in the SE when compared with the original benchmark scenario. The best performance is obtained by the small cell radius of 50 m.
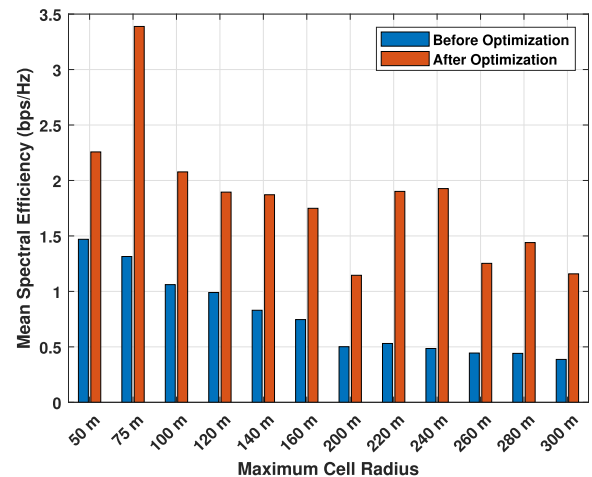
In Fig. 10, the lowest interference for Scenario 2 is for the cell with the radius equal to 100 m. However, this cell radius does not get the highest spectral efficiency because FBSs are deployed in such a way that the desired signal is not so strong when compared with cell radii of 50 and 75 m.

C. ANALYSIS OF SCENARIO 3

In this scenario, the SE indicates that the cell with a radius equal to 160 m obtains an SE of 4.5 bps/Hz, which is the best
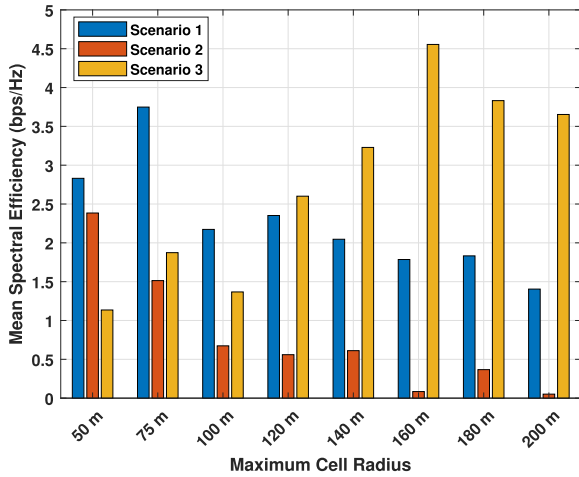
FIGURE 9. Mean spectral efficiency performance of different cell radii deployed for Scenarios 1, 2, and 3.
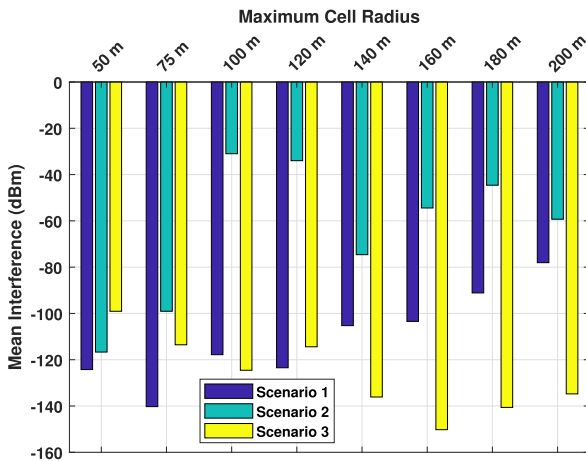

FIGURE 10. System interference of different cell radii deployed for Scenarios 1, 2, and 3.

performance when compared with the previous scenarios. The interference in Fig. 10 indicates that in all cell radii of this scenario, the interference is above -100 dBm, which represents a less aggressive scenario in terms of interference when compared with the previous scenarios.

