IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON INNOVATIVE TRENDS IN 6G ECOSYSTEMS

Received 8 January 2024, accepted 31 January 2024, date of publication 8 February 2024, date of current version 16 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3364349

RESEARCH ARTICLE

Energy Efficiency and Latency Optimization for IoT URLLC and mMTC Use Cases

OSAMA ELGARHY[1], LUCA REGGIANI[2], MUHAMMAD MAHTAB ALAM[1], (Senior Member, IEEE), AHMED ZOHA[3], (Senior Member, IEEE), RIZWAN AHMAD[4], AND ALAR KUUSIK[1], (Member, IEEE)

[1]Thomas Johann Seebeck Department of Electronics, Tallinn University of Technology, 19086 Tallinn, Estonia
[2]Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, 20133 Milan, Italy
[3]James Watt School of Engineering, University of Glasgow, G12 8QQ Glasgow, U.K.
[4]School of Electrical Engineering and Computer Science, National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan

Corresponding author: Osama Elgarhy (osama.elgarhy@taltech.ee)

This work was supported in part by the European Union's Horizon 2020 Research and Innovation Program under Grant 668995 and Grant 951867, in part by the European Union Regional Development Fund in the Framework of the Tallinn University of Technology Development Program (2016–2022), and in part by the Estonian Research Council through the Center of Excellence "EXCITE IT" under Grant PRG667 and Grant TAR16013.

ABSTRACT The Internet of Things (IoT) is an essential part of 5G, Beyond-5G (B5G), and 6G systems; it has several applications in two of the principal 5G use cases, namely ultra-reliable low-latency communications (URLLC) and massive machine-type communications (mMTC), and in their successors within B5G and 6G: extreme ultra-reliable low-latency communication (eURLLC) and ultra-massive machine-type communication (umMTC). IoT systems, which are characterized by narrow bandwidths, have stringent requirements owing to the specific nature of their applications and use cases. The purpose of this study is to investigate and jointly optimize the energy efficiency (EE) and latency through resource allocation for IoT cellular systems. With regard to the contributions, in this study we investigated the optimization of EE in narrowband IoT systems, compared resource unit configurations (RUCs), jointly formulated the optimization of EE and latency, and introduced a suboptimal but efficient algorithm. More precisely, as the EE performance of various resource unit configurations has not been exhaustively investigated in the current state of the art, we analyzed and compared the EE of RUCs. The results show vast differences in performance between RUCs. For example, in terms of EE, the best RUC has an EE more than 80 times higher than the worst, which illustrates the importance of this investigation. We then proposed a scheduler based on the shortest job first (SJF) for minimum latency allocation, and another scheduler based on a joint evaluation of EE and latency. With respect to conventional techniques, these schedulers achieve a better trade-off between latency reduction and gain in terms of EE for a wider range of parameter configurations in multi-cellular layouts. The study demonstrates that in the presence of repetitions, algorithms that achieve high EE will mostly achieve low latency.

INDEX TERMS Energy efficiency, latency, resource allocation, uplink scheduling.

I. INTRODUCTION

Currently, the Internet of Things (IoT) is one of the main components of modern wireless communication paradigms. There are three main 5G services: massive machine-type communication (mMTC), ultra-reliable low-latency communications (URLLC), and enhanced mobile broadband (eMBB). Two of these services are related to IoT: mMTC, which is for IoT devices and high-density connections, and URLLC, which serves industrial and mission-critical IoT. These services will not only remain in Beyond-5G (B5G) and 6G, but will also evolve into ultra-massive machine-type communication (umMTC) and extreme ultra-reliable low-latency communication

The associate editor coordinating the review of this manuscript and approving it for publication was Pavlos I. Lazaridis.

(eURLLC). In addition to the requirements of these services, IoT devices often impose additional power consumption requirements. Therefore, the resource allocation should be designed to target multiple performance indicators. This is in line with the trend for B5G and 6G, because these systems are expected to have a mix of the three services rather than only one service at a time [1]; for example, merging the requirements of reliable low latency communications and machine-type communication in a unique service. In this study, we aim to achieve energy efficiency (EE) and latency optimization owing to the importance of energy consumption in IoT devices and the ever-increasing importance of energy savings and latency reduction in 5G, B5G, and 6G applications. In several IoT scenarios, power is limited, and the main performance targets are a low data rate, long range, and narrow bandwidth. Thus, the narrowband IoT (NB-IoT) is a suitable technology for these scenarios. Since its introduction in Release 13 [2], NB-IoT has been present in all successive releases of the 3GPP standard specifications, and it has been recognized as one of the main building blocks for IoT in 5G IoT low-power wide area (LPWA) communication.

In order to summarize the main requirements of mMTC and URLLC, we can list the following KPI (Key Performance Indicators) [3], which are also the expression of the wide range of verticals in the 5G and B5G application scenarios:

- mMTC: Here, the connectivity is extended to extremely high densities, that is, 10^6 devices per km^2 in an urban environment, a target for UE battery life beyond 10 years, assuming a stored energy capacity of 5 Wh.
- URLLC: Here the focus is on the reliability, and the general reliability requirement for one transmission of a packet is $1-10^{-5}$ for 32 bytes with a user plane latency of 1 ms and end-to-end latency of less than 10 ms. For eV2X (enhanced Vehicle to Everything), reliability is expected to be $1-10^{-5}$, and user plane latency $3-10$ ms when the packet is relayed via BS.

Looking at the KPIs, it is immediately clear that it is crucial to understand and develop all the techniques that enable a wide range of requirements and combinations of reliability, latency, and EE. Moreover, when we look at the URLLC and mMTC requirements within an IoT device, the typical requirements and use case characteristics are

- Sporadic or periodic traffic, with data rates that can achieve maximum levels of approximately $1-2$ Mbps.
- Extremely variable reliability levels: from 99% to 99.9999% according to the application.
- The importance of battery life: consequently, EE remains a specific design and operational requirement.
- The user plane latency is in the range $1-5$ ms for applications with more stringent requirements.

Among the examples of IoT use cases where EE and low latency play a relevant design role, we can mention (i) smart city and environment monitoring systems for public safety and (ii) systems for Industrial IoT (IIoT), such as critical monitoring systems, factory automation, and location/motion control. For example, wireless sensors in an industrial context—specifically safety-related sensors—have stringent requirements in terms of battery life and latency, that is, a battery life measured in years and 5–10 ms latency [4]. Additionally, as reported in [5],although mMTC is characterized by a long battery life, there are applications such as IIoT orchestration or automation that are also characterized by latency requirements in the range of 10-50 ms. The main motivation of this study starts from the need of optimizing EE, a major objective in current and future networks, in which the goal is to reduce the carbon footprint and work towards green communication. On the other hand, reducing latency to the order of milliseconds is critical for several applications, such as safety-related ones, augmented reality, and vehicular ones. Moreover, as previously mentioned [1], in the future, there will be a need to achieve both reduced latency and EE in several applications and use cases, such as, ''factories of the future''-related applications [6]. Therefore, we found a need to investigate the potential trade-off between EE and latency within a design that integrates and considers their joint optimization. In this study, the goals of the mathematical analysis and proposed algorithms are to maximize EE and reduce latency for IoT devices, with an emphasis on NB-IoT systems using resource allocation. Several possible resource unit configurations can be allocated to devices, and each of these resource unit configurations has distinct characteristics. The EE performance of these resource unit configurations should be investigated and understood exhaustively. Moreover, often in the literature only one resource unit configuration has been chosen in the allocation approaches, and this work starts with an analytical study to investigate the EE performance of all possible resource unit configurations. We then propose an optimization problem and practical algorithms that optimize EE and latency.

The remainder of this paper is organized as follows. Related works and novel contributions are presented in Sect. II. Sect. III briefly introduces EE concepts and various aspects of the problem. Sect. IV presents the optimization problem and mathematical formulation. Sect. V presents the proposed suboptimal solutions and the related algorithms. Sect. VI presents numerical results. Finally, conclusions are in Sect. VII.

II. RELATED WORK AND NOVEL CONTRIBUTIONS

As already mentioned, one of the main design aspects of IoT systems is EE, which is at the center of this research article. Many research works in the literature either target or claim to target this aspect. A relevant portion of the literature deals with energy consumption rather than EE. In [7], [8], [9], and [10], energy consumption was calculated for a range of scenarios. Based on [11] and [12], the average energy consumption and average delay were calculated using a queuing model to consider the average values of the synchronization, resource reservation, and transmission processes. Parameters such as the average traffic, average transmission

time, and average rate were also considered. However, energy consumption was investigated without considering the impact of various resource unit configurations (RUCs) and resource allocation on the transmission process. In [13], a semi-Markov model of radio resource control (RRC) connected and idle states (particularly the discontinuous reception and the power saving mode) was proposed to finalize energy savings. Similarly, the authors of [14] analyzed the NB-IoT downlink power consumption model and used the particle swarm algorithm for energy consumption optimization. In [15], the focus was on the resource allocation process and energy consumption for a range of RUCs, and the minimum transmitted power was derived for all possible combinations of repetitions, modulation and coding schemes, and resource units (RUs). In [16], a survey of techniques that target energy consumption models in two specific examples (agriculture and health) was presented. A different approach was presented in [17], where channel coding was used to improve energy consumption and rate. For a D2D system, the authors of [18] used reinforcement learning aided by federated learning to minimize the power consumption and maximize the sum rate through resource allocation. The adopted system model comprises one cell, no repetitions, and no latency optimization. For an NB-IoT single-cell system, the authors of [19] propose a scheduling algorithm for search space periods, not for downlink or uplink resource allocation, which optimizes energy consumption by reducing idle time and blind decoding. Finally, in [20], the effects of semi-persistent scheduling and transmission time length on the latency of narrowband devices were investigated.

In the context of EE, several studies have focused on optimizing the resource allocation in B5G and 6G systems. The authors of [21] proposed a resource allocation algorithm for a system in which delta-orthogonal multiple access was used. The algorithm comprises two phases: Base station selection and power allocation. A multi-cellular scenario with imperfect successive interference cancellations was assumed. The latency and repetitions were not considered. In [22], the authors considered a system comprising cognitive radio and nonorthogonal multiple access. They also proposed a resource allocation optimization problem to maximize EE. This problem was solved by using the Dinkelbach algorithm. A single cell was considered without repetitions or latency optimization. Resource allocation for reconfigurable intelligent surfaces (RIS) was the focus of [23]. EE maximization is achieved through the allocation of power and the RIS phase shift. The system was a single cell, and repetitions were not considered. Finally, [24] provides a survey of resource allocation and power control techniques specifically for EE in cell-free ultra-dense systems that exploit massive MIMO.

Regarding EE in the NB-IoT context, the following studies provided varying views on this problem. The authors of [25] researched EE optimization for NB-IoT. The main idea is to cluster users and split them into three types: (1) cluster heads, (2) users belonging to clusters (but not cluster heads), and (3) nonclustered users. The authors assumed a single-tone resource configuration, no repetitions, and a single-cell scenario. The resource allocation problem is based on allowing nonclustered and clustered users to use the same resources. In [26], the EE was calculated in the presence of interference between narrowband physical uplink shared and random access channels (NPUSCH and NPRACH). In [27], an EE problem formulation for the simultaneous transmission of cellular 5G devices and IoT devices without considering sub-carriers (SCs) and repetitions was introduced. In [28], an analytical M/D/H/K queue model was formulated to investigate the EE and improve battery life. Although the concept of the four RUs is explained, the effect of RU types on EE has not been studied, and latency has not been considered. The authors of [29] proposed an NB-IoT uplink scheduler that performs resource allocation and link adaptation to optimize the EE. Repetitions, delay constraints, and various RUCs were considered. However, the impact of each RUC on EE was not shown. Furthermore, the optimization problem was formulated only for EE in single-cell systems.

