Dynamic Non-Orthogonal Multiple Access (NOMA) and Orthogonal Multiple Access (OMA) in 5G Wireless Networks

Mina Baghani[*], Saeedeh Parsaeefard[†], Mahsa Derakhshani[‡], and Walid Saad [§]

[*] Faculty of Technical and Engineering, Imam Khomeini International University, Qazvin, Iran
[†] Communication Technologies & Department, ITRC, Tehran, Iran
[‡] Wolfson School of Mechanical, Electrical & Manufacturing Engineering, Loughborough University, UK
[§] Wireless@VT, Electrical and Computer Engineering Department, Virginia Tech, VA, USA

Abstract—In this paper, a novel dynamic multiple access technology selection among orthogonal multiple access (OMA) and non-orthogonal multiple access (NOMA) techniques is proposed. For this setup, a joint resource allocation problem is formulated in which a new set of access technology selection parameters along with power and subcarrier are allocated for each user based on each user's channel state information. Here, a novel utility function is defined to take into account the rate and costs of access technologies. This cost reflects both the complexity of performing successive interference cancellation and the complexity incurred to guarantee a desired bit error rate. This utility function can inherently capture the tradeoff between OMA and NOMA. Due to non-convexity of the proposed resource allocation problem, a successive convex approximation is developed in which a two-step iterative algorithm is applied. In the first step, called access technology selection, the problem is transformed into a linear integer programming problem, and then, in the second step, a nonconvex problem, referred to power allocation problem, is solved via the difference-of-convex-functions (DC) programming. Moreover, the closed-form solution for power allocation in the second step is derived. For diverse network performance criteria such as rate, simulation results show that the proposed new dynamic access technology selection outperforms single-technology OMA or NOMA multiple access solutions.

Index Terms—Orthogonal multiple access (OMA), non-orthogonal multiple access (NOMA), technology selection.

I. INTRODUCTION

The anticipated exponential growth in the demand for wireless access is expected to strain the capacity and coverage of existing wireless cellular networks [1]–[4]. In particular, the fixed multiple access techniques of yesteryears, such as time division multiple access (TDMA), code division multiple access (CDMA), and frequency division multiple access (FDMA), which guarantee the orthogonality in time, code and frequency, respectively, will no longer be able to sustain this growing demand for wireless access. In order to address this challenge in the fifth generation (5G) of cellular systems, several new techniques for multiple access have recently emerged based on the concept of non-orthogonal multiple access (NOMA) as discussed in [5]–[10]. In power-domain NOMA, multiple users can share each subcarrier and the diversity on that subcarrier is obtained by allocating different power levels to the users. The basic principle of NOMA is to exploit the difference in channel gains among users in order to offer multiplexing gains. For example, in a two-user

NOMA case, the lower power level is allocated to the user with higher channel gain (the first user) compared to the user with lower channel gain (the second user). Then, the information of different users is superimposed and transmitted.

Despite the proven benefits of NOMA [11]–[13], several practical challenges must be addressed before NOMA can be effectively deployed. One such challenge is to analyze the sensitivity of NOMA to the accuracy of channel state information (CSI) [14]. Another major challenge is the complexity of transceivers. Indeed, a typical NOMA transceiver requires the use of superposition coding and successive interference cancellation (SIC). Moreover, the performance of NOMA can be substantially limited when the difference in the channel gains of the involved users is not sufficiently significant. Clearly, these practical issues make it challenging to solely rely on NOMA, particularly, when the wireless users experience somewhat similar channel gains.

In particular, the performance of NOMA degrades when the difference in channel gains among the wireless users is small. Therefore, the need for a more complex receiver coupled with the higher probability of error imposed by SIC might limit the practicality of using NOMA in 5G under all network conditions. One promising approach to overcome this issue is to leverage the software-defined nature of 5G systems [15], in order to implement a dynamic approach for multiple access selection depending on the network state, e.g. CSI. This motivates the development of new multiple access solutions that can dynamically select between NOMA and OMA such as orthogonal frequency division multiple access (OFDMA) [16].

The main contribution of this paper is a new framework for multiple access technology selection that can enable a 5G system to flexibly decide on whether to use NOMA or OMA depending on the state of its users. This programmable structure can be implemented in practical systems by using the inherent software defined structure of 5G and beyond [15]. Although the rate of transmission is a very critical factor in the performance of a communication system, the cost of implementation complexity is another important factor in system design which has been ignored in prior works. The problem is formulated as an optimization problem whose objective captures the tradeoff between the achievable rate and a processing cost of using each access technology. Considering the processing cost of NOMA, the access technology selection between OMA and NOMA for each subcarrier is then used as an example of a dynamic, selective access scheme. As a

result, a new utility function is proposed and defined as the total rate of network minus the cost of performing NOMA for the allocated subcarriers.

A. Related Works

There exists a large body of works that addressed the resource allocation problems for OFDMA [17]–[20] and NOMA [21]–[25] techniques in 5G. In OFDMA setup, the proposed resource allocation problems are consisting of subcarrier and power allocation. Meanwhile, in NOMA, one needs to optimize user pairing along with power and sub-carrier allocation [22]. For both NOMA and OFDMA, network optimization is often posed using non-deterministic polynomial-time (NP)-hard problems [21], [26].

The authors in [24] studied the problem of analytically characterizing the optimal power allocation for NOMA, considering various objective functions and constraints. To reduce the computational complexity, a new user pairing and power allocation scheme is proposed in [25]. Meanwhile, the works in [27]–[29] have studied advanced approaches that combine NOMA with other emerging transmission techniques, such as full-duplex communications and multiple-input, multiple-output (MIMO) systems and heterogeneous systems.

This idea of utilizing a hybrid of OMA and NOMA in 5G has been recently studied in [30] and [31]. In [30], considering three different regions in the cell based on the distance, the access technology is chosen according to the region of users without any optimization and it is predetermined for each region. In particular, in [30], the access technology selection is not dynamically optimized taking into account the instantaneous CSI.

In [31], a heterogeneous network in which OMA and NOMA coexist, is considered. In such a network, four generic pairing methods for NOMA with a heuristic pairing cost function are studied. When those methods cannot achieve a suitable performance level OMA will be used for that subcarrier.

B. Motivation and Contribution

In this paper, in contrast to [31], we propose a joint resource allocation problem in which access technology selection, user pairing (for NOMA) or user selection (for OMA) and power allocation are jointly determined. Also, the requirements of different services in 5G are very diverse and such diverse requirements must be considered in the resource allocation [32]. As shown in [13] and [33], NOMA can achieve a higher data rate compared to OMA. However, the required processing of NOMA can be higher than OMA due to the need for SIC at the receiver. Thus, to address this tradeoff, we formulate these two conflicting design aspects together by defining a new utility function for resource allocation. In the defined utility, the imposed cost of extra processing of NOMA is subtracted from the achievable data rate. The cost of performing NOMA is modeled using two components, one representing the complexity of performing SIC in NOMA and the other one expressing the cost of employing complex designs to combat the error propagation. The complexity of

SIC receivers is a function of the number of users who share a single subcarrier. Since we consider NOMA with two users, this cost component will be constant. The second component of the cost is an increasing function of the requested bit error rate (BER) for SIC [34]. Note that the BER is inversely proportional to the experienced signal-to-interference-plus-noise-ratio (SINR). Thus, we model the second term of the NOMA cost as a logarithmic function of the inverse of SINR which is a concave increasing function. We define a new set of optimization variables for access technology selection for each subcarrier.

Furthermore, to enhance infrastructure utilization, we consider a virtualized network in which each service provider (SP) has its own quality-of-service (QoS) requirements, which should be guaranteed via effective resource allocation [35].

The formulated problem is then shown to be nonconvex and complex to solve. To address this challenge, we propose a two-step iterative algorithm. To this end, the variables of the resource allocation problem are divided into two groups and are optimized iteratively in two steps to alleviate the problem complexity [36]. The problem of allocating the first group of variables is transformed into a linear integer programming problem. To solve the power allocation in the second problem, DC programming is applied [37]. For the power allocation strategy, the closed-form solution is also derived. The obtained expression sheds light on the effects of the NOMA processing costs on the power allocation strategy. One of the the main technical challenges to develop the proposed dynamic OMA-NOMA scheme is the need for optimized technology access selection jointly with user pairing and power allocation which are required for conventional NOMA systems. Thus, resource allocation is more complex for dynamic access selection, compared to the classical NOMA case.

To study the performance of our proposed resource allocation strategy, three criteria are investigated in simulation section, i.e., rate, utility and outage probability. The outage probability is defined as the probability of not meeting constraints of optimization problem simultaneously. Simulation results show that the proposed dynamic multiple access selection approach yields significant performance gains in terms of achievable rate, utility value, and outage probability compared to single-technology OMA or NOMA multiple access techniques. In particular, when the maximum power limitation or minimum required rate of SPs is a dominant constraint of feasibility, our proposed scheme achieves 20% higher utility performance compared to the cases in which pure OMA and NOMA are adopted.

The rest of this paper is organized as follows. First, Section II describes the system model and problem formulation. Section III introduces the proposed resource allocation approach. Simulation results and analysis are presented in IV. Finally, Section V concludes this paper.

II. System Model and Problem Formulation

Consider a cellular network with a single base station (BS) that services a set $\mathcal{K}$ of K users in its own, specific coverage area. In this system, a set $\mathcal{S}$ of S different SPs, provide service

to their users sharing the BS. Each user belongs to one SP. Hence, we define $\mathcal{K}_s \subset \mathcal{K}$ as the subset of K_s users subscribed to SP $s \in \mathcal{S}$. The purpose of introducing multiple SPs is to enable service customization as each SP has a minimum rate requirement for its own subscribed users.

The total available bandwidth is partitioned into a set $\mathcal{N}$ of N subcarriers. We assume that the BS can switch between two access technologies (i.e., OMA and NOMA) for each subcarrier. β_n represents the access technology selection parameter and is defined as follows:

$$\beta_n = \begin{cases} 0, & \text{if OMA is selected for subcarrier } n, \\ 1, & \text{if NOMA is selected for subcarrier } n. \end{cases} \quad (1)$$

Here, we consider that only two users can share a single subcarrier in NOMA [14]. Let $h_{k,n}$ be the wireless channel gain from BS to the user k on subcarrier n, and σ^2 be the additive white Gaussian noise (AWGN) variance. Therefore, the downlink rate for a transmission over subcarrier $n \in \mathcal{N}$ will be given by:

$$R_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{P}) = \sum_{k \in \mathcal{K}} \alpha_{k,k,n} \left[\log(1 + \frac{p_{k,n} h_{k,n}}{\sigma^2}) \right. \quad (2)$$
$$\left. + \beta_n \sum_{\substack{k_2 \in \mathcal{K} \\ k_2 \neq k}} \log(1 + \frac{p_{k_2,n} h_{k_2,n}}{p_{k,n} h_{k_2,n} + \sigma^2}) \alpha_{k,k_2,n} \right],$$

where the first term captures the rate of a user in OMA or the first user in NOMA, while the second term represents the rate of the second user that accounts for the interference of the first user in the NOMA case. When $\beta_n = 0$, the second term is equal to zero and OMA is selected. In (2), $p_{k,n}$ represents the power allocated by the BS to user k over subcarrier n, and $\alpha_{k,k_2,n}$ is a binary variable given by:

$$\alpha_{k,k_2,n} = \begin{cases} 1, & \text{if } k \text{ and } k_2 \text{ are the first and second users} \\ & \text{of NOMA, respectively, on subcarrier } n, \\ & \quad \forall k, k_2, \; k \neq k_2, \\ 1, & \text{if } k_2 \text{ is the first user of NOMA or user} \\ & \text{of OMA on subcarrier } n, \; \forall k, k_2, \; k = k_2, \\ 0 & \text{otherwise.} \end{cases}$$

The three-dimensional matrix $\boldsymbol{\alpha} = [\alpha_{k,k_2,n}]$, the vector $\boldsymbol{\beta} = [\beta_n]$, and the two-dimensional matrix $\boldsymbol{P} = [p_{k,n}]$ are optimization variables that capture the problem of resource allocation in our system.