D. ANALYSIS OF THE RAN SLICING NETWORK PERFORMANCE

The proposed RAN slicing framework enables the analysis of the SE performance of individual RAN slices. We considered that the cellular network supports three different RAN slices following the 5G requirements defined in [34]. One slice supports eMBB, the second supports URLLC, and the third supports mMTC. To show the performance of this RAN slicing framework, we considered the cell radius equal to 50 m in Scenario 2.

In Fig. 11, we show the improvement in the time domain of the policy generated by the DRL model to prioritize the slice that supports URLLC data transmission over the other RAN slices. Based on the spectral efficiency of the previous figure, Fig. 12 shows the performance of each RAN slice in terms of delay. Here, the performance is improved only after the
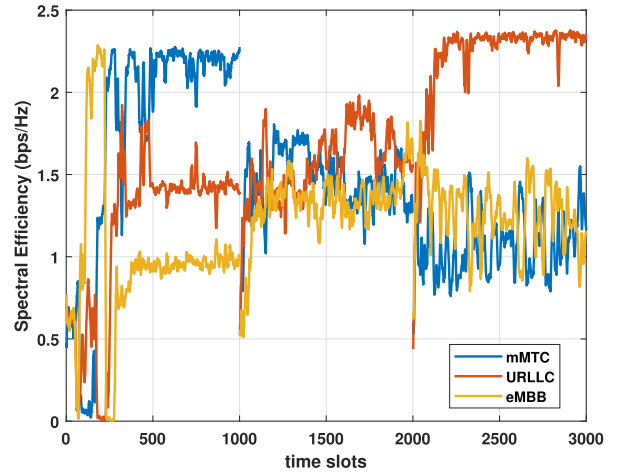

FIGURE 11. Spectral efficiency performance of three RAN slices deployed in the cell with a radius of 50 m in Scenario 2.
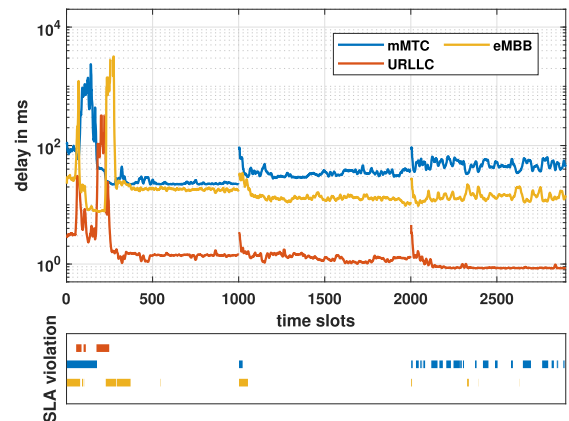

FIGURE 12. Delay of three RAN slices in the scenario with a cell radius of 50 m in Scenario 2.

time slot 400, which means that the DRL policy has enough understanding of the RAN slicing environment. In the same Fig. 12, we present another way to visualize the performance of each RAN slice though SLA violation. Here, we can verify that the SLA violation of the URLLC slice only happens when the DRL model is starting to improve the policies defined by the optimizer. The other RAN slices still present frequent SLA violation, especially the slice related to mMTC, which is mainly used for noncritical applications. However, this SLA violations can be improved if the RAN slicing target is relaxed, which is feasible in non-critical applications.

VII. CONCLUSION

In this paper, we have proposed a RAN slicing framework that enables the allocation of radio resources (slices) carrying specific data services since it can achieve the diverse performance requirements introduced by 5G wireless systems. Our proposed framework aims to accommodate a diversity of services over a single shared 5G infrastructure and lays the foundation for fine-grained service management in FBS networks. In particular, we have demonstrated that the DRL model with the proposed RAN slicing approach is suitable for improving the SE of a predefined location distribution of FBSs. Three scenarios were considered to evaluate the

performance of the proposed framework. These scenarios were generated by modifying key parameters, such as the number of FBSs and cell radius, and the optimization target of the predefined FBSs location. In all cases, the SE performance was improved when compared with the benchmark performance. However, the proposed methodology was more suitable for Scenario 3 because it presented a wireless network setup with low intercell interference.