The main objective of this study is to investigate and determine the characteristics of good schedulers for joint EE and latency optimization in systems with repetitions and, where present, RUCs. As seen in the NB-IoT literature review, various RUCs are not generally considered for EE optimization, and only a single-tone RUC is used, although we show that this is the one with the lowest EE. In addition, other typical characteristics of IoT,such as the impact of multi-cellular scenario and repetitions are not considered, even though repetitions play a significant role in increasing coverage and reducing the block error rate [30]. Because these elements are all related to the latency and EE of the system, the joint latency and EE optimization of the system performance are explored here based on all these elements.

Therefore, in this study, we examine EE and its relationship to latency (they are not independent). We formulate a joint optimization problem for maximizing the EE and minimizing the latency and propose and investigate suboptimal solutions for the problem. We begin by formulating the EE optimization problem for a range of RUCs, find the best RUC analytically, and update the optimization problem accordingly. Finally, we use the best strategy to jointly optimize latency and EE. The novel contributions of this study are summarized as follows:

- We propose and formulate a joint EE and latency optimization problem that includes a range of RUCs and minimum rate constraints. The impact of repetitions is also included.
- We compare and analyze the performance of RUCs in terms of EE to determine which (if any) is the best. A comparison is performed mathematically, and the simulation results are provided. The problem formulation is then simplified and reformulated for the best RUC with the smallest number of subframes and the best EE.
- We propose suboptimal solutions for the formulated joint optimization problem:

TABLE 1. Summary of notations.

Notation	Definition
C	The set of the cells in the system, that is, the Base Stations (BS)
c	Index of a cell belonging to C
U_c	The set of devices within the cell c
u	Index of a device belonging to a generic U_c
T	The set of available subframes (SFs)
t	Index of a given SF belonging to T
S	The set of the available sub-carriers (SCs)
s	Index of a given SC belonging to S
SR	The number of available SCs in a given RUC
G	The set of available RUCs
g	Index of the generic RUC
$x^{(u)}_{c,t,s}$	Binary allocation variable (0/1) for the device u in cell c at slot t and SC s
$x^{(u)}_{c,g}$	Binary allocation variable for device u in cell c and RU g
T_S	Duration of the OFDMA symbol time
N_S	Number of OFDMA symbols in each SF
N_U	Total number of devices
$R^{(u)}_{c,t,s}$	The rate of device u
$r^{(u)}_c$	Number of repetitions of the same message for each user
$P^{(u)}_{c,t,s}$	The transmission power of device $u \in c$ at t and s
P_T	Total transmission power per device
P_C	The circuit power
$CH^{(u)}_{c,c,t,s}$	The channel gain of device u with respect to to its BS c
$CH^{(u)}_{a,c,t,s}$	The channel gain of device u, affiliated to BS c, with respect to to another BS $a \in C$

  - The shortest job first (SJF) strategy: SJF is exploited here as an allocation and scheduling strategy for minimizing the latency in a system with repetitions.
  - Proposed joint EE and latency scheduler: The scheduler is denoted as the "*best score*," which is a hybrid scheduler that allocates resource units to users according to latency and EE.
  - A generalized multi-dimensional form of the "*best score*" scheduler.

Finally, the proposed solutions were compared with other benchmarks, and their performances were assessed through simulations of the NB-IoT application case.

III. ENERGY EFFICIENCY

In current and future wireless communication systems, the two fundamental objectives are a high data rate and EE. As decreasing energy consumption by sacrificing the data transmission rate or vice versa is not a satisfactory solution, there has always been a need to optimize both the energy consumption and data rate. One of the main goals in the IoT context is the optimization of EE; that is, the number of bits transmitted with a single Joule (bit/J) [31]. In a communication system, EE is the ratio between the data transmission rate and consumed power. Therefore, to derive EE, we must express the rate, possibly through the Shannon capacity bound. Power is typically composed of two parts: transmitted power (P) and consumed circuit power (P_C). There are other variants of the definition of EE, for example, taking into account various characteristics of the communication system, such as the use of P_C values depending on the transmitted power and rate [32], but these variations are usually suited to specific systems, scenarios and optimization procedures.

Here, we focus on OFDMA waveforms, which are the basis of 4G and 5G mobile systems, and adopt the conventional definition of EE as the ratio between the rate and consumed power [33], [34].

It is challenging to optimize the EE of a mobile system, particularly in the case of an interference scenario, where the numerator of the EE represents the well-known non-convex sum rate of the interference cells. The EE comprises rate, circuit power, and transmission power, each of which affects it differently. In addition, these elements can have different impacts depending on the scenario, channel conditions, and the interference state. In general, we observed the following.

- Regarding the relationship between the rate and signal-to-interference-and-noise ratio (SINR).
  - For one user, as the SINR increases, the rate clearly increases, i.e., maximizing the SINR is equivalent to maximizing the rate.
  - For multiple users, maximizing the sum of the SINRs is not equivalent to maximizing the sum of the rates, because of the log relation in the rate equation. This means that in the case of interfering users, the sum rate is not optimized by maximizing the sum SINR.
- Regarding the relation between rate and transmission power.
  - In the absence of interference, increasing the power leads to an increase in the SINR and, consequently, the rate.
  - In the case of interference, when rate optimization is performed for a single user in a cell independently from the other external cells and interfering users, increasing the power will lead to an increase in the rate for this main/target user, but may have a negative impact on the others.
  - In the case of interference and rate maximization for the entire system, i.e., jointly for all users, we achieve the well-known sum throughput maximization problem, which is a non-convex problem. The optimal rate can be achieved at a very high or very low transmission power, according to the scenario and power setup for the users (e.g., if the power control setup and power control parameters are the same for all users in each cell, or if the system is interference-limited or noise-limited). In some scenarios, the maximum sum rate is achieved at the maximum transmit power or very close to it, in others at low transmit power, and in yet others at power values adjusted to compensate for path loss [35].

In the literature, many aspects of the relationship between EE, rate, transmission power, and circuit power have been investigated, and a summary is provided below.

- EE and circuit power: P_C is an important factor for EE behavior—more than what can be expected from a superficial analysis, as can be seen in [31], [33], [34], [36], and [37]. Moreover, it has been proven that adding or removing P_C from the EE equation affects the way the

TABLE 2. Summary of the main relations between EE, SE, and P_T while considering the presence of P_C and interference.

	Case	Relation
EE - P_T	no interference - no P_C	EE decreases as P_T increases
	no interference - P_C	EE has a unique maximum
	interference	non-convex
SE - P_T	no interference	SE increases as P_T increases
	fixed interference	SE of the target user increases as its P_T increases
	interference	non-convex
EE - SE	no interference - no P_C	EE decreases as SE increases
	no interference - P_C	EE has a unique maximum
	interference	non-convex

EE changes with respect to the rate and power, at least in the abs of interference [37].

- EE and transmission power (P_T):
  - In the case of no interference: Either EE decreases as P_T increases when we remove P_C from the EE equation, or EE has a unique maximum when P_C is included. In general, in the case of no interference, a minimum P_T does not mean a maximum EE, unless P_C is neglected.
  - In the case of interference: As P_T affects both the rate—which is a non-convex function of P_T—and EE, the relationship between P_T and EE becomes more complicated, and in general, we cannot say that the minimum P_T corresponds to the maximum EE.
- EE and rate (or spectral efficiency): The relationship between the EE and Spectral Efficiency (SE) can be obtained by writing the transmit power and EE as a function of the spectral efficiency.
  - In the case of no interference: Either EE decreases as spectral efficiency increases if we remove P_C from the EE equation, or EE has a unique maximum when P_C is included [37], [38].
  - In the case of interference: Because of the nonconvexity of the sum rate, the EE is non-convex, and maximizing the rate does not mean maximizing the EE, as can be seen in the examples in [39].

A summary of the main relations is presented in Table 2.

In general, in the case of interference scenarios, the EE optimization problem is difficult to solve and analyze, and one of the main reasons is the presence of the sum rate at the EE numerator. Adding binary scheduling variables further complicates the problem, making it combinatorial [34], whereas adding OFDM allocation constraints makes it NP-hard [40]. Another aspect is the allocation constraints, constituted by the so-called RUCs, which are the typical predefined sets of SCs and SFs allowed for device resource allocation. These RUCs are defined in the specifications and designed to provide a limited but important degree of flexibility in the occupation of the frequency/time-domain resource grid.

Several methods can be used to represent the EE of the entire system. The representation also depends on the final objective and constraints, i.e., the optimization of the global performance of the system or fairness for types/groups of users. We show two approaches to EE formulation [41]: the first is used to calculate the EE for the entire system, not distinguishing between groups of users with differing EE requirements, and it aims to maximize the overall performance of the system. The simplified EE formula has the following structure:

$$EE = \frac{\sum_u R^{(u)}}{\sum_u (P_C + P^{(u)})}, \tag{1}$$

where u is the index of the generic users, $R^{(u)}$ is the rate of user u, and $P^{(u)}$ is the transmission power. The second approach is used to distinguish between EE for each group of users and/or separate EE priorities and requirements while considering the total system performance, and is the sum (or average) of the individual EEs. Therefore, the EE can be expressed as follows:

$$EE = \sum_u \frac{R^{(u)}}{P_C + P^{(u)}}. \tag{2}$$

It is important to highlight that the work presented in this paper is focused on the uplink of OFDM mobile systems because we are interested in optimizing the EE of the IoT devices.

IV. MODEL AND PROBLEM FORMULATION

The system is characterized by a set C of seven hexagonal cells served by the respective Base Stations (BSs, such as eNBs or gNBs) and the network layout is shown in Fig. 1. The single cell is split into three 120 degrees sectors, and each sector comprises a set of devices U_c, transmitting in the same band, which constitutes the most common sectorization used in the design of mobile systems, as it assures an advantageous trade-off between frequency reuse and interference reduction. In the numerical results, a cell with a single 360 degrees sector is considered to simulate a condition with high interference. The cell size is defined by the cell radius R, which is also related to the inter-site distance (ISD, the minimum distance between two BSs) by the geometric relation $R = 0.577 \cdot ISD$. The resources are composed of SCs in the S domain and SFs in T domain. The S domain or band is shared by all seven sectors, generating mutual interference: the performance is always measured in the center cell, which is in the worst interfering condition, and all the parameters are equal in all the cells. Table 1 lists all variables and notations.

The optimization problem contains the joint optimization of the EE and latency and is expressed by the maximization of the weighted difference between the EE function $f_E(.)$ and latency function $f_L(.)$ [31], [42], that is,

$$f_{OPT}(x^{(u)}_{c,t,s}) = w_1 \cdot f_E(x^{(u)}_{c,t,s}) - w_2 \cdot f_L(x^{(u)}_{c,t,s}) \tag{3}$$

where w_1 and w_2 are two positive weights that also include different scaling owing to the unit of measurements, and the binary (0/1) allocation variable $x^{(u)}_{c,t,s}$ determines the assignment of each Resource Element (RE) (t, s) (t is the SF within the scheduling interval T and s the SC in the available

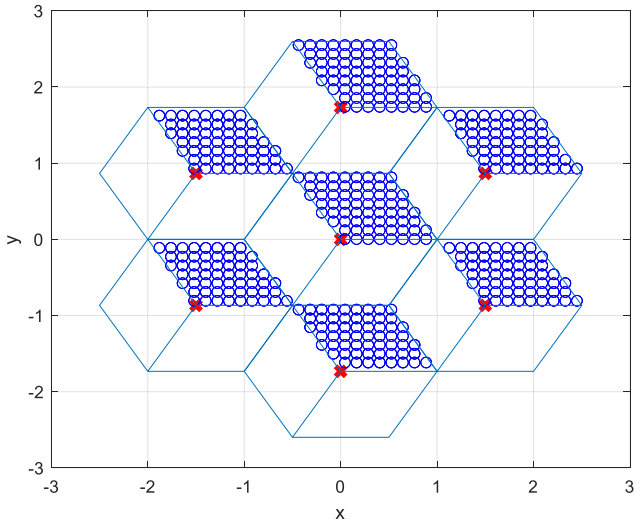

FIGURE 1. Layout of the cells, each one comprising three sectors. The red *X* markers represent the base stations, while the blue circles represent the active devices in the simulations.

set S) in each cell c to device u and it is the variable on which we will maximize the EE optimization function $f_E(x^{(u)}_{c,t,s})$.