The achievable rate of NOMA is higher than OMA [13], [33]. However, since additional processing is required at the receiver for SIC when NOMA is selected, this additional complexity can be considered as a cost of using this access technology. Such cost calls for optimizing the multiple access technology that is chosen by the wireless network at a given time. In recent works, such as [38], [39], and [32], multi-objective optimization is used to optimize multiple, conflicting objective functions. To do so, scalarization, which consists of creating a linear combination of the different objectives, has emerged as a very popular technique to deal with multi-objective optimization [40]. Thus, we use scalarization to consider two conflicting design criteria which are the achievable rate and the implementation complexity cost by defining our

utility as a weighted subtraction of these two objectives. Note that weighting is needed in this combination to normalize the units of the two objectives and to assign them the desired priority levels. Thus, we define the utility which incorporates the processing cost of access technology for each subcarrier n, as follows:

$$U_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{P}) = R_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{P}) - w\beta_n F_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{P}), \; \forall n \in \mathcal{N}, \quad (3)$$

where w is a normalizing factor to harmonize the cost and rate functions to the desired priority levels of these two parameters in designing the system and $F_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{P})$ represents the total processing cost of NOMA.

For the case of NOMA, SIC receivers are required and this will incur an extra cost to the system as SIC receivers are computationally more complex. Depending on the structure of NOMA as well as some practical considerations, in order to define $F_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{P})$, we focus on two major components that contribute to the NOMA complexity:

- *SIC processing:* Primarily, the complexity of SIC receivers is a function of the number of code layers in superposition coding scheme [41]. In our setup, we assume that a maximum of two users can share one subcarrier in NOMA, to keep the SIC complexity low, which is a common assumption in the NOMA literature, e.g., see [42] and [23]. Consequently, two layers of superposition coding schemes will be used in this setup which leads to a constant cost in $F_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{P})$ which is represented by parameter A' hereinafter.

- *Error propagation:* SIC receivers suffer from propagation errors. To alleviate such errors, more complex designs have been proposed such as multiple decision aided SIC receivers [34]. In [34], it is shown that the required complexity is an increasing function of the requested BER for SIC. On the other hand, hybrid automatic repeat request (HARQ) is necessary to have reliable communications when there is an error in signal detection. Compared to OMA, NOMA encounters higher HARQ probability and its HARQ design is more challenging [43] as it requires more processing, increases complexity, and imposes extra cost. In traditional SIC receivers, the achievable BER corresponding to the detection of a strong signal in presence of interference from a weaker signal, is inversely proportional to the experienced SINR. Therefore, to capture these effects, our considered NOMA cost will also include a component that is an increasing function of the inverse SINR.

Therefore, we propose a NOMA cost function that properly captures these two components as follows:

$$F_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{P}) = A' + \quad (4)$$
$$\sum_{k \in \mathcal{K}} \sum_{k_2 \in \mathcal{K}, k_2 \neq k} \left[\alpha_{k,k_2,n} \alpha_{k,k,n} G\left(\frac{h_{k,n} p_{k_2,n} + \sigma^2}{h_{k,n} p_{k,n}} \right) \right],$$

where A' denotes the constant cost of considering two users per subcarrier in our setup and $G(.)$ is an increasing function of the inverse SINR. $G(\cdot)$ represents the additional complexity

required to alleviate the error propagation to a desired level. By increasing the difference between the power of the desired signal and the interference induced from other users over the same subcarrier, proper interference cancellation can be performed with lower complexity. For tractability of the analysis, we consider a concave increasing logarithmic function for $G(.)$ such that $G(v) = V' \log(v)$ where V' is a positive scalar. This is a reasonable choice in terms of tractability for optimization purposes. Moreover, since the relation between BER and SINR generally follows the complementary error function, the logarithmic function of the inverse of the SINR is a suitable candidate with similar characteristics. Considering (3) and (4), the final utility function is

$$U_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{P}) = R_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{P}) - \beta_n \Big(A \tag{5}$$
$$+ \sum_{k \in \mathcal{K}} \sum_{k_2 \in \mathcal{K}, k_2 \neq k} \left[\alpha_{k,k_2,n} \alpha_{k,k,n} V \log \left(\frac{h_{k,n} p_{k_2,n} + \sigma^2}{h_{k,n} p_{k,n}} \right) \right] \Big),$$
$$\forall n \in \mathcal{N},$$

where $V = V'w$ and $A = A'w$ hereinafter.

Considering virtualization at the radio access unit and assuming that multiple SPs share the BS, the isolation between SPs should be provided for resource allocation purposes [44]. This isolation requirement is represented as a rate requirement for each SP as follows [45],

$$\sum_{k \in \mathcal{K}_s} \sum_{n \in \mathcal{N}} \left[\alpha_{k,k,n} \log(1 + \frac{p_{k,n} h_{k,n}}{\sigma^2}) + \beta_n \right. \tag{6}$$
$$\sum_{\substack{k_1 \in \mathcal{K} \\ k_1 \neq k}} \log(1 + \frac{p_{k,n} h_{k,n}}{p_{k_1,n} h_{k,n} + \sigma^2}) \alpha_{k_1,k,n} \alpha_{k_1,k_1,n} \Big] > R_s, \forall s \in \mathcal{S},$$

where R_s is the minimum required rate for SP $s \in \mathcal{S}$. In (6), the second term belongs to the case in which user k is the second user of NOMA and shares the subcarrier n with user k_1.

If subcarrier $n \in \mathcal{N}$ is selected for NOMA transmission, SIC should be applied at the receiver side. For performing SIC, the difference in power levels of NOMA users should be larger than a specific lower bound [46]. This implementation constraint can be written as follows

$$\beta_n \sum_{k \in \mathcal{K}} \frac{h_{k,n}}{\sigma^2} \alpha_{k,k,n} \Big(\sum_{\substack{k_2 \in \mathcal{K} \\ k_2 \neq k}} p_{k_2,n} \alpha_{k,k_2,n} - p_{k,n} \Big) > \beta_n P_{\mathrm{d}}, \tag{7}$$
$$\forall n \in \mathcal{N},$$

where P_{d} represents the minimum required difference of the received power levels between two NOMA users.

Due to the fact that, in NOMA, the second user should have lower CSI compared to the first user, we should have $\alpha_{k,k_2,n} = 0$ when $h_{k_2,n} > h_{k,n}$, which can be represented as

$$\left(\frac{h_{k_2,n}}{h_{k,n}} \right) \alpha_{k,k_2,n} \leq 1, \quad \forall k, k_2, k \neq k_2 \in \mathcal{K}, n \in \mathcal{N}. \tag{8}$$

Moreover, since only one user could be selected as the second user in NOMA, we need to have,

$$\sum_{k \in \mathcal{K}} \sum_{k_2 \in \mathcal{K}, k_2 \neq k} \alpha_{k,k_2,n} \leq 1, \quad \forall n \in \mathcal{N}. \tag{9}$$

When a given user is selected as the second user, the associated subcarrier should use NOMA. Moreover, only one user should be chosen as the second user, which imposes the following constraint

$$\beta_n = \sum_{k \in \mathcal{K}} \sum_{k_2 \in \mathcal{K}, k_2 \neq k} \alpha_{k,k_2,n} \alpha_{k,k,n}, \quad \forall n \in \mathcal{N}. \tag{10}$$

Furthermore, the features of OMA and NOMA impose three additional constraints on the resource allocation problem. First, each subcarrier is assigned to only one user in OMA and only one user should be selected as the first user in NOMA. Therefore, we have,

$$\sum_{k \in \mathcal{K}} \alpha_{k,k,n} \leq 1, \quad \forall n \in \mathcal{N}. \tag{11}$$

When user k is not assigned to subcarrier $n \in \mathcal{N}$, its allocated power should be equal to zero. The mathematically expression of this practical consideration is

$$p_{k,n} - \sum_{k' \in \mathcal{K}} \alpha_{k',k,n} P_{\max} \leq 0, \quad \forall k \in \mathcal{K}, n \in \mathcal{N}. \tag{12}$$

Moreover, the transmit power limitation of BS is controlled by

$$\sum_{k \in \mathcal{K}} \sum_{n \in \mathcal{N}} \alpha_{k,k,n} (p_{k,n} + \sum_{\substack{k_2 \in \mathcal{K} \\ k_2 \neq k}} \alpha_{k,k_2,n} p_{k_2,n}) \leq P_{\max} \tag{13}$$

Finally, by using (3), we can pose the following dynamic multiple access technology selection problem:

$$\max_{\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{P}} \sum_{n \in \mathcal{N}} R_n - w \sum_{n \in \mathcal{N}} \beta_n F_n \tag{14}$$
$$\text{s.t. } (6) - (13),$$

where β_n and $\alpha_{k,k_2,n} \in \{0, 1\}$, and $p_{k,n} \geq 0$. Problem (14) is a mixed integer assignment programming problem whose objective function is not concave and, hence, it is an NP-hard optimization problem. To solve this problem, next, we propose an iterative resource allocation algorithm.

III. PROPOSED ALGORITHM AND PERFORMANCE ANALYSIS

To solve (14), we categorize the variables into two groups. The first group is discrete variables, including technology selection ($\beta_n, n \in \mathcal{N}$) and subcarrier allocation or user pairing for OMA or NOMA ($\alpha_{k,k_2,n}, k, k_2 \in \mathcal{K}, n \in \mathcal{N}$) parameters. The second group of variables, that includes the power levels $p_{k,n}, k \in \mathcal{K}, n \in \mathcal{N}$, is continuous. A two-step iterative algorithm is proposed to allocate the variables. In the first step, at iteration t, the values of $\alpha_{k,k_2,n}^t$ and β_n^t are optimized considering the previous optimal value of $p_{k,n}^{t-1}$ at iteration $(t-1)$. In the second step, $p_{k,n}^t$ is optimized for a fixed value of $\alpha_{k,k_2,n}^t$ and β_n^t obtained from the first step. The mathematical expression of total iterative optimization procedure is as

$$\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0 \to \boldsymbol{P}^0 \to \boldsymbol{\beta}^1, \boldsymbol{\alpha}^1, ..., \boldsymbol{\beta}^t, \boldsymbol{\alpha}^t \to \boldsymbol{P}^t \to \boldsymbol{\beta}^*, \boldsymbol{\alpha}^* \to \boldsymbol{P}^* \tag{15}$$

This iteration continues until the algorithm converges and the following conditions are held

$$\left\|\boldsymbol{\beta}^t - \boldsymbol{\beta}^{t-1}\right\| \leq \varepsilon_\beta, \left\|\boldsymbol{\alpha}^t - \boldsymbol{\alpha}^{t-1}\right\| \leq \varepsilon_\alpha, \left\|\boldsymbol{P}^t - \boldsymbol{P}^{t-1}\right\| \leq \varepsilon_p, \tag{16}$$

where $0 < \varepsilon_\beta, \varepsilon_\alpha, \varepsilon_p \ll 1$.

