

New Optimal and Suboptimal Resource Allocation Techniques for Downlink Non-orthogonal Multiple Access

Marie-Rita Hojeij^{1,2} · Joumana Farah³ ·
Charbel Abdel Nour² · Catherine Douillard²

Published online: 10 May 2015
© Springer Science+Business Media New York 2015

Abstract This paper investigates several new strategies for the allocation of radio resources (bandwidth and transmission power) using a non-orthogonal multiple access (NOMA) scheme with successive interference cancellation (SIC) in a cellular downlink system. In non-orthogonal access with SIC, the same subband is allocated to multiple users, which requires elaborate multiuser scheduling and subband assignment techniques, compared to orthogonal multiplexing. While taking into account various design issues, we propose and compare several optimum and suboptimum power allocation schemes. These are jointly implemented with multiple user scheduling strategies. Besides, a minimization of the total amount of used bandwidth is targeted. Also, to increase the total achieved system throughput, a hybrid orthogonal-non orthogonal scheme is introduced. This hybrid scheme enables a dynamic switching to orthogonal signaling whenever the non-orthogonal cohabitation in the power domain does not improve the achieved data rate per subband. Extensive simulation results show that the proposed strategies for resource allocation can improve both the spectral efficiency and the cell-edge user throughput, especially when compared to previous schemes employing either orthogonal signaling or NOMA with static

✉ Joumana Farah
joumanafarah@ul.edu.lb

Marie-Rita Hojeij
marieritahojeij@usek.edu.lb

Charbel Abdel Nour
charbel.abdelnour@telecom-bretagne.eu

Catherine Douillard
catherine.douillard@telecom-bretagne.eu

¹ Department of Telecommunications, Faculty of Engineering, Holy-Spirit University of Kaslik, P.O. Box 446, Jounieh, Lebanon

² Department of Electronics, Lab-STICC - UMR 6285 Technopole Brest Iroise, Telecom Bretagne, CS 83 818, 29238 Brest Cedex, France

³ Department of Electricity and Electronics, Faculty of Engineering, Lebanese University, Roumieh, Lebanon

inter-subband power allocation. They also prove to be robust in the context of crowded areas.

Keywords Non-orthogonal multiple access · User pairing · Power allocation · Waterfilling

1 Introduction

With the proliferation of Internet usages, future communication networks will have to face by 2020 a mobile traffic volume 500 times larger than today's [1], while supporting a multitude of services with a more efficient utilization of the spectrum. These challenges are pushing the limits of the actual generation, and pointing toward a need for a 5th generation of cellular technology. In this sense, new designs of radio access technology (RAT) in terms of spectrum management and multiple access techniques become essential to accommodate such requirements [2, 3].

In the 3.9 and 4th Generation (4G) of mobile communication systems, such as Long-Term Evolution (LTE) [4] and LTE-Advanced [5, 6], Orthogonal Multiple Access (OMA) based on OFDM (Orthogonal Frequency Division Multiplexing) or Single Carrier Frequency Division Multiple Access (SC-FDMA) has been adopted, respectively for the downlink and uplink transmissions. Orthogonal multiple access techniques have proven their ability to achieve good system-level throughput performance in packet-domain services, as well as their viability in practical implementations, thanks to a simplified receiver design.

Beside the mobile traffic increase, allowing an efficient cohabitation of very different classes of traffic/service has been lately of a great concern. To address this challenge, a flexible air interface, together with efficient resource allocation techniques, is needed in order to satisfy the various requirements of those different traffic classes. To this end, spectrum optimization and dynamic resource allocation have been the core of a large number of studies in the context of multiuser OFDM signaling [7–9]. For instance, in [10], an iterative algorithm for subband allocation in OFDM was proposed in such a way to maximize the smallest user capacity, where subbands are allocated an equal amount of power. In [11], an iterative subband assignment followed by waterfilling for power allocation has been applied in order to maximize the sum-rate capacity. In [7], several allocation techniques are proposed for optimizing spectrum allocation in such a way to minimize the total amount of allocated bandwidth, under the constraints of requested services by multiple users and total transmission power.

In the aim of further leveraging the spectral efficiency of orthogonal multiplexing techniques, Non-Orthogonal Multiple Access (NOMA) has recently emerged as a promising candidate solution for systems beyond 4G, within the Future Radio Access (FRA). In this technique, multiple users are multiplexed in the power-domain, at the transmitter side, on top of the OFDM layer, and multi-user signal separation is conducted at the receiver side, based on Successive Interference Cancellation (SIC) [12–19].

Since dynamic resource allocation techniques have shown their efficiency in spectrum and power optimization within OFDM systems, and since a flexible radio interface is required in future 5G systems, we aim throughout this work to propose new strategies for the optimization of resource allocation based on NOMA. By combining the potential gain

achieved by NOMA through power multiplexing, with a proper optimization of the amount of used bandwidth, we seek to boost the spectral efficiency, while respecting a set of traffic requirements. However, using NOMA within such strategies for resource optimization is not straightforward. This is mainly due to the fact that, in order to respect the resource requirements of a set of users, each one of the latter needs to be multiplexed on several subcarriers, with different cohabiting users and in various multiplexing orders within each subband. For this purpose, several design aspects will be taken into consideration throughout this study, such as multi-user frequency scheduling, power allocation among subbands, power repartition between scheduled users within a subband, concentration and prioritization of users, etc., as well as the interaction of these different design issues.

So far, the majority of existing works dealing with NOMA have investigated the system-level performance in terms of system capacity and cell-edge user throughput, compared to previous systems adopting orthogonal signaling as a multiple access scheme.

In [17], the NOMA concept using a SIC decoder as a baseline receiver scheme is presented and assessed. System-level evaluation is done with respect to OMA (i.e. when a subband is orthogonally divided, in bandwidth and in power, among scheduled users). Results show that the use of NOMA improves both the system capacity and the cell-edge user throughput. Furthermore, a possible extension of NOMA including a MIMO scheme is investigated in order to achieve further capacity gain.

In [20], the benefits of NOMA over OMA are discussed under practical considerations. In this sense, the impact of error propagation within the SIC receiver is studied and is shown to be negligibly small, even in high mobility conditions. Furthermore, the cell throughput achieved by NOMA is shown to be higher by more than 30 % compared to OMA, under multiple configurations and setups.

In [21], the performance of NOMA is evaluated in a downlink system with randomly deployed mobile users, where two different situations are considered. In the first situation, each user has a targeted data rate based on its quality of service, and outage probability is considered as the metric for performance evaluation. In the second situation, user resources are allocated opportunistically according to encountered channel conditions (i.e. without specifying target data rates), and the ergodic sum rate achieved by NOMA is assessed. It is shown that, in both situations, NOMA can achieve better performance compared to orthogonal multiple access, under the condition that the users targeted data rates and the total allocated power are carefully chosen.

In [22], the performance of downlink NOMA with wideband and subband frequency scheduling is evaluated under wide-area cellular system configurations. Power allocation is done such that the total transmit power per subband is common to all subbands, and intra-subband power repartition between each pair of scheduled users is done as a subsequent stage using fractional transmit power allocation (FTPA) [12].

In [23], the throughput improvement using non-orthogonal superposition of users on top of an uplink OFDMA-based system is studied. All mobile users are considered to be transmitting signals at the same power level. Optimized scheduling techniques are proposed in such a way to maximize system throughput. In this sense, a cost function is assigned to each possible pair of users and the Hungarian method is used to solve the problem of user pairing. Cost functions are chosen such as to optimize either the sum-rate or the weighted sum-rate. These two scheduling techniques are shown to provide significant improvements in sum-rates and cell-edge rates compared to orthogonal signaling.

In [24], a non-orthogonal multiuser beamforming system is proposed for improving the system capacity. Degradation in the sum capacity due to inter-cluster interference and inter-user interference is assessed, leading to a clustering and power allocation algorithm

that aims at reducing interference and improving capacity. Clustering is done by selecting two users to be paired together, in the same beamforming vector, as having a large gain difference and a high correlation. In addition, power is equally divided between clusters, and then allocated among scheduled users in each cluster, in such a way to maximize the sum capacity, while guaranteeing a minimum capacity for the weakest user.

In [16], the system-level throughput performance of non-orthogonal access with a SIC in downlink is investigated, using a scheduling algorithm based on the proportional fair scheduler. Power is equally partitioned among subbands, and an optimal power allocation strategy is proposed in order to allocate power iteratively among scheduled users within the same subband. The proposed scheduling algorithm followed by intra-subband optimal power allocation has shown to achieve a good tradeoff between total and cell-edge user throughputs. In addition, it allows enhancing system-level throughput compared to orthogonal access. However, although the optimal power allocation strategy shows good performance, it is computationally complex. For this purpose, two simplified suboptimal intra-subband power allocation techniques are also proposed to reduce the complexity of the optimal scheme.

Altogether, as far as power allocation is concerned, the majority of the listed works have considered an equal repartition of power between subbands, followed by an intra-subband power allocation scheme [16, 20, 22, 24]. The aim of applying equal power allocation was to circumvent the problem of optimally distributing the power between subbands supporting two or more users with different channel amplitudes on the same subbands. In this context, the present work aims at proposing several promising solutions that significantly enhance the power allocation by a proper unequal inter-subband allocation.

On the other hand, in [17, 20], it is assumed that all participating users have access to the whole spectrum (i.e. to all frequency subbands), which avoids the need for complicated scheduling algorithms, but restricts the study to very low numbers of simultaneous users.

