Energy-Efficient and Radio Resource Control State Aware Resource Allocation with Fairness Guarantees

Alba Jano*, Rakash Sivasiva Ganesan†, Fidan Mehmeti*, Serkut Ayvaşik*, Wolfgang Kellerer*
*Technical University of Munich, Chair of Communication Networks
{alba.jano, fidan.mehmeti, serkut.ayvaşik, wolfgang.kellerer}@tum.de
†Nokia Bell Labs
rakash.sivasivaganesan@nokia-bell-labs.com

***Abstract*—In the next-generation wireless networks, energy efficiency (EE) is a fundamental requirement due to the limited battery power and the deployment of various devices in hardly accessible areas. While a plethora of approaches have been proposed to increase users' EE, there are still many unresolved issues stemming mainly from the limited wireless resources. In this paper, we investigate the energy-efficient resource allocation, taking into account users' radio resource control (RRC) state. We aim to achieve max-min fairness among users in an uplink orthogonal frequency-division multiple access (OFDMA) system while fulfilling data rate requirements and transmit power constraints. In particular, we avoid waste of the energy through unnecessary state transitions when no network resources are available. We study the impact of the RRC Resume procedure on users' EE and propose allocating resources while users are in their current RRC Connected or RRC Inactive state. The solution is obtained from a constrained optimization problem, whose output is max-min fair and energy-efficient. To that end, we use generalized fractional programming and the Lagrangian dual decomposition approach to allocate the radio resources and transmission power iteratively. Using extensive realistic simulations with input parameters from measurement data, we compare the results of our approach against benchmark models and show the performance improvements RRC state awareness brings. Specifically, using our approach, the users' EE increases by at least 10% on average.**

***Index Terms*—Energy efficiency, RRC state awareness, max-min fairness.**

I. Introduction

The enhancement of spectral and energy efficiency in next-generation wireless networks is an important topic, attracting the attention of both academic and industry research [1], [2]. Energy efficiency (EE) has become a key performance indicator for sustainable 5G networks due to the growth of mobile devices connected to the network with energy-preserving requirements. In that regard, industrial wireless sensors, wearables, and video surveillance use cases, which require low complexity, low-cost internet of things (IoT) devices, with battery life requirements reaching up to 10 years [3] are in our focus. The relevance of EE is on the base station (BS) side and the device side. In both cases the EE is defined as the ratio of the achieved data rate and the energy consumed. The limited battery life of devices has increased the importance of EE on the device side. To address the EE issue, research is ongoing on device power-saving techniques and battery lifetime enhancement [4] while meeting the quality of service (QoS) requirements.

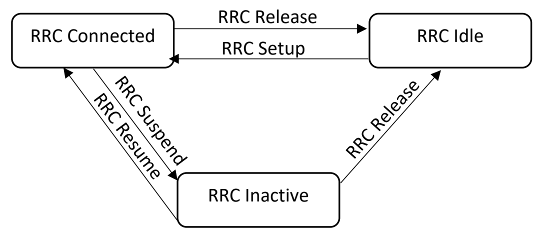


Fig. 1: 5G state diagram for RRC states and signalling.

A. New device state: RRC Inactive

Numerous methods are proposed to increase the EE of devices, such as bandwidth adaption, energy harvesting, DRX mechanism, and wake-up signaling [5]. To reduce further the energy consumption in 5G, an energy-efficient state named RRC Inactive is introduced, as shown in Fig. 1. RRC Inactive state decreases users' power consumption, latency, and signaling overhead during the resumption of the connected state, compared to the resumption from RRC Idle state [6]. Although RRC Inactive state is energy efficient, data transmission is performed mainly in the connected state, except in small data transmission cases supported by the inactive state [7]. Therefore, RRC Inactive is insufficient for energy-efficient devices as transition to the connected state is inevitable, thus decreasing the EE. The number of transitions to the connected state increases for bursty and small data packet applications.

B. Contribution: RRC state aware resource allocation

The impact of transition among states increases further when accounting for the limited radio resources, which can not support all users' QoS requirements. Consequently, resuming the connection of devices containing uplink data to transmit before resource allocation gives additional energy consumption. This is caused by the RRC Resume procedure and listening of the physical downlink control channel (PDCCH) while in the connected state. This energy is wasted when users do not receive resources and cannot transmit uplink data.

ISBN 978-3-903176-49-2 © 2022 IFIP

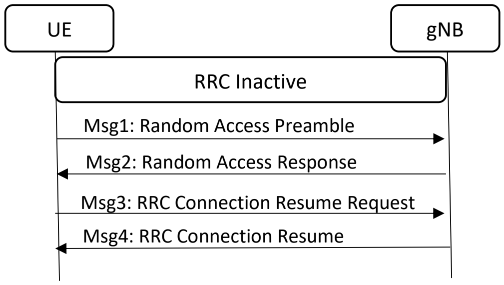


Fig. 2: RRC Connection Resume procedure [6].

When this is the case, a data inactivity timer expiration triggers the transition of the device back to the inactive state [6]. As a result, the device consumes considerable energy [8] without transmitting the uplink data. To overcome this problem, in this paper, we consider the impact of device transition among RRC states and propose a state aware resource allocation approach, assigning resources while devices are in their current inactive or connected state to avoid the redundant resume procedure completion for devices that do not get resources during a period. We introduce a resource allocation mechanism, executed after message 3 of the resume procedure (shown in Fig. 2). The algorithm runs after message 3 since devices transmit the information about the number of packets stored in their buffer during this message. We develop a mathematical representation of the EE in a multi-RRC state scenario, focusing on user-based EE maximization while satisfying the transmit rate and power constraints. The type of EE fairness we consider in this work is max-min fairness. Due to the non-convexity of our problem, we utilize generalized fractional programming theory and Lagrangian dual decomposition to allocate the scarce subcarriers and power jointly. To verify the results, we use extensive realistic simulations from 5G channel gain measurements. We compare our approach with a baseline procedure, specified in 3GPP [6], where the resources are assigned only in the connected state after RRC Resume execution, concluding an improvement of the user's EE by up to 30% for simulations.