A. Technology Selection and Subcarrier Assignment Problem

The optimization problem at iteration t for the fixed value of $p_{k,n}^{t-1}$ from the previous iteration is

$$\max_{\boldsymbol{\beta},\boldsymbol{\alpha}} \sum_{n\in\mathcal{N}} \sum_{k\in\mathcal{K}} \alpha_{k,k,n}\left(Y_1(p_{k,n}^{t-1}) + \right. \tag{17}$$

$$\beta_n \sum_{k\in\mathcal{K}, k_2 \neq k} Y_2(p_{k,n}^{t-1}, p_{k_2,n}^{t-1})\alpha_{k,k_2,n}\Big) - \sum_{n\in\mathcal{N}} \beta_n\Big(A +$$

$$V \sum_{k\in\mathcal{K}} \sum_{k_2\in\mathcal{K}, k_2\neq k} \alpha_{k,k_2,n}\alpha_{k,k,n}Y_3(p_{k,n}^{t-1}, p_{k_2,n}^{t-1})\Big),$$

$$\text{s.t. } (6), (8) - (11),$$

where $Y_1(p_{k,n}^{t-1}) = \log\left(1 + \frac{p_{k,n}^{t-1}h_{k,n}}{\sigma^2}\right)$ and $Y_2(p_{k,n}^{t-1}, p_{k_2,n}^{t-1}) = \log\left(1 + \frac{p_{k_2,n}^{t-1}h_{k_2,n}}{p_{k,n}^{t-1}h_{k_2,n}+\sigma^2}\right)$ are the rate of users according to the optimal allocated power in the previous iteration. Also, $Y_3(p_{k,n}^{t-1}, p_{k_2,n}^{t-1}) = \log(\frac{p_{k_2,n}^{t-1}h_{k,n}+\sigma^2}{h_{k,n}p_{k,n}^{t-1}})$ represents the second part of the processing cost model which is a function of the optimal allocated power derived in the second step of optimization problem. Since Y_1, Y_2, and Y_3 only depend on the power allocation, they are assumed to be constant in the first step of resource allocation, i.e., technology selection and subcarrier assignment problem.

To convert this problem into linear optimization, we use an auxiliary variable $u_{k,k_2,n} = \beta_n\alpha_{k,k,n}\alpha_{k,k_2,n}$. Note that, the defined variable $u_{k,k_2,n}$ depends on $\alpha_{k,k,n}$. As a result, a new constraint should be added in the optimization problem to prevent the unacceptable values for these dependent variables. In essence, $u_{k,k_2,n}$ cannot be equal to one when $\alpha_{k,k,n} = 0$. Therefore, one of the key additional constraints is to have $\alpha_{k,k,n} - u_{k,k_2,n} \geq 0$. Similarly, since $u_{k,k_2,n}$ cannot be equal to one when $\beta_n = 0$, the following constraint should be satisfied $\beta_n - u_{k,k_2,n} \geq 0$.

The optimization problem should be reformulated according to the new auxiliary variable. For this purpose, by multiplying both sides of (10) by β_n, we obtain:

$$\beta_n^2 = \sum_{k\in\mathcal{K}} \sum_{k_2\in\mathcal{K}} u_{k,k_2,n}. \tag{18}$$

Since the variable β_n is binary, we have $\beta_n^2 = \beta_n$. Therefore, (18) is used in (17). Also, $u_{k,k_2,n}$ is equal to zero when $\alpha_{k,k_2,n}$ is zero. Thus, $\alpha_{k,k_2,n}$ in (8) and (9) can be replaced by $u_{k,k_2,n}$. Finally, (17) is transformed into

$$\max_{\boldsymbol{\beta},\boldsymbol{u},\boldsymbol{\alpha}} \sum_{n\in\mathcal{N}} \sum_{k\in\mathcal{K}} \alpha_{k,k,n}Y_1(p_{k,n}^{t-1}) + \sum_{\substack{k\in\mathcal{K}\\k_2\neq k}} Y_2(p_{k,n}^{t-1}, p_{k_2,n}^{t-1})u_{k,k_2,n}$$

$$- \sum_{n\in\mathcal{N}} \left(\beta_n A + V \sum_{k\in\mathcal{K}} \sum_{k_2\in\mathcal{K}, k_2\neq k} u_{k,k_2,n}Y_3(p_{k,n}^{t-1}, p_{k_2,n}^{t-1})\right), \tag{19a}$$

$$\text{s.t. } \alpha_{k,k,n} - u_{k,k_2,n} \geq 0 \quad \forall k, k_2 \in \mathcal{K}, n \in \mathcal{N}, \tag{19b}$$

$$\beta_n - u_{k,k_2,n} \geq 0 \quad \forall k, k_2 \in \mathcal{K}, n \in \mathcal{N}, \tag{19c}$$

$$\sum_{n\in\mathcal{N}} \sum_{k\in\mathcal{K}_s} \alpha_{k,k,n} Y_1(p_{k,n}^{t-1}) +$$

$$\sum_{\substack{k_1\in\mathcal{K}\\k_1\neq k}} Y_2(p_{k_1,n}^{t-1}, p_{k,n}^{t-1})u_{k_1,k,n} > R_s, \forall s \in \mathcal{S}, \tag{19d}$$

$$\left(\frac{h_{k_2,n}}{h_{k,n}}\right) u_{k,k_2,n} \leq 1, \quad \forall k, k_2 \in \mathcal{K}, n \in \mathcal{N}, \tag{19e}$$

$$\sum_{k\in\mathcal{K}} \sum_{k_2\in\mathcal{K}, k_2\neq k} u_{k,k_2,n} \leq 1, \quad \forall n \in \mathcal{N}, \tag{19f}$$

$$\beta_n = \sum_{k\in\mathcal{K}} \sum_{k_2\in\mathcal{K}, k_2\neq k} u_{k,k_2,n}, \tag{19g}$$

$$\sum_{k\in\mathcal{K}} \alpha_{k,k,n} \leq 1, \quad \forall n \in \mathcal{N}. \tag{19h}$$

This problem is a linear integer programming. There are different approaches to solve linear optimization, among them, the interior-point method has gained much more attention due to its simplicity. Here the linear integer programming problem of (19) is solved using CVX [47], which uses the interior-point method.

B. Power Allocation Problem

In Step 2, given to the best access selection and subcarrier allocation for users derived in Step 1 (β_n^t and $\alpha_{k,k,n}^t$, $u_{k,k_2,n}^t$), the BS should decide on its power allocation across subcarriers. In fact, by substituting the derived values for the optimization variables in (14) and omitting the constraints which do not depend on the power allocation, the optimization problem for power allocation becomes

$$\max_{\boldsymbol{P}} \sum_{n\in\mathcal{N}} \sum_{k\in\mathcal{K}} \alpha_{k,k,n}^t \log(1 + \frac{p_{k,n}h_{k,n}}{\sigma^2}) \tag{20}$$

$$+ \sum_{n\in\mathcal{N}} \sum_{k\in\mathcal{K}} \sum_{\substack{k\in\mathcal{K}\\k_2\neq k}} u_{k,k_2,n}^t \big(\log(p_{k,n}h_{k_2,n} + \sigma^2 + p_{k_2,n}h_{k_2,n})$$

$$- \log(p_{k,n}h_{k_2,n} + \sigma^2)\big) - V \sum_{n\in\mathcal{N}} \sum_{k\in\mathcal{K}} \sum_{\substack{k\in\mathcal{K}\\k_2\neq k}} u_{k,k_2,n}^t$$

$$\big(\log(p_{k_2,n}h_{k,n} + N_o) - \log(p_{k,n}h_{k,n})\big)$$

$$\text{s.t. } (6), (7), (12), (13)$$

where $\boldsymbol{P}$ is a $K \times N$ matrix in which each element at row k and column n is equal to $p_{k,n}$.

we can approximate the negative logarithmic terms in the objective function at iteration t_2 with affine functions as $f(x) = f(x_{t_2-1}) + f'(x_{t_2-1})(x - x_{t_2-1})$ where x_{t_2-1} is the optimal solution of $t_2 - 1$ iteration.

By applying the DC algorithm [37], we can approximate the negative logarithmic terms in the objective function at iteration t_2 with affine functions as $f(x) = f(x_{t_2-1}) + f'(x_{t_2-1})(x - x_{t_2-1})$ where x_{t_2-1} is the optimal solution of $t_2 - 1$ iteration. Therefore,

$$
\begin{aligned}
J(\boldsymbol{P}) = & \sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{K}}\alpha_{k,k,n}^t \log(1+\frac{p_{k,n}h_{k,n}}{\sigma^2}) \\
& + \sum_{\substack{k\in\mathcal{K}\\k_2\neq k}} u_{k,k_2,n}\Big(\log(p_{k,n}h_{k_2,n}+\sigma^2+p_{k_2,n}h_{k_2,n}) \\
& -\big(\log(p_{k,n}^{t_2-1}h_{k2,n}+\sigma^2)+\frac{(p_{k,n}-p_{k,n}^{t_2-1})h_{k_2,n}}{p_{k,n}^{t_2-1}h_{k_2,n}+\sigma^2}\big)\Big) \\
& +V\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{K}}\sum_{k_2\in\mathcal{K},k_2\neq k} u_{k,k_2,n}\Big(-\big(\log(p_{k_2,n}^{t_2-1}h_{k,n}+\sigma^2) \\
& +(p_{k_2,n}-p_{k_2,n}^{t_2-1})\frac{h_{k,n}}{p_{k_2,n}^{t_2-1}h_{k,n}+\sigma^2}\big)+\log(p_{k,n}h_{k,n})\Big). \quad (21)
\end{aligned}
$$

The same approach is used for constraint (6)

$$
\begin{aligned}
& \sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{K}_S}\Bigg(\alpha_{k,k,n}^t\log(1+\frac{p_{k,n}h_{k,n}}{\sigma^2})+ \\
& \sum_{\substack{k_1\in\mathcal{K}\\k_1\neq k}} u_{k_1,k,n}\Big(\log(p_{k_1,n}h_{k,n}+\sigma^2+p_{k,n}h_{k,n}) \\
& -(\log(p_{k_1,n}^{t_2-1}+\sigma^2)+\frac{(p_{k_1,n}-p_{k_1,n}^{t_2-1})h_{k,n}}{p_{k_1,n}^{t_2-1}h_{k,n}+\sigma^2})\Big)\Bigg)> R_s, \forall s\in\mathcal{S}. \\
& \hspace{10cm}(22)
\end{aligned}
$$

Finally, the transformed power allocation optimization problem to the convex one, is

$$
\max_{\boldsymbol{P}} J(\boldsymbol{P}) \quad (23)
$$
$$
\text{s.t.} \quad (22),(7),(12),(13).
$$