The SE performance obtained in Scenarios 2 and 3 indicates that the deployment of the proposed framework in real scenarios can consider both approaches, and the specific cell radius should be chosen based on the network scenario. For instance, when there is a high density of users, Scenario 2 with a cell radius of 50 m is the most suitable setup to improve the SE system performance. On the other hand, if there is a moderate or low user density, Scenario 3 with a cell radius of 160 m will yield a better performance.

Potential future work that we have identified after the elaboration of this paper could include but is not limited to running a simulation on a setup with different radii on each FBS; adding another dimension to the analysis, based on the flying level or altitude of each FBS; and generating a full optimization of FBS location considering a cost function based on the system SE.

REFERENCES

[1] R. Shakeri, M. A. Al-Garadi, A. Badawy, A. Mohamed, T. Khattab, A. K. Al-Ali, K. A. Harras, and M. Guizani, "Design challenges of multi-UAV systems in cyber-physical applications: A comprehensive survey and future directions," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3340–3385, 2019.

[2] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Mobile unmanned aerial vehicles (UAVs) for energy-efficient Internet of Things communications," *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7574–7589, Nov. 2017.

[3] A. Fekete and S. Sandholz, "Here comes the flood, but not failure? Lessons to learn after the heavy rain and pluvial floods in Germany 2021," *Water*, vol. 13, no. 21, p. 3016, Oct. 2021.

[4] A. F. Ganazhapa, N. Palmieri, and M. I. C. Pushug, "Drones for connectivity support in disaster recovery," *Proc. SPIE*, vol. 11425, Apr. 2020, Art. no. 114250M.

[5] M. Mozaffari, A. T. Z. Kasgari, W. Saad, M. Bennis, and M. Debbah, "Beyond 5G with UAVs: Foundations of a 3D wireless cellular network," vol. 18, no. 1, pp. 357–372, Jan. 2019.

[6] G. Amponis, T. Lagkas, M. Zevgara, G. Katsikas, T. Xirofotos, I. Moscholios, and P. Sarigiannidis, "Drones in B5G/6G networks as flying base stations," *Drones*, vol. 6, no. 2, p. 39, Feb. 2022. [Online]. Available: https://www.mdpi.com/2504-446X/6/2/39

[7] P. Mach, Z. Becvar, and M. Najla, "Joint association, transmission power allocation and positioning of flying base stations considering limited backhaul," in *Proc. IEEE 92nd Veh. Technol. Conf. (VTC-Fall)*, Nov. 2020, pp. 1–7.

[8] J. Plachy, Z. Becvar, P. Mach, R. Marik, and M. Vondra, "Joint positioning of flying base stations and association of users: Evolutionary-based approach," *IEEE Access*, vol. 7, pp. 11454–11463, 2019.

[9] M. S. Hossain and Z. Becvar, "Soft frequency reuse with allocation of resource plans based on machine learning in the networks with flying base stations," *IEEE Access*, vol. 9, pp. 104887–104903, 2021.

[10] J. Qin, Z. Wei, C. Qiu, and Z. Feng, "Edge-prior placement algorithm for UAV-mounted base stations," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2019, pp. 1–6.

[11] T. Kimura and M. Ogura, "Distributed collaborative 3D-deployment of UAV base stations for on-demand coverage," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Jul. 2020, pp. 1748–1757, doi: 10.1109/INFOCOM41043.2020.9155283.

[12] H. Wang, H. Zhao, W. Wu, J. Xiong, D. Ma, and J. Wei, "Deployment algorithms of flying base stations: 5G and beyond with UAVs," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 10009–10027, Dec. 2019.

[13] J. Pokorny, P. Seda, M. Seda, and J. Hosek, "Modeling optimal location distribution for deployment of flying base stations as on-demand connectivity enablers in real-world scenarios," *Sensors*, vol. 21, no. 16, p. 5580, Aug. 2021.