Let us consider the expression of the overall system *EE* [bit/J] as the first step in building the optimization function in (3).

$$EE = f_E(x^{(u)}_{c,t,s}) = \frac{\sum_{c\in C}\sum_{u\in U_c}\sum_{t\in T}\sum_{s\in S} x^{(u)}_{c,t,s} R^{(u)}_{c,t,s}}{\sum_{c\in C}\sum_{u\in U_c}(P_C + \sum_{t\in T}\sum_{s\in S} x^{(u)}_{c,t,s} P_{c,t,s})^{(u)}}. \quad (4)$$

The total rate assigned to each device in the numerator is subject to a constraint on the minimum rate $R^{(u)}_{min,c}$, which is usually assumed in this type of problem:

$$\sum_{t\in T}\sum_{s\in S} x^{(u)}_{c,t,s} R^{(u)}_{c,t,s} \geq R^{(u)}_{min,c}, \qquad \forall u \in U_c,\ \forall c \in C. \quad (5)$$

In the problem formulation, the rate of device u in cell c at SF t and SC s is computed by Shannon's bound in [bit/s] and is reduced by the number of repetitions $r^{(u)}_c$ for each user (the total transmission time of the bits is expanded from T_S to $r^{(u)}_c \cdot T_S$), as

$$R^{(u)}_{c,t,s} = \frac{1}{r^{(u)}_c \cdot T_S} \cdot log_2 \times \left(1 + \frac{P^{(u)}_{c,t,s} CH^{(u)}_{c,c,t,s} T_S}{\sum_{a\in C, a\neq c}\sum_{j\in U_a} P^{(j)}_{a,t,s} CH^{(j)}_{a,c,t,s} x^{(j)}_{a,t,s} T_S + kT_E}\right), \quad (6)$$

where the terms $P^{(u)}_{c,t,s}$ and $P^{(j)}_{a,t,s}$ denote the transmitted power of the considered and interfering devices, respectively (for each cell a excluding c because it is not admitted intracell interference), and T_E is the BS receiver total equivalent noise temperature. In addition, $CH^{(u)}_{c,c,t,s}$ is the channel gain between device u and its BS in cell c at the given RE (t,s) and $CH^{(j)}_{a,c,t,s}$ is the channel gain between device j belonging to cell a and BS c, which configures the role of the channel in the uplink SINR. The repetitions have a clear impact on the rate, reducing it, because the same message is assumed to be repeated to increase its reliability at the receiver. However, they contribute in proportion to the increase in energy consumption.

An alternative formulation of the EE function is expressed by the average EE per device (denoted as EE_S) instead of the global EE, which differs in the expansion of the sums at the denominator and numerator of (4) in the single EEs [bit/J] of each device, that is,

$$EE_S = f_{ES}(x^{(u)}_{c,t,s}) = \frac{1}{N_U}\sum_{c\in C}\sum_{u\in U_c} \frac{\sum_{t\in T}\sum_{s\in S} x^{(u)}_{c,t,s} R^{(u)}_{c,t,s}}{P_C + \sum_{t\in T}\sum_{s\in S} x^{(u)}_{c,t,s} P^{(u)}_{c,t,s}}, \quad (7)$$

where N_U is the total number of devices.

In optimization function (5) and (6), it is clear that the problem has two fundamental parts: scheduling (allocation of resources in the grid, or $x^{(u)}_{c,t,s}$) and power allocation. Considering the powers $P^{(u)}_{c,t,s}$ and $P^{(u)}_{a,t,s}$, we make an important assumption of uniform power allocation for the subsequent steps of the formulation. Therefore, the transmitted power of each device user is divided equally among the SCs of the allocated RU because we are interested in optimizing EE through RUC allocation. The impact of this perspective is clarified by the suboptimal solution proposed in Section V.

In addition, the EE function is associated with the following constraints. non-overlapping devices in the same cell, that is, no intracell interference, or

$$\sum_{u\in U_c} x^{(u)}_{c,t,s} \leq 1, \qquad \forall s \in S,\ \forall t \in T,\ \forall c \in C, \quad (8)$$

This guarantees that resource elements are used by only one device within the same cell.

A. THE NB-IoT CASE: SPECIFIC CONSTRAINTS

In the following section, we present NB-IoT-specific constraints and analyze the optimal RUC. As the resources consist of SCs and SFs organized in RUs, in NB-IoT and technologies with predefined RUCs, we introduce the following additional constraints:

1) The allocation and shape of the RUs force the scheduler to choose one of the predefined RUCs or none [43]. In the NB-IoT case, there are four RUCs, as listed in Table 3. However, these relations can be easily extended to any other case in which the RUCs have a rectangular shape. These constraints can be written according to the variables of our formulation and are reported in the appendix. These are not reported here because they become redundant after the choice of a single RUC, as explained in section IV-B.
2) Number of repetitions, from minimum to maximum, according to the specifications of the considered technology. In the case of the NB-IoT, the number of

TABLE 3. Main RUC parameters (in the time and frequency domains) of the LTE NB-IoT.

	RUC1	RUC2	RUC3	RUC4
Total size	8	12	12	12
$z_c^{(u)}$ [SFs]	8	4	2	1
$q_c^{(u)}$ [SCs]	1	3	6	12

repetitions $r_c^{(u)}$ should be a power of 2 to a maximum of 128. The relations

$$r_c^{(u)} = 1\, h_0^{(u)} + 2\, h_1^{(u)} + 4\, h_2^{(u)} + 8\, h_3^{(u)} + \cdots, \quad (9)$$

and

$$\sum_{q=0}^{7} h_q^{(u)} = 1, \quad (10)$$

use the binary (0, 1) allocation variables $h_q^{(u)}$ for $q = \{0, 1, \cdots, 7\}$ for indicating 2^q the allocated number of repetitions of the same message at each device.

To express the latency function $f_L(.)$ in (3), we extend the formulation in [43] by including the impact of the repetitions; in fact the different components of the latency are the (1) scheduling waiting time, (2) transmission time of the assigned RUC, (3) time due to the repetitions of the transmission for increasing the SINR at the receiver. The transmission of the SFs allocated to a device, including possible repetitions, occurs consecutively and it can be expressed as

$$f_L(x_{c,t,s}^{(u)}) = -\sum_{c\in C}\sum_{u\in U_c}\sum_{t\in T} t \cdot \frac{\sum_{s\in S}(x_{c,t,s}^{(u)} - x_{c,t-1,s}^{(u)} - |x_{c,t,s}^{(u)} - x_{c,t-1,s}^{(u)}|)}{2N_{c,t-1}^{(u)} + \epsilon}, \quad (11)$$

where $N_{c,t-1}^{(u)}$ is the number of SCs allocated to user (u) at time $(t-1)$ and cell c (see also the appendix). In (11) we have the sum of all the times in which each device transmission is completed (ϵ is a constant arbitrarily small to avoid the denominator in (11) being equal to zero). In the equation, the latency is controlled by the allocation variable x for user u and cell, SF and SC, (c, t, s) respectively: the last allocated time slot gives the only term different from zero in the sum and determines the scheduling and transmission time, including repetitions. If we look at a more general model, like that in [44], (11) expresses, as a function of the allocation variable, a latency model equivalent to

$$L = t_{SQ} + t_T \quad (12)$$

where L is the total latency, $t_S Q$ is the scheduling or queuing latency, and t_T is the transmission latency, including repetitions.

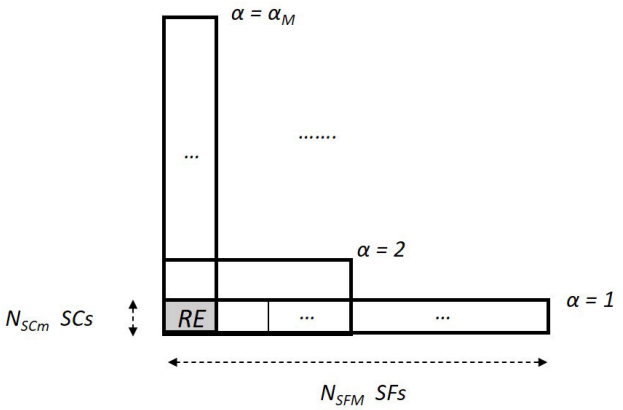

FIGURE 2. The generic set of RUCs as a function of the parameter α.

B. COMPARING THE RUCs IN TERMS OF EE

We have seen that, in relevant cases such as the NB-IoT, the RUCs have an adaptive rectangular shape. As shown in Table 3, the different RUCs respond to different layouts in the SC/SF grid, keeping the total number of occupied resource elements (1 SC × 1 SF), which are 12 in the LTE NB-IoT case (except for RUC1 with eight). In this section, we investigate whether there are differences in EE contributions from different RUCs. Furthermore, the analysis helps identify which RUC contributes better to the optimization of cost function (3), which further allows the simplification of the scheduling process.

To compare the different RUCs, we generalize the research by introducing a parameter α for generating a generic set G or RUCs characterized by the same number of RE, equal to $N_{SCm} \times N_{SFM}$, as shown in Fig. 2, where N_{SCm} is the minimum number of SCs (1 in Table 3), N_{SFM} is the maximum number of SFs, and generic RUCα is characterized by $N_{SCm} \times \alpha$ SCs and N_{SFM}/α SFs for $\alpha = \{1, 2, \cdots, \alpha_M\}$. We also make the following assumptions:

1) For each device, the channel gain is the same in all REs that compose each RUCα, including small-scale fading, which means that the coherence bandwidth and time of the channel are larger than the maximum bandwidth and maximum duration of the RUCs, respectively. This assumption was made to simplify the analysis; however, it did not affect the final results.
2) Power is allocated uniformly in the RUC; therefore, the total transmission power per device P_T is divided among the $N_{SCm} \times \alpha$ SCs. This assumption is consistent with what is already specified in the problem formulation, as we are interested in the optimization with respect to the allocation strategy through the variables $x_{c,t,s}^{(u)}$.
3) All devices in the network used the same α. This assumption follows the idea that in the presence of an α that is optimal with respect to EE, all networks should share this choice to improve the overall result.