The remainder of this paper is structured as follows. Section II provides a comprehensive state-of-the-art overview of EE resource allocation. In Section III, we describe the system model and problem formulation. We elaborate on the iterative subcarrier assignment and power allocation in Section IV. The performance of the proposed approach is evaluated in Section V, whereas the practical aspects are discussed in section VI. Finally, Section VII concludes the paper.

II. Related Work

The problem of energy efficiency in the wireless network has received substantial attention in the last couple of years due to the rapid increase of connected IoT devices. Relevant work regarding the power saving techniques in next-generation networks is found in [5]. From the wireless resource management perspective, resource allocation aims at increasing the overall EE of the OFDMA system [9], [10] or the EE of individual users [11], [12], [13]. The authors in [11] address the drawbacks of increasing the overall system EE instead of improving the individual users' EE, concluding that fairness is a necessary metric for the EE problem. From the fairness perspective, various optimization approaches ranging from logarithmic EE [14], weighted sum EE [11], to max-min EE [12], [13] are utilized for an efficient allocation of the wireless resources among users. From the mathematical perspective, the generalized fractional programming [15] theory and the Lagrangian dual decomposition [16] are applied to transform the non-convex problem and jointly allocate subcarriers and power iteratively to the users. While the former works are beneficial and efficiently allocate the scarce radio resources to the users, they do not consider different RRC states of devices where the transition among states has to be addressed carefully during the resource assignment procedure, which we do in this work.

Similar to our work, there is previous state of the art considering the energy-efficient resource allocation problem with a fairness guarantee. In [11], the weighted-sum approach and max-min approach are investigated using generalized fractional programming and Lagrangian dual decomposition. However, the allocation of the resources takes place only for RRC Connected users, not addressing the impact of transitioning the device in the connected state. Under minimum user throughput and power constraints, [12] investigates a max-min energy-efficient problem in uplink OFDMA systems, ensuring fairness among users. In [13], a similar problem is formulated as an EE maximization problem with fairness among users. The authors solve the problem using iterative joint subcarrier and power allocation algorithms or separate subcarrier and power allocation algorithms, which differ from algorithms proposed in [12]. Yet, they do not evaluate the energy efficiency deterioration caused by the unnecessary transitions of the device to the connected state to compete for the limited radio resources.

While the aforementioned works [9] - [13] can be sufficient for scenarios only with connected users, they might not be optimal for scenarios with a portion of users in an RRC Inactive state. Although data transmission consumes the majority of the energy, the RRC Resume procedure, presenting the transition of the devices from inactive to connected state, increases the overall energy consumption of the device, which is the reason we consider it in this paper.

III. System Model and Problem Formulation

We consider an OFDMA uplink scenario of a cellular network consisting of a single BS. The system is time-slotted, where the slot duration is defined by a transmission time interval (TTI) of 1 ms for 15 kHz subcarrier spacing. The BS serves a total of N users, which are in RRC Connected and RRC Inactive states at a given slot. Upcoming uplink data trigger the RRC Resume, and on the other hand, the expiration of data inactivity timer triggers the RRC Release [6]. The spectrum is divided into R physical resource blocks (PRBs), with a bandwidth of W. A PRB can be assigned to at most one user to avoid intra-cell interference. Therefore, we use a binary variable $\rho_{i,j}$, set to 1 to indicate the assignment of PRB j to user i, or 0 otherwise. Let $h_{i,j}$ denote the channel gain of user i on PRB j. The transmit power of user i on PRB j is $p_{i,j}$, whereas N_0 represents the background noise at user. The transmit rate of user i on PRB j (in units bits/s/Hz) is

$$r_{i,j} = \log_2\left(1 + \frac{p_{i,j}|h_{i,j}|^2}{N_0 W}\right) = \log_2\left(1 + p_{i,j} g_{i,j}\right), \quad (1)$$

where $g_{i,j} = |h_{i,j}|^2/(N_0 W)$. Therefore, the total transmit rate and the total transmit power of user i are calculated as

$$r_i(\rho, p) = \sum_{j=1}^{R} \rho_{i,j} r_{i,j}, \ \forall i, \quad (2)$$

$$p_i(\rho, p) = \sum_{j=1}^{R} \rho_{i,j} p_{i,j}, \ \forall i. \quad (3)$$

The inefficiency of the power amplifier $\xi_i \geq 1$, circuit power consumption $p_i^C \geq 0$ [12], and RRC Resume power consumption $p_i^T \geq 0$ [8], depicting the state awareness component, are used to calculate the total power consumption of user i

$$P_i(\rho, p) = \xi_i p_i(\rho, p) + p_i^C + p_i^T, \quad \forall i. \quad (4)$$

The total transmission rate and power consumption of the network are obtained by summing the transmission rate and power consumption of all the users, as follows

$$R_T(\rho, p) = \sum_{i=1}^{N} r_i(\rho, p), \quad (5)$$

$$P_T(\rho, p) = \sum_{i=1}^{N} P_i(\rho, p). \quad (6)$$

Finally, the EE of user i and the EE of the network are

$$\eta_i^{EE}(\rho, P) = \frac{r_i(\rho, p)}{P_i(\rho, p)}, \quad (7)$$

$$\eta_T^{EE}(\rho, P) = \frac{R_T(\rho, p)}{P_T(\rho, p)}. \quad (8)$$

A. Problem formulation

To achieve the EE max-min fairness among RRC state aware users, while always satisfying a minimum transmit rate and maximum power consumption constraints, we jointly optimize the radio resource assignment and power allocation across PRBs. The optimization problem is defined as