[14] J. Yao and J. Xu, "Joint 3D maneuver and power adaptation for secure UAV communication with CoMP reception," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6992–7006, Oct. 2020.

[15] M. Mozaffari, A. T. Z. Kasgari, W. Saad, M. Bennis, and M. Debbah, "Beyond 5G with UAVs: Foundations of a 3D wireless cellular network," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 357–372, Jan. 2019.

[16] M. Kenyeres and J. Kenyeres, "Average consensus over mobile wireless sensor networks: Weight matrix guaranteeing convergence without reconfiguration of edge weights," *Sensors*, vol. 20, no. 13, p. 3677, Jun. 2020.

[17] B. Wang, R. Zhang, C. Chen, X. Cheng, L. Yang, H. Li, and Y. Jin, "Graph-based file dispatching protocol with D2D-enhanced UAV-NOMA communications in large-scale networks," *IEEE Internet Things J.*, vol. 7, no. 9, pp. 8615–8630, Sep. 2020.

[18] M. Kenyeres and J. Kenyeres, "Distributed mechanism for detecting average consensus with maximum-degree weights in bipartite regular graphs," *Mathematics*, vol. 9, no. 23, p. 3020, Nov. 2021.

[19] M. R. Garey and D. S. Johnson, *Computers and Intractability*, vol. 174. San Francisco, CA, USA: Freeman, 1979.

[20] C. ReVelle, C. Toregas, and L. Falkson, "Applications of the location set-covering problem," *Geographical Anal.*, vol. 8, no. 1, pp. 65–76, Sep. 2010.

[21] R. Church and C. ReVelle, "The maximal covering location problem," *Papers Regional Sci. Assoc.*, vol. 32, no. 1, pp. 101–118, Dec. 1974.

[22] Q. Zhang, W. Saad, M. Bennis, X. Lu, M. Debbah, and W. Zuo, "Predictive deployment of UAV base stations in wireless networks: Machine learning meets contract theory," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 637–652, Jan. 2021.

[23] *Technical Specification Group Services and System Aspects; Management and orchestration; Architecture Framework*, document TS 28.533, 3GPP, Technical Specification, Mar. 2020.

[24] V. Francois-Lavet, P. Henderson, R. Islam, M. G. Bellemare, and J. Pineau, "An introduction to deep reinforcement learning," 2018, *arXiv:1811.12560*.

[25] M. Kenyeres and J. Kenyeres, "Comparative study of distributed consensus gossip algorithms for network size estimation in multi-agent systems," *Future Internet*, vol. 13, no. 5, p. 134, May 2021.

[26] M. Morales, *Grokking Deep Reinforcement Learning*. Shelter Island, NY, USA: Manning Publications, 2020.

[27] E. Krijestorac, S. Hanna, and D. Cabric, "UAV access point placement for connectivity to a user with unknown location using deep RL," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2019, pp. 1–6.

[28] J. Qiu, J. Lyu, and L. Fu, "Placement optimization of aerial base stations with deep reinforcement learning," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2020, pp. 1–6.

[29] Q. V. Do, Q.-V. Pham, and W.-J. Hwang, "Deep reinforcement learning for energy-efficient federated learning in UAV-enabled wireless powered networks," *IEEE Commun. Lett.*, vol. 26, no. 1, pp. 99–103, Jan. 2021.

[30] U. Challita, W. Saad, and C. Bettstetter, "Interference management for cellular-connected UAVs: A deep reinforcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2125–2140, Apr. 2019.

[31] P. Luong, F. Gagnon, L.-N. Tran, and F. Labeau, "Deep reinforcement learning-based resource allocation in cooperative UAV-assisted wireless networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7610–7625, Nov. 2021.

[32] X. Liu, Y. Liu, Y. Chen, and L. Hanzo, "Trajectory design and power control for multi-UAV assisted wireless networks: A machine learning approach," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 7957–7969, Aug. 2019.