The EE for a single device transmitting on α can be written as

$$EE(\alpha) = \frac{N_{SCm} \cdot \alpha \cdot log_2(1 + SINR(\alpha))}{(P_C + N_{SCm} \cdot \alpha \cdot P_{SC}(\alpha))T_S} \tag{13}$$

where $P_{SC}(\alpha)$ is the power allocated to a single SC, that is $P_{SC}(\alpha) = P_T/(N_{SCm} \cdot \alpha)$. The SINR, which is responsible for the rate bound, as shown in (6), is expressed as a function of α as follows.

$$\begin{aligned} SINR(\alpha) &= \frac{P_{SC}(\alpha)T_S CH}{kT_E + \sum_j T_S P_{SC}(\alpha) CH_I^{(j)}} \\ &= \frac{P_T T_S CH/(N_{SCm} \cdot \alpha)}{kT_E + T_S P_T \overline{CH}_I/(N_{SCm} \cdot \alpha)}, \end{aligned} \tag{14}$$

With respect to (6), we have omitted indices u and c for the sake of clarity: CH is the signal power channel gain, while $CH_I^{(j)}$ is the interference power channel gain, resulting in a total term $\overline{CH}_I$. Considering the derivative,

$$\frac{d\ EE(\alpha)}{d\ \alpha} = \frac{d}{d\ \alpha}\left[\frac{N_{SCm} \cdot \alpha \cdot log_2(1 + SINR(\alpha))}{(P_C + N_{SCm} \cdot \alpha \cdot P_{SC})T_S}\right], \tag{15}$$

denoting $N_{SCm} \cdot \alpha \cdot P_{SC}$ as P_T and $(P_C + N_{SCm} \cdot \alpha \cdot P_{SC})T_S$ as c and by including $d\ SINR(\alpha)/d\ \alpha$, we get

$$\begin{aligned} \frac{d\ EE(\alpha)}{d\ \alpha} =& \frac{1}{c}\Bigg(N_{SCm} \cdot \alpha \cdot log_2(1 + SINR(\alpha)) \\ &+ -\frac{CH \cdot T_S \cdot P_T \cdot KT_E \cdot N_{SCm}^2 \cdot \alpha}{D(\alpha)} \\ &\cdot \frac{1}{D(\alpha) + CH \cdot T_S \cdot P_T}\Bigg) \end{aligned} \tag{16}$$

where $D(\alpha) \approx P_T \cdot T_S \bar{CH}_T$ in the interference-limited scenario and $D(\alpha) \approx KT_E \cdot N_{SCm}\alpha$ in the noise-limited scenario. Therefore, for an interference-limited scenario, we have:

$$SINR(\alpha) \approx \frac{CH}{\overline{CH}_I}. \tag{17}$$

$$\begin{aligned} \frac{d\ EE(\alpha)}{d\ \alpha} \approx& \frac{N_{SCm}}{c}\Bigg(log_2(1 + \frac{CH}{CH_1}) \\ &- \alpha \frac{CHKT_E N_{SCm}}{\bar{CH}_T(\bar{CH}_T + CH)}\Bigg) \end{aligned} \tag{18}$$

and, for a noise-limited scenario,

$$SINR(\alpha) \approx \frac{P_T T_S CH}{kT_E N_{SCm} \cdot \alpha}, \tag{19}$$

$$\frac{d\ EE(\alpha)}{d\ \alpha} \approx \frac{N_{SCm}}{c}\left(log_2(1 + SINR) - \frac{1}{1 + (1/SINR)}\right). \tag{20}$$

Therefore, we can make the following remarks:

- $SINR(\alpha)$ decreases linearly with α in the noise-limited region; that is, the associated rate decreases logarithmically with α. In interference-limited regions, $SINR(\alpha)$ and the rate do not depend on α.

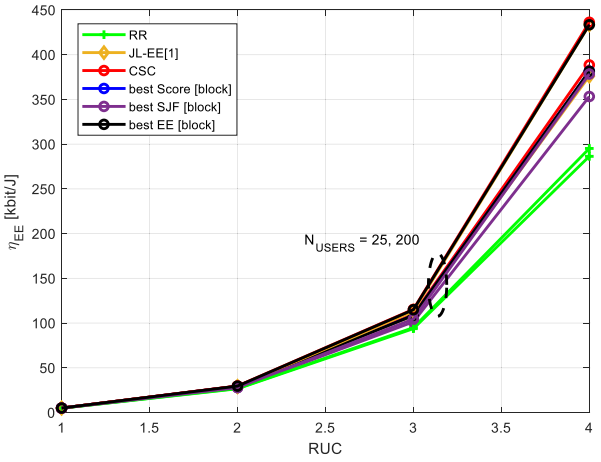

FIGURE 3. Energy efficiency comparison between the 4 different RUCs defined for NB-IoT using the different schedulers.

- The energy consumed by one RUCα transmission is equal to

$$\frac{P_T T_S N_{SCm} \cdot \alpha \cdot N_{SFM}/\alpha}{N_{SCm} \cdot \alpha} = \frac{P_T T_S N_{SFM}}{\alpha} \tag{21}$$

  and it decreases linearly with α, that is, a good reason for increasing α.
- From (18), we observe that $\alpha \frac{CHKT_E N_{SCm}}{\bar{CH}_T(\bar{CH}_T + CH)} \ll P_T T_S \ll 1$ and the derivative is positive, indicating that EE increases with α.
- From (20), we observe that, for $SINR > 1$, the derivative is always positive, indicating that EE increases with α.

Therefore, we observe that increasing α produces an advantage from the EE perspective, because the potential rate decreases only logarithmically and only in the noise-limited case, whereas the consumed energy decreases linearly.

We also simulated seven cells of the layout for different numbers of devices per cell to compare the performance of the four specific RUCs of the NB-IoT, as shown in Fig. 3 (see also Sect. VI). It is clear that the EE of RUC4 is higher than those of the other RUCs. For example, RUC4 EE is 3.7, 14 and 86 times higher than that of RUC3, RUC2, and RUC1, respectively. Moreover, the worst scheduling strategy of RUC4, that is, the round robin (RR), is better than any strategy of the other RUCs.

1) USE OF A SINGLE RUC

We have seen that, in the presence of uniform power and RUC allocation, the RUC with the highest α (RUCα_M, 4 in the LTE NB-IoT case) is the best from the point of view of a single EE. Moreover, in the case of uniform power allocation, we observe that there is no difference between the system EE in (4) and the average of the single EEs in (7).

We also note that using the same RUC can reduce the flexibility of allocation in terms of space occupation. However, if the grid of available SCs $\times$ SFs can be partitioned into a set of RUCα_M without loss owing to the particular RUC shape, this choice does not affect the function optimization.

Under these conditions, we can reformulate the optimization problem by removing the constraints concerning the RUCs and, instead of indexes s and t for the SCs and SFs, using only the index g to indicate the selected RU among the set G of available ones; in fact the entire grid SCs $\times$ SFs is partitioned into the set G of RUC without shape loss. Therefore, we can rewrite the EE as

$$f_E(x_{c,g}^{(u)}) = \frac{\sum_{c\in C}\sum_{u\in U_c}\sum_{g\in G} x_{c,g}^{(u)} R_{c,g}^{(u)}}{\sum_{c\in C}\sum_{u\in U_c}\sum_{g\in G} x_{c,g}^{(u)}(P_C + P_T)}, \tag{22}$$

subject to

$$\sum_{g\in G} x_{c,g}^{(u)} R_{c,g}^{(u)} \geq R_{min,c}^{(u)}, \ \forall u \in U_c, c \in C,$$

and the latency function as

$$f_L(x_{c,g}^{(u)}) = -\frac{N_{SFM}}{\alpha_M} \cdot \sum_{c\in C}\sum_{g\in G}\sum_{u\in U_c} g \cdot \frac{(x_{c,g}^{(u)} - x_{c,g-1}^{(u)} - |x_{c,g}^{(u)} - x_{c,g-1}^{(u)}|)}{2}, \tag{23}$$

where we assume that for $\alpha = \alpha_M$, the corresponding RUC occupies all the available SCs, as in the specific case of the LTE NB-IoT, which means that the index g corresponds to a new equivalent time index or $t = g \cdot N_{SFM}/\alpha_M$.

Then the rate is given by

$$R_{c,g}^{(u)} = \frac{1}{r_c^{(u)} T_S} \cdot log_2\left(1 + \frac{P_{c,g}^{(u)} T_S CH_{c,c,g}^{(u)}}{E_I + kT_E N_{SCm}\alpha_M}\right), \tag{24}$$

with the interference term

$$E_I = \sum_{a\neq c, a\in C}\sum_{j\in U_a} P_T T_S CH_{a,c,g}^{(j)} x_{a,g}^{(j)}$$

and the constraint on non-overlapping devices in the same cell can be written as

$$\sum_{u\in U_c} x_{c,g}^{(u)} \leq 1, \quad \forall g \in G, \ c \in C.$$

Finally, the repetition constraint can be written as in (9) and (10), as follows:

$$r_c^{(u)} = 1\, h_0^{(u)} + 2\, h_1^{(u)} + 4\, h_2^{(u)} + 8\, h_3^{(u)} + 16\, h_4^{(u)} + \cdots, \tag{25}$$

and

$$\sum_{q=0}^{7} h_q^{(u)} = 1, \tag{26}$$

V. THE PROPOSED SOLUTIONS FOR THE JOINT EE-LATENCY OPTIMIZATION PROBLEM

This section presents suboptimal implementations for a feasible approximation of (3) and numerical validation of the problem properties that we want to emphasize. Notably, the algorithm is suitable for a system that allocates one RB or RU (such as RUC4 in the case of the NB-IoT). The first is a heuristic algorithm for EE optimization, followed by a solution for the latency compatible with EE maximization, and a novel algorithm, named "*best score*." Finally, we present a generalized multi-dimensional version of the "*best score*" algorithm.

A. ENERGY EFFICIENCY MAXIMIZATION

The algorithm is simple: the EE for each device is calculated for a specific RU, the devices are ordered according to their achieved single EE; and finally, the RU is assigned to the device with the highest EE. This algorithm is part of the "*best score*," as shown in Table 4. Because the total power for each device is fixed, maximizing EE is equivalent to maximizing the rate or SINR for each device.

The sub-optimality of the algorithm is related to interference optimization, which is neglected in this study. It is clear from (6) that there is a particular configuration of mutual interference terms that optimizes the rate in the allocation process. Here, each cell allocates the RUs independently without considering their impact on the others.

B. LATENCY MINIMIZATION

To minimize latency, we must minimize the scheduling waiting time and total transmission time of the RU, including consecutive repetitions. Given that the scheduling time depends on the queue formed by allocating the list of devices along the resource grid, the total transmission time is minimized if the number of repetitions is also minimized. Simultaneously, the number of repetitions depends on the SINR; consequently, maximizing the SINR, as expressed for the EE, is strictly coherent with the minimization of the number of repetitions.

Therefore, we formulate a process for achieving overall latency minimization, including scheduling waiting time:

1) For each device, we calculated the total transmission time for the available RU, as was done for EE.
1) We[2)] apply a known theorem from queue theory, that is, SJF [45], [46]: the scheduling order of the device is selected according to its transmission time.

The second step minimizes the average waiting time for devices in a queue by first serving the devices with the shortest service or transmission times.

The shortcoming of this method, which makes it sub-optimal in this context, is that it is not guaranteed that the SINR or rate will be the same for a device along the entire queue or resource grid according to the coherence time of the channel. Therefore, the order with respect to the lowest transmission time (that is, according to the best SINR) could change when the queue is served. Nevertheless, as we will see in the next section, the approach realizes a best-effort strategy in which the device is immediately served with the current instantaneous SINR, achieving a coherent trade-off between the best rate (best EE), minimum number of repetitions, and minimum average scheduling time at each SF.

TABLE 4. Procedure for suboptimal resource allocation: "*best score*".

Resource allocation procedure
1) Set rates and transmitted powers to 0
2) Take the first available RU
3) Loop on the unscheduled devices
4) Calculate rate, number of repetitions, EE according to (22)
5) Calculate latency (including repetitions)
6) Repeat from (3) for all unscheduled devices
7) Order the devices according to the maximum EE
8) Order the devices according to SJF
9) Select the device according to the joint, weighted criterion, and schedule to it the current RU.
10) Repeat from (2) until all the devices are scheduled

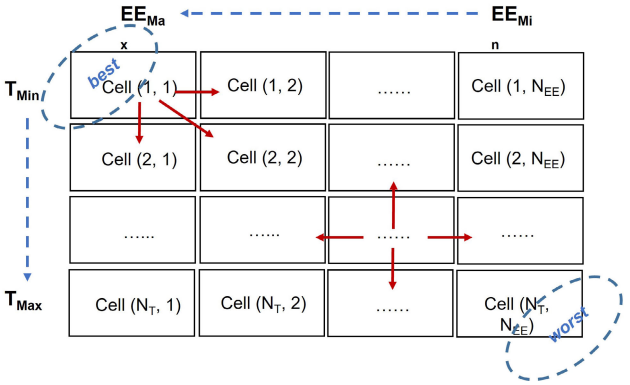

FIGURE 4. The T-EE 2D array and next cell selection (red arrows).