We write the Lagrange function of the convex problem (23) by considering the Lagrange multipliers $\boldsymbol{\lambda}$, $\boldsymbol{\gamma}$, $\boldsymbol{\zeta}$, and $\boldsymbol{\eta}$ for constraints in (22), (7), (12), and (13), respectively. Thus, the optimization problem is encapsulated in single term as

$$
\begin{aligned}
L(\boldsymbol{P},\boldsymbol{\lambda},\boldsymbol{\gamma},\boldsymbol{\zeta},\boldsymbol{\eta}) = & \sum_{k}\sum_{n}\alpha_{k,k,n}^t\log(1+\frac{p_{k,n}h_{k,n}}{\sigma^2}) \\
& + \sum_{k_2,k_2\neq k} u_{k,k_2,n}^t\Big(\log(p_{k,n}h_{k_2,n}+p_{k_2,n}h_{k_2,n}+\sigma^2) \\
& -\frac{h_{k_2,n}}{p_{k,n}^{t_2-1}h_{k_2,n}+\sigma^2}(p_{k,n}-p_{k,n}^{t_2-1})-\log(p_{k,n}^{t_2-1}h_{k_2,n}+\sigma^2)\Big) \\
& +V\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{K}}\sum_{k_2\in\mathcal{K},k_2\neq k} u_{k,k_2,n}^t\Big(-\big(\log(p_{k_2,n}^{t_2-1}h_{k,n}+\sigma^2)
\end{aligned}
$$

$$
\begin{aligned}
& +\frac{(p_{k_2,n}-p_{k_2,n}^{t_2-1})h_{k,n}}{p_{k_2,n}^{t_2-1}h_{k,n}+\sigma^2}\big)+\log(p_{k,n}h_{k,n})\Big) \\
& +\sum_{s\in\mathcal{S}}\lambda_s\Big(\sum_{k\in\mathcal{K}s}\sum_n\alpha_{k,k,n}^t\log(1+\frac{p_{k,n}h_{k,n}}{\sigma^2}) \\
& +\sum_{k_1,k_1\neq k} u_{k_1,k,n}^t\Big(\log(p_{k,n}h_{k,n}+p_{k_1,n}h_{k,n}+\sigma^2) \\
& -\log(p_{k_1,n}^{t_2-1}h_{k,n}+\sigma^2)-\frac{h_{k,n}}{p_{k_1,n}^{t_2-1}h_{k,n}}(p_{k_1,n}-p_{k_1,n}^{t_2-1})\big)-R_s\Big) \\
& +\sum_{k_1,k_1\neq k} u_{k_1,k,n}^t\Big(\log(p_{k,n}h_{k,n}+p_{k_1,n}h_{k,n}+\sigma^2) \\
& -\log(p_{k_1,n}^{t_2-1}h_{k,n}+\sigma^2)-\frac{h_{k,n}}{p_{k_1,n}^{t_2-1}h_{k,n}}(p_{k_1,n}-p_{k_1,n}^{t_2-1})\big)-R_s\Big) \\
& +\sum_n\gamma_n\Big(\sum_k\frac{h_{k,n}}{\sigma^2}\big(\sum_{k_2,k_2\neq k}p_{k_2,n}u_{k,k_2,n}^t-\beta_n^t p_{k,n}\alpha_{k,k,n}^t\big)-\beta_n^t p_d\Big) \\
& -\sum_{k_2}\sum_n\zeta_{k_2,n}(p_{k_2,n}-\sum_k u_{k,k_2,n}^t P_{\max})-\eta\Big(\sum_k\sum_n\alpha_{k,k,n}^t p_{k,n} \\
& +\sum_{k_2,k_2\neq k} u_{k,k_2,n}^t p_{k_2,n}-P_{\max}\Big). \quad (24)
\end{aligned}
$$

In this method, the primal (maximizing $L(\boldsymbol{P})$ to find $\boldsymbol{P}$) and dual (minimizing $L(\boldsymbol{\lambda},\boldsymbol{\gamma},\boldsymbol{\zeta},\boldsymbol{\eta})$ to find Lagrange multipliers) problems are solved iteratively until changes in the variables are negligible and the iterative algorithm converges to a fixed point. Since the optimization problem (23) is convex, this fixed point will be the optimal.

Next, in Proposition 1, for fixed access technology selections, we derive closed-form expressions for the optimal power allocation strategy as a function of the NOMA processing cost coefficient (V), channel gains, and Lagrange multipliers. The Lagrangian multipliers (dual variables) are obtained by solving the dual problem using the gradient method.

Proposition 1. *Given the access technology selection for subcarrier n, the optimal power allocation will be given by:*

* *If $\beta_n = 0$ and $\alpha_{k,k,n} = 1$,*

$$
p_{k,n} = \left[-\frac{1+\lambda_{s'}}{D_{k,n}}-\frac{\sigma^2}{h_{k,n}}\right]^+ \quad (25)
$$

where $D_{k,n} = -\zeta_{k,n}-\eta$.

* *If $\beta_n = 1$, $\alpha_{k_1,k_1,n} = 1$, and $\alpha_{k_1,k_2,n} = 1$,*

$$
p_n = \left[-\frac{1+\lambda_{s'}}{D_{k_2,n}}-\frac{\sigma^2}{h_{k_2,n}}\right]^+ \quad (26)
$$

and

$$
p_{k_1,n} = \frac{-(h_{k_1,n}(1+\lambda_{s'}+V)+\sigma^2 Q)}{2Qh_{k_1,n}} \quad (27)
$$
$$
\pm\frac{\sqrt{(h_{k_1,n}(1+\lambda_{s'}+V)+\sigma^2 Q)^2-4V\sigma^2 Qh_{k_1,n}}}{2Qh_{k_1,n}}
$$

where

$$
p_n = p_{k_1,n}+p_{k_2,n}, \forall k_1\in\mathcal{K}_{s'},\forall k_2\in\mathcal{K}_{s''}, \quad (28)
$$
$$
Q = \frac{-D_{k_2,n}(1+\lambda_{s''})}{1+\lambda_{s'}}+D_{k_1,n},
$$

$$D_{k_1,n} = -(1 + \lambda_{s''}) \frac{-h_{k_2,n}}{p_{k_1,n}^{t_2-1} h_{k_2,n} + \sigma^2} + \gamma_n \frac{h_{k_1,n}}{\sigma^2}$$
$$- \zeta_{k_1,n} - \eta, \; and$$
$$D_{k_2,n} = -V \frac{h_{k_2,n}}{p_{k_2,n}^{t_2-1} h_{k_2,n} + \sigma^2} + \gamma_n \frac{h_{k_1,n}}{\sigma^2} - \zeta_{k_2,n} - \eta.$$

Proof. See Appendix A. □

From (25), the allocated power to an OMA user is based on a pseudo water filling algorithm as $p_{k,n} = \left[-\frac{1+\lambda_{s'}}{D_{k,n}} - \frac{\sigma^2}{h_{k,n}} \right]^+$ where $D_{k,n} = -\zeta_{k,n} - \eta$. Similarly, in (26), the power allocation for NOMA transmission $p_n = \left[-\frac{1+\lambda_{s'}}{D_{k_2,n}} - \frac{\sigma^2}{h_{k_2,n}} \right]^+$ where $D_{k_2,n} = -V \frac{h_{k_2,n}}{p_{k_2,n}^{t_2-1} h_{k_2,n} + \sigma^2} + \gamma_n \frac{h_{k_1,n}}{\sigma^2} - \zeta_{k_2,n} - \eta$ is in a water filling format. For NOMA power allocation, the channel gain of the second NOMA user and the processing gain play key roles in the total power allocated to both NOMA users in one subcarrier.

We first compare the amount of power allocated to NOMA and OMA subcarriers depending on the NOMA processing cost. For instance, assume that NOMA and OMA are selected, respectively, for subcarriers n and n' during the technology selection phase. By comparing (25 and (26, we can observe that if $|D_{k_2,n}| < |D_{k,n'}|$, then more power is allocated to NOMA subcarriers compared to OMA subcarriers when $\frac{h_{k,n'}}{\sigma^2}, \frac{h_{k_2,n}}{\sigma^2} >> 1$. To satisfy the condition of $|D_{k_2,n}| < |D_{k,n'}|$, $\left(-V \frac{h_{k_2,n}}{p_{k_2,n}^{t_2-1} h_{k_2,n} + \sigma^2} + \gamma_n \frac{h_{k_1,n}}{\sigma^2} \right)$ should be positive or equivalently, $V < h_{k_1,n} \gamma_n \left(\frac{p_{k_2,n}^{t_2-1}}{\sigma^2} + \frac{1}{h_{k_2,n}} \right)$. In other words, when the NOMA processing cost V is lower than a certain threshold, higher power levels are allocated to NOMA subcarriers compared to OMA subcarriers. Note that since user k_2 and k are assigned to subcarriers n and n', respectively, the allocated power to these users will not be equal to zero which leads to $\zeta_{k,n} = \zeta_{k_2,n} = 0$.

Next, we consider the case in which the noise power is very low, i.e., $\sigma^2 \to 0$. In this case, $\frac{h_{k_i,n}}{\sigma^2}, \forall i \in \{1, 2\}$ tends to infinity and this leads to $D_{k_i,n} \to +\infty, \forall i \in \{1, 2\}$. Thus, the total allocated power to the NOMA subcarriers will be zero according to (26), since $-\frac{1+\lambda_{s'}}{D_{k_2,n}} - \frac{\sigma^2}{h_{k_2,n}}$ is a small negative value. On the other hand, in this case, the allocated power to an OMA user over subcarrier n is obtained by $\frac{1+\lambda_{s'}}{\zeta_{k,n}+\eta}$. Therefore, we can conclude that, whenever the noise power is very low, the power is allocated only to the OMA subcarriers and no power is allocated to the NOMA subcarriers.

We further consider the case when P_{max} is high, and hence, constraints (7), (12), and (22) are inactive, not causing any limitation on the objective functions. According to the Lagrangian multiplier method, when the constraints are satisfied by inequality, the associated Lagrange multipliers will be equal to zero. As a result, the Lagrange multipliers λ, γ and ζ are equal to zero. In this scenario, the allocated power to the OMA users will be $p_{k,n} = \left[\frac{1}{\eta} - \frac{\sigma^2}{h_{k,n}} \right]^+$ which is exactly the same as the water-filling algorithm. On the other hand, the allocated power to the NOMA users will be

$$p_n = \left[\frac{1}{\eta + V \frac{h_{k_2,n}}{p_{k_2,n}^{t_2-1} h_{k_2,n} + \sigma^2}} - \frac{\sigma^2}{h_{k_2,n}} \right]^+. \quad (29)$$

At a high SNR regime, when $\frac{p_{k_2,n}^{t_2-1}}{\sigma^2} \gg 1$, we have $p_n = \left[\frac{1}{\eta} - \frac{\sigma^2}{h_{k_2,n}} \right]^+$. This means the NOMA power allocation will follow the water filling algorithm depending on the channel gains of the second users. However, the allocated power is a decreasing function of V when $\frac{p_{k_2,n}^{t_2-1}}{\sigma^2} \ll 1$ as $p_n = \left[\frac{1}{V \frac{h_{k_2,n}}{\sigma^2}+\eta} - \frac{\sigma^2}{h_{k_2,n}} \right]^+$.

Finally, to study the effects of V on the power allocation of OMA and NOMA users, we assume the Lagrange multipliers to be fixed. By increasing V, the absolute value of $D_{k_2,n}$ increases which leads to a lower value for the first term in (26). As a result, the total power allocated to the NOMA users decreases. Considering a fixed total power P_{max} for the BS, decreasing the total power allocated to the NOMA users leads to a higher allocated power to the OMA users. In other words, Proposition 1 shows that the optimal power allocated to OMA increases (thus the allocated power to NOMA decreases) when the cost of NOMA grows.

In a nutshell, the following engineering insights can be observed from **Proposition 1**:

- If $V < h_{k_1,n} \gamma_n \left(\frac{p_{k_2,n}^{t_2-1}}{\sigma^2} + \frac{1}{h_{k_2,n}} \right)$ for all subcarriers allocated to NOMA, the allocated power to the NOMA subcarriers is larger than the allocated power to the OMA subcarriers when channel gains normalized by noise power are sufficiently high.
- When noise power is very low, all power is allocated to the subcarriers that OMA is selected for them.
- When satisfying the constraints of resource allocation (except for the maximum power constraint) is not imposing any limitation on the objective function, the allocated power to OMA is

$$p_{k,n} = \left[\frac{1}{\eta} - \frac{\sigma^2}{h_{k,n}} \right]^+. \quad (30)$$

In this case, for high SNRs, the allocated power to the NOMA users is a decreasing function of V as

$$p_n = \left[\frac{1}{V \frac{h_{k_2,n}}{\sigma^2} + \eta} - \frac{\sigma^2}{h_{k_2,n}} \right]^+. \quad (31)$$

However, for low SNRs, the allocated power to NOMA users will be independent of V as

$$p_n = \left[\frac{1}{\eta} - \frac{\sigma^2}{h_{k_2,n}} \right]^+. \quad (32)$$

- By increasing V, the total allocated power to the NOMA users decreases.