As in the previous works dealing with NOMA, we aim throughout this work to improve the spectral efficiency and the cell-edge user throughput. Additionally, our proposed algorithm for the dynamic assignment of available subbands seeks to achieve two goals. The first one is to reduce the amount of used bandwidth. In fact, the optimization of the amount of occupied bandwidth is of a great interest for the upcoming generations of mobile communications, in order to cope with the increasing scarcity in spectrum and, at the same time, meet the exploding demands of higher data rates. One application of particular interest in this context is cognitive radio [7–9], since the adaptivity in assigning spectral bands to users in a non-selfish way will allow freeing an optimized amount of bandwidth that may be used by other cognitive systems or other mobile operators. The second goal of our work is to improve system capacity in such a way to provide each user with a requested service data rate. Our proposed allocation technique also takes into consideration the fact that, in certain situations and on specific subbands, NOMA may not constitute the appropriate solution. Therefore, a dynamic switching between NOMA and a classical OFDMA (i.e. OFDMA without the additional non-orthogonal signaling layer) is allowed under certain conditions. Note that, in [20], a dynamic switching from NOMA to OMA is mentioned, such that the proposed algorithm for resource allocation can be either entirely based on NOMA or entirely on OMA, for the sake of performance comparison. In other words, all subbands bear the same signaling scheme, either OMA or NOMA, and simulation results are separately presented for the two cases. Moreover, no criterion is proposed to enable this switching in a dynamic way or to vary the signaling scheme within subbands between NOMA and OMA.

Finally, the proposed framework supports a large number of users and therefore shows a great robustness in the context of crowded areas, in terms of the occupied bandwidth and the probability of success (i.e. the probability of succeeding to respect required data rates by all users).

This paper is organized as follows: Section 2 gives a general description of NOMA with SIC. Then, Sect. 3 details the proposed iterative methods for spectrum optimization. Section 4 gives a brief comparison with the allocation system based on OMA. Simulation results are given in Sect. 5, as well as the computational load of the different proposed allocation techniques, while Sect. 6 concludes the paper.

2 Basic Description of Noma with SIC

In this section, we describe the general concept of NOMA including user multiplexing at the base station (BS) transmitter and signal separation at the user terminal.

We assume throughout this paper a downlink system with a single transmitter and single antenna receiver. We consider K users per cell, and a frequencyselective scheduling, where system bandwidth B is divided into S subbands.

We assume that the multiuser scheduler selects, among the K users, a set of users $N_s = \{k_1, k_2, \dots, k_n, \dots, k_{n(s)}\}$, to be scheduled at a frequency subband s , ($1 \leq s \leq S$). Term k_n indicates the n -th ($1 \leq n \leq n(s)$) user scheduled at subband s , and $n(s)$ denotes the number of scheduled users at frequency subband s . At the BS transmitter side, the information sequence of each scheduled user at subband s is independently coded and modulated. The coded and modulated symbol of the n -th scheduled user at subband s is x_{s,k_n} . Thus, the transmitted signal, x_s , by the BS, at a certain subband s , is a simple power multiplexing of the symbols generated by the $n(s)$ scheduled users:

$$x_s = \sum_{n=1}^{n(s)} x_{s,k_n}, \text{ with } E \left[|x_{s,k_n}|^2 \right] = P_{s,k_n} \tag{1}$$

where P_{s,k_n} is the power allocated to user k_n at subband s . The received signal vector of user k_n at subband s , y_{s,k_n} , is represented by:

$$y_{s,k_n} = h_{s,k_n} x_{s,k_n} + w_{s,k_n} \tag{2}$$

where h_{s,k_n} is the channel coefficient between the n -th user, k_n , at subband s and the BS. w_{s,k_n} represents the received Gaussian noise plus inter-cell interference by user k_n on subband s . Let P_{max} be the maximum allowable transmit power by the BS. The sum power constraint is therefore written as follows:

$$\sum_{s=1}^S \sum_{n=1}^{n(s)} P_{s,k_n} = P_{max} \tag{3}$$

Multi-user signal separation is conducted at the receiver side using the SIC process [25]. The optimal order for user decoding is in the increasing order of the users' channel gains normalized by the noise and inter-cell interference $h_{s,k_n}^2/n_{s,k_n}$, where h_{s,k_n}^2 is the equivalent channel gain and n_{s,k_n} the average power of w_{s,k_n} . Therefore, any user can correctly decode signals of other users whose decoding order comes before that user. In other words, user k_n at subband s can remove the inter-user interference from the j -th user, k_j , at subband s ,

whose $h_{s,k_j}^2/n_{s,k_j}$ is lower than $h_{s,k_n}^2/n_{s,k_n}$, and treats the received signals from other users whose $h_{s,k_j}^2/n_{s,k_j}$ is higher as noise [13, 26].

Assuming successful decoding and no error propagation, and supposing that inter-cell interference is randomized such that it can be considered as white noise [16, 23], the throughput of user k_n , at subband s , R_{s,k_n} , is given by:

$$R_{s,k_n} = \frac{B}{S} \log_2 \left(1 + \frac{h_{s,k_n}^2 P_{s,k_n}}{\sum_{j \in N_s, \frac{h_{s,k_n}^2}{n_{s,k_n}} < \frac{h_{s,k_j}^2}{n_{s,k_j}}} h_{s,k_n}^2 P_{s,k_j} + n_{s,k_n}} \right) \tag{4}$$

For the sake of simplicity, we assume throughout this paper that only two users are selected from K to be scheduled at each subband s ($1 \leq s \leq S$). In this case, considering user 2 having a lower channel gain than user 1 ($h_{s,k_2}^2 < h_{s,k_1}^2$), user 2 does not perform SIC since it comes first in the decoding order, but rather considers the signal designated to user 1 as noise. On the contrary, user 1 first decodes the signal x_{s,k_2} designated to user 2 and subtracts its component from the received signal x_s . Then, it decodes its own signal without interference from x_{s,k_2} . Based on Eq. (4), the throughput of user k_n at subband s , R_{s,k_n} ($1 \leq n \leq 2$), can be written as:

$$R_{s,k_1} = \frac{B}{S} \log_2 \left(1 + \frac{h_{s,k_1}^2 P_{s,k_1}}{N_0 \frac{B}{S}} \right) \tag{5}$$

$$R_{s,k_2} = \frac{B}{S} \log_2 \left(1 + \frac{h_{s,k_2}^2 P_{s,k_2}}{h_{s,k_2}^2 P_{s,k_1} + N_0 \frac{B}{S}} \right) \tag{6}$$

where N_0 is the power spectral density of the additive white Gaussian noise, including inter-cell interference, and assumed to be constant over all subbands. Note that, in most papers dealing with resource allocation in downlink NOMA [17, 20, 24, 26], the number of users per subband was set to two in order to limit the SIC complexity in the mobile receiver, except for [16, 27] where this number can respectively reach 3 and 4. However, in the last two cases, static power allocation is assumed, which simplifies the power allocation step but degrades throughput performance.

It can be seen from (5) and (6) that the choice of the multiplexed users over subband s , as well as the amount of power allocated to each user, significantly affects user throughput performance. Therefore, a flexible radio interface allowing an optimized resource allocation, depending on transmission conditions and user pairing strategies, is needed to explore the gain experienced with NOMA towards orthogonal-based multi-user allocation schemes, like OMA or OFDMA.

3 Description of the Proposed Iterative Method for Resource Allocation

3.1 Formulation of the Resource Allocation Problem

In addition to maximizing system throughput, as performed in the majority of existing literature on NOMA, this work targets minimizing the amount of used bandwidth. In other words, the

proposed allocation technique tends to provide each user with its requested data rate with the minimum number of subbands, under the constraint of a maximum allowed transmit power.

Let S_A be the actual number of available subbands ($1 \leq S_A \leq S$), i.e. $S - S_A$ subbands are supposed to be occupied by another system. $R_{k,requested}$ is the data rate requested by user k from the BS, $1 \leq k \leq K$, $P_{s,k}$ the transmit power of user k over subband s ($P_{s,k} \neq 0$ if k is scheduled on s), $R_{s,k}$ the achieved data rate by user k over subband s , and S_k the set of all subbands allocated to user k . The optimization problem can be formulated as follows:

$$\underset{P_{s,k}}{\text{minimize}} \sum_{k=1}^K \text{card}(S_k) \tag{7}$$

Subject to

$$\sum_{s \in S_k} R_{s,k} = R_{k,requested}, \quad \forall k, 1 \leq k \leq K \tag{8}$$

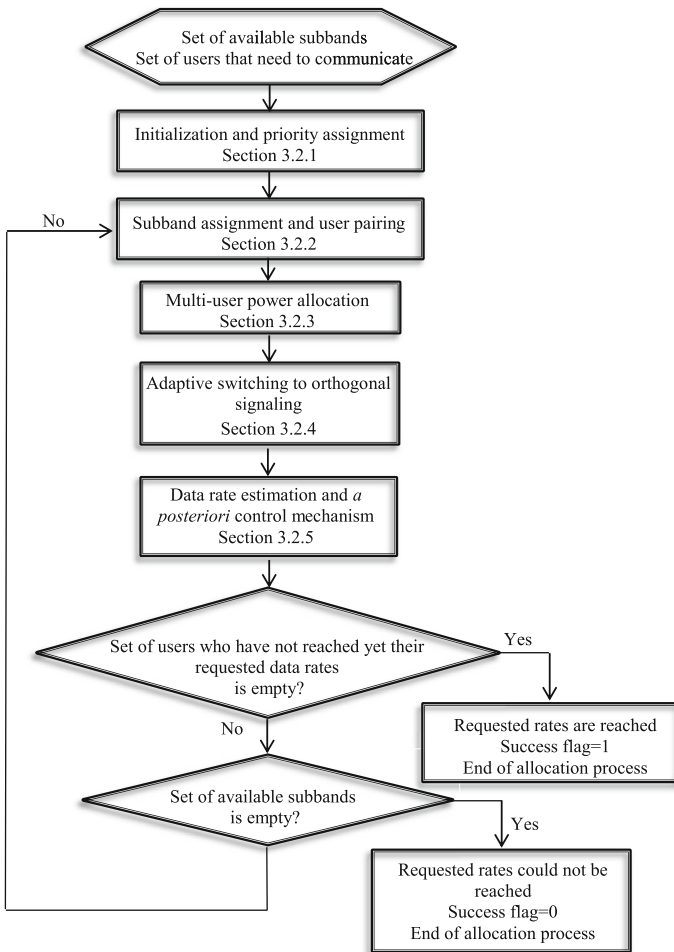


Fig. 1 Proposed allocation algorithm

$$\sum_{k=1}^K \left(\sum_{s \in S_k} P_{s,k} \right) \leq P_{\max} \tag{9}$$

$$P_{s,k} \geq 0, \quad \forall s \in S_k, \quad 1 \leq k \leq K \tag{10}$$

where $\text{card}(S_k)$ represents the cardinality of the set of subbands allocated to user k . If user k has a channel gain over s that allows him to perform SIC, his data rate $R_{s,k}$ is computed based on Eq. (5). Otherwise, it is computed based on Eq. (6).