C. SCORE MAXIMIZATION

This suboptimal solution is a combination of EE maximization and latency minimization according to (3). The algorithm, which is reported in Table 4, operates as follows: the scheduler starts from the first available RU in the resource grid according to index g and computes the current SINR, rate, EE, and latency for all unscheduled devices. Two lists are created: in the former, the devices are ordered according to their EE from the highest; in the latter, the devices are ordered according to their latency from the lowest. We then search for the device in which the weighted sum between EE and latency is the maximum, according to weights w_1 and w_2 in (3). Therefore, the current RU is allocated to the device with the highest "*best score*" value. The scheduler moves to the next available RU and recomputes the EE and latency, if necessary, according to the channel variations, and the operation is repeated until all devices are scheduled.

D. MULTI-DIMENSIONAL ALGORITHM

In this section, we propose a generalized form of the "*best score*" algorithm, denoted as JL-EE (Joint Latency - Energy Efficiency). The proposed algorithm performs the scheduling function and is based on a two-dimensional matrix or array, denoted T-EE. Each dimension of the matrix contains the quantized N_T levels of the associated parameters, latency L in the interval $\{T_{Min}, T_{Max}\}$ for the rows, and the N_{EE} levels of EE in the interval $\{EE_{Min}, EE_{Max}\}$ for the columns (Fig. 4). Thus, T-EE creates a set of N_{EE} $times N_T$ cells, each associated with two (T, EE) values. Note that the intervals are ordered in the matrix from best to worst; that is, from T_{Min} for latency and EE_{Max} for EE.

The algorithm has the following parameters:

- The number N_T of levels of L, which means that the interval $\{T_{Min}, T_{Max}\}$ is divided into N_T uniform intervals ΔT.
- The number N_{EE} of levels of EE, which means that the interval $\{EE_{Min}, EE_{Max}\}$ is divided into N_{EE} uniform intervals ΔEE.
- The weights w_1, associated with T, and w_2, associated to EE, for computing the weighted performance values.

The algorithm has the following inputs:

- Set of devices $i = 1, \ldots N_U$, which represents the operations to be scheduled.
- Set of Resource Units (RU) $k = 1, \ldots N_(L)$. The set of RUs represents the available resources, and without loss of generality, we can imagine it as a set of non-overlapping elements, each constituting a possible resource choice for the scheduler. For instance, these RUs can all be RUCα with maximum α.
- For each i-th device and k-th RU, the set of parameters T_{ik} and EE_{ik}. These values are typically computed before the scheduling process on a block basis, according to the SNR.

The algorithm is presented here as a *block-based scheduler*; that is, it allocates the entire set of RUs to the N_U devices and then moves to a new block of RUs and devices. According to the well-known strategies exploited in this field, devices that have not been allocated to one of the blocks can be moved to the next. Furthermore, it is straightforward to extend the algorithm to a *continuous modality* by adding new devices and resources when available and updating the corresponding inputs.

For each device, the algorithm outputs the allocation to one of the available RUs, or no allocation.

The algorithm steps (Fig. 5) can be described as follows.

1) Each T-EE cell (n, m) was filled with the couples (T_{ik}, EE_{ik}) which belong to the corresponding performance parameters intervals.
2) The algorithm begins with the upper-left cell ($n = 1$, $m = 1$), which then becomes the current cell.
3) The $N_{C,nm}$ couples present in the current cell (n, m) are ordered with respect to the weighted sum of the corresponding performance parameters:
$$w_j = w_1 T_{ik} + w_2 EE_{ik}, \tag{27}$$
for $j = 1, \cdots, N_{C,nm}$.
4) According to the ordered list in the current cell (n, m), the corresponding RU (index k in couple (T_{ik}, EE_{ik})) is allocated to the associated device (index i in couple (T_{ik}, EE_{ik})). When a device is allocated, the other

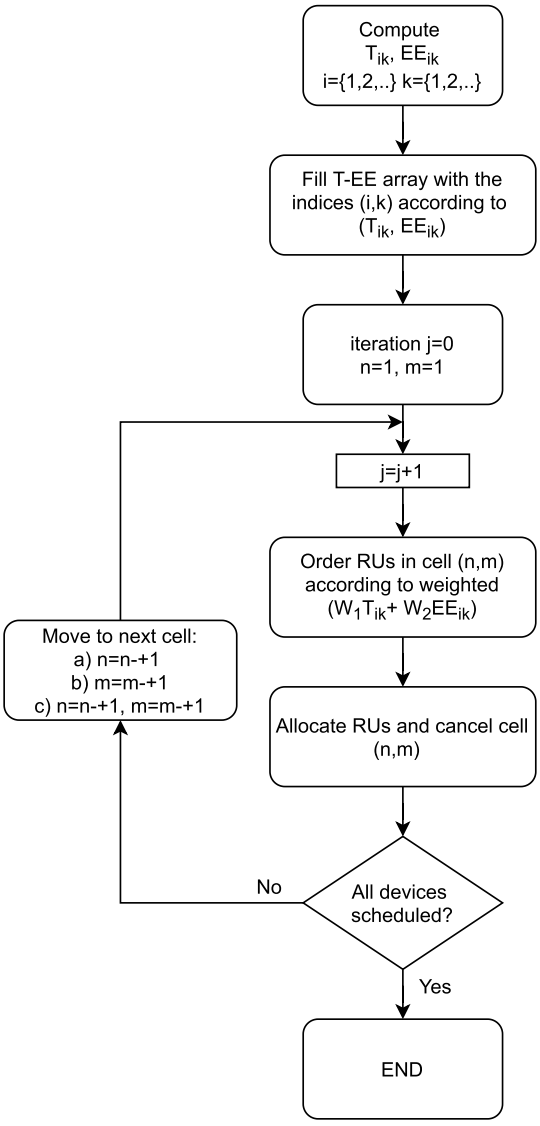

FIGURE 5. Flowchart of the proposed algorithm.

possible couples with the same index i can be erased from the list if the algorithm is programmed to allocate only one RU to each device.

5) The current cell is cancelled from the matrix, and the algorithm checks whether all devices (or other constraints) are allocated. If so, the algorithm ends and it moves to the next resources/device block. If no, the *next cell* (n', m') is selected in the matrix (moving in the right and/or down direction, Fig. 4), and the algorithm performs a new iteration from point (3).

The next cell selection deserves separate discussion because it is one of the characteristic features of the algorithm. This step is denoted as the *path strategy* of the algorithm in the T-EE array. We can distinguish at least three basic path strategies, but this choice can be easily extended to more elaborate approaches.

- Fixed path: A predefined path among the cells in the matrix is set before the algorithm begins.
- The next cell is the closest cell, which has the maximum number of devices, to increase the algorithm speed and efficiency.
- The next cell is the closest cell, which has the minimum average weighted sum computed by averaging all values w_j (27) for the couples in each neighboring cell.

We can observe the following for characterizing this algorithm with respect to the state-of-the-art.

- If $N_{EE} = N_T = 1$, the algorithm reduces to the computation of an ordered list with respect to the weighted sums of all possible allocation couples.
- If $N_{EE} = 1$, the algorithm reduces to a pure SJF approach.

E. COMPLEXITY ANALYSIS

In this section, we present the complexity analysis of the proposed algorithms, that is, the best score and multi-dimensional version of the algorithm. First, a complexity analysis of the best score algorithm (Sect. V-C), Table 4 shows two main nested loops, an outer loop and an inner loop. The outer loop runs a number of times equal to the number of devices within the cell; thus, $O(n)$. The inner loop represents the number of unscheduled devices, which decreases by one device at each iteration of the outer loop; thus, its complexity is $O(n/2)$. Therefore, the overall complexity of the algorithm is $O(n \cdot n/2)$, or $O(n^2)$. Regarding the multi-dimensional algorithm (Sect. V-D) a flowchart of the algorithm is presented in Fig. 5. First, to fill the 2D array (Fig. 4), the complexity is $O(n \cdot m)$ where n represents the number of devices, and m is the number of RUs per scheduling frame. Second, there is a nested loop similar to that used in the best score algorithm and its complexity is $O(n^2)$. Therefore the overall complexity turns out to be $O(m \cdot n + n^2)$.

VI. NUMERICAL RESULTS

The proposed solutions were simulated based on the channel model, parameters and assumptions listed in Table 5. Simulations were carried out for a seven-cell urban macro scenario, and the cells caused inter-cell interference. In addition, a full-buffer model was assumed.

All schedulers are block-based schedulers, which means that the scheduling decision is made after analysing a block of RUs; for each RU in the block, the EE, number of repetitions, and transmission latency for each device are computed according to its SINR (assumed known without errors) and then the algorithms operate their decisions according to the corresponding scores and strategies. The scheduling algorithms are described in Sect. V ('best score', 'SJF', 'best EE', 'JL-EE'), with the addition of the standard RR used as a benchmark. In the T-EE matrix of the JL-EE, the $N_T = 8$ levels of latency correspond to the eight options for the number of repetitions (1, 2, 4, 8, 16, 32, 64, 128) and the

TABLE 5. Simulation parameters.

Channel model	
Attenuation (d = distance [m], f = frequency [GHz])	$(44.9-6.5\cdot log_{10}(32))\cdot log_{10}(d)+34.46+5.83\cdot log_{10}(32)+23\cdot log_{10}(f/5)$
Shadowing	Log-normal with $\sigma = 8$ dB
Multipath	SCM model, Urban macro
Radio network	
Cell layout	7 hexagonal cells (in the set C)
Cell radius	$R = 4000, 2000, 1000, 500$ m
Bandwidth	1 RB = 180 KHz, 12 SCs in S
Maximum transmit power (P_T)	23 dBm
Numerology μ (SC spacing)	0 (15 KHz)
Scheduling	"*best score*", Best EE, SJF, RR, JL-EE ($N_T = 8$, $N_{EE} = 8$), CSC
eNodeB	Three sectors or one sector with omni-directional antenna
Number of active devices per sector	from 5 to 200

$N_{EE} = 8$ levels of the EE are obtained by dividing the range of EE values into eight uniform levels. The weights w_1 and w_2 of 'best Score' and 'JL-EE' are normalised to balance the values obtained from the transmission latency and EE, i.e., they are equal to $0.5/T_{Max}$ and $0.5/EE_{Max}$ respectively. To enhance the comparison with the proposed algorithms, we also implemented the algorithm proposed in [29], named here the combination-based SCheduler (CSC); this algorithm operates in the same block as the others, selecting the best allocation according to the score function proposed in [29].

It is note worthy that all schedulers are subject to a minimum rate constraint. Moreover, the transmission power of all the devices is set to the maximum power, or is subject to automatic transmit power control (ATPC). In all the figures, the simulations are run for 100 independent channel realisations and device positions. In addition, specific care has been taken to realistically consider inter-cell interference. For each of the 100 runs, the seven cells are subjected to iterative repetition of the simulation until convergence of the allocation process is achieved, as described below.

- In the initial step, a random allocation of resources allows the computation of the interference at each device for each RU and, consequently, of the initial SINR.
- At the end of each step, the current allocation is used to update the values of the transmit powers of all devices and, consequently, the value of the interference.
- The simulation is repeated until the average EE measured in the centre cell achieves a stable convergence. The final values of EE, latency, power, and rate measured in this cell for all devices are used as performance results.