The flow chart of resource allocation algorithm is shown in Fig. 1. Also, pseudo-code of resource allocation algorithm is presented in Algorithm 1.
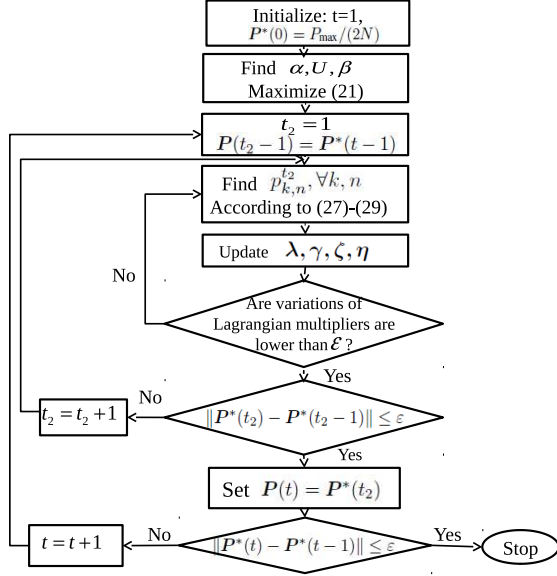
Fig. 1. A flow chart of the proposed approach.

Algorithm 1: Pseudo-code of Resource Allocation Algorithm

Initialization: Set $t := 1$ and initialize $\boldsymbol{P}^*(0) = P_{\max}/(2N)$
  Repeat
    Step 1: Derive $\boldsymbol{\beta}^*(t)$, $\boldsymbol{u}^*(t)$, and $\boldsymbol{\alpha}^*(t)$ to maximize (19) with considering fixed value of $\boldsymbol{P}^*(t-1)$
    Step 2: Set $t_2 := 1$ and $\boldsymbol{P}(t_2 - 1) = \boldsymbol{P}^*(t-1)$
    Repeat
    Step I:
      repeat
      Step A: find $p_{i,j}^{t_2} i \in \mathcal{K}, j \in \mathcal{N}$:
        for $n \in \mathcal{N}$
          if $\beta_n^*(t) == 0$
            for $k \in \mathcal{K}$
              if $\alpha_{k,k,n}^*(t) == 1$
                find $p_{k,n}^{t_2}$ from (25)
              else
                $p_{k,n}^{t_2} = 0$
          end
          else
            Set $p_{i,n}^{t_2} = 0$, $i \in \mathcal{K}$
            for $k_1 \in \mathcal{K}$ and $k_2 \in \mathcal{K}$
              if $\alpha_{k_1,k_1,n}^*(t) == 1$ and $\alpha_{k_1,k_2,n}^*(t) == 1$
                find $p_{k_1,n}^{t_2}$ from (26), and find $p_{k_2,n}^{t_2}$ from (27)
          end
         end
        update $\boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \boldsymbol{\eta}$ by derived value of $p_{i,j}^{t_2}$ $i \in \mathcal{K}, j \in \mathcal{N}$
        if variation of lagrangian multipliers are higher than ε
          go to **Step A**
        else
          Stop *repeat*
      if $\|\boldsymbol{P}^*(t_2) - \boldsymbol{P}^*(t_2 - 1)\| \le \varepsilon$
        Set $\boldsymbol{P}(t) = \boldsymbol{P}^*(t_2)$ and **Stop Repeat**
      else
        set $t_2 := t_2 + 1$ and go to **Step I**
    Step 3: if $\|\boldsymbol{P}^*(t) - \boldsymbol{P}^*(t-1)\| \le \varepsilon$
      Stop Repeat.
    else
      Set $t := t + 1$ and go to **Step 1**.

Two other important aspects to evaluate the performance of the proposed resource allocation are its convergence and computational complexity. Regarding the convergence, the proposed iterative algorithm uses the block coordinate descent (BCD) method in which one group of variables is optimized and the others are assumed to be fixed. In [48], it is shown that the convergence of BCD is guaranteed when the variable groups are updated by a successive sequence of approximations of the objective function like strictly convex local approximations. Therefore, applying successive convex approximation, the convergence of algorithm is guaranteed. However, the convergence is guaranteed to a local optimum, which may not be the global optimum.

Regarding the computational complexity, in a primal-dual interior point method for linear programming of the first step of the proposed algorithm, $O(\sqrt{n_v}n_s)$ (where n_v is the number of variables and n_s is the size of the problem data) iterations are required in the worst case to obtain a solution that can be transformed easily into an optimal basic feasible solution [49]. The major computation in each iteration of the primal-dual interior point method is the construction and Cholesky factorization of a symmetric and positive definite matrix of size m by m, where m is the number of linear equality constraints. The computational complexity of Cholesky factorization is $O(m^3)$ in the worst case [50]. The complexity of the Lagrangian method for power allocation is $O(n_v/\epsilon_{\text{sub}}^2)$ where $1/\epsilon_{\text{sub}}^2$ is the number of iterations of sub-gradient method to find a ϵ_{sub}-suboptimal point. On the other hand, the convergence of DC method is achieved by complexity of $O(\log(1/\epsilon_{\text{dc}}))$ where ϵ is the stopping criterion [51]. Consequently, the total complexity of power allocation algorithm is $O(n_v/(\epsilon_{\text{sub}}^2\epsilon_{\text{dc}}))$. This proves that the complexity of our proposed scheme only grows polynomially with the number of variables, which is a considerable improvement over direct search methods with exponential complexities.

IV. SIMULATION RESULTS

For our simulations, we consider a network with 10 subcarriers and 20 users. Two SPs seek to provide service to their subscribed users. Users are uniformly distributed in a square area with a unit length. The channel gains of users are modeled by assuming large and small scale fading as $h_{k,n} = d_k^{-\alpha}s_{k,n}$, where d_k is the distance between the BS and user k, $\alpha = 3$ and $s_{k,n}$ has an exponential distribution with unit variance in Rayleigh fading channel. In our simulation setup, we consider normalized noise, i.e., $\sigma^2 = 1$, and $P_{\max} = 20$ dB, and the required minimum power difference for SIC is set equal to $P_{\text{d}} = 0.01$. The cost of NOMA processing is set to $A = V = 2$. In our simulation results, when the constraints of the optimization problem cannot be satisfied simultaneously for a given CSI, the utility is set to zero. All statistical results are averaged over a large number of independent runs. To show the benefits of dynamic multiple access technology selection, the achieved utility of the proposed algorithm is compared to
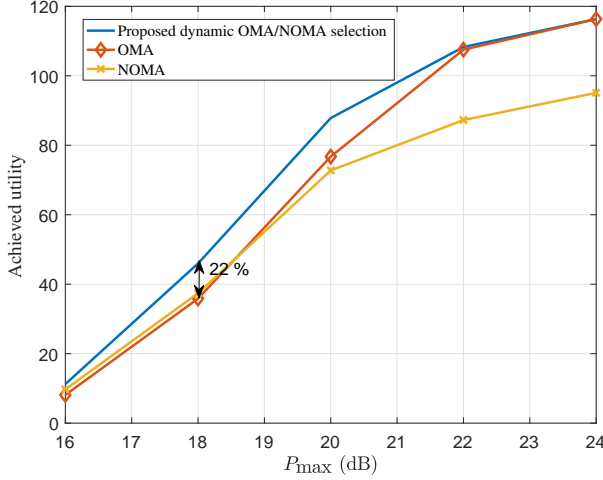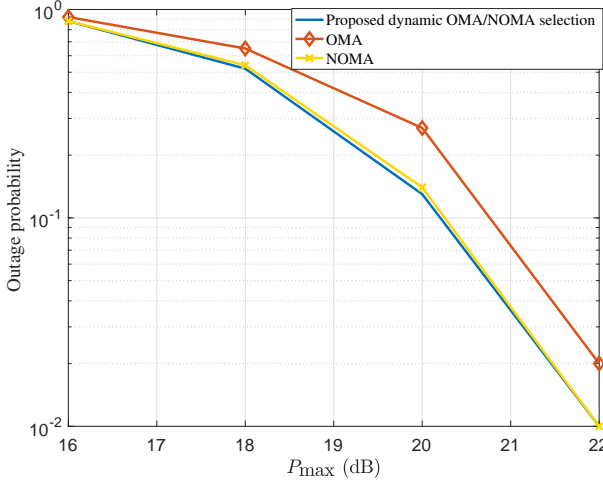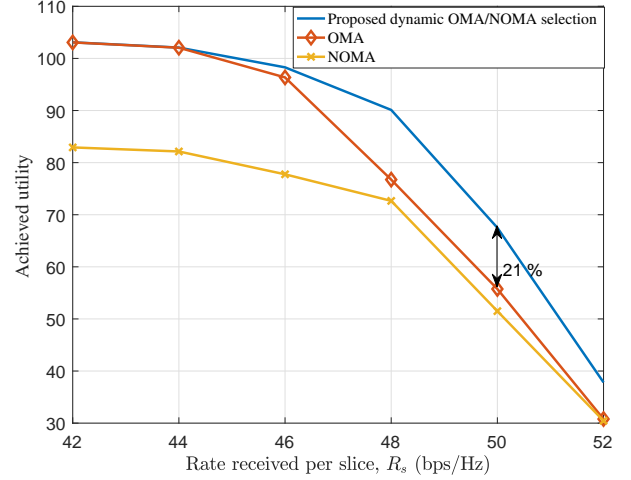
Fig. 2. (a) Utility performance and (b) Outage probability of different access technologies versus different values for maximum transmit power.

Fig. 3. (a) Utility performance and (b) Outage probability of different access technologies versus different values for QoS of SPs.
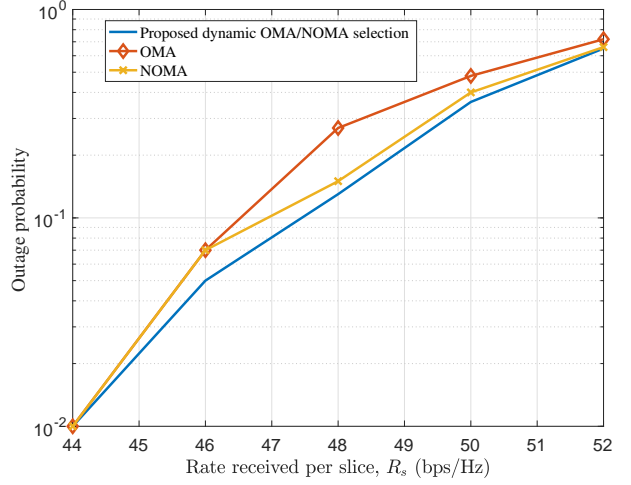
the NOMA and OMA cases. The optimization of OMA and NOMA, respectively, can be obtained by setting $\beta_n = 0$ and $\beta_n = 1$ in (19) and (23).

In Fig. 2(a), the value of the proposed objective function which is a difference between the rate and the cost of NOMA processing is shown for different values of maximum transmit power where $R_s = 48$ bps/Hz. As can be seen from Fig. 2(a), our proposed scheme outperforms the OMA and NOMA technologies. NOMA can support more than one user in each subcarrier and thus can facilitate meeting the QoS requirements of SPs compared to OMA. Therefore, it can be observed that NOMA enhances the performance for low power ranges despite of its processing cost. Note that in 18 dB point where the feasibility of (6) is very sensitive to the resource allocation strategy due to the dominant constraint (13), the performance of our proposed scheme is approximately 22% higher than NOMA.

We define the outage probability in (6) as the probability

that the rate of at least one SP is lower than its minimum required rate. This performance metric is studied in Fig. 2(b). As expected, our proposed scheme offers the lowest outage probability due to the flexibility in access technology selection. Based on Fig. 2(b), for $R_s = 48$ bps/Hz, $P_{\max} = 16$ dB is not enough to satisfy the rate requirements for SPs and the optimization problem will be mostly infeasible which explains the very high outage at those values. As can be seen for $P_{\max} = 20$ dB, our proposed strategy achieves approximately half of the outage probability of NOMA and OMA. The flexibility offered by our proposed scheme to select between OMA and NOMA leads to the higher probability to satisfy the minimum rate of SPs, and, thus, lower outage probability compared to the OMA and NOMA.