Equation (7) represents the main design function. It aims at minimizing the number of allocated subbands under the rate and power constraints expressed by Eqs. (8–10). This optimization problem is combinatorial and cannot be resolved to yield closed-form solutions. Besides, an exhaustive search for the values of $P_{s,k}$ and S_k is impractical, because of the large number of parameters and constraints involved. Moreover, compared to previous orthogonal-based allocation techniques [7–9], resource allocation for the non-orthogonal system must also consider the following additional design constraints: the choice of user pairing scheme, the power distribution between allocated subbands, and the power division between paired users within a subband. Therefore, the allocation technique proposed in Fig. 1 aims to efficiently take all of these design constraints into consideration, by dividing the overall optimization problem into several steps that will be detailed in the sequel.

3.2 Description of the Proposed Algorithm for Resource Allocation

3.2.1 Initialization and Priority Assignment

In order to properly set user priorities, the BS relies on the actual transmission rates of all considered users, as well as on their Channel State Information (CSI), i.e. the channel gains between mobile terminals and the BS. Therefore, it is assumed, in this work, that CSI information is periodically transmitted from mobile users to the BS on dedicated control channels. The channel gains are grouped in a matrix \mathbf{H} with dimensions $S_A \times K$, where $h_{s,k}$ is the channel gain experienced by user k on subband s .

At the beginning of the allocation process, transmit powers $P_{s,k}$ and user rates $R_{s,k}$ are all set to zero. Therefore, at this point, priorities for channel allocation are defined at the BS based on the channel gain matrix \mathbf{H} represented in Fig. 2:

Fig. 2 Channel gain matrix \mathbf{H}

$$\mathbf{H} = \begin{bmatrix} h_{1,1} & \cdots & h_{1,k} & \cdots & h_{1,K} \\ \vdots & & \vdots & & \vdots \\ h_{s_{\text{best},1}} & \ddots & \vdots & & \vdots \\ \vdots & & \vdots & & \vdots \\ h_{s,1} & & h_{s,k} & & h_{s,K} \\ \vdots & & \vdots & & \vdots \\ \vdots & & h_{s_{\text{best},k}} & \ddots & \vdots \\ \vdots & & \vdots & & \vdots \\ h_{S_A,1} & \cdots & h_{S_A,k} & \cdots & h_{S_A,K} \end{bmatrix}$$

- For each user k , select the highest channel gain $h_{\text{best},k}$ among the elements in the k th column of matrix \mathbf{H} (denoted by a circle in Fig. 2).
- The user with highest (resp. lowest) priority is the one having the lowest (resp. highest) best channel gain among circled elements.

3.2.2 Subband Assignment and User Pairing

During the iterative process depicted in Fig. 1, users to be paired together over an assigned subband s_f are identified by applying the following steps:

Step 1: User selection

Select user k_1 among the set of users that need to communicate, and that have not reached their target rate yet. Selection is based on one of the following criteria:

- While there exists at least 2 users whose data rates are zero, select user k_1 based on the priority constraints defined in Sect. 3.2.1.
- Once the above criteria is no longer verified, i.e. data rates of all users are non-zero (each user has at least one attributed subband, with a non-zero transmission power on this subband), or it still exists only one user whose data rate is equal to zero, select user k_1 as the one showing the largest rate distance (or gap) towards its requested service data rate.

Step 2: Subband assignment

Attribute the most favorable subband (the one with the highest channel gain), denoted by s_f , to user k_1 . s_f is then removed from the set of available subbands.

Step 3: User pairing

Select user k_2 to be multiplexed in the power domain with user k_1 on the current subband s_f . User pairing can be done in different ways. In this work, we have evaluated two options:

Pairing 1:

User k_2 is chosen as the user having the next lowest channel gain over s_f , when compared to the one of k_1 .

Pairing 2:

User k_2 is chosen as the user having the worst channel gain over s_f .

In both pairing options, the channel gain of user k_2 is lower than that of k_1 . Therefore, user k_2 does not perform SIC. Instead, his corresponding receiver considers the signal of user k_1 as interfering noise with as the interfering term. In the case where we extend the number of scheduled users per subband to be more than two, steps 1 and 2 are kept the same, and step 3 is modified such as $n(s_f) - 1$ users are chosen to be multiplexed with k_1 on the current subband s_f , with $n(s_f)$ the number of scheduled users on s_f . User scheduling can still be performed in two ways:

- The $n(s_f) - 1$ other users are chosen as those having channel gains strictly less than that of k_1 .
- Users having channel gains lower than that of k_1 are divided into $n(s_f) - 1$ groups, and users to be scheduled on s_f are chosen as those having the worst channel gain in each group.

Step 4: Inverting roles

If, during the allocation process, it happens that user k_1 has the lowest channel gain on its attributed subband s_f (i.e. all other users have higher gains on this subband, compared to

k_1), user k_2 is then chosen as the user having the highest gain on this subband, if pairing 2 is adopted in step 3. Otherwise, i.e. if pairing 1 is used, k_2 is chosen as the user having the next highest channel gain over s_f , when compared to the one of k_1 .

3.2.3 Multi-user Power Allocation

In order to distribute power among users, several power allocation techniques are proposed in this section:

3.2.3.1 Optimum Waterfilling-Based Power Allocation In [22, 26], static power allocation is used for NOMA, where the total transmit power is identically divided between subbands. However, it is stated that the resulting achievable throughput is penalized since the waterfilling principle [28, 29] is not applied. For this reason, we propose to apply a waterfilling-based optimal subband power allocation that takes into consideration the channel gains of the two paired users within each subband. It is described by the following optimization problem:

At each stage of the allocation process, maximize the total achieved throughput for users that have not yet reached their requested data rate, under the constraint of the total remaining power:

$$\underset{\{P_{s,k_1}, P_{s,k_2}, \forall s \in S_u\}}{\text{maximize}} \sum_{s \in S_u} (R_{s,k_1} + R_{s,k_2}) \tag{11}$$

Subject to

$$\sum_{s \in S_u} (P_{s,k_1} + P_{s,k_2}) = P_{rem} \tag{12}$$

S_u is the set of subbands attributed to users whose target data rates have not been reached so far (those users constitute a set U), and P_{rem} denotes the remaining transmit power to be distributed between subbands, at the current stage of the allocation technique.

Solving this optimization problem using Lagrange multipliers leads to the following formulation of the objective function, where λ is the Lagrange multiplier:

$$J = \sum_{s \in S_u} \frac{B}{S} \log_2 \left(1 + \frac{P_{s,k_1} h_{s,k_1}^2}{N_0 \frac{B}{S}} \right) + \sum_{s \in S_u} \frac{B}{S} \log_2 \left(1 + \frac{P_{s,k_2} h_{s,k_2}^2}{P_{s,k_1} h_{s,k_2}^2 + N_0 \frac{B}{S}} \right) + \lambda \left(P_{rem} - \sum_{s \in S_u} (P_{s,k_1} + P_{s,k_2}) \right) \tag{13}$$

Generally, power multiplexing in NOMA is done such that the highest power is given to the user with the weakest channel (user k_2 in our case) [4, 12, 16]. Therefore, we adjust the power allocation ratio between P_{s,k_1} and P_{s,k_2} by setting a parameter β_s such that:

$$P_{s,k_2} = \frac{1 - \beta_s}{\beta_s} P_{s,k_1}, \quad \text{with} \quad \frac{1}{2} \leq \beta_s \leq 1 \tag{14}$$

By substituting (14) in (13), then differentiating J with respect to and λ , and by setting the resulting expressions to zero, we obtain a non-linear system described in (15) by equations with $N_u + 1$ unknowns, P_{s,k_1} and λ , N_u being the current number of elements in S_u .

$$\left\{ \begin{aligned} & \frac{\frac{h_{s,k_1}^2}{N_0} \frac{B}{S}}{1 + \frac{h_{s,k_1}^2}{N_0} \frac{B}{S} P_{s,k_1}} + \frac{\frac{1 - \beta_s}{\beta_s} \frac{h_{s,k_2}^2}{N_0} \frac{B}{S}}{\left(1 + \frac{h_{s,k_2}^2}{N_0} \frac{B}{S} P_{s,k_1}\right) \left(1 + \frac{1}{\beta_s} \frac{h_{s,k_2}^2}{N_0} \frac{B}{S} P_{s,k_1}\right)} = \lambda \ln 2 \frac{1}{\beta_s} \frac{S}{B}, s \in S_u \quad (15) \\ & \sum_{s \in S_u} \frac{1}{\beta_s} P_{s,k_1} = P_{rem} \end{aligned} \right.$$