At the reference cell radius $R = 1000$ m with the device transmit power P_T fixed at the maximum value (no ATPC), the average EE achieved by each device is presented in Fig. 6, and the resulting latency is shown in Fig. 7. We can observe that all of the proposed schedulers have a clear advantage

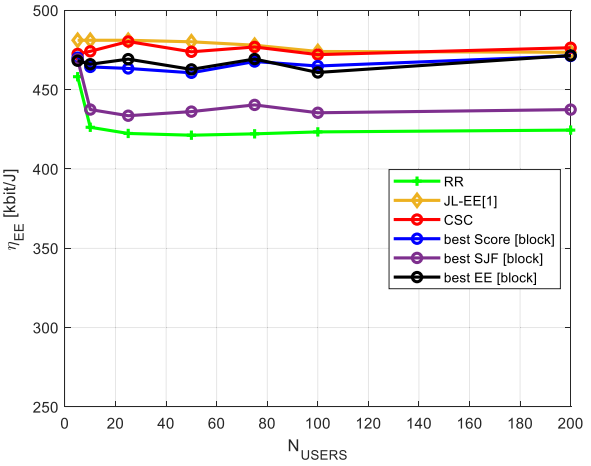

FIGURE 6. Average EE per user for RUC4 with radius $R = 1000$ m and maximum transmission power.

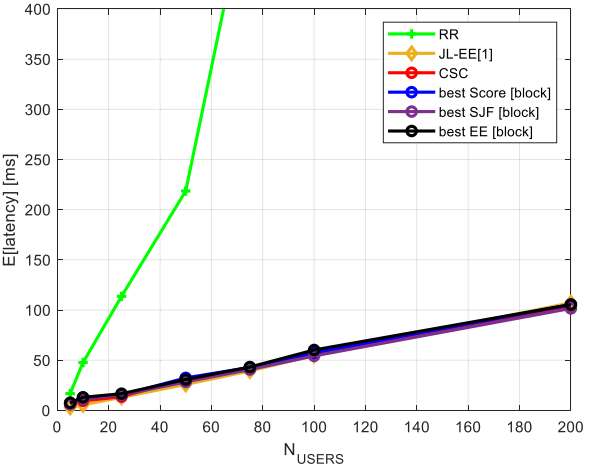

FIGURE 7. Average latency per user for RUC4 with radius $R = 1000$ m and maximum transmission power.

over RR, especially for latency. One of the key factors for understanding these results is the role of repetitions. For the same RU, if one device uses fewer repetitions than other devices, this means that (i) this device has a better channel in this RU than the others, (ii) it consumes less total energy for transmitting all repetitions, and (iii) it will have better latency. From the results, this is confirmed by the fact that the performances of the simulated algorithms is similar, except for the RR. From another point of view, this is because of the correlation between the solutions for optimal EE and optimal latency, making solutions for EE optimization or latency minimization close to each other, even if designed initially with differing objectives. When the SNR for a device is low, the EE is usually low, and the number of repetitions increases to achieve the required performance, thereby increasing the transmission latency. In addition, these devices are usually postponed in the transmission queue because this contributes to the minimization of the overall latency (see the principle of SJF in Sect. V).

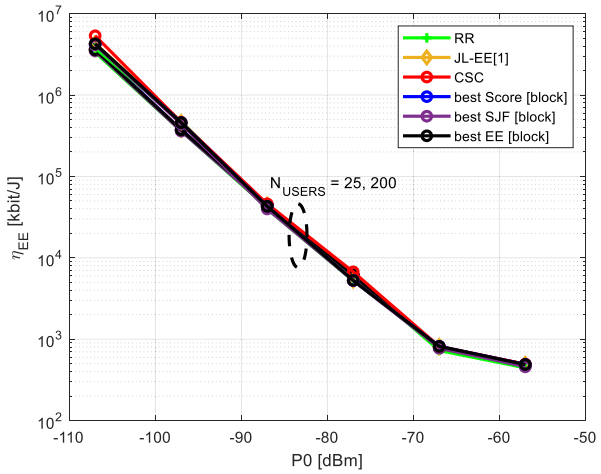

FIGURE 8. Average EE per user for RUC4, $R = 1000$ m, and number of users $N_{users} = 25, 200$. Power control with different values of P_0 and $\alpha = 1$.

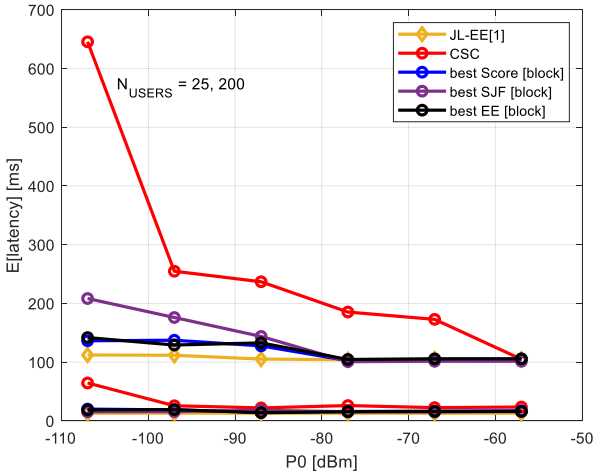

FIGURE 9. Average latency per user for RUC4, $R = 1000$ m, and number of users $N_{users} = 25, 200$. Power control with different values of P_0 and $\alpha = 1$.

To investigate the impact of the ATPC, we consider the following simple formulation for power control:

$$P_{Tx} = P_0 + \alpha P_L, \tag{28}$$

where P_{Tx} is the transmission power, P_0 is a parameter that indicates the desired received power, and $0 \leq \alpha \leq 1$ is the compensation factor for the path loss P_L. First, we investigate the impact of P_0 for the minimum and maximum number of users in the simulations (25 and 200), and the results are presented in Figs. 8 and 9, where it can be observed that as P_0 increases, both the EE and latency decrease even if in the latency the difference among the algorithms is clearly more relevant. We then investigate the impact of the number of users for two values of P_0 that achieve good latency performance and reasonable EE, that is, -67 dBm and -77 dBm. Figs. 10 and 11 show the resulting EE and latency, respectively, with ATPC. With regard to EE, the results corresponding to the two key values of P_0 clearly show that the algorithms are similar and P_0 has a significant impact on EE. At the same time, comparing this EE to the values in Fig. 6, it is clear that the ATPC has a positive impact on EE. However, there is a noticeable impact of each scheduler on the latency, as the RR performance is clearly poor and the proposed algorithm achieves a latency reduction of approximately 40% with respect to the benchmark CSC. It should be noted that both values of P_0 achieve similar latencies and only one value is presented.

As explained in the system model, a three-sector antenna is used as the reference system. We now present the results obtained using an omni-directional antenna (single sector). It can be seen in Figs. 12 and 13 that there is a decrease in the EE and, in particular, a substantial increase in the latency; the reason for this is the significant increase in interference in this type of layout.

Another relevant layout parameter is obviously the cell radius R. In Figs. 14 and 15, we present the results when R is

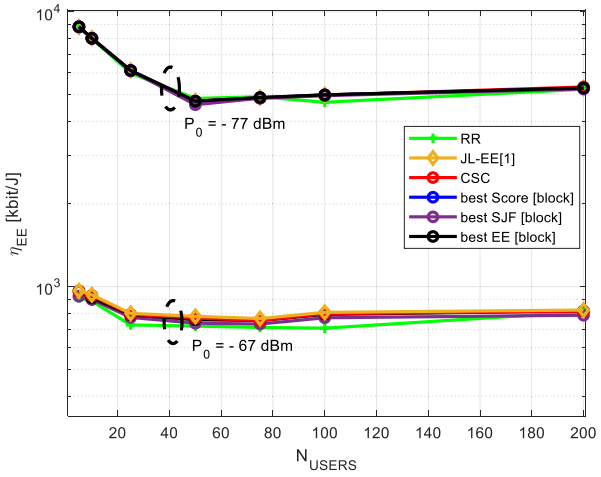

FIGURE 10. Average EE per user for RUC4 and $R = 1000$ m. Power control with $P_0 = -67, -77$ dBm and $\alpha = 1$.

doubled to 2000 m. It can be observed that the overall trend and behaviour of the schedulers do not change because the SINR does not change, and the same is true for $R = 4000$ m.

Finally, with respect to the requirements of EE and latency compatibility with IoT cases, as discussed in Sect. I, we can observe that:

- EE can increase considerably reducing the parameter $P0$ in the ATPC, and it turns out to be very similar for all scheduling techniques. However, reducing P_0 increases the average latency owing to the reduction in the SNR. However, the JL-EE algorithm provides more robustness with respect to latency degradation as $P0$ decreases (Figs. 8 and 9, Figs. 12 and 13).
- With the maximum power transmission, the latency is similar for most of the techniques (except RR) and the JL-EE and CSC algorithms assure, at the same time, also the best EE (Figs. 6 and 7, Figs. 14 and 15).

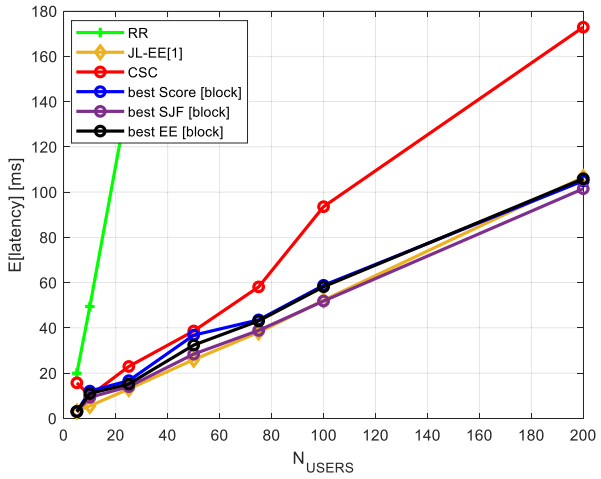

FIGURE 11. Average Latency per user for RUC4 and $R = 1000$ m. Power control with $P_0 = -67$ dBm and $\alpha = 1$.

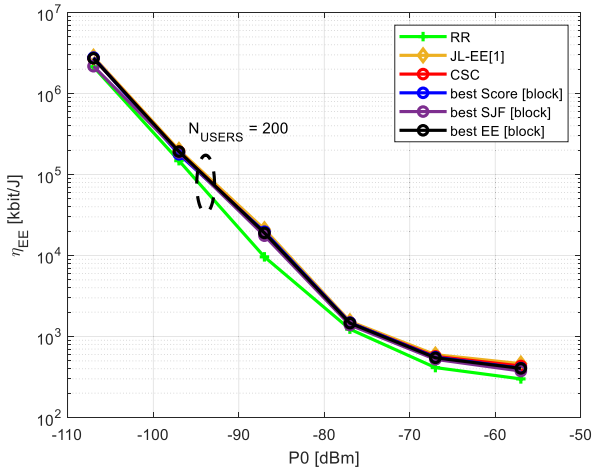

FIGURE 12. Average EE per user for RUC4 with one-sector antenna, $R = 1000$ m, and maximum transmission power.

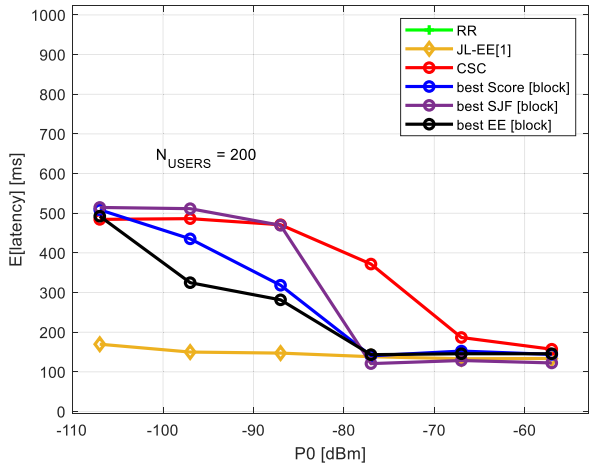

FIGURE 13. Average latency per user for RUC4 with one-sector antenna, $R = 1000$ m, and maximum transmission power.