$$\max_{\rho,p} \min_{i} \quad \eta_i^{EE}(\rho, p) \quad (9)$$

$$\text{s.t.} \sum_{j=1}^{R} \rho_{i,j} r_{i,j} \geq r_i^{min}, \ \forall i \quad (9a)$$

$$\sum_{j=1}^{R} \rho_{i,j} p_{i,j} \leq p_i^{max}, \ \forall i \quad (9b)$$

$$p_{i,j} \geq 0, \forall i, \ \forall j \quad (9c)$$

$$\sum_{i=1}^{N} \rho_{i,j} \leq 1, \ \forall j \quad (9d)$$

$$\rho_{i,j} \in \{0, 1\}, \ \forall i, \ \forall j. \quad (9e)$$

The optimization problem (9) maximizes the EE of the worst-case user, i.e., provides max-min fairness. The parameters r_i^{min} and p_i^{max} are the minimum required transmit rate and the maximum transmission power for each user, respectively. Constraint (9a) ensures the minimum guaranteed rate for all users. Constraint (9b) guarantees that the transmit power does not exceed the maximum transmit power of each user, while (9c) captures the fact that the power cannot be negative. Constraints (9d) and (9e) guarantee the orthogonal allocation of resources.

IV. Radio resource Assignment and Power allocation algorithm

The optimization problem (9) is a non-convex mixed-integer nonlinear programming problem because of the nonlinear function in the objective, denoted as a fractional function of $\rho_{i,j}$ and $p_{i,j}$ and constraint (9b) indicated by the multiplication of $\rho_{i,j}$ by $p_{i,j}$ [17]. We use generalized fractional programming and Lagrangian optimization as efficient tools to solve this problem.

A. Problem transformation with fractional programming

We apply generalized fractional programming and Lagrangian dual decomposition to transform the problem into an equivalent convex problem [17] because the objective is defined as several fractional functions, defining users' energy efficiency. The algorithm that is most often used to solve the generalized fractional programming is the generalized Dinkelbach's algorithm [15], which converges to an optimal solution with limited complexity [17]. In [15], it is denoted that the sequence of fractional problems is solved with a linear convergence rate. Generalized Dinkelbach's algorithm reformulates the objective function to decouple the nominator $r_i(\rho, p) > 0$ and the denominator $P_i(\rho, p) > 0$, using an iterative updated variable η (shown in Algorithm 1), which results into

$$\eta_{EE}^{it} = \max_{\rho,p} \min_{i} \frac{r_i(\rho, p)}{P_i(\rho, p)} = \min_{i} \frac{r_i(\rho^{it}, p^{it})}{P_i(\rho^{it}, p^{it})}, \quad (10)$$

where subscript *it* represents the iteration index, ρ^{it} the radio resource assignment and p^{it} the power allocation for problem (9) at iteration *it*. After applying the generalized Dinkelbach's algorithm [15], the objective function transforms into

$$\begin{aligned} &\max_{\rho,p} \min_{i} \quad [r_i(\rho, p) - \eta P_i(\rho, p)] \\ &\text{s.t.} \quad \text{(9a), (9b), (9c), (9d), (9e).} \end{aligned} \quad (11)$$

The optimal solution of (9) represented by ρ^*, p^* and the value of η_{EE}^*, achieved from the optimal solution, are unknown in advance and defined by iteratively calculating the suboptimal radio resource assignment and power allocation until the optimal values are achieved, or the maximum number of iterations is reached. The procedure for achieving the optimal resource assignment and power allocation is defined in Algorithm 1 and explained in the following sections. Algorithm 1 contains two iteration loops which converge to the optimal solutions. The inner iteration loop updates iteratively the radio resource assignment, power allocation and Lagrange multipliers in order to calculate the optimal EE solutions. Meanwhile the outer loop updates iteratively the variable η until convergence of the algorithm is achieved or the iteration index it has reached the maximum value.

Since the objective function is formed by multiple users' fractional ratios of energy efficiency, a variable φ is introduced to smoothen the objective function [12]. It is defined that $\varphi \geq 0$ for $0 \leq \eta \leq \eta^*_{EE}$ and feasible values of ρ, p, while (11) is transformed as follows after incorporating φ

$$\begin{aligned} \max_{\rho,p,\varphi} \quad & \varphi \\ \text{s.t.} \quad & \text{(9a), (9b), (9c), (9d), (9e),} \\ & R_i(\rho,p) - \eta P_i(\rho,p) \geq \varphi, \ \forall i. \end{aligned} \tag{12}$$

B. Radio resource assignment and power allocation

The Lagrangian dual decomposition approach is applied to (12) to assign resources and allocate power [11]–[13]. Due to the nonconvex constraint (9b), the binary variable $\rho_{i,j} \in \{0,1\}$ is relaxed in the interval $[0,1]$. Consequently, the transmitted power is $p_{i,j} = 0$ if subcarriers are not allocated to users , i.e., $\rho_{i,j} = 0$. Otherwise, if $\rho_{i,j} = 1$, the transmitted power is $p_{i,j} \geq 0$. Accordingly, to relax the assignment process we use a variable $s_{i,j} = \rho_{i,j} p_{i,j}$, where $s_{i,j} \in S$ defines the power allocation for user i on PRB j. Meanwhile, the constrained optimization (12) is transformed into (13),

$$\max_{\rho,S,\varphi} \quad \varphi \tag{13}$$

$$s.t. \sum_{j=1}^{R} \rho_{i,j} \log_2\left(1 + \frac{s_{i,j} g_{i,j}}{\rho_{i,j}}\right) \geq r_i^{min}, \ \forall i \tag{13a}$$

$$\sum_{j=1}^{R} s_{i,j} \leq p_i^{max}, \ \forall i \tag{13b}$$

$$s_{i,j} \geq 0, \ \forall i, \ \forall j \tag{13c}$$

$$\sum_{i=1}^{N} \rho_{i,j} \leq 1, \ \forall j \tag{13d}$$

$$0 \leq \rho_{i,j} \leq 1, \ \forall i, \ \forall j \tag{13e}$$

$$\begin{aligned} & \sum_{j=1}^{R} \rho_{i,j} \log_2\left(1 + \frac{s_{i,j} g_{i,j}}{\rho_{i,j}}\right) - \\ & \eta \left(\xi_i \sum_{j=1}^{R} s_{i,j} + p_i^C + p_i^T\right) \geq \varphi, \ \forall i, \end{aligned} \tag{13f}$$

which is convex in S, ρ, p.