Note that, in the particular case where $P_{s,k_2} = 0$ ($\beta_s = 1$), we get a linear system:

$$\frac{\log_2(e) \frac{B}{S}}{P_{s,k_1} + \frac{N_0 B}{h_{s,k_1}^2}} = \lambda, s \in S_u \quad (16)$$

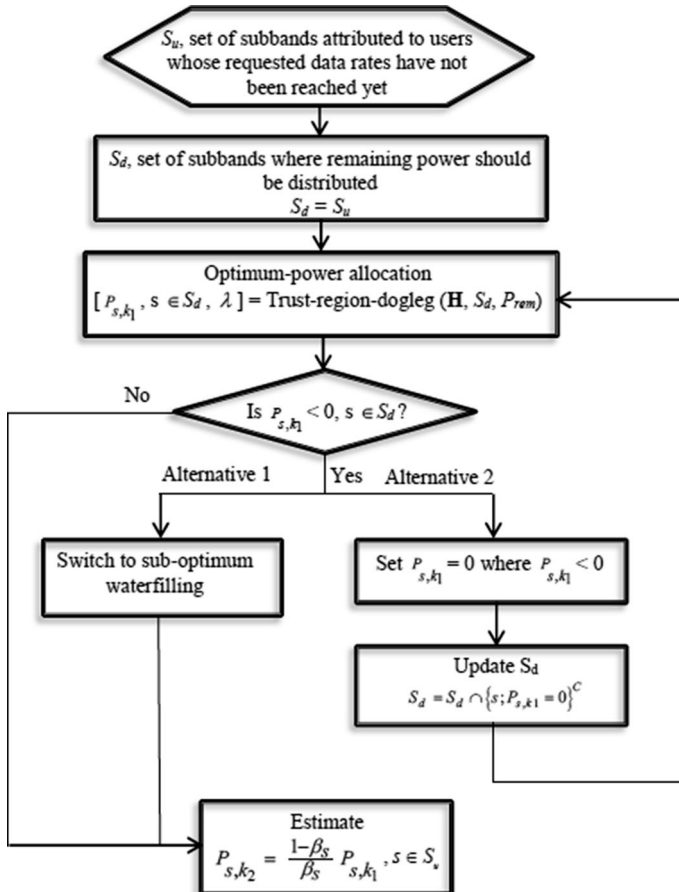


Fig. 3 Optimum waterfilling-based power allocation

This linear system corresponds to the solution of the optimization problem using the classical OFDM with orthogonal signaling as a multiple access technique (i.e. with no user cohabitation).

In order to solve the above non-linear system of equations described in (15), we use a numerical solver, the trust-region method based on the Dogleg algorithm [30, 31]. This algorithm determines the values of the $N_u + 1$ unknowns P_{s,k_1} and λ . However, it does not always guarantee non-negative solutions. To overcome this problem, we propose two alternatives that can be applied when at least one negative power is found at a certain stage of the iterative allocation process:

Alternative 1: switch, at this stage of the iterative process, to a sub-optimum solution for power allocation such as the one subsequently described in Sect. 3.2.3.2

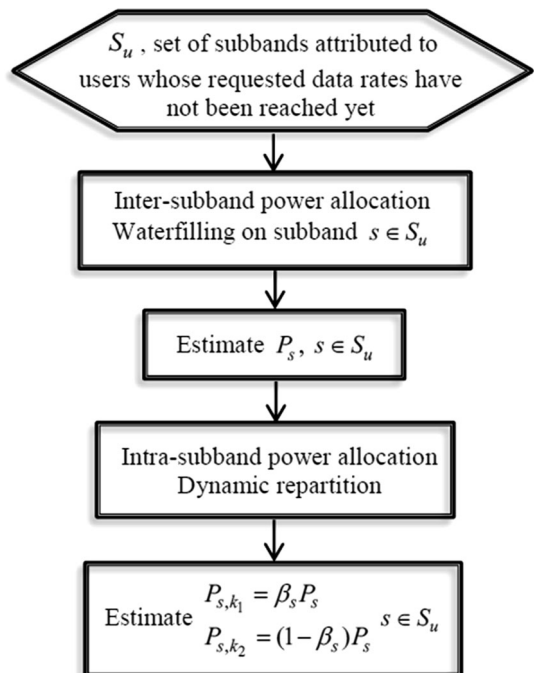
Alternative 2: substitute the negative powers by zeros, and re-distribute the remaining power P_{rem} to the set of subbands where power was found to be positive. This redistribution is done using the same trust-region method, and the process is iterated until only positive solutions are found.

This optimum power allocation strategy, including both alternatives, is depicted in Fig. 3.

To extend the number of scheduled users per subband to more than two, Eqs. (11) and (12) are to be respectively replaced by Eqs. (17) and (18), with the appropriate rate expressions expressed using Eq. (4):

$$\underset{\{P_{s,k_n}, \forall k_n \in N_s, 0 \leq s \leq S_u\}}{\text{maximize}} \sum_{s \in S_u} \sum_{n=1}^{n(s)} R_{s,k_n} \tag{17}$$

Fig. 4 Sub-optimum waterfilling-based power allocation



Subject to

$$\sum_{s \in S_u} \sum_{n=1}^{n(s)} P_{s,k_n} = P_{rem} \tag{18}$$

3.2.3.2 Sub-optimum Waterfilling-Based Power Allocation The optimum solution, described in the preceding section, performs a waterfilling-based allocation that takes into consideration the channel gains of all paired users. This solution reveals to be rather complex to implement. Therefore, we also propose a sub-optimum solution, where the power is allocated among users in two consecutive stages, inter-subband and intra-subband allocation, as shown in Fig. 4.

Stage 1: Inter-subband power allocation

In this first step, we propose to consider only the highest channel gain (i.e. for user k_1) within each subband. In other words, the channel gain of k_1 on subband s determines the total amount of power, P_s , that will be attributed to s , using a waterfilling process, and that will be subsequently partitioned between the two paired users on s . The waterfilling process is performed in an iterative way as in [9]. Even though this allocation technique represents a sub-optimum solution, it is expected to perform better than static power allocation that equally partitions power between subbands.

Stage 2: Intra-subband power allocation

Power is now to be partitioned between paired users within each subband. Intra-subband repartition can be done in a static way, according to a fixed threshold, or in a dynamic way, based on paired users channel gains.

- *Static intra-subband power allocation: Fixed Power Allocation (FPA)*

The repartition is done in a static way over all subbands, where the total transmit power P_s , allocated in stage 1 to subband s , is divided between paired users according to $(\beta P_s, (1 - \beta)P_s)$, with β ($0 \leq \beta \leq 0.5$) being a constant parameter over all subbands. The user with the highest channel gain will be given βP_s and the paired user will be given the rest.

- *Dynamic intra-subband power allocation: Fractional Transmit Power Allocation (FTPA)*

The repartition is done in a dynamic way, similarly to the algorithm in [22] which is based on the channel gains of the two multiplexed users, such that β_s in (14) is given by:

$$\beta_s = \frac{h_{s,k_1}^{-2\alpha}}{h_{s,k_1}^{-2\alpha} + h_{s,k_2}^{-2\alpha}} \tag{19}$$

where α ($0 \leq \alpha \leq 1$) is a decay factor that accounts for the amount of power attributed to user k_2 (this amount is increased with α). α is kept constant over the subbands and is determined a priori via computer simulations, such that the achieved spectral efficiency is maximized.

If the number of multiplexed users is to be extended to more than 2, inter-subband power allocation can be performed similarly to the case of two multiplexed users. In other words, the highest channel gain among the $n(s)$ ($s \in S_u$) scheduled users is considered in the waterfilling process. As for the intra-subband power allocation, power is allocated to each scheduled user k_n ($n \in N_s$), within each subband s , based on FTPA, such that:

$$P_{s,k_n} = P_s \frac{(h_{s,k_n}^2)^{-\alpha}}{\sum_{j \in N_s} (h_{s,k_j}^2)^{-\alpha}}, s \in S_u \tag{20}$$

3.2.3.3 Optimum Weighted Waterfilling-Based Power Allocation Similarly to the optimum solution proposed in Sect. 3.2.3.1, we also propose to apply a weighted waterfilling-based power allocation. The weighted version still takes into consideration the channel gains of the two paired users within each subband, but it also adds a proper weight to each user. It is described by the following optimization problem:

$$\text{maximize}_{\{P_{s,k_1}, P_{s,k_2}\}} \sum_{s \in S_u} (a_{s,k_1} R_{s,k_1} + a_{s,k_2} R_{s,k_2}) \tag{21}$$

Subject to

$$\sum_{s \in S_u} (P_{s,k_1} + P_{s,k_2}) = P_{rem} \tag{22}$$

where a_{s,k_1} and a_{s,k_2} are positive weights that can be chosen using differing criteria, and that verify:

$$\sum_{s \in S_u} (a_{s,k_1} + a_{s,k_2}) = 1 \tag{23}$$

Solving this optimization problem using Lagrange multipliers leads to the formulation of the corresponding objective function J such as:

$$J' = \sum_{s \in S_u} a_{s,k_1} \frac{B}{S} \log_2 \left(1 + \frac{P_{s,k_1} h_{s,k_1}^2}{N_0 \frac{B}{S}} \right) + \sum_{s \in S_u} a_{s,k_2} \frac{B}{S} \log_2 \left(1 + \frac{P_{s,k_2} h_{s,k_2}^2}{P_{s,k_1} h_{s,k_2}^2 + N_0 \frac{B}{S}} \right) + \lambda \left(P_{rem} - \sum_{s \in S_u} (P_{s,k_1} + P_{s,k_2}) \right) \tag{24}$$

As in (13), after differentiating J with respect to P_{s,k_1} and λ , and setting the results to zero, we obtain a non-linear system of $N_u + 1$ equations with $N_u + 1$ unknowns and λ . We also solve this non-linear system using the trust-region Dogleg algorithm, in the same way as was done in Sect. 3.2.3.1. The problem of negative solutions is resolved using the second alternative (substituting the negative powers by zero and re-distributing the remaining power). This choice will be justified by the simulation results.