- The latency achieves the minimum values with maximum power or relatively high values of $P0$ as the SNR increases, thus reducing the repetitions. Under these conditions, it is possible to achieve the minimum transmission latency of a single slot (here, 1 ms), and the linear increase in the latency as a function of the number of devices (approximately all the results for the three sectors cells, except RR) shows that the latency increase is due only to the queue latency. Therefore, the correct management of packets priorities would guarantee the most stringent low latency requirements.

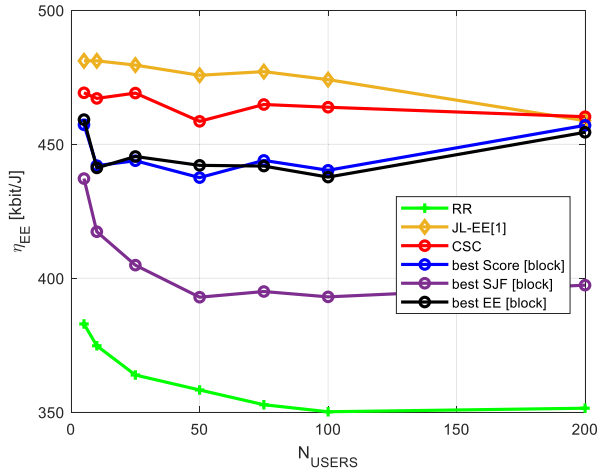

FIGURE 14. Average EE per user for RUC4 with $R = 2000$ m, and maximum transmission power.

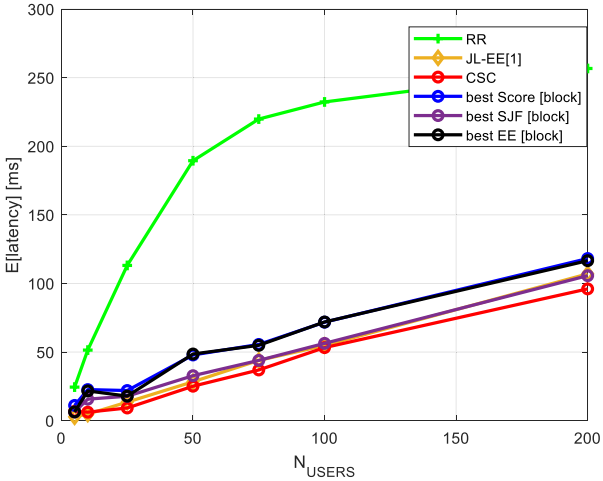

FIGURE 15. Average latency per user for RUC4 with $R = 2000$ m and maximum transmission power.

VII. CONCLUSION

In this work, we studied the relationship between EE and latency in multi-cellular scenarios, with an emphasis on IoT technologies. We proposed an optimization problem based on the joint optimization of energy efficiency and latency, and discussed the mutual relations between these key performance factors. When multiple rectangular resource unit configurations are prescribed by the standard,

as in the NB-IoT, we investigated their energy efficiency and provided some guidelines for simplifying the initial problem formulation. Then in the theoretical framework, we incorporated an approach based on the "shortest job first" from queuing theory and proposed a scheduler for minimising latency in the presence of repetitions. Moreover, we designed a sub-optimal joint scheduler to show that the resulting performance outperforms the round robin scheduler, providing the best results for all the simulated options. Furthermore, the proposed scheduler was compared with a benchmark algorithm. The results also confirm the expected high correlation between EE and latency but also reveal the space for performance enhancement and trade-offs in the presence of a multi-cellular interference-limited scenario. From the results, the impact of the power control and its parameters on the EE became clear.

APPENDIX
FORMULATION OF THE SHAPE CONSTRAINTS

As part of the optimization problem described in Section IV. In the following, we provide the constraints that force the scheduler to allocate the RUs according to the specific shapes provided by the 3GPP. Hence, we write constraints for both the time and frequency domains to force the predefined values that constitute the RU shape.

In the time dimension, for device u at a given SF t,

$$\sum_{s\in S} x_{c,t,s}^{(u)} = N_{c,t}^{(u)}, \quad \forall t \in T,\ u \in U_c,\ c \in C \tag{29}$$

This implies that the number of SCs allocated to device is equal to $N_{c,t}^{(u)}$. Subsequently, to guarantee that the allocated SCs are consecutive,

$$\sum_{s=1}^{end-1} |(x_{c,t,s}^{(u)} - x_{c,t,s+1}^{(u)})| \leq 2, \quad \forall t \in T,\ c \in C. \tag{30}$$

In the frequency dimension, for device u at a given SC s,

$$\sum_{t\in T} x_{c,t,s}^{(u)} = M_{c,s}^{(u)}, \quad \forall s \in S,\ c \in C \tag{31}$$

The above constraint guarantees that the number of SFs allocated to the device is $M_{c,s}^{(u)}$. Then, the consecutive SFs constraint can be written as:

$$\sum_{t=1}^{end-1} |(x_{c,t,s}^{(u)} - x_{c,t+1,s}^{(u)})| \leq 2, \quad \forall s \in S,\ c \in C. \tag{32}$$

In addition, by introducing the binary RUC allocation variable $v_{c,q}^{(u)}$ ($q = \{1, 2, 3, 4\}$ is the index of the four RUCs). Because each device has only one assigned RUC, we can state that

$$\sum_{q} v_{c,q}^{(u)} = 1, \tag{33}$$

Moreover, the values $N_{c,t}^{(u)}$ and $M_{c,s}^{(u)}$ defined in (29) and (31) can assume, at each SF t or SC s, only two values, that is, 0 if the SF or SC is not allocated to that device or the values $r_c^{(u)} \cdot z_c^{(u)}$, $q_c^{(u)}$, which are the corresponding sides of the RUCs at time (including the repetitions) and frequency domains, respectively (Table 3). Consequently, we should also force the number of allocated SCs per SF to be equal to the value $r_c^{(u)} \cdot z_c^{(u)}$ or zero for each device, that is,

$$N_{c,t}^{(u)} \cdot (N_{c,t}^{(u)} - r_c^{(u)} \cdot z_c^{(u)}) = 0, \quad \forall t \in T,\ \forall c \in C, \tag{34}$$

and the number of allocated SFs per SC is equal to the value q_c^u or 0 for each device, that is,

$$M_{c,s}^{(u)} \cdot (M_{c,s}^{(u)} - q_c^{(u)}) = 0, \quad \forall s \in S,\ \forall c \in C. \tag{35}$$

Therefore, considering just 1 RUC per device, we have that

$$z_c^{(u)} = 1 \cdot v_{c,1}^{(u)} + 3\, v_{c,2}^{(u)} + 6 v_{c,3}^{(u)} + 12\, v_{c,4}^{(u)}, \tag{36}$$

$$q_c^{(u)} = 8\, v_{c,1}^{(u)} + 4\, v_{c,2}^{(u)} + 2 v_{c,3}^{(u)} + 1 \cdot v_{c,4}^{(u)}. \tag{37}$$

Equations (33) and (37) characterize the constraints on the four RUCs: (33) guarantees that only one value of $v_{c,q}^{(u)}$ is 1, whereas the values from (36) and (37) force the total numbers of SCs and SFs per device to be equal to one of the predefined RUC shapes.

ACKNOWLEDGMENT

This material reflects only the authors' view and the EC Research Executive Agency is not responsible for any use that may be made of the information it contains.

REFERENCES

[1] M. Rasti, S. K. Taskou, H. Tabassum, and E. Hossain, "Evolution toward 6G wireless networks: A resource management perspective," 2021, *arXiv:2108.06527*.
[2] Y.-P. E. Wang, X. Lin, A. Adhikary, A. Grovlen, Y. Sui, Y. Blankenship, J. Bergman, and H. S. Razaghi, "A primer on 3GPP narrowband Internet of Things," *IEEE Commun. Mag.*, vol. 55, no. 3, pp. 117–123, Mar. 2017.
[3] *Study on Scenarios and Requirements for Next Generation Access Technologies (Release 17)*, document TR38.913, V17.0.0, 3GPP, ETSI, Sophia Antipolis, France, 2022.
[4] *Study on Support of Reduced Capability NR Devices (Release 17)*, document TR38.875, V17.0.0, 3GPP, ETSI, Sophia Antipolis, France, Mar. 2021.
[5] S. R. Pokhrel, J. Ding, J. Park, O.-S. Park, and J. Choi, "Towards enabling critical mMTC: A review of URLLC within mMTC," *IEEE Access*, vol. 8, pp. 131796–131813, 2020.
[6] *Study on Communication for Automation in Vertical Domains (Release 16)*, document TR22.804, V16.3.0, 3GPP, ETSI, Sophia Antipolis, France, Jul. 2020.
[7] G. Tsoukaneri, F. Garcia, and M. K. Marina, "Narrowband IoT device energy consumption characterization and optimizations," in *Proc. Int. Conf. Embedded Wireless Syst. Netw. (EWSN)*, 2020, pp. 1–12.
[8] E. Migabo, K. Djouani, and A. Kurien, "An energy-efficient and adaptive channel coding approach for narrowband Internet of Things (NB-IoT) systems," *Sensors*, vol. 20, no. 12, p. 3465, Jun. 2020.
[9] Y. Guo and M. Xiang, "Multi-agent reinforcement learning based energy efficiency optimization in NB-IoT networks," in *Proc. IEEE Globecom Workshops*, Dec. 2019, pp. 1–6.
[10] D. Di Lecce, A. Grassi, G. Piro, and G. Boggia, "Boosting energy efficiency of NB-IoT cellular networks through cooperative relaying," in *Proc. IEEE 29th Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2018, pp. 1–5.
[11] A. Azari, G. Miao, C. Stefanovic, and P. Popovski, "Latency-energy tradeoff based on channel scheduling and repetitions in NB-IoT systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–7.
[12] A. Azari, C. Stefanovic, P. Popovski, and C. Cavdar, "On the latency-energy performance of NB-IoT systems in providing wide-area IoT connectivity," *IEEE Trans. Green Commun. Netw.*, vol. 4, no. 1, pp. 57–68, Mar. 2020.