The Lagrangian function (14) and Lagrangian dual function (15) are defined in order to relax the problem. The Lagrangian function of our optimization problem is formed by augmenting the objective function with the constraint functions [16] using the Lagrange multipliers. We define these multipliers as vectors $\beta = (\beta_1, \beta_2 ... \beta_i) \succeq 0$, $\mu = (\mu_1, \mu_2 ... \mu_i) \succeq 0$, $\nu = (\nu_1, \nu_2 ... \nu_i) \succeq 0$, $\gamma = (\gamma_1, \gamma_2 ... \gamma_i) \succeq 0$ which are associated with constraints (13a), (13b), (13d) and (13f), respectively. Hence, we have the Lagrangian function defined as

$$\begin{aligned} \mathcal{L}(\rho,S,\varphi,\beta,\mu,\nu,\gamma) = \varphi + & \sum_{i=1}^{N} \beta_i \left[\sum_{j=1}^{R} \rho_{i,j} \log_2\left(1 + \frac{s_{i,j} g_{i,j}}{\rho_{i,j}}\right)\right. \\ \left. - r_i^{min}\right] + & \sum_{i=1}^{N} \mu_i \left(p_i^{max} - \sum_{j=1}^{R} s_{i,j}\right) + \sum_{j=1}^{R} \nu_j \left(1 - \sum_{j=1}^{R} \rho_{i,j}\right) \\ + \sum_{i=1}^{N} \gamma_i \Bigg\{ & \sum_{j=1}^{R} \left[\rho_{i,j} \log_2\left(1 + \frac{s_{i,j} g_{i,j}}{\rho_{i,j}}\right) - \eta \xi_i s_{i,j}\right] \\ - \eta(p_i^C + p_i^T) - \varphi & \Bigg\}. \end{aligned} \tag{14}$$

Moreover, the Lagrangian dual function used to find the Lagrange multipliers, is derived by the primal problem (13) and defined by the maximum value of Lagrangian function (14) with respect to ρ, s, φ for given β, μ, ν, γ as

$$D(\beta,\mu,\nu,\gamma) = \max_{\rho,S,\varphi} \mathcal{L}(\rho,S,\varphi,\beta,\mu,\nu,\gamma). \tag{15}$$

The optimal values of the problem are denoted as $\rho^*_{i,j}$, $p^*_{i,j}$ and φ^* for given values of Lagrange multipliers, assuming a zero duality gap. First, the optimal values are calculated from initial Lagrange multiplier values, and afterwards, the Lagrange multipliers are derived from the optimal values in an iterative process. The Karush-Kuhn-Tucker conditions [17] $\frac{\partial \mathcal{L}}{\partial s^*_{i,j}} = 0$ and $\frac{\partial \mathcal{L}}{\partial \rho^*_{i,j}} = 0$ are used to solve the optimization problem with the inequality constraints. To find the optimal power allocation and subcarrier assignment, we need to consider the Lagrangian function defined in (14) and solve the first derivative of the function with respect to s and ρ while setting them to zero as above. This yields the optimal power allocation for the given radio resource assignment ρ as

$$p^*_{i,j} = \frac{s^*_{i,j}}{\rho_{i,j}} = \left[\frac{\beta_i + \gamma_i}{(\mu_i + \eta \xi_i \gamma_i) \ln 2} - \frac{1}{g_{i,j}}\right]^+, \ \forall i, j, \tag{16}$$

where $[y]^+ = \max[0, y]$.

Since $\rho_{i,j}$ is a binary variable, the resource j is allocated to user i with the maximum positive partial derivative of the Lagrangian function with respect to $\rho_{i,j}$ [12], denoted as

$$\begin{aligned} \max_{1 \leq i \leq N} \quad (\beta_i + \gamma_i) \Bigg\{ & \left[\log_2\left(g_{i,j} \frac{\beta_i + \gamma_i}{(\mu_i + \eta \xi_i \gamma_i) \ln 2}\right)\right]^+ \\ & - \frac{1}{\ln 2} \left[1 - \frac{1}{g_{i,j} \frac{\beta_i + \gamma_j}{(\mu_i + \eta \xi_i \gamma_i) \ln 2}}\right]^+ \Bigg\}. \end{aligned} \tag{17}$$

To find the optimal φ^*, we derive the Lagrangian function (14) with respect to φ, apply the constraint (13f) and obtain

$$\begin{aligned} & \max \left(1 - \sum_{i=1}^{N} \gamma_i\right) \varphi \\ & \text{s.t.}\ 0 \leq \varphi \leq \sum_{j=1}^{R} \rho_{i,j} \log_2\left(1 + \frac{s_{i,j} g_{i,j}}{\rho_{i,j}}\right) \\ & \qquad - \eta \left(\xi_i \sum_{j=1}^{R} s_{i,j} + p_i^C + p_i^T\right). \end{aligned} \tag{18}$$

Since $\varphi \geq 0$, from (18) we obtain that the optimal value is $\varphi^* = 0$, if $\sum_{i=1}^{N} \gamma_k > 1$. Otherwise, the optimal value is

$$\min_k \left\{ \sum_{j=1}^{R} \rho_{i,j}^* \log_2\left(1 + \frac{s_{i,j}^* g_{i,j}}{\rho_{i,j}^*}\right) - \eta \left(\xi_i \sum_{j=1}^{R} s_{i,j}^* + p_i^C + p_i^T\right) \right\}. \tag{19}$$