As for the choice of the weight attributed to every user, two possible schemes are proposed in the sequel.

Weights based on the actual rates of considered users

In order to give importance to users who are far from reaching their requested data rates, a weight is assigned to each user based on the quadratic distance between its actual achieved throughput and its requested data rate, such as:

$$a_{s,k_1} = \frac{(|R_{k_1,requested} - R_{k_1,tot}|)^{2\eta}}{\sum_{s \in S_u, i=1,2} (|R_{k_i(s),requested} - R_{k_i(s),tot}|)^{2\eta}}, 0 \leq a_{s,k_1} \leq 1 \tag{25}$$

$$a_{s,k_2} = \frac{(|R_{k_2,requested} - R_{k_2,tot}|)^{2\eta}}{\sum_{s \in S_u, i=1,2} (|R_{k_i(s),requested} - R_{k_i(s),tot}|)^{2\eta}}, 0 \leq a_{s,k_2} \leq 1 \tag{26}$$

where $\eta(\eta \leq 0)$ is a control parameter to be determined a priori via simulations such that the achieved spectral efficiency is maximized. $R_{k_1,tot}$ (resp. $R_{k_2,tot}$) is the actual total data rate of user k_1 (resp. k_2), and $R_{k_1,requested}$ (resp. $R_{k_2,requested}$) is the requested data rate by user k_1 (resp. k_2). Results in Sect. 5 will provide insight on the sensitivity of the system performance to the value of η . In general, increasing η will increase the power difference between users being far and users being near to reach their requested data rates.

Weights based on the position of users within the cell

In this option, weights are based on the geographical distance between each mobile user and the base station:

$$a_{s,k_1} = \frac{r_{s,k_1}^{2\phi}}{\sum_{s \in S_u, i=1,2} r_{s,k_i}^{2\phi}}, 0 \leq a_{s,k_1} \leq 1 \tag{27}$$

$$a_{s,k_2} = \frac{r_{s,k_2}^{2\phi}}{\sum_{s \in S_u, i=1,2} r_{s,k_i}^{2\phi}}, 0 \leq a_{s,k_2} \leq 1 \tag{28}$$

where r_{s,k_1} (resp. r_{s,k_2}) is the Euclidean distance between user k_1 (resp. k_2) and the base station. $\phi(\phi \geq 0)$ is a parameter to be determined offline via simulations such that the cell-edge user throughput and/or the spectral efficiency is maximized.

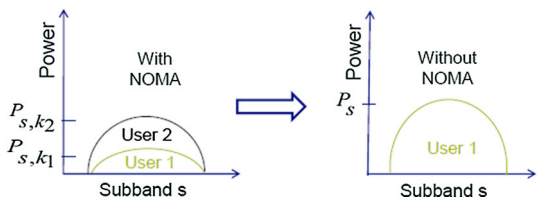
3.2.3.4 Power Allocation According to the Actual Achieved Throughput In this technique, power is allocated to users in two successive steps. First, it is partitioned among subbands in a proportional way to the distance between the actual achieved throughput and the requested data rate, using:

$$P_s = \frac{d_{s,k_1}^2 + d_{s,k_2}^2}{\sum_{s \in S_u} d_{s,k_1}^2 + \sum_{s \in S_u} d_{s,k_2}^2} \times P_{rem}, s \in S_u \tag{29}$$

where d_{s,k_1} (resp. d_{s,k_2}) is the distance between the actual throughput of user k_1 (resp. k_2) over subband s and its requested data rate (Fig. 5).

Then, power is divided within each subband based on the FTPA algorithm used in Sect. 3.2.3.2.

Fig. 5 Adaptive switching from NOMA to OS



3.2.3.5 Static Power Allocation In the majority of existing works related to NOMA, static power allocation is used, where the total transmit power is identically divided between subbands. In order to test a static power allocation scheme within the proposed resource allocation framework, we propose to equally distribute, at each stage of the iterative process, the remaining power P_{rem} among subbands containing users that have not reached yet their requested data rates. The power allocated to subband s is computed as:

$$P_s = \frac{P_{rem}}{N_u}, s \in S_u \quad (30)$$

Then, P_s is distributed between paired users on subband s using FTPA.

3.2.4 Adaptive Switching to Orthogonal Signaling

Improvement in spectral efficiency thanks to NOMA is not systematic. Indeed, sometimes the loss in data rate experienced by user k_1 , when sharing its subband with user k_2 , is greater than the data rate gain achieved by k_2 on this subband. In this case, NOMA is not the appropriate solution; in such case, we propose to allocate this subband to user k_1 alone.

The decision to switch to orthogonal signaling (OS) can be made by testing the following condition:

$$\gamma(R_s - R_{s,k_1}) > R_{s,k_2} \quad (31)$$

with

$$R_s = \frac{B}{S} \log_2 \left(1 + \frac{P_s h_{s,k_1}^2}{N_0 \frac{B}{S}} \right) \quad (32)$$

R_s is the data rate achieved on subband s without NOMA. When condition (31) is satisfied, the resource allocation technique automatically switches to orthogonal signaling, for the current subband s . $\gamma(0 < \gamma < 1)$ is a control parameter to be determined a priori via simulations such as to maximize the achieved spectral efficiency. In the simulation results, we will study the influence of γ on the system performance. In general, with increasing values of γ , the allocation process tends to switch more often to orthogonal signaling.

3.2.5 Data Rate Estimation and Control Mechanism

At the end of each subband assignment stage, with its power allocation to users k_1 and k_2 (as well as the other users from the set S_u), the algorithm re-estimates the data rates for users in the set S_u (Fig. 1). Then, it verifies if user k_1 (who has been attributed a subband during the latest stage of the algorithm) has reached its requested data rate, that is if the actual total data rate of k_1 , $R_{k_1,tot}$, is equal to $R_{k_1,requested}$. In such a case, user k_1 is removed from the set S_u , and the amounts of allocated power on subbands assigned to user k_1 (for k_1 and for the paired users on those subbands) are kept unvaried for the rest of the allocation process. In other words, such subbands will no longer participate in the power distribution step within the allocation process. Such subbands are then removed from the set S_u and their allocated powers are subtracted from the remaining power P_{rem} .

If it happens that the actual data rate is higher than the requested data rate ($R_{k_1,tot} > R_{k_1,requested}$), the total amount of power allocated to user k_1 should be reduced in such a way to reach the exact requested rate. Among the subbands allocated to user k_1 that

remain modifiable (i.e. on which k_1 is not paired with a user that has reached its requested rate), we adjust the power on subband s_a having the least channel amplitude. A similar procedure is also applied on user k_2 if it reaches its requested data rate.

When adjusting the transmit power of user k_1 on subband s_a , we encounter two cases:

The first case occurs when user k_1 exhibits the highest channel gain over s_a . The adjustment is then done as follows:

First, the transmission rate of k_1 over s_a is estimated using:

$$R_{s_a,k_1} = \frac{B}{S} \log_2 \left(1 + \frac{P_{s_a,k_1} h_{s_a,k_1}^2}{N_0 \frac{B}{S}} \right) \tag{33}$$

Then, this rate is subtracted from the actual total rate of user k_1 , yielding:

$$R_{rem} = R_{k_1,tot} - R_{s_a,k_1} \tag{34}$$

Now, the necessary data rate on s_a is estimated as:

$$R_{k_1,requested} - R_{rem} \tag{35}$$

The power of user k_1 over s_a is modified in such a way to yield the above estimated data rate:

$$P_{s_a,k_1} = \frac{2^{(R_{k_1,requested} - R_{rem}) \frac{S}{B}} - 1}{h_{s_a,k_1}^2} N_0 \frac{B}{S} \tag{36}$$

Since the power of user k_1 over s_a has been modified, the power of the collocating user should also be reduced according to Eq. (14) in order to maintain the same power ratio $(1 - \beta_{s_a}) / \beta_{s_a}$.

For the second case, when user k_1 exhibits the lowest channel gain over s_a , i.e. not performing SIC over s_a , power adjustment is done by modifying Eq. (33) using Eq. (6), and keeping Eqs. (34) and (35) unchanged. Eq. (36) is then replaced by Equation (37) using Eqs. (14) and (6).

$$P_{s_a,k_1} = \frac{\left(2^{(R_{k_1,requested} - R_{rem}) \frac{S}{B}} - 1 \right) N_0 \frac{B}{S}}{\left(1 - \left(\left(2^{(R_{k_1,requested} - R_{rem}) \frac{S}{B}} - 1 \right) \frac{\beta_{s_a}}{1 - \beta_{s_a}} \right) \right) h_{s_a,k_2}^2} \tag{37}$$

Sometimes, when trying to adjust the power of user k_1 over s_a , it can happen that R_{rem} is still greater than $R_{k_1,requested}$. In this case, another subband, having a channel gain higher than that of s_a , is chosen for power adjustment. The same kind of power adjustment is performed for user k_2 in case its total actual power is higher than its target data rate: $R_{k_2,tot} > R_{k_2,requested}$.