[13] E. Rastogi, M. Kumar Maheshwari, A. Roy, N. Saxena, and D. Ryeol Shin, "Energy efficiency analysis of narrowband Internet of Things with auxiliary active cycles for small data transmission," *Trans. Emerg. Telecommun. Technol.*, vol. 33, no. 1, pp. 1–16, Jan. 2022, Art. no. e4376.
[14] S. A. Manzar, S. Verma, and S. H. Gupta, "Analysis and optimization of downlink energy in NB-IoT," *Sustain. Comput., Informat. Syst.*, vol. 35, Sep. 2022, Art. no. 100757.
[15] J.-M. Liang, K.-R. Wu, J.-J. Chen, P.-Y. Liu, and Y.-C. Tseng, "Energy-efficient uplink resource units scheduling for ultra-reliable communications in NB-IoT networks," *Wireless Commun. Mobile Comput.*, vol. 2018, pp. 1–17, Jul. 2018.
[16] S. Popli, R. K. Jha, and S. Jain, "A survey on energy efficient narrowband Internet of Things (NBIoT): Architecture, application and challenges," *IEEE Access*, vol. 7, pp. 16739–16776, 2019.
[17] E. Migabo, K. Djouani, and A. Kurien, "Energy efficient data rate enhancement channel coding technique for narrowband Internet of Things (NB-IoT)," in *Proc. IEEE AFRICON*, Sep. 2021, pp. 1–6.
[18] Q. Guo, F. Tang, and N. Kato, "Federated reinforcement learning-based resource allocation in D2D-enabled 6G," *IEEE Netw.*, pp. 1–7, Sep. 2022.
[19] Y.-J. Yu and C.-L. Wu, "Energy-efficient scheduling for search-space periods in NB-IoT networks," *IEEE Syst. J.*, vol. 17, no. 3, pp. 3974–3985, Sep. 2023.
[20] Z. Amjad, A. Sikora, J.-P. Lauffenburger, and B. Hilt, "Latency reduction in narrowband 4G LTE networks," in *Proc. 15th Int. Symp. Wireless Commun. Syst. (ISWCS)*, Aug. 2018, pp. 1–5.
[21] F. D. O. Torres, V. A. S. Júnior, D. B. D. Costa, D. L. Cardoso, and R. C. L. Oliveira, "Radio resource allocation in a 6G D-OMA network with imperfect SIC: A framework aided by a bi-objective hyper-heuristic," *Eng. Appl. Artif. Intell.*, vol. 119, Mar. 2023, Art. no. 105830.
[22] H. Al-Obiedollah, H. Bany Salameh, S. Abdel-Razeq, A. Hayajneh, K. Cumanan, and Y. Jararweh, "Energy-efficient opportunistic multi-carrier NOMA-based resource allocation for beyond 5G (B5G) networks," *Simul. Model. Pract. Theory*, vol. 116, Apr. 2022, Art. no. 102452.
[23] N. Moosavi, A. Zappone, P. Azmi, and M. Sinaie, "Delay-aware and energy-efficient resource allocation for reconfigurable intelligent surfaces," *IEEE Commun. Lett.*, vol. 27, no. 2, pp. 605–609, Feb. 2023.
[24] A. L. Imoize, H. I. Obakhena, F. I. Anyasi, and S. N. Sur, "A review of energy efficiency and power control schemes in ultra-dense cell-free massive MIMO systems for sustainable 6G wireless communication," *Sustainability*, vol. 14, no. 17, p. 11100, Sep. 2022.
[25] S. Zhu, W. Wu, L. Feng, P. Zhao, F. Zhou, P. Yu, and W. Li, "Energy-efficient joint power control and resource allocation for cluster-based NB-IoT cellular networks," *Trans. Emerg. Telecommun. Technol.*, vol. 30, no. 4, pp. 1–22, Apr. 2019, Art. no. e3266.
[26] Z. Shuang, Z. Ningbo, and K. Guixia, "Energy efficiency for NPUSCH in NB-IoT with guard band," *ZTE Commun.*, vol. 16, no. 4, pp. 46–51, 2020.
[27] X. Liu and X. Zhang, "Rate and energy efficiency improvements for 5G-based IoT with simultaneous transfer," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 5971–5980, Aug. 2019.
[28] R. Barbau, V. Deslandes, G. Jakllari, and A.-L. Beylot, "An analytical model for evaluating the interplay between capacity and energy efficiency in NB-IoT," in *Proc. Int. Conf. Comput. Commun. Netw. (ICCCN)*, Jul. 2021, pp. 1–9.
[29] F. Yassine, M. El Helou, S. Lahoud, and O. Bazzi, "Energy-efficient uplink scheduling in narrowband IoT," *Sensors*, vol. 22, no. 20, p. 7744, Oct. 2022.
[30] A. T. AbuSabah, M. A. Rahman, R. Oliveira, and A. Flizikowski, "The importance of repetitions in ultra-dense NB-IoT networks," *IEEE Commun. Lett.*, vol. 26, no. 5, pp. 1199–1203, May 2022.
[31] A. Zappone and E. Jorswieck, "Energy efficiency in wireless networks via fractional programming theory," *Found. Trends® Commun. Inf. Theory*, vol. 11, nos. 3–4, pp. 185–396, 2015.
[32] K. J. Kim, N. Wattanapongsakorn, and N. Joukov, *Mobile and Wireless Technology 2016*, vol. 391. Cham, Switzerland: Springer, 2016.
[33] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2349–2360, Sep. 2005.
[34] S. Khakurel, L. Musavian, and T. Le-Ngoc, "Energy-efficient resource and power allocation for uplink multi-user OFDM systems," in *Proc. IEEE 23rd Int. Symp. Pers., Indoor Mobile Radio Commun. - (PIMRC)*, Sep. 2012, pp. 357–361.
[35] O. Elgarhy, "Hybrid power control for LTE uplink," M.S. thesis, Dept. Electron., Inf. Bioeng., Politecnico di Milano, Milan, Italy, 2016.
[36] H. Pervaiz, L. Musavian, and Q. Ni, "Joint user association and energy-efficient resource allocation with minimum-rate constraints in two-tier HetNets," in *Proc. IEEE 24th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Sep. 2013, pp. 1634–1639.
[37] G. Y. Li, Z. Xu, C. Xiong, C. Yang, S. Zhang, Y. Chen, and S. Xu, "Energy-efficient wireless communications: Tutorial, survey, and open issues," *IEEE Wireless Commun.*, vol. 18, no. 6, pp. 28–35, Dec. 2011.
[38] C. Xiong, G. Y. Li, S. Zhang, Y. Chen, and S. Xu, "Energy- and spectral-efficiency tradeoff in downlink OFDMA networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 11, pp. 3874–3886, Nov. 2011.
[39] O. Elgarhy, L. Reggiani, and R. Ferrari, "Hybrid power control for multi-carrier systems," in *Proc. 10th IFIP Wireless Mobile Netw. Conf. (WMNC)*, Sep. 2017, pp. 1–5.
[40] S. Hayashi and Z.-Q. Luo, "Spectrum management for interference-limited multiuser communication systems," *IEEE Trans. Inf. Theory*, vol. 55, no. 3, pp. 1153–1175, Mar. 2009.
[41] L. Venturino, A. Zappone, C. Risi, and S. Buzzi, "Energy-efficient scheduling and power allocation in downlink OFDMA networks with base station coordination," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 1–14, Jan. 2015.
[42] M. Ehrgott, *Multicriteria Optimization*, vol. 491. Cham, Switzerland: Springer, 2005.
[43] O. Elgarhy, L. Reggiani, H. Malik, M. M. Alam, and M. A. Imran, "Rate-latency optimization for NB-IoT with adaptive resource unit configuration in uplink transmission," *IEEE Syst. J.*, vol. 15, no. 1, pp. 265–276, Mar. 2021.
[44] S. R. Sabuj, A. Ahmed, Y. Cho, K. Lee, and H. Jo, "Cognitive UAV-aided URLLC and mMTC services: Analyzing energy efficiency and latency," *IEEE Access*, vol. 9, pp. 5011–5027, 2021.
[45] M. Harchol-Balter, *Performance Modeling and Design of Computer Systems: Queueing Theory in Action*. Cambridge, U.K.: Cambridge Univ. Press, 2013.
[46] A. Cobham, "Priority assignment in waiting line problems," *J. Oper. Res. Soc. Amer.*, vol. 2, no. 1, pp. 70–76, Feb. 1954.

OSAMA ELGARHY received the B.Sc. degree in electrical and communication engineering, in 2009, and the M.Sc. and Ph.D. degrees in telecommunication engineering from Politecnico di Milano, Milan, Italy, in 2016 and 2020, respectively. From 2009 to 2013, he was a Teaching Assistant with the Pyramids Higher Institute of Engineering, Cairo, Egypt. In 2016, he spent a research period with Azcom Technology, Milan. He is currently a Postdoctoral Researcher with the Thomas Johann Seebeck Department of Electronics, Tallinn University of Technology, Tallinn, Estonia. His current research interests include resource allocation for cellular systems, NB-IoT, and device-to-device communications.

LUCA REGGIANI received the Ph.D. degree in electronics and communications engineering from Politecnico di Milano, Italy, in 2001. He has collaborated with several industries and universities in the field of wireless communications and magnetic recording, as a consultant or within Italian and European research programs. He is the co-founder of two startup companies in the ICT field. His research interests include mobile systems, wireless sensor networks, and UAV applications.

MUHAMMAD MAHTAB ALAM (Senior Member, IEEE) received the M.Sc. degree in electrical engineering from Aalborg University, Denmark, in 2007, and the Ph.D. degree in signal processing and telecommunication from the University of Rennes1 France (INRIA Research Center), in 2013. He was a Postdoctoral Researcher with the Qatar Mobility Innovation Center, Qatar, from 2014 to 2016. In 2016, he joined as the European Research Area Chair and an Associate Professor with the Thomas Johann Seebeck Department of Electronics, Tallinn University of Technology, where he was elected as a Professor, in 2018, and a Tenured Full Professor, in 2021. Since 2019, he has been the Communication Systems Research Group Leader. He has over 15 years of combined academic and industrial multinational experiences while working in Denmark, Belgium, France, Qatar, and Estonia. He has several leading roles as the PI in multimillion Euros international projects funded by the European Commission (Horizon Europe LATEST-5GS, 5G-TIMBER, H2020 5G-ROUTES, NATOSPS (G5482), Estonian Research Council (PRG424), and Telia Industrial Grant). He is the author and coauthor of more than 100 research publications. He is actively supervising several Ph.D. students and postdoctoral researchers. He is also a contributor to two standardization bodies (ETSI SmartBAN and IEEE-GeeenICT-EECH), including ''Rapporteur'' of work item: DTR/SmartBAN-0014. His research interests include wireless communications–connectivity, mobile positioning, and 5G/6G services and applications.

AHMED ZOHA (Senior Member, IEEE) received the M.Sc. degree in communication engineering from Chalmers University, Sweden, and the Ph.D. degree in electrical and electronics engineering from the 6G/5GIC Centre, University of Surrey (UniS), U.K., in July 2014. His research interests include artificial intelligence (AI) and machine learning, advanced signal processing, and cutting-edge self-learning strategies. He has accumulated over 13 years of experience designing intelligent applications and algorithms for 5G and beyond wireless communication systems, connected healthcare, the Internet of Everything, and smart energy monitoring systems. His research contributions have garnered recognition from national and international organizations, regulatory bodies, and media outlets. He was a technical program committee member of top-tier conferences. Additionally, he has been honored with three IEEE best paper awards. He was a U.K. Exceptional Talent endorsed by the Royal Academy of Engineering, which is awarded to early-career, world-leading innovators, and scientists. Furthermore, he regularly chairs and serves as the track chair for special sessions.

RIZWAN AHMAD received the M.Sc. degree in communication engineering and media technology from the University of Stuttgart, Stuttgart, Germany, in 2004, and the Ph.D. degree in electrical engineering from Victoria University, Melbourne, Australia, in 2010. From 2010 to 2012, he was a Postdoctoral Research Fellow with Qatar University on a QNRF Grant. He is currently a Professor and the HoD of the School of Electrical Engineering and Computer Science, National University of Sciences and Technology, Pakistan. His research interests include public safety networks, medium access control protocols, spectrum and energy efficiency, energy harvesting, and performance analysis for wireless communication and networks. He has published and served as a reviewer for IEEE journals and conferences. He was a recipient of the prestigious International Postgraduate Research Scholarship from the Australian Government.

ALAR KUUSIK (Member, IEEE) received the Ph.D. degree in IT from the Tallinn University of Technology (TalTech), Estonia, in 2001. He is currently a Senior Research Scientist with the T. J. Seebeck Institute of Electronics, TalTech, focusing on the IoT, data acquisition, and networking technologies. He has been involved with several international research and innovation projects related to smart environments, smart cities, and wearable technologies. He has published more than 50 peer-reviewed articles and is the author of five patent families. He is the Vice-Coordinator of the GAC of IEEE Computer Society in Region 8.