Fig. 3(a) and Fig. 3(b) analyze the impact of the SPs' QoS requirement parameter R_s on the achieved utility and outage probability. As expected, the proposed algorithm improves the performance in terms of both utility function and outage
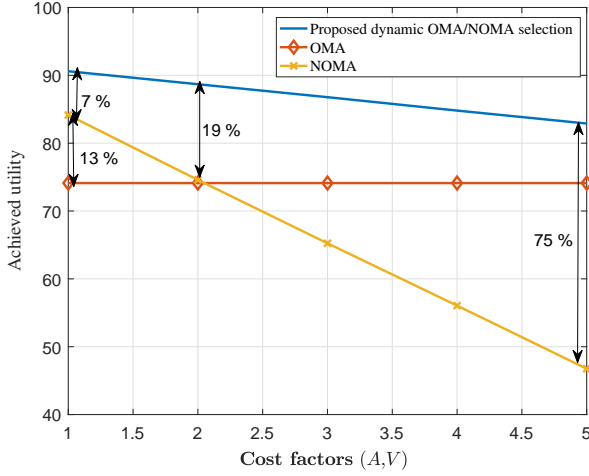
Fig. 4. Utility performance of different access technologies versus different values for A and V for $R_s = 48$.

Fig. 5. Utility performance of different access technologies versus number of users.

probability compared to the pure OMA or NOMA cases. As can be seen in Fig. 3(a), for low ranges of R_s, the proposed scheme will mostly choose OMA since the probability of infeasibility is very low. In these cases, OMA is the best choice since the cost is lower. By increasing the minimum rate requirement, the feasibility region is shrunk in any access technology and outage probability increases. Meanwhile, in OMA the outage probability increases more quickly compared to the NOMA case since there are more restrictions as each subcarrier can be allocated to only one user. In other words, the ability of NOMA to allocate two users to each subcarrier results in a higher opportunity to satisfy the minimum required rate of SPs. Compared to the pure OMA and NOMA cases, our proposed strategy for resource allocation achieves lower outage probability since dynamic access technology selection expands the feasibility region. Also, the ability of the proposed scheme to choose between OMA with no processing cost as well as NOMA leads to performance improvement in the terms of defined utility compared to the pure NOMA technology. For example, when the feasibility of (6) is the dominant constraint (e.g. for $R_s = 50$), the proposed strategy for dynamic access technology selection yields utility gains of approximately 21% compared to OMA.

In Fig. 4, we evaluate the effects of the processing cost values on the system performance by presenting the value of the utility function for different values of A and V. From Fig. 4, by increasing the value of these parameters, the proposed scheme can enhance the performance up to 75% compared to the NOMA scenario. For low values of A and V, NOMA achieves up to 13% utility improvement compared to the OMA case. In fact, when taking into account both the rate and processing cost, on their own, neither NOMA nor OMA will be suitable for all use cases. As can be seen, our proposed scheme with flexible access technology improves the utility from 7% to 19%.

In Fig. 5, we study effect of the number of users on the performance of proposed method. 5. In this figure, the other parameters are similar to those used in Fig. 2(a) where
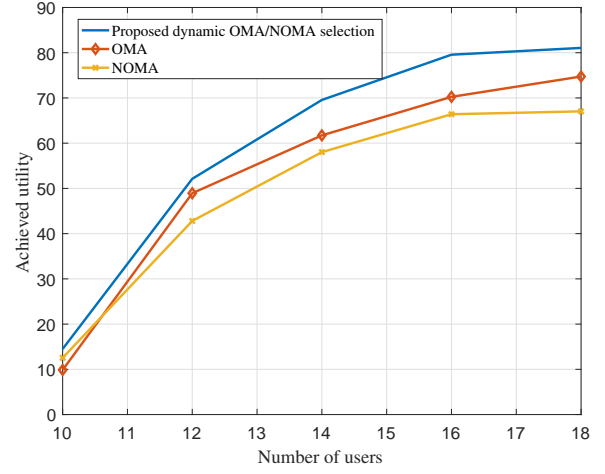
the maximum power is set to the 20 dB. In general, by increasing the number of users in the system, the utility of all schemes increases due to the users' diversity gain. When $N < 11$, NOMA has higher utility compared to OMA despite the processing cost. This is because NOMA can expand the feasibility region as two users can be supported in each subcarrier. For $N \geq 11$, the performance of OMA improves owing to the users' diversity gain and up to 10% higher performance is achieved compared to NOMA. This means the processing cost for NOMA influences its effectiveness compared to OMA. However, our proposed dynamic scheme outperforms both the NOMA and OMA cases for any number of users in the network.

In Fig. 6 and Fig. 7, we study the effect of the distribution density of users in the coverage region of the BS on the achievable utility and total achievable rate, respectively. The total achievable rate is calculated as $\sum_{n \in \mathcal{N}} R_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{P})$ where $R_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{P})$ is defined in (2). For this purpose, we divide the coverage region into central and edge regions. The central region is the square with the half length of total coverage area, which is located at the center of the coverage region, and the rest is the edge region. The users in the edge region are called cell-edge users. In these figures, the x-axis represents the percentage of users in the cell-edge region. Other parameters are set as in Fig. 2(a) where $P_{\max} = 20$ dB. In Fig. 6, when no users are located in the cell-edge region, NOMA achieves the lowest performance among the three schemes in the terms of achieved utility, in presence of NOMA cost. Furthermore, it is shown that, the utility performance of OMA is equal to the utility performance of NOMA when 80% of users are distributed at the cell edge, even though NOMA imposes an extra processing cost. As a result, we can conclude that the performance of OMA is severely degraded when all users are at the cell edge (point 100%). However, NOMA can provide a better performance since it can support more users, and consequently, the probability to satisfy the isolation constraint will increase and the feasibility region of the optimization problem is expanded. When the number of
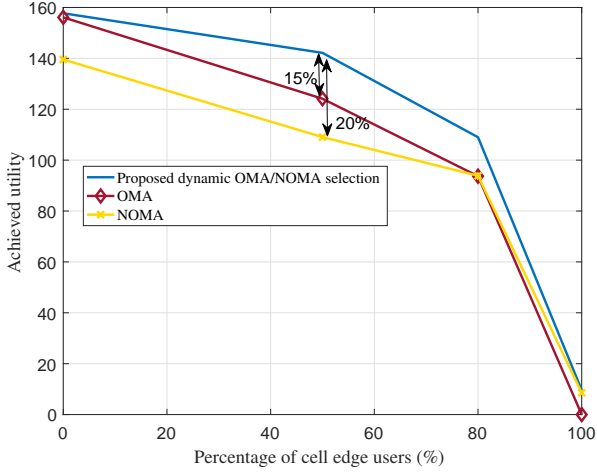
Fig. 6. Utility of different access technologies versus percentage of cell edge users.
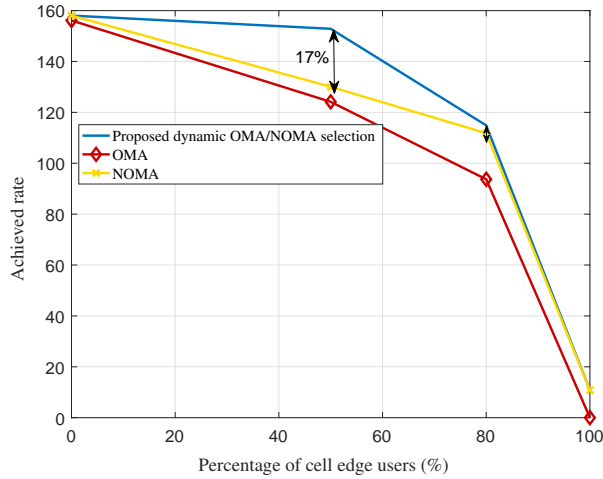

Fig. 7. Rate performance of different access technologies versus percentage of cell edge users.

cell-edge and central users in the network are equal, i.e., point 50% in Fig. 6, our proposed scheme achieves 15% and 20% higher utility performance compared to the OMA and NOMA cases, respectively, due to its capability to choose the best technology depending on users CSI and cost of technology implementation.

From Fig. 7, it is evident that all access methods have similar performance in terms of the total rate achieved by the users when all users are in the central region. Comparing the two points at 50% and 80% in Fig. 7, it can be observed that the gap between the total rate of our proposed scheme and NOMA decreases. In fact, when the number of users at cell edge and central region is equal, (i.e., point 50%), the flexible access technology selection of our proposed scheme leads to 17% improvement in the total rate.

To study the gap between our proposed scheme and the optimal resource allocation, we use the exhaustive search to find the optimal solution. First, we note that, the complexity of technology selection and user assignment of exhaustive search

is $\left(\binom{K}{2} + K \right)^{N}$. Consequently, the exhaustive search suffers from huge computational complexity at large values of N and K. This forces us to focus on $K = 6$ and $N = 3$ and 4 for this simulation. We also set $P_{\max} = 20$ dB and $R_s = 12$ bps/Hz. The results are summarized in Table I, which reveals that the gap between the optimal resource allocation and our proposed algorithm is negligible when N and K are small, while this gap expands when network size grows. However, this gap does not exceed 7% as seen from Table I.

V. CONCLUSION

In this paper, we have introduced a new approach for access technology that can dynamically choose a suitable technology based on the instantaneous CSI. As an example, two different access technologies (OMA and NOMA) have been considered as options of access technology selection. We have considered a multi-user multi-carrier single cell downlink communication system, assuming a set of the SPs, each of which having a set of its own users and a minimum QoS requirement. We have then proposed a novel algorithm that can allocate resources including subcarriers, power and technology selection variables. We have defined a novel utility function which reflects the tradeoff between the achievable rate and the imposed cost for NOMA processing. To efficiently solve the proposed resource allocation problem, we have developed a two-step iterative algorithm. In the first step, by introducing auxiliary variables, the subcarrier assignment and technology selection problem is transformed and solved using linear integer programming. Subsequently, in the second step, the power allocation is solved by applying DC programming. Simulation results highlight that higher utility and lower outage probability can be achieved via the proposed dynamic access technology selection. Future work can extend to the case where more than two users can share one subcarrier in NOMA.