4 Comparison with OMA

In the majority of existing works dealing with NOMA, the system level-performance is mostly evaluated with respect to OMA [17, 20, 22], i.e. when a subband is orthogonally divided in bandwidth and in power between scheduled users. When we assume that OMA signaling is used instead of NOMA within our framework, the bandwidth of subband s is

partitioned between users k_1 and k_2 using two factors δ and $1 - \delta$ respectively, where ($0 < \delta < 1$). In addition, power allocation is performed in two stages: inter-subband power allocation based on the waterfilling principle (sub-optimum technique), followed by intra-subband power allocation based on an equal division of the power between collocating users (i.e. $P_{s,k_1} = P_{s,k_2}$). Since no interference between collocating users occurs in this case, the throughput of user k_i , ($i = 1, 2$) is computed by:

$$R_{s,k_1} = \frac{B}{S} \delta \log_2 \left(1 + \frac{P_{s,k_1} h_{s,k_1}^2}{\delta N_0 \frac{B}{S}} \right) \quad (38)$$

$$R_{s,k_2} = \frac{B}{S} (1-\delta) \log_2 \left(1 + \frac{P_{s,k_2} h_{s,k_2}^2}{(1-\delta) N_0 \frac{B}{S}} \right) \quad (39)$$

In the sequel, OMA signaling will be tested when substituting it to NOMA within our allocation technique and comparison towards NOMA will be done for the two pairing cases described in Step 3 within Sect. 3.2.2.

5 Numerical Results

5.1 Performance Evaluation

In this paper, we mainly consider four system-level performance indicators: the achieved system capacity, the amount of used bandwidth, the probability of success, and the cell-edge user throughput. The first two indicators can be merged into a single metric, the spectral efficiency, calculated as:

$$\text{Spectral_Efficiency} = \frac{\text{Achieved system capacity}}{\text{Amount of used bandwidth}} \quad (40)$$

Several techniques with different combinations of user pairing and multi-user power allocation schemes are evaluated and compared. The following acronyms will be used to refer to the main studied methods:

- NO_O_Opt_Alternative1: Combination of NOMA and OS, with the pairing 2 scheme and optimum waterfilling-based power allocation using alternative 1.
- NO_O_Opt_Alternative2: Combination of NOMA and OS, with the pairing 2 scheme and optimum waterfilling-based power allocation using alternative 2.
- NO_O_Weighted_Opt_rate_distance: Combination of NOMA and OS, with the pairing 2 scheme and optimum weighted waterfilling-based power allocation, where weights are based on distances to target rates of considered users.
- NO_O_Weighted_Opt_position: Combination of NOMA and OS, with the pairing 2 scheme and optimum weighted waterfilling-based power allocation, where weights are based on the geographical position of users in the cell.
- NO_O_WF: Combination of NOMA and OS with pairing 2 and sub-optimum waterfilling.
- NO_O_EP: Combination of NOMA and OS with pairing 2 and a static inter-subband power allocation scheme where power is equally divided among subbands (followed by FTPA for intra-subband power distribution).

- NO_O_rate_distance: Combination of NOMA and OS with pairing 2, the power being allocated based on the actual achieved throughput, as was described in Sect. 3.2.3.4.
- NO_WF: This technique refers to the case when switching to OS (as described in Sect. 3.2.4) is not allowed (i.e. $\gamma = 0$ in Eq. (31)). In this case, the allocation process is purely based on NOMA. Pairing 2 and sub-optimum waterfilling are used.
- O_WF: Only OS (classical OFDM) is applied and non-orthogonal cohabitation is not allowed; sub-optimum waterfilling is used.
- O_rate_distance: Classical OFDM is applied, the power being allocated based on the actual achieved throughput in the same way as it is done in NO_O_rate_distance, except that cohabitation is not allowed (k_2 is not present).
- OMA_WF: This technique corresponds to the case when OMA is used instead of NOMA. Bandwidth is equally divided between paired users ($\gamma = 0.5$). Waterfilling is applied as an inter-subband power allocation technique, followed by an equal repartition of power between paired users within each subband (as described in Sect. 4).

In our simulation setup, K users are randomly positioned following a uniform distribution in a 10 km radius cell, with a maximum path loss difference of 20 dB between users. K varies between 5 and 20. The system bandwidth B is 100 MHz and the maximum number of available subbands S is 128. The total transmit power of the BS is 1000 mW, and the requested data rate is set to 5 Mbps for each user. The noise power spectral density is 4.10^{-18} W/Hz. The transmission medium is modeled by a frequency-selective Rayleigh fading channel with a root mean square delay spread of 500 ns. Perfect knowledge of the channel gains of all users (i.e. perfect CSI) by the BS is assumed in this study.

5.2 Simulation Results

Simulations were first performed to validate the choices of different design parameters within our framework, in terms of user pairing, multi-user power allocation, and adaptive

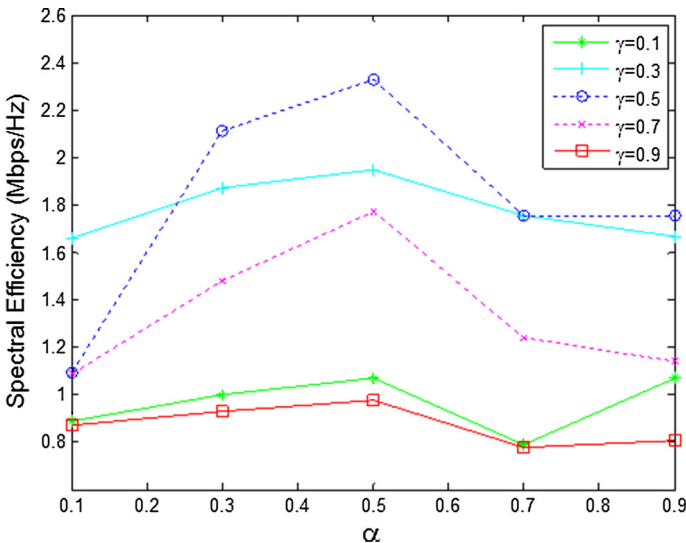


Fig. 6 Spectral efficiency of NO_O_WF for different values of α and γ

switching to OS. Then, robustness of the proposed system is evaluated in the context of communication in crowded areas.

First, we start by identifying the optimal values of the FTPA decay factor α and of the adaptive switching to OS parameter γ . Figure 6 shows the obtained spectral efficiency when NO_O_WF is evaluated for different α and γ values with $K = 10$ and the actual number of available subbands S_A is equal to 128 ($S_A = S$).

Spectral efficiency is maximized for $\alpha = 0.5$ and $\gamma = 0.5$. Similar optimal values were observed for different values of K and S_A . Therefore, these values of α and γ are adopted in the remainder of the study.

Now, the impact of user pairing and intra-subband power allocation strategies on system performance is evaluated, for $S_A = 128$. Figure 7 shows the spectral efficiency of NO_O_WF when FTPA and FPA are used. The effect of the two pairing techniques presented in Sect. 3.2.2 is also shown on the same graph. We notice that the combination between FTPA and pairing 2 outperforms FPA for different values of β , with a gain ranging from 16 % when the number of users is high, up to 40 % when the number of users per cell is equal to 5. This is due to the fact that FTPA allocates powers dynamically, by taking into account the encountered channel states by all users, whereas static repartition within a subband may be inconvenient for some paired users. In addition, the use of pairing 2, as implemented within the proposed allocation technique, is a reasonable choice since the performance gain of NOMA compared to orthogonal signaling increases with the difference in channel gains between scheduled users [16]. Therefore, using pairing 2, together with FTPA, shows the best performance.

In fact, when users having the largest possible gain difference within each subband are paired together (pairing 2), the power difference between their received signals should be large too thanks to the application of FTPA (see Eq. 19). Therefore, the amount of inter-user interference experienced by user k_2 (i.e. $P_{s,k_1} h_{s,k_2}^2$ the term in the denominator of R_{s,k_2})

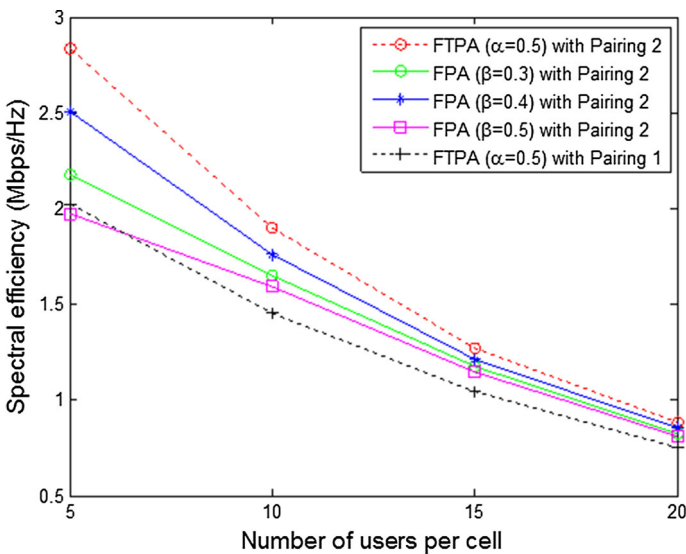


Fig. 7 Spectral efficiency of NO_O_WF for dynamic and fixed intra-subband power allocation schemes, and for different channel gain differences between paired users

is reduced, not only due to the choice of user k_2 (by reducing h_{s,k_2}^2) but also because the signal power P_{s,k_1} of user k_1 is lowered.

Next, in order to identify the optimal value of η that should be used as a design parameter, we evaluate the spectral efficiency of NO_O_Weighted_Opt_rate_distance for different values of η , when the number of users per cell varies between 5 and 20, and for an actual number of available subbands equal to 128. In fact, increasing η increases the amount of power attributed to the users that are far from reaching their target, at the expense of decreasing the amount of power allocated to the users whose data rates are close to the target.

Figure 8 shows that the incidence of η on system performance is especially significant when the number of users per cell is small, and that the spectral efficiency is maximized when $\eta = 1$; therefore, this value is adopted in the remainder of the study.