Algorithm 1 Iterative Subcarrier Assignment and Power Allocation Algorithm

Initialization, set outer loop iteration index $it = 0$ and iteration energy efficiency $\eta_{EE}^{it} = 0$
Set the maximum outer iteration number it_{max} and outer loop error threshold ϵ
repeat
  Set inner loop iteration index $t = 0$, maximum inner iteration number t_{max}, inner loop error threshold δ, $\beta, \mu, \gamma, a, b, c$
  repeat
    Allocate transmit power $p_{i,j}^*$, assign subcarriers $\rho_{i,j}^*$
    Select φ^*
    Update β, μ, γ and $t = t + 1$
    if $\parallel \beta(t+1) - \beta(t) \parallel_2 < \delta$, $\parallel \mu(t+1) - \mu(t) \parallel_2 < \delta$ and $\parallel \gamma(t+1) - \gamma(t) \parallel_2 < \delta$ **then**
      break
    end if
  until $t > t_{max}$
  if $\left| \min_i \left[r_i(\rho^{it}, p^{it}) - \eta_{EE}^{it} P_i(\rho^{it}, p^{it}) \right] \right| < \epsilon$ **then**
    $\eta_{EE}^* = \min_i \frac{r_i(\rho^*, p^*)}{P_i(\rho^*, p^*)}$
    break
  else
    $\eta_{EE}^{it+1} = \min_i \frac{r_i(\rho^{it}, p^{it})}{P_i(\rho^{it}, p^{it})}$
    $it = it + 1$
  end if
until $it > it_{max}$

C. Lagrange multipliers update

The optimal power allocation and radio resource assignment depend on the optimal values of Lagrange multipliers. These values are derived iteratively from the dual optimization problem (20), which is always convex because the objective function is concave and the Lagrangian multipliers constraints are convex [16], as follows

$$\begin{aligned} \min \quad & D(\beta, \mu, \nu, \gamma) \\ \text{s.t.} \quad & \beta \succeq 0, \mu \succeq 0, \nu \succeq 0, \gamma \succeq 0. \end{aligned} \tag{20}$$

To solve the dual optimization problem, we take the derivative of the dual function with respect to Lagrangian multipliers and use the projected subgradient method [18] to minimize the convex function. The projected subgradient method starts from a random feasible initial point which is updated by a step of subgradient descent and then projects the solution on the constraint set. The procedure is repetitive until the convergence is achieved. The subgradients of $D(\beta, \mu, \nu, \gamma)$ are given by

$$\nabla \beta_i = \sum_{j=1}^{R} \rho_{i,j}^* \log_2 \left(1 + \frac{s_{i,j}^* g_{i,j}}{\rho_{i,j}^*}\right) - r_i^{min}, \ \forall i \tag{21}$$

$$\nabla \mu_i = p_i^{max} - \sum_{j=1}^{R} s_{i,j}^*, \ \forall i \tag{22}$$

$$\nabla \gamma_i = \sum_{j=1}^{R} \left[\rho_{i,j}^* \log_2 \left(1 + \frac{s_{i,j}^* g_{i,j}}{\rho_{i,j}^*}\right) - \eta \xi_i s_{i,j}^* \right] - \eta(p_i^C + p_i^T) - \varphi. \tag{23}$$

Let t denote the inner loop iteration index, where the radio resource assignment, power allocation and Lagrange multipliers are updated. Let a_t, b_t and c_t represent the projected subgradient method step size at iteration t [18], set to $\frac{0.01}{t}$ [12]. The projected subgradient method updates the Lagrange multipliers at iteration t based on the subgradient of dual function (21), (22), (23), as follows

$$\beta_i^{(t+1)} = [\beta_i^{(t)} - a_t \nabla \beta_i^{(t)}]^+, \tag{24}$$
$$\mu_i^{(t+1)} = [\mu_i^{(t)} - b_t \nabla \mu_i^{(t)}]^+, \tag{25}$$
$$\gamma_i^{(t+1)} = [\gamma_i^{(t)} - c_t \nabla \gamma_i^{(t)}]^+. \tag{26}$$

The updated Lagrange multipliers at iteration t are updated at Lagrangian function and used to calculate the radio resource assignment and power allocation at iteration $t + 1$, until convergence or maximum iteration index is achieved. The detailed procedure is described in Algorithm 1.

V. Numerical and Measurement Evaluation

Considering that energy efficiency is the focus of our work, we demonstrate why RRC state transition energy consumption needs to be included in the resource assignment to extend the battery life of reduced capability devices. Then we show the improvements of our RRC state aware resource allocation compared to the baseline approach. The latter is described in the 3GPP documentation [6] and allocates the resources only to the devices which are in RRC Connected state or have transitioned to the connected state. In that regard, the inactive users with uplink traffic should perform the RRC Resume procedure, shown in Fig. 2, before the BS determines the wireless resource allocation. Given that, cases, when the users execute the RRC Resume procedure but afterwards do not receive resources, are not excluded. Further, we compare the performance of the proposed and baseline method with the approach analyzed by the authors in [12], which considers all the users being in the RRC Connected state. Since it is not the focus of our work, we do not consider during our evaluations the energy consumed due to continuous listening of PDCCH channels in the connected state.

We apply the max-min fairness optimization to the considered methods to ensure a fair comparison of the performance. Even though the initial state of the users is identical for all methods, as RRC Connected, the subsequent state of each user is affected based on the applied resource allocation method. In that regard, the users in [12] are always in the RRC Connected state even though there is no uplink data to be transmitted at that time or network resources are not allocated. Meanwhile, for the RRC state aware and baseline approaches, the users' RRC state changes over time, depending on the allocation of resources, uplink traffic pattern, and the data inactivity timer. In the remainder of the paper, we refer to our proposed approach as *RRC state aware*, the operation described in 3GPP as the *baseline*, and the method in [12] as *RRC only method*.