APPENDIX A
POWER ALLOCATION STRATEGY

In the primal problem, we want to find the variables of matrix $\boldsymbol{P}$ with dimension $K \times N$ by considering fix value for $\boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \boldsymbol{\eta}$. Because each user can be first or second user of NOMA, we take derivative of general formula (24) with respect to $p_{k',n}$

$$\frac{dL(\boldsymbol{P}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \boldsymbol{\eta})}{dp_{k',n}} = \underbrace{\alpha^t_{k',k',n} \frac{h_{k',n}}{\sigma^2 + p_{k',n} h_{k',n}}}_{a_1} +$$

$$\sum_{k_2 \in \mathcal{K}, k_2 \neq k'} \Big(\underbrace{\frac{h_{k_2,n}}{p_{k',n} h_{k_2,n} + p_{k_2,n} h_{k_2,n} + \sigma^2}}_{a_2}$$

$$- \underbrace{\frac{h_{k_2,n}}{p^{t_2-1}_{k',n} h_{k_2,n} + \sigma^2}}_{a_3} \Big) u^t_{k',k_2,n}$$

$$+ \underbrace{\sum_{k_1 \in \mathcal{K}, k_1 \neq k'} \frac{h_{k',n} u^t_{k_1,k',n}}{p_{k_1,n} h_{k',n} + p_{k',n} h_{k',n} + \sigma^2}}_{a_4}$$

TABLE I
ACHIEVED UTILITY OF THE PROPOSED RESOURCE ALLOCATION ALGORITHM AND OPTIMAL RESOURCE ALLOCATION

	Exhaustive search for OMA/NOMA	Our proposed algorithm for OMA/NOMA
Achieved Utility ($N = 3$, $K = 6$)	30.5326	29.5603
Achieved Utility ($N = 4$, $K = 6$)	44.2093	41.0337

$$-V\Bigg(\underbrace{\sum_{k_1 \in \mathcal{K}, k_1 \neq k'} u_{k_1,k',n}^t \frac{h_{k',n}}{p_{k',n}^{t_2-1} h_{k',n} + \sigma^2}}_{a_5}$$

$$-\underbrace{\sum_{k_2 \in \mathcal{K}, k_2 \neq k'} u_{k',k_2,n}^t \frac{1}{p_{k',n}}}_{a_6}\Bigg) + \lambda_{s'}(\underbrace{\alpha_{k',k',n}^t \frac{h_{k',n}}{\sigma^2 + p_{k',n} h_{k',n}}}_{a_7}$$

$$+ \underbrace{\sum_{k_1 \in \mathcal{K}, k_1 \neq k'} \frac{h_{k',n}}{p_{k',n} h_{k',n} + p_{k_1,n} h_{k',n} + \sigma^2} u_{k_1,k',n}^t)}_{a_8}$$

$$+ \sum_{s \in \mathcal{S}, s \neq s'} \lambda_s \sum_{k_2 \in \mathcal{K}_s} u_{k',k_2,n}^t \underbrace{\Bigg(\frac{h_{k_2,n}}{p_{k_2,n} h_{k_1,n} + p_{k',n} h_{k_2,n} + \sigma^2}}_{a_9}$$

$$- \underbrace{\frac{h_{k_2,n}}{p_{k',n}^{t_2-1} h_{k_2,n} + \sigma^2}}_{a_{10}}\Bigg) + \gamma_n \Bigg(\underbrace{\frac{-h_{k',n} \beta_n^t \alpha_{k',k',n}^t}{\sigma^2}}_{a_{11}}$$

$$+ \underbrace{\sum_{k_1 \in \mathcal{K}, k_1 \neq k'} u_{k_1,k',n}^t \frac{h_{k_1,n}}{\sigma^2}}_{a_{12}}\Bigg) - \zeta_{k',n} - \eta\Bigg(\alpha_{k',k',n}^t + \sum_{k_1 \in \mathcal{K}} u_{k_1,k,n}^t\Bigg),$$

$$\forall k', n. \quad (33)$$

In (33), the first three terms (a_1, a_2, and a_3) capture the case in which user k' is the first user of NOMA or is the OMA user. Note that when k' is the first user of NOMA, its power causes interference to the second user of NOMA, which leads to terms a_2 and a_3.

The term a_4 is related to the case in which k' is the second user of NOMA. In (33), since we do not replace $\alpha_{k',k',n}^t$ with its derived values from the previous step (considering the general format), the summation over k_1 appears when user k' is the second user of NOMA. a_5 and a_6 are the derivations of the processing cost of NOMA in the objective function when k' is the second user of NOMA and the first user of NOMA, respectively. In a_5, we have a summation on k_1 similar to the a_4. Similar to a_2 and a_3, a_6 has a summation on k_2.

The terms a_7, a_8, a_9, and a_{10} are the Lagrangian terms related to the constraint (22). Let assume user k' belongs to SP s' The term a_7 is related to the case in which user k' is the first user of OMA or NOMA. Again, when user k' is the second user of NOMA, its rate should be considered in the rate constraint of SP s', which leads to a_8. When the user k' is the first user of NOMA and its power is considered as the interference for the user belong to the other SPs ($s \neq s'$) leads to the a_9 and a_{10} terms. The a_{11} and a_{12} terms are due to the Lagrangian term related to constraint (7). Similar to a_4 and a_8, we have summation in a_{12}.

All terms of $\frac{dL}{dp_{k',n}}$ that are not function of $p_{k,n}, k =$

$1, ..., K$, in the above equation are represented by $D_{k',n}$ as,

$$D_{k',n} = \beta_n^t \sum_{k_2, k_2 \neq k'} \Bigg(-\frac{h_{k_2,n}}{p_{k',n}^{t_2-1} h_{k_2,n} + \sigma^2}\Bigg) u_{k',k_2,n}^t$$

$$-V\Bigg(\sum_{k_1 \in \mathcal{K}, k_1 \neq k'} u_{k_1,k',n}^t \frac{h_{k',n}}{p_{k',n}^{t_2-1} h_{k',n} + \sigma^2}\Bigg)$$

$$+ \sum_{s \in \mathcal{S}, s \neq s'} \lambda_s \sum_{k_2 \in \mathcal{K}_s} u_{k',k_2,n}^t \Bigg(-\frac{h_{k_2,n}}{p_{k',n}^{t_2-1} h_{k_2,n} + \sigma^2}\Bigg)$$

$$+ \gamma_n \Bigg(\frac{-h_{k',n} \beta_n^t \alpha_{k',k',n}^t}{\sigma^2} + \sum_{k_1, k_1 \neq k'} u_{k_1,k',n}^t \frac{h_{k_1,n}}{\sigma^2}\Bigg)$$

$$- \zeta_{k',n} - \eta\Bigg(\alpha_{k',k',n}^t + \sum_{k_1} u_{k_1,k',n}^t\Bigg) \quad (34)$$

Therefore, the relation between the optimum values of $p_{k,n}, k = 1, .., K$ is

$$\alpha_{k',k',n}^t \frac{h_{k',n}}{p_{k',n} h_{k',n} + \sigma^2}(1 + \lambda_{s'})$$

$$+ \sum_{k, k \neq k'} \frac{h_{k,n}}{p_{k',n} h_{k,n} + p_{k,n} h_{k,n} + \sigma^2} u_{k',k,n}^t(1 + \sum_{s \in S, k \in \mathcal{K}_s} \lambda_s)$$

$$+ \sum_{k, k \neq k'} \frac{h_{k',n} u_{k,k',n}^t}{p_{k,n} h_{k',n} + p_{k',n} h_{k',n} + \sigma^2}(1 + \lambda_{s'})$$

$$+ V\Bigg(\sum_{k, k \neq k'} u_{k',k,n}^t \frac{1}{p_{k',n}}\Bigg) + D_{k',n} = 0. \quad (35)$$

Note that by allocation in step 1 and considering only two users for NOMA at most, there are two equations and two variables for each subcarrier. Next, we consider two cases.

In the first case, OMA is selected for subcarrier n and the user k uses it. Thus, $\beta_n^t = 0$, $u_{k_1,k_2,n}^t = 0$, $\forall k_1, k_2$ and $\alpha_{k,k,n}^t = 1$. So, by substituting these value in (35), the allocated power to the user k on subcarrier n is obtained according to (25).

For subcarrier n, when NOMA is selected and if user k_1 and k_2 are the first and second users, respectively, we have $\beta_n^t = 1$, $\alpha_{k_1,k_1,n}^t = 1$, $u_{k_1,k_2,n}^t = 1$, and other variables, i.e., $\boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \boldsymbol{\eta}$, are equal to zero. By assuming $p_n = p_{k_1,n} + p_{k_2,n}$ we have two equations and two variables

$$\frac{h_{k_1,n}}{p_{k_1,n} h_{k_1,n} + \sigma^2}(1 + \lambda_{s'}) + \frac{h_{k_2,n}}{p_n h_{k_2,n} + \sigma^2}(1 + \lambda_{s''})$$

$$+ \frac{V}{p_{k_1,n}} + D_{k_1,n} = 0 \quad (36)$$

$$\frac{h_{k_2,n}}{p_n h_{k_2,n} + \sigma^2}(1 + \lambda_{s'}) + D_{k_2,n} = 0 \quad (37)$$

where user k_1 and k_2 are in $\mathcal{K}_{s'}$ and $\mathcal{K}_{s''}$, respectively. The quadratic equation for obtaining $p_{k_1,n}$ is as

$$(\frac{-D_{k_2,n}(1+\lambda_{s''})}{1+\lambda_{s'}} + D_{k_1,n})h_{k_1,n}p_{k_1,n}^2 + (h_{k_1,n}(1+\lambda_{s'})$$

$$+\sigma^2(\frac{-D_{k_2,n}(1+\lambda_{s''})}{1+\lambda_{s'}} + D_{k_1,n}) + Vh_{k_1,n})p_{k_1,n} + V\sigma^2 = 0$$

$$(38)$$

As a result, the power of user k_1 and k_2 on the subcarrier n are obtained by (26).

REFERENCES

[1] Report ITU-R M.2370-0, "IMT traffic estimates for the years 2020 to 2030," *Technical report*, 2008.

[2] I. Cisco, "Cisco visual networking index: Forecast and methodology, 20112016," *CISCO White paper*, 2012.

[3] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Unmanned aerial vehicle with underlaid device-to-device communications: Performance and tradeoffs," *IEEE Transactions on Wireless Communications*, vol. 15, no. 6, pp. 3949–3963, Jun. 2016.

[4] W. Saad, M. Bennis, and M. Chen, "A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems," 2019, arXiv:1902.10265.

[5] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *Proc. IEEE Vehicular Technology Conference*, Dresden, Germany, Jun. 2013, pp. 1–5.

[6] M. El-Bamby, M. Bennis, W. Saad, M. Debbah, and M. Latva-aho, "Resource optimization and power allocation in full duplex non-orthogonal multiple access (FD-NOMA) networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 12, pp. 2860–2873, Dec. 2017.

[7] T. Park, G. Lee, and W. Saad, "Message-aware uplink transmit power level partitioning for non-orthogonal multiple access (NOMA)," in *Proc. IEEE Global Communications Conference (GLOBECOM), Next Generation Networking and Internet Symposium*, Abu Dhabi, Dec. 2018.

[8] H. Nikopour and H. Baligh, "Sparse code multiple access," in *in Proc. IEEE Int. Symp. Personal, Indoor, and Mobile Commun. (PIMRC)*, London, UK, Sep. 2013, pp. 332–336.

[9] S. Chen, B. Ren, Q. Gao, S. Kang, S. Sun, and K. Niu, "Pattern division multiple access (PDMA) - A novel non-orthogonal multiple access for 5G radio networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 4, pp. 3185–3196, Jul. 2016.

[10] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, "System level performance evaluation of downlink nonorthogonal multiple access (NOMA)," in *Proc. IEEE Int. Symposium on Personal, Indoor and Mobile Radio Commun.*, London, UK, Sep. 2013.

[11] W. Han, J. Ge, J. Men, "Performance analysis for NOMA energy harvesting relaying networks with transmit antenna selection and maximal-ratio combining over Nakagami-m fading," *IET Communication*, vol. 10, no. 18, pp. 2687–2693, Dec. 2016.

[12] B. He, A. Liu, N. Yang, V. K. N. Lau, "On the design of secure non-orthogonal multiple access systems," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 10, pp. 2196–2206, Jul. 2017.

[13] S. Islam, M. Zeng, O. A. Dobre, and K.-S. Kwak, "Resource allocation for downlink NOMA systems: Key techniques and open issues," *IEEE Wireless Communications*, vol. 25, no. 2, pp. 40–47, Apr. 2018.

[14] Z. Wei, D.W.K. Ng, J. Yuan, and H.M. Wang, "Optimal resource allocation for power-efficient MC-NOMA with imperfect channel state information," *IEEE Transactions on Communications*, vol. 65, no. 9, pp. 3944–3961, Sep. 2017.

[15] H. H. Cho, "Integration of SDR and SDN for 5G," *IEEE Access*, vol. 2, pp. 1196–1204, Sep. 2014.

[16] H. Ekstrom, A. Furuskar, J. Karlsson, M. Meyer, S. Parkvall, J. Torsner, and M. Wahlqvist, "Technical solutions for the 3G long-term evolution," *IEEE Communications Magazine*, vol. 44, no. 3, pp. 38–45, Mar. 2006.

[17] S. Sadr, A. Anpalagan, and K. Raahemifar, "Radio resource allocation algorithms for the downlink of multiuser OFDM communication systems," *IEEE Communications Surveys and Tutorials*, vol. 11, no. 3, pp. 92–106, Aug. 2009.