[33] T. Bai, C. Pan, J. Wang, Y. Deng, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Dynamic aerial base station placement for minimum-delay communications," *IEEE Internet Things J.*, vol. 8, no. 3, pp. 1623–1635, Feb. 2020.

[34] M. J. Marcus, "5G and 'IMT for 2020 and beyond' [spectrum policy and regulatory issues]," *IEEE Wireless Commun.*, vol. 22, no. 4, pp. 2–3, Aug. 2015.

DICK CARRILLO MELGAREJO (Member, IEEE) received the B.Eng. degree (Hons.) in electronics and electrical engineering from the National University of San Marcos, Lima, Peru, in 2004, and the M.Sc. degree in electrical engineering from the Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, Brazil, in 2008. He is currently pursuing the Ph.D. degrees in electrical engineering with the Lappeenranta–Lahti University of Technology. From 2008 to 2010, he contributed to WIMAX (IEEE 802.16m) Standardization. From 2010 to 2018, he worked with the design and implementation of cognitive radio networks and projects based on 3GPP technologies. Since 2018, he has been a Researcher at the Lappeenranta–Lahti University of Technology. Since 2022, he has been a Senior Standardization Specialist at Nokia Bell Labs, where he is contributing on shaping the 3GPP release 18 standard (5G-Advanced). His research interests include mobile technologies beyond 5G, energy harvesting, intelligent meta-surfaces, cell-free mMIMO, and RAN slicing.

JIRI POKORNY received the M.Sc. degree in communications and informatics from the Brno University of Technology (BUT), Czech Republic, in 2016. He is currently pursuing the dual Ph.D. degree with the Department of Telecommunications, BUT, and Tampere University, Finland. He has published five journals articles and more than ten papers in international conferences. His main task is the location optimization of flying base stations in crowded scenarios. His past experience involves research on LPWA networks, mainly on the NB-IoT technology. His research interests include wireless communication networks and unmanned aerial vehicles.

PAVEL SEDA received the first M.Sc. degree in communications and informatics from the Brno University of Technology, Brno, and the M.Sc. degree in applied informatics from Masaryk University. From 2014 to 2018, he worked as a Java Developer at IBM. His research interests include cybersecurity, technologies, optimization, and cyber-security education.

ARUN NARAYANAN (Member, IEEE) received the B.E. degree in electrical engineering from the Visvesvaraya National Institute of Technology, Nagpur, India, in 2002, and the M.Sc. and Ph.D. degrees in energy technology from the School of Energy Systems, Lappeenranta University of Technology (LUT), Lappeenranta, Finland, in 2013 and 2019, respectively.

He is currently a Postdoctoral Researcher at the Cyber-Physical Systems Group, LUT. He focuses on applying optimization, computational concepts, and artificial intelligence techniques to renewable electrical energy problems. His research interests include renewable energy-based smart microgrids, electricity distribution and markets, demand-side management, energy management systems, and information and communications technology.

PEDRO H. J. NARDELLI (Senior Member, IEEE) received the B.S. and M.Sc. degrees in electrical engineering from the State University of Campinas, Brazil, in 2006 and 2008, respectively, and the dual Ph.D. degrees from the University of Oulu, Finland, and the State University of Campinas, in 2013. He was an Academy of Finland Research Fellow with a project called Building the Energy Internet as a Large-Scale IoT-Based Cyber-Physical System that manages the energy inventory of distribution grids as discretized packets via machine-type communications (EnergyNet). He leads the Cyber-Physical Systems Group at the Lappeenranta University of Technology (LUT), Lappeenranta, Finland, and a Project Coordinator at the CHIST-ERA European Consortium Framework for the Identification of Rare Events via Machine Learning and IoT Networks (FIREMAN). He is also an Adjunct Professor at the University of Oulu in the topic of "communications strategies and information processing in energy systems." He is currently an Associate Professor (Tenure Track) in IoT in energy systems at LUT. His research interests include wireless communications particularly applied in industrial automation and energy systems. He received the Best Paper Award of the IEEE PES Innovative Smart Grid Technologies Latin America in the track "Big Data and Internet of Things," in 2019.