Now that the main system parameters have been chosen, the performance gain of the different proposed techniques is investigated in the context of a congested area for two different setups:

Case 1: The number of users per cell is equal to 10 and the actual number of available subbands ranges from 16 to 128, with a fixed subband bandwidth at 100/128 MHz.

Case 2: The actual number of available subbands is 128 ($S_A = S$) and the number of users per cell is varied between 5 and 20.

Figure 9 compares the spectral efficiency of the simulated methods, for the two simulation setups 1 (left) and 2 (right).

Let us first compare the four proposed allocation schemes NO_O_WF, NO_WF, NO_O_EP, and O_WF. In both setup cases, NO_O_WF outperforms the other 3 simulated methods. This gain in spectral efficiency is due to several reasons:

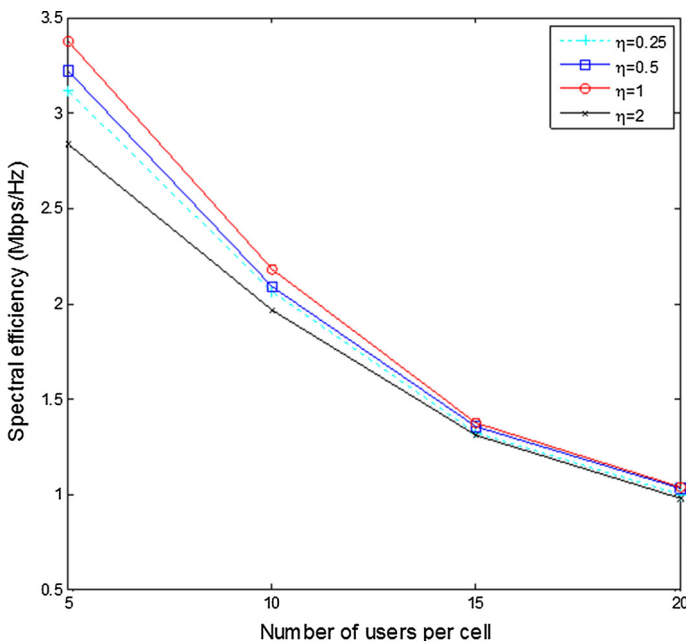


Fig. 8 Spectral efficiency of NO_O_Weighted_Opt_rate_distance for different values of η

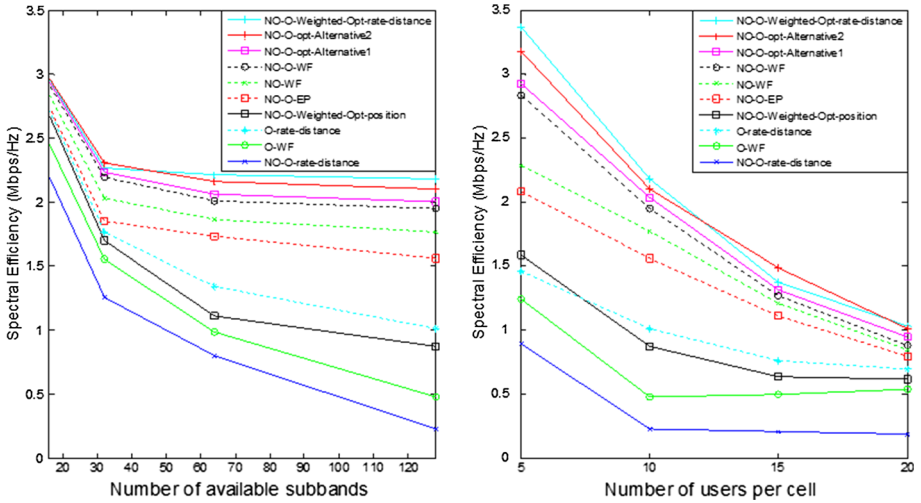


Fig. 9 Spectral efficiency of the proposed methods in terms of number of available subbands (*left*), and number of users per cell (*right*)

- The reduction in the amount of used bandwidth due to non-orthogonal cohabitation in the power domain makes NO_WF and NO_O_WF clearly outperform O_WF; i.e., a smaller total number of subbands is needed to achieve the data rate targets.
- The improvement in system capacity due to the waterfilling process helps NO_WF and NO_O_WF outperform NO_EP. This capacity is expressed by the total achieved

throughput $\sum_{k=1}^K \sum_{s \in S_k} R_{s,k}$ at the end of the allocation process, which is also the sum of target data rates by all users in case of success.

- The use of a dynamic adaptive switching to orthogonal-based system improves NO_O_WF performance with respect to NO_WF.

For instance, with 32 available subbands, and for a number of users per cell equal to 10, NO_O_WF has a spectral efficiency of 2.2 Mbps/Hz compared to 2.02, 1.85, and 1.5 Mbps/Hz with NO_WF, NO_EP, and O_WF, respectively. When the number of available subbands decays, the advantages of NO_O_WF, in terms of the achieved total data rate and the reduced amount of necessary bandwidth, are maintained. For example, when this number drops to 16, the observed spectral efficiency remains in favor of NO_O_WF and is respectively 2.9, 2.8, 2.7, and 2.4 Mbps/Hz for NO_O_WF, NO_WF, NO_EP, and O_WF. This shows the efficiency of non-orthogonal signaling in congested areas.

The two proposed optimal techniques, NO_O_Opt_Alternative1 and NO_O_Opt_Alternative2 show a significant improvement in the system performance with respect to the sub-optimum technique NO_O_WF, especially when the second alternative is applied, i.e., when negative values are replaced by zero followed by a power re-distribution. This result is due to the fact that these two techniques allow a joint inter and intra power distribution by resolving the optimization problem formulated in (11), rather than dividing the allocation process into two separate inter and intra stages, as in NO_O_WF and NO_WF. By doing so, the channel gains of all paired users are jointly taken into consideration when distributing power between subbands, instead of considering only the channel gain of the user with the highest gain in each subband.

When weights are introduced, the performance in the spectral efficiency of NO_O_Weighted_Opt_rate_distance is improved with respect to NO_O_Opt_Alternative1 and NO_O_Opt_Alternative2. Taking into account the distance between the actual throughput and the requested data rate for each user, in the power distribution, reveals to be more efficient than the non-weighted solutions, since it allows giving more priority and thus more power to users that are far from their requested data rates. At the same time, users weighted throughputs are being maximized using the waterfilling process. The gain obtained with NO_O_Weighted_Opt_rate_distance with respect to NO_O_Opt_Alternative2 is observed for a low number of users and a high number of available subbands.

As for NO_O_Weighted_Opt_position, this method presents an important loss in the average spectral efficiency, compared to NO_O_WF, due to the use of geographical position as weight. This weighted version tends to give more priority in the power allocation to cell-edge users. In fact, users being far from the BS require a high level of power in order to reach their requested data rates, since their channel gains on their allocated subbands generally tend to be low.

In this case, increasing their priority will cause the remaining users to be given low levels of power; therefore, those users will need more subbands in order to reach their requested data rates. At the same time, the amount of power taken from the close users does not significantly increase the data rate of far users (because of their low channel gains), leading to a decrease in the overall spectral efficiency. However, as will be shown later, the advantage of this technique resides in a slightly increased cell-edge user throughput, compared to the other allocation techniques.

On the other hand, when the power is directly distributed according to actual rates of users, without maximizing the total achieved throughput as done in NO_O_rate_distance, we fail to outperform O_WF. In NOMA, allocating power to a user in a proportional way to the distance between its actual throughput and its target data rate, without using the waterfilling principle, does not directly guarantee a maximization of the worst user throughput neither the total user throughput. This is mainly due to the collocation principle: Suppose k_1 is the least privileged user (the farthest from his target), at a certain stage of the algorithm. Since its distance d_{s,k_1}^2 in Eq. (29) is high, this user should get a higher amount of power on its allocated subbands, compared to other users. However, since the weight on any of its subbands also takes into consideration the distance of the collocating user d_{s,k_2}^2 , the amount of power attributed to a certain subband of k_1 will be lower than necessary if d_{s,k_2}^2 is low. This leads to a power allocation far from being optimal. However, when applying this allocation technique to OS (O_rate_distance), i.e. without user cohabitation, this strategy shows to be of a great interest for classical OFDM, since it outperforms the widely used waterfilling technique (O_WF). This is explained by the fact that our target in this work is to allow users reach their requested data rates, and not only to maximize the average throughput, which is commonly achieved by the waterfilling process. For classical OFDM, this target is reached much more quickly (i.e. with a lower number of subcarriers) with a power allocation proportional to rate distances, than with waterfilling.

Since our goal throughout this study is to allow users reach their requested data rates, while maximizing the total achieved throughput, the proposed techniques should also be compared based on their abilities to reach this goal. When users reach their requested data rates, the success flag of the proposed allocation technique is set to 1, otherwise, it is set to zero (Fig. 1). For a high number of conducted simulations, Fig. 10 shows the average

success probability of the proposed techniques for different numbers of users per cell, and an actual number of available subbands equal to 128.

Although the success rate decreases when the number of users per cell increases, the proposed techniques based on NOMA greatly outperform O_WF for a large numbers of users. For instance, starting from a number of users of 15, O_WF fails to provide a solution for the allocation problem, whereas the other proposed techniques based on NOMA still succeed with a probability between 23 % (for $K = 20$) and 50 % (for $K = 15$).

The cell-edge user throughput is an important fairness evaluator for the allocation process. Figure 11 shows this metric as a function of the number of users per cell, where the number of available subbands is fixed to 128.

The cell-edge user throughput when using NOMA is always higher than that of orthogonal signaling. The gain can reach 21 % approximately with NO_O_Weighted_Opt_position, with respect to O_WF. The proposed technique for power allocation based on user position NO_O_Weighted_Opt_position shows the best performance. This is due to the fact that the rates in the optimized sum-rate function are weighted proportionally to the distance of users towards the BS. In this case, cell-edge users are given a high level of power, in order to increase their chance to reach the requested data rates.