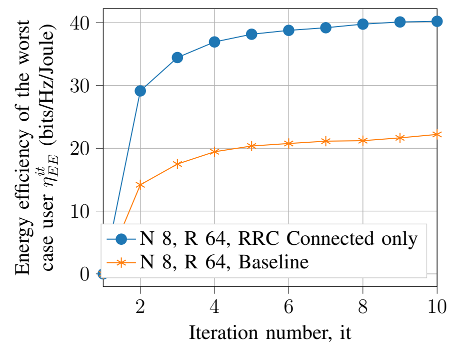
A. Simulation parameters

Our results evaluate the behavior of the RRC state aware resource allocation using data generated from simulations and measurements. While the simulated wireless channel is modeled as a frequency-selective channel using Clarke's flat fading model [12], the measurement datasets include time-varying channel gains generated for a mobile user during various periods of the day and on different days. These values are generated in our lab's 5G network, and the measurement data correspond to a bandwidth of 20 MHz, subcarrier spacing of 15 kHz, and a total number of PRBs of 106. The channel gain measurements, performed in a small room with low transmit power and varying distance between the device and BS, have small values in comparison with simulation, which are also reflected in the achieved EE. We vary the number of devices N and the number of subcarriers R to study the behavior of the methods in different scenarios. Table I summarizes the simulation parameters for channel gains achieved from the simulations and the measurements. The majority of the algorithm's parameters are the same for simulated and measurement channel gains scenarios. To test our algorithm,

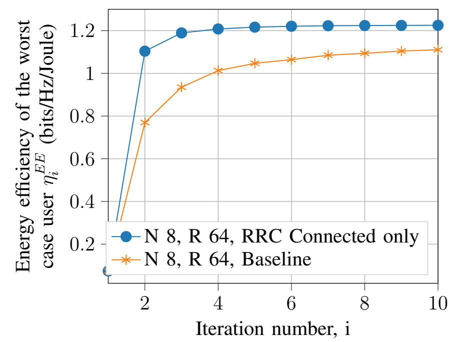
TABLE I: Simulation and measurement parameters

Parameters	Value
System bandwidth per BS	1 MHz (simulation) 20 MHz (measurement)
Subcarrier spacing	15 kHz
Average noise power	$N_0 = 1.1565 \times 10^{-8}$ W/Hz
Transmission Time Interval	1 ms
Device minimum data rate	$r_i^{min} = 15$ bits/s/Hz (simulation) $= 3.5$ bits/s/Hz (measurement)
Device maximum transmit power	$p_i^{max} = 0.2$ W
Inefficiency of power amplifier	$\xi_i = 18$
Circuit power consumption	$p_i^C = 0.4$ W
RRC Resume power consumption	$p_i^T = 2.5425$ W
Data inactivity timer	1 ms
Algorithm convergence error threshold	$\epsilon = 0.01$ $\delta = 0.01$
Number of BSs	1
Number of users	[8, 16]
Number of radio resources	[64, 128]
Simulated channel model	Frequency selective channel using Clarke's flat fading model
Measurement channel model	Channel gain values in 5G

we assume a minimal data inactivity timer value of 1 ms [6] to suspend the user connection when detecting a lack of allocated resources, even though users have uplink data to transmit. The generated results are averaged over time. We choose a different number of timeslots to average our results achieved from simulation and measurements to test the performance of the proposed algorithm and benefit from the diversity gains depending on the frequency-selective wireless channel. Accordingly, each of the figures from the simulated wireless channel is based on the averaged values of 50 timestamps. The results from the measured wireless channel are generated by averaging over 500 timestamps to include more diverse data in our algorithm. To guarantee a fair comparison among the RRC state aware, baseline and RRC Connected only methods, we use the seed concept in our simulation. Consequently, we assure the generation of the same channel characteristics, uplink traffic, and user initial state values.



(a) The scenario with simulation channel



(b) The scenario with measurement channel

Fig. 3: Comparison of the worst-case user's energy efficiency during 10 outer loop iterations, for RRC Connected-only and baseline approach for $N = 8$ users and $R = 64$ PRBs generated using the simulated and measurement channel.

B. Worst-case user EE performance

The initial results of our work illustrated in Fig. 3 identify the possibility of improving the users' EE by reducing the transitions among states. Intending to demonstrate the effect of connection resume, we compare the results from the max-min optimization applied to the RRC Connected only [12] and baseline method. The comparison is performed on the EE achieved from the worst-case user, based on simulated and measured channel gains. The x-axis represents the outer loop iterations index *it* used in Algorithm 1 to achieve the optimal users' EE (η_{EE}^*), or until the maximum iteration index is reached $it = it_{max}$. The results do not account for continuous PDCCH listening, which would significantly degrade the EE of the RRC Connected only case because the user listens to the PDCCH every TTI. The degradation is less for the baseline case since the user listens to the PDCCH for short periods before sleeping in the inactive state. Therefore, we consider the RRC Connected only as a corner case to show the maximum achievable EE concerning state changes, as illustrated in Fig. 3 because users do not perform

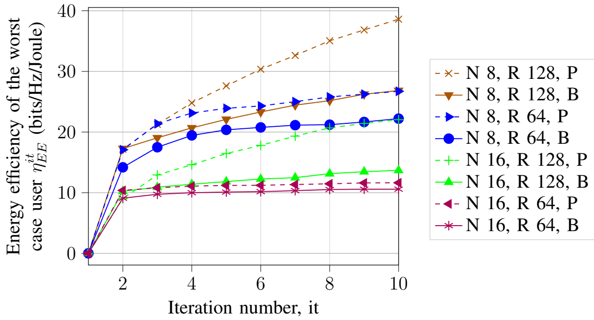
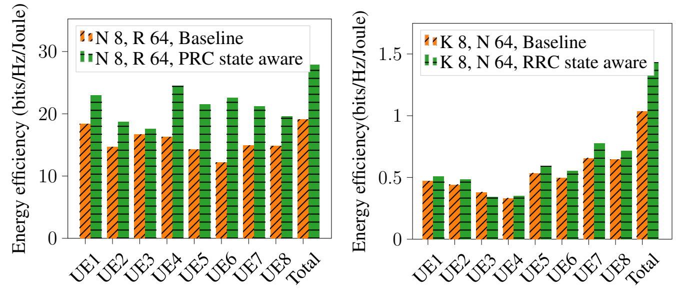


Fig. 4: Comparison of the worst-case user's energy efficiency during 10 outer loop iterations, for baseline (B) and RRC state aware (P) approaches under different N users and R PRBs for the simulated channel gains.