MEHDI RASTI (Senior Member, IEEE) received the B.Sc. degree from Shiraz University, Shiraz, Iran, in 2001, and the M.Sc. and Ph.D. degrees in electrical engineering from Tarbiat Modares University, Tehran, Iran, in 2003 and 2009, respectively. From November 2007 to November 2008, he was a Visiting Researcher at the Wireless@KTH, Royal Institute of Technology, Stockholm, Sweden. From September 2010 to July 2012, he was with the Shiraz University of Technology, Shiraz, Iran. From June 2013 to August 2013, and from July 2014 to August 2014, he was a Visiting Researcher with the Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, MB, Canada. He is currently an Associate Professor at the Department of Computer Engineering, Amirkabir University of Technology, Tehran. He is also a Visiting Researcher at the Lappeenranta-Lahti University of Technology (LUT), Lappeenranta, Finland. His current research interests include radio resource allocation in the IoT and beyond 5G and 6G wireless networks.

JIRI HOSEK (Senior Member, IEEE) received the M.S. and Ph.D. degrees in electrical engineering from the Faculty of Electrical Engineering and Communication, Brno University of Technology (BUT), Czech Republic, in 2007 and 2011, respectively. He is currently working as an Associate Professor and the Deputy Vice Head for research and development and international relations at the Department of Telecommunications, BUT. He has (co)authored more than 130 research works on networking technologies, wireless communications, quality of service, user experience, and the IoT applications. His research interests include industry-oriented research and development projects in the area of future mobile networks, the Internet of Things, and home automation services.

MILOS SEDA received the M.Sc. and Ph.D. degrees in cybernetics from the Brno University of Technology (BUT) and the M.Sc. degree in mathematical computer science from Masaryk University. From 1976 to 1986, he was a System Designer at První Brněnská Strojírna. He is currently a Professor and a Senior Researcher with the Institute of Automation and Computer Science, BUT. His current research interests include meta-heuristic algorithms, stochastic systems, and graph algorithms. He has (co)authored over 150 research works on those topics.

DEMÓSTENES Z. RODRÍGUEZ (Senior Member, IEEE) received the B.S. degree in electronic engineering from the Pontifical Catholic University of Peru, Peru, and the M.S. and Ph.D. degrees from the University of São Paulo, in 2009 and 2013, respectively. He is currently an Adjunct Professor with the Department of Computer Science, Federal University of Lavras, Brazil. He has a solid knowledge in telecommunication systems and computer science based on 15 years of professional experience in major companies. His research interests include QoS and QoE in multimedia services, architect solutions in telecommunication systems, artificial intelligence algorithms, and online social networks.

YEVGENI KOUCHERYAVY received the Ph.D. degree from the Tampere University of Technology (TUT), Finland. He is currently a Professor at the Unit of Electrical Engineering, Tampere University. He is the author of numerous publications in the field of advanced wired and wireless networking and communications. His current research interests include various aspects in heterogeneous wireless communication networks and systems, the Internet of Things and its standardization, and nanocommunications.

GUSTAVO FRAIDENRAICH was born in Pernambuco, Brazil, in 1977. He graduated in electrical engineering from the Federal University of Pernambuco (UFPE), Brazil. He received the M.Sc. and Ph.D. degrees from the State University of Campinas, (UNICAMP), Brazil, in 2002 and 2006, respectively. From 2006 to 2008, he worked as a Postdoctoral Fellow at the Star Laboratory Group, Stanford University, Stanford, CA, USA. He is currently an Assistant Professor at UNICAMP. He has published more than 50 international journals articles and more than 100 conference papers of the first line. His research interests include multiple antenna systems, cooperative systems, radar systems, and wireless communications. He was a recipient of the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) Young Researcher Scholarship, in 2009. He has been an Associated Editor of the *European Transactions on Telecommunications* (ETT) journal for many years. He is the President of the Technical Board of Venturus Company, a branch from Ericsson Company.

• • •