It should also be noted that, when the number of users per cell is limited, waterfilling-based power allocation (especially NO_O_Weighted_Opt_position) shows a slightly higher cell-edge user throughput compared to equal power allocation. However, when the number of users per cell becomes large, the success rate of all allocation strategies generally decreases (Fig. 10). In such conditions, adopting a uniform power allocation or allocating powers while taking into consideration the geographical distribution of users yields similar cell-edge user throughput, which is higher than that of other allocation techniques. This is due to the fact that non-weighted waterfilling-based algorithms generally optimize the average throughput and may not give the best fairness to the cell-edge user, especially for large values of K . Therefore, we can conclude that when it comes to optimizing the cell-edge user throughput in crowded areas, equal power repartition may perform as good as a weighted optimized repartition, and better than a repartition that does

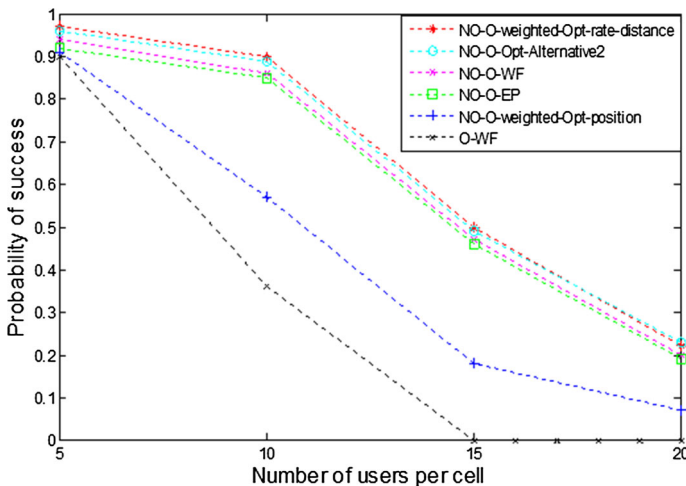


Fig. 10 Probability of success of the proposed methods in terms of the number of users per cell

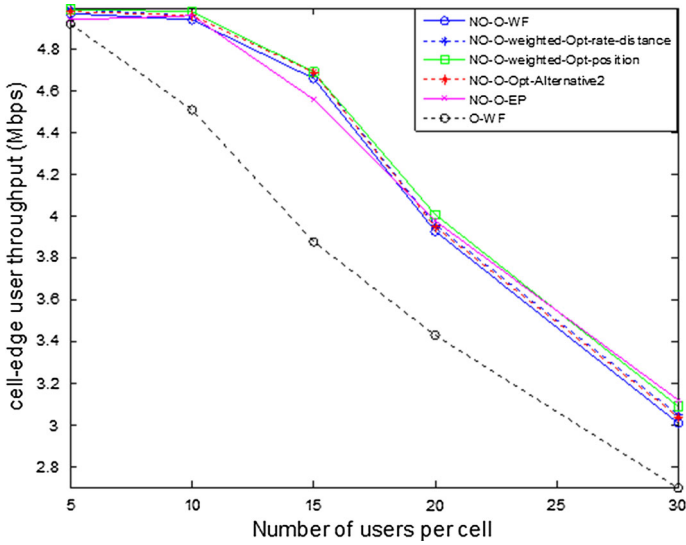


Fig. 11 Cell-edge user throughput as a function of the number of users per cell

not consider geographical positions. Nevertheless, the proposed non-weighted approaches still present important gains when compared to orthogonal signaling, not only in terms of spectral efficiency, but also in the degree of fairness.

In order to compare the proposed allocation techniques based on NOMA to the case where OMA is used within our allocation framework (Sect. 4), Fig. 12 shows the

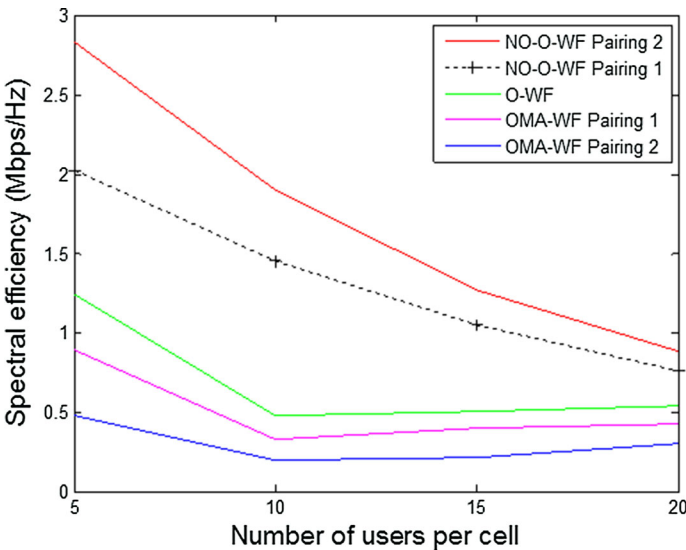


Fig. 12 Comparison of the spectral efficiency of NOMA and OMA for different channel gain differences between paired users

comparative results between NO_O_WF, O_WF, and OMA_WF, with the two pairing options described in Step 3 within Sect. 3.2.2.

NO_O_WF, with the two pairing options, shows significantly better performance than both O_WF and OMA_WF. On the other hand, as opposed to NOMA, OMA_WF based on pairing 1 gives better performance than pairing 2, because in OMA, there is no interference between users scheduled on the same subband, since the collocation is done by dividing both power and bandwidth. Therefore, the channel gain of the second collocating user does not need to be minimized, with respect to the first user, as is done in NOMA. Furthermore, it is shown that O_WF gives better performance than OMA_WF. This is due to the fact that the total throughput achieved per subband, when using OMA, is often lower than when assigning the subband to the user with the best channel gain on that subband. For instance, consider the practical example given in [16], where two users were assumed per subband with user 1 a cell-interior user, and user 2 a cell-edge user, such that $h_1^2 = 10h_2^2$. If power and bandwidth are equally allocated among the two users using OMA, the user rates are found in [16] as 3.33 and 0.50 bps, respectively for user 1 and user 2. On the other hand, using NOMA, when the power is allocated as $P_1 = P_2/4$, user rates are found to be 4.39 and 0.74 bps, for user 1 and user 2 respectively. In this case, NOMA provides higher sum rate than OMA. However, what the authors did not consider in this example is the case where the subband is allocated to only 1 user, for instance the user with the highest channel gain on this subband. In such case, its achieved rate would be equal to 6.66 bps, which is higher than 5.13 bps, the total rate achieved by the two users when NOMA is used, and also higher than 3.83 bps, the total rate achieved by the two users when OMA is used. In this case, orthogonal signaling without subband division is the most appropriate solution. This explains the superiority of O_WF on OMA_WF in our work, and also the importance of integrating an adaptive switching from NOMA to orthogonal signaling in the allocation technique, in order to account for such situations.

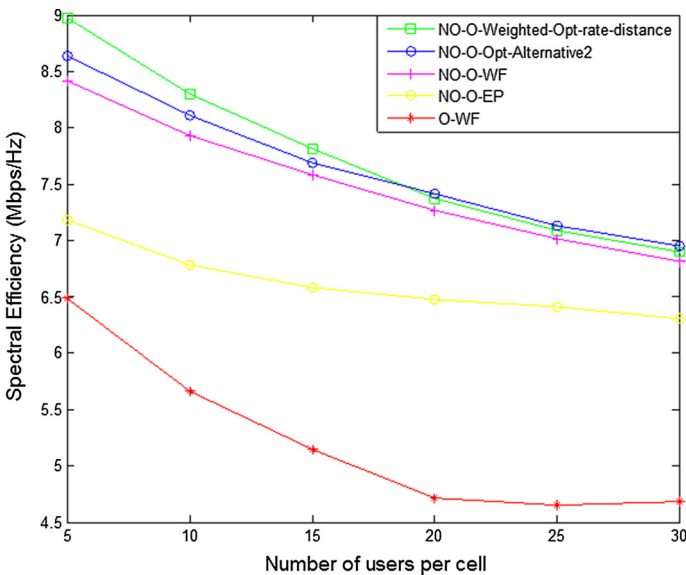


Fig. 13 Spectral Efficiency of the proposed methods in terms of the number of users per cell for a higher amount of transmit power and a lower amount of bandwidth

The power delivered by the base station is as critical to system performance as system bandwidth. For this sake, we propose to study the influence of increasing the total transmit power P_{max} and decreasing system bandwidth B . As an example, we take $B = 50$ MHz and $P_{max} = 40$ W (46 dBm). Figure 13 shows the spectral efficiency for different numbers of users per cell and an actual number of available subbands equal to 128. The proposed optimal algorithms NO_O_Opt_Alternative2 and NO_O_Weighted_Opt_rate_distance still show significantly better performance than sub-optimum techniques. Besides, a higher gain in performance is noticed when compared to NO_O_EP and O_WF, with respect to the case where $B = 100$ MHz and $P_{max} = 1$ W, especially for a high number of users.

In the aim of assessing the implementation feasibility of the different proposed allocation techniques, we measured the computational load of the main resource allocation methods to be integrated at the Base Station. At the downlink receiver side, we estimated the average number of SIC procedures that are needed for a user to recover its useful information. The second measurement allows us to gain insight into the complexity increase, in mobile receivers, with respect to a resource allocation based on classical orthogonal signaling. Measurements, reported in Table 1, were conducted using Matlab, run under windows 8, on an intel core i3 CPU, for the case where $K = 10$ users and $S_A = 128$ available subbands. The results