(a) Simulation channel (b) Measurement channel

Fig. 5: Comparison of the individual users' and network's energy efficiency for baseline and RRC state aware approaches for $N = 8$ users and $R = 64$ PRBs.

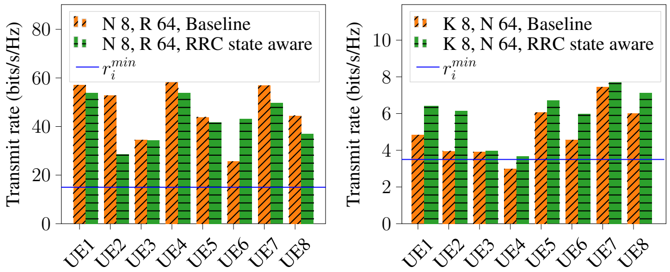
transitions. However, in reality for EE for low energy devices that we target here, the users' transition among RRC states is inevitable, so staying RRC Connected is not a solution for EE. Although the network, where the resource allocation algorithms run, knows that radio resources are not available, users have to transition to the connected state to compete for radio resources, causing the decrease in EE evaluated above. We would like to point out that the important conclusion for us from Fig. 3 is that reducing the number of executed resume procedures can improve the user's EE achieved in the baseline scenario by more than 40% in simulations and more than 15% in measurements, depending on the iteration number.

Therefore, we propose our RRC state aware mechanism to improve the energy efficiency of the baseline approach. In the following, we show numerical results comparing our proposed RRC state aware resource allocation algorithms with the baseline approach. Both consider multi-RRC state users in the scenarios. Fig. 4 illustrates the decrease of the RRC Resume procedure effects on EE when the allocation of resources is performed following the RRC state aware algorithm, which allocates the radio resources while devices are in the inactive or connected state. Given the lack of resource allocation for a device, the network suggests that the device should remain in the RRC Inactive state. As a result, the worst-case device's EE increases compared to the baseline scenario because the user does not consume the transition power among states. Due to space limitations, in Fig. 4 we show only the results pertaining to the simulated channel gains.

In order to demonstrate the effectiveness of our approach in different system configurations, we vary the number of users and radio resources. The results are depicted in Fig. 4, where the RRC state aware allocation of resources performs better for any number of users or radio resources in comparison with the baseline approach. Given a constant number of users, the increasing number of radio resources improves further the worst-case user EE of the RRC state aware algorithm related to the baseline. After users reach the minimum throughput requirement specified in Table I, the remaining radio resources are distributed based on the channel conditions. In that regard, the users with good channel conditions will receive most of the resources. In contrast, the users with bad channel conditions are kept in the inactive state without receiving resources. On the contrary, the EE improvement is smaller when the number of users increases, given a constant number of radio resources, because the data rate requirements of all users need to be satisfied, causing more frequent transitions even though the channel conditions are not good.

C. Energy efficiency improvement

In this subsection, we conduct a comparison between our proposed RRC state aware resource allocation with the baseline approach, focusing on the energy efficiency improvements for individual users and fairness. As before, we verify the performance of our approach with simulated data and more importantly, with the measurement data. We present the results of one system configuration, consisting of 8 users and 64 radio resources. Moreover, we add to the analysis the results corresponding to the total EE achieved from the network. Fig. 5 illustrates the EE achieved by individual users. Results demonstrate that the RRC Resume power consumption does not impact just the worst-case user EE but all users in the system due to the limited number of radio resources. Thus, the overall network's EE has an evident increase, supporting further the idea of allocating radio resources while the devices are in the current RRC state, such as RRC Inactive and RRC Connected state. As can be observed by Fig. 5, individual users have a balanced EE, where the difference between the worst-case user and best-case user EE is not considerable. As result, the EE fairness is maintained by the RRC state aware approach, in a similar way as the baseline approach. Our proposed approach increases users' EE by 5%- 30% on average compared to the baseline approach, when using the measurement or simulated data. The results illustrate an increase in measurement data, even though the channel gains are small. Further investigations are conducted to demonstrate the fulfilment of the minimum required transmit rate constraint by individual users, besides the increase of the users' EE. Fig. 6 shows that the achieved transmit rate with our approach surpasses the user's minimum required rate with a value of $r_{min} = 15$ bits/s/Hz, for the simulations and $r_{min} = 3.5$ bits/s/Hz, for the measurements. The transmit rate of some users decreases when applying the RRC state aware



(a) Simulation channel (b) Measurement channel

Fig. 6: Comparison of the individual users' transmit rate for baseline and RRC state aware approaches for $N = 8$ users, $R = 64$ PRBs, where $r_i^{min} = 15$ bits/s/Hz for simulated channel and $r_i^{min} = 3.5$ bits/s/Hz for measured channel gains.

resource allocation because the radio resources are assigned to maximize the EE instead of the throughput. This behaviour is described in [19]. We notice in Fig. 6a an increased transmit rate for user 6 to increase the achieved EE as indicated in Fig. 5. Meanwhile, in Fig. 6b, the RRC state aware not only achieves a higher transmit rate in comparison with the baseline approach, but fulfills the throughput requirement for user 4, for which the baseline approach fails. Therefore, we observe that the proposed energy-efficient and RRC state aware resource allocation approach increases users' and network EE while fulfilling the constraints of the optimization problem.

VI. Practical aspects

The proposed approach in this work can be adopted by the standardization for the reduced capability (RedCap) devices investigated by the 3GPP for the 5G network in Release 17 [3]. Redcap devices are defined as mid-range IoT devices with requirements between ultra-reliable low latency communications (URLLC) and massive machine-type communications (mMTC) features. The industrial wireless sensors, video surveillance in smart cities, and wearable devices are defined as use cases of the RedCap devices. While RedCap devices can serve a wide range of use cases with device lifetime requirements reaching up to 10 years for industrial sensors, they have a low device complexity, low cost, small size, and often run applications solely supported by batteries. Due to medium data rate requirements reaching up to 150 Mbps for downlink and 50 Mbps for the uplink scenario, the devices have to resume the connection and can not apply the small data transmission technique, which saves energy. At this point, our energy-efficient RRC state aware resource allocation mechanism is a useful power-saving technique for RedCap devices to fulfill the battery lifetime requirements.

VII. Conclusion

This work studies the energy-efficient resource allocation with fairness guarantees for an uplink OFDMA system. Considering the multiple device RRC states, we introduced the power consumption during state transition in the EE definition. While the state-of-the-art approaches perform well in the scenario with only RRC Connected users, our results demonstrate that the allocation is not optimal for systems with devices in multiple RRC states. We emphasize that the RRC state awareness of resource allocation algorithm avoids the RRC Resume procedure execution on battery-powered devices that lack sufficient resource allocation. We have addressed the problem by using the max-min optimization problem for fairness, generalized fractional programming, and Lagrangian dual decomposition for the iterative allocation of power and radio resources. In that regard, we formulate the optimization problem as individual users' EE maximization with QoS constraints. Our results demonstrate the improvement of users' EE when applying our proposed RRC state aware approach to simulations and measurement datasets. We plan to consider the impact of varying the system parameter in the achieved EE in the future.

References

[1] Y. Qian, "Spectrum efficiency and energy efficiency in wireless communication networks," *IEEE Wireless Communications*, 2020.
[2] G. Wikström *et al.*, "Challenges and technologies for 6G," in *2020 2nd 6G wireless summit (6G SUMMIT)*, IEEE, 2020.
[3] R. Ratasuk, N. Mangalvedhe, G. Lee, and D. Bhatoolaul, "Reduced capability devices for 5G IoT," in *Proc. of IEEE PIMRC*, 2021.
[4] Ericsson, "New SID on support of reduced capability NR devices," Technical Document (Tdoc) RP-193238, (3GPP), 12 2019.
[5] Y.-N. R. Li, M. Chen, J. Xu, L. Tian, and K. Huang, "Power saving techniques for 5G and beyond," *IEEE Access*, 2020.
[6] 3GPP, "NR; Radio Resource Control (RRC); Protocol specification," Technical Specification (TS) 36.331, (3GPP), 10 2018. Version 15.3.0.
[7] Z. Corporation, "Work Item on NR small data transmissions in INACTIVE state," Technical Document (Tdoc) RP-193252, (3GPP), 12 2019.
[8] A. Khlass, D. Laselva, and R. Jarvela, "On the flexible and performance-enhanced radio resource control for 5G NR networks," in *Proc. of IEEE VTC)*, pp. 1–6, 2019.
[9] X. Gao, K. Yang, J. Wu, Y. Zhang, and J. An, "Energy-efficient resource allocation and power control for downlink multi-cell OFDMA networks," in *Proc. of IEEE GLOBECOM*, 2017.
[10] Y. Zhang, J. An, K. Yang, X. Gao, and J. Wu, "Energy-efficient user scheduling and power control for multi-cell ofdma networks based on channel distribution information," *IEEE Transactions on Signal Processing*, vol. 66, no. 22, pp. 5848–5861, 2018.
[11] L. Xu, G. Yu, and Y. Jiang, "Energy-efficient resource allocation in single-cell ofdma systems: Multi-objective approach," *IEEE Transactions on Wireless Communications*, vol. 14, no. 10, pp. 5848–5858, 2015.
[12] Y. Li, M. Sheng, C. W. Tan, Y. Zhang, Y. Sun, X. Wang, Y. Shi, and J. Li, "Energy-efficient subcarrier assignment and power allocation in ofdma systems with max-min fairness guarantees," *IEEE Transactions on Communications*, vol. 63, no. 9, pp. 3183–3195, 2015.
[13] Z. Mohammadian, M. J. Dehghani, and M. Eslami, "Efficient resource allocation algorithms for high energy efficiency with fairness among users in ofdma networks," *Engineering Science and Technology, an International Journal*, vol. 23, no. 5, pp. 982–988, 2020.
[14] W. Jing, X. Wen, Z. Lu, Z. Hu, and T. Lei, "Proportional-fair energy-efficient radio resource allocation for ofdma smallcell networks," *Wireless Networks*, vol. 24, no. 3, pp. 695–707, 2018.
[15] J.-P. Crouzeix and J. A. Ferland, "Algorithms for generalized fractional programming," *Mathematical Programming*, 1991.
[16] B. G. Lee, D. Park, and H. Seo, *Wireless communications resource management*. John Wiley & Sons, 2009.
[17] A. Zappone and E. Jorswieck, "Energy efficiency in wireless networks via fractional programming theory," *Foundations and Trends in Communications and Information Theory*, vol. 11, no. 3-4, pp. 185–396, 2015.
[18] S. Boyd, L. Xiao, and A. Mutapcic, "Subgradient methods," *lecture notes of EE392o, Stanford University, Autumn Quarter*, 2003.
[19] I. P. Chochliouros *et al.*, "Energy efficiency concerns and trends in future 5G network infrastructures," *Energies*, vol. 14, no. 17, p. 5392, 2021.