[18] A.S. Hamza, S.S. Khalifa, H.S. Hamza, "A survey on inter-cell interference coordination techniques in OFDMA-based cellular networks," *IEEE Communications Surveys and Tutorials*, vol. 15, no. 4, pp. 1642–1670, Fourth Quarter 2013.

[19] M. Naeem, A. Anpalagan, M. Jaseemuddin, and D. C. Lee , "Resource allocation techniques in cooperative cognitive radio networks," *IEEE Communications Surveys and Tutorials*, vol. 16, no. 2, pp. 729–744, Nov. 2014.

[20] R. Afolabi, A. Dadlani, and K. Kim, "Multicast scheduling and resource allocation algorithms for OFDMA-based systems: A survey," *IEEE Communications Surveys and Tutorials*, vol. 15, no. 1, pp. 240–254, Feb. 2013.

[21] L. Lei, D. Yuan, C. K. Ho, and S. Sun, "Power and channel allocation for non-orthogonal multiple access in 5G systems: Tractability and computation," *IEEE Transactions on Wireless Communications*, vol. 15, no. 12, pp. 8580–8594, Dec. 2016.

[22] B. Di, L. Song, and Y. Li, "Sub-channel assignment, power allocation, and user scheduling for non-orthogonal multiple access networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 11, pp. 7686–7698, Nov. 2016.

[23] F. Fang, H. Zhang, J. Cheng, and V. C. M. Leung, "Energy-efficient resource allocation for downlink non-orthogonal multiple access network," *IEEE Transactions on Communications*, vol. 64, no. 9, pp. 3722–3732, Sep. 2016.

[24] J. Zhu, J. Wang, Y. Huang, S. He, X. You, and L. Yang, "On optimal power allocation for downlink non-orthogonal multiple access systems," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 12, pp. 2744–2757, Jul. 2017.

[25] J. He, and Z. Tang, "Low-complexity user pairing and power allocation algorithm for 5G cellular network non-orthogonal multiple access," *Electronics Letters*, vol. 53, no. 9, pp. 626–627, May 2017.

[26] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 10, pp. 1747–1758, Oct. 1999.

[27] Y. Sun, D. W. K. Ng, Z. Ding, and R. Schober, "Optimal joint power and subcarrier allocation for full-duplex multicarrier non-orthogonal multiple access systems," *IEEE Transactions on Communications*, vol. 65, no. 3, pp. 1077–1091, Mar. 2017.

[28] S. Ali, E. Hossain, and D.I. Kim, "Non-orthogonal multiple access (NOMA) for downlink multiuser MIMO systems: User clustering, beamforming, and power allocation," *IEEE Access*, vol. 5, pp. 565–577, Dec. 2017.

[29] J. Zhao, Y. Liu, K. K. Chai, A. Nallanathan, Y. Chen, and Z. Han, "Low-complexity user pairing and power allocation algorithm for 5G cellular network non-orthogonal multiple access," *IEEE Transactions on Wireless Communications*, vol. 16, no. 9, pp. 5825–5837, May 2017.

[30] A. J. Morgado, K. M. S. Huq, J. Rodriguez, C. Politis, and H. Gacanin, "Hybrid resource allocation for millimeter-wave NOMA," *IEEE Wireless Communications Magazine*, vol. 24, no. 5, pp. 23–29, Oct. 2017.

[31] A. S. Marcano, and H. L. Christiansen, "Impact of NOMA on network capacity dimensioning for 5G HetNets," *IEEE Access*, vol. 6, pp. 13 587–13 603, Feb. 2018.

[32] M. Baghani, S. Parsaeefard, and T. Le-Ngoc, "Multi-objective resource allocation in density-aware design of C-RAN in 5G," *IEEE Access*, vol. 6, pp. 45 177–45 190, Aug. 2018.

[33] A. Benjebbour, K. Saito, A. Li, Y. Kishiyama, and T. Nakamura, "Nonorthogonal multiple access (NOMA): Concept, performance evaluation and experimental trials," in *Proc. IEEE International Conference on Wireless Networks and Mobile Commun.*, Marrakech, Morocco, Oct. 2015.

[34] B. Ling, C. Dong, J. Dai, and J. Lin, "Multiple decision aided successive interference cancellation receiver for NOMA systems," *IEEE Wireless Communications Letter*, vol. 6, no. 4, pp. 498–501, May 2017.

[35] H. Wen, P. K. Tiwary, and T. Le-Ngoc, *Wireless virtualization*. Cham: Springer, 2013.

[36] D. Bertsekas, *Nonlinear Programming*. 2nd ed. , Belmont, Massachusetts: Athena Scientific, 1999.

[37] S. Boyd, L. Xiao, A. Mutapic, and J. Mattingley, "Sequential convex programming," in *Stanford University, Stanford, CA, USA Technical Report EE364b*, 2007.

[38] E. Bedeer, O. A. Dobre, M. H. Ahmed, and K. E. Baddour, "A systematic approach to jointly optimize rate and power consumption for OFDM systems," *IEEE Transactions on Mobile Computing*, vol. 15, no. 6, pp. 1305–1317, Jun 2016.

[39] R. Han, Y. Gao, C. Wu, and D. Lu, "An effective multi-objective optimization algorithm for spectrum allocations in the cognitive-radio-based internet of things," *IEEE Access*, vol. 6, pp. 12 858–12 867, Jan. 2018.

[40] E. Bjornson, E. A. Jorswieck, M. Debbah, and B. Ottersten, "Multiobjective signal processing optimization: The way to balance conflicting

metrics in 5G systems," *IEEE Signal Processing Magazine*, vol. 31, no. 6, pp. 14–23, Nov. 2014.

[41] L. Dai, B. Wang, Y. Yuan, S. Han, C.-L. I, Z. Wang, "Non-orthogonal multiple access for 5G: Solutions challenges opportunities and future research trends," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 74–81, Sep. 2015.

[42] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 6010–6023, Aug. 2016.

[43] A. Li, A. Benjebbour, X. Chen, H. Jiang, and H. Kayama, "Investigation on hybrid automatic repeat request (HARQ) design for NOMA with SU-MIMO," in *IEEE 26th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Hong Kong, China, Sep. 2015, pp. 590–594.

[44] A. Taleb Zadeh Kasgari and W. Saad, "Stochastic optimization and control framework for 5G network slicing with effective isolation," in *IEEE 52nd Annual Conference on Information Sciences and Systems (CISS)*, Princeton, NJ, USA, Mar. 2018.

[45] R. Kokku, R. Mahindra, H. Zhang, and S. Rangarajan, "NVS: A substrate for virtualizing wireless resources in cellular networks," *Society for Industrial and Applied Mathematics Journal on Optimization*, vol. 20, no. 5, pp. 1333–1346, Oct. 2012.

[46] M. S. Ali, H. Tabassum, and E. Hossain, "Dynamic user clustering and power allocation in non-orthogonal multiple access (NOMA) systems," *IEEE Access*, vol. 4, pp. 6325–6343, Aug. 2016.

[47] M. Grant and S. Boyd, "CVX: MATLAB software for disciplined convex programming," in *Version 2.1. [Online]. Available: http://cvxr.com/cvx*, Mar. 2017.

[48] M. Razaviyayn, M. Hong, and Z. Q. Lue, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *Society for Industrial and Applied Mathematics Journal on Optimization*, vol. 23, no. 2, pp. 1126–1153, Jun. 2013.

[49] S. J. Wright, *Primal-Dual Interior-Point Methods*. Philadelphia:SIAM: PA, 1997.

[50] B. Borchers and J. Young, "Implementation of a primal-dual method for SDP on a shared memory parallel architecture," *Computational Optimization and Applications*, vol. 37, no. 3, pp. 355–369, Jul. 2007.

[51] A. L. Yuille and A. Rangarajan, "The concave-convex procedure," *Neural Computation*, vol. 15, no. 4, pp. 915–936, Apr. 2003.

Mina Baghani received the B.Sc. and M.Sc. degrees from Shahed University, Tehran, Iran, in 2008 and 2011, respectively and the Ph.D. degree in electrical engineering from Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, in 2017. In 2018, she worked as a researcher at the Iran Telecommunication Research Center (ITRC). Since 2019, she is an assistant professor of the Faculty of Technical and Engineering, Imam Khomeini International University, Qazvin, Iran. Her current research interests include multilayer coding, nonlinear optimization methods, cognitive radio networks, and resource allocation in wireless networks.

Saeedeh Parsaeefard (S09M14) received the Ph.D. degree in electrical and computer engineering from Tarbiat Modares University, Tehran, Iran, in 2012. From November 2010 to October 2011, she was a visiting Ph.D. student with the University of California, Los Angeles, Los Angeles, CA, USA. She was a Postdoctoral Research Fellow with the Department of Electrical and Computer Engineering, McGill University, Montreal, QC, Canada. She is currently a Faculty Member with the Iran Telecommunication Research Center, Tehran, Iran and visiting faculty member in University of Toronto. She received the IEEE Iran Section Women in Engineering awards in 2018. Her research interests span the overall systems architecture of wireless networks, including the physical, medium access, and networking layers, and applications of optimization theory and game theory in system designs.

Mahsa Derakhshani (S10, M13) received the Ph.D. degree in electrical engineering from McGill University, Montreal, Canada, in 2013. She was a Post-Doctoral Research Fellow with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, Canada, a Research Assistant with the Department of Electrical and Computer Engineering, McGill University, from 2013 to 2015, and an Honorary NSERC Post-Doctoral Fellow with the Department of Electrical and Electronic Engineering, Imperial College London, from 2015 to 2016. She is currently a Lecturer (Assistant Professor) in digital communications with the Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, U.K. Her research interests include non-orthogonal multiple access (NOMA), ultra-reliable low-latency communications (URLLC), software-defined wireless networking, and the applications of optimization and machine learning for radio resource management. She has received several awards and fellowships, including the John Bonsall Porter Prize, the McGill Engineering Doctoral Award, the Fonds de Recherche du Quebec Nature et Technologies (FRQNT), and the Natural Sciences and Engineering Research Council of Canada (NSERC) Post-Doctoral Fellowships. She currently serves as an Associate Editor for IET Signal Processing Journal.

Walid Saad (S07, M10, SM15, F19) received his Ph.D degree from the University of Oslo in 2010. He is currently a Professor at the Department of Electrical and Computer Engineering at Virginia Tech, where he leads the Network Science, Wireless, and Security laboratory. His research interests include wireless networks, machine learning, game theory, security, unmanned aerial vehicles, cyber-physical systems, and network science. Dr. Saad is a Fellow of the IEEE and an IEEE Distinguished Lecturer. He is also the recipient of the NSF CAREER award in 2013, the AFOSR summer faculty fellowship in 2014, and the Young Investigator Award from the Office of Naval Research (ONR) in 2015. He was the author/co-author of seven conference best paper awards at WiOpt in 2009, ICIMP in 2010, IEEE WCNC in 2012, IEEE PIMRC in 2015, IEEE SmartGridComm in 2015, EuCNC in 2017, and IEEE GLOBECOM in 2018. He is the recipient of the 2015 Fred W. Ellersick Prize from the IEEE Communications Society, of the 2017 IEEE ComSoc Best Young Professional in Academia award, and of the 2018 IEEE ComSoc Radio Communications Committee Early Achievement Award. From 2015-2017, Dr. Saad was named the Stephen O. Lane Junior Faculty Fellow at Virginia Tech and, in 2017, he was named College of Engineering Faculty Fellow. He currently serves as an editor for the IEEE Transactions on Wireless Communications, IEEE Transactions on Mobile Computing, IEEE Transactions on Cognitive Communications and Networking, and IEEE Transactions on Information Forensics and Security. He is an Editor-at-Large for the IEEE Transactions on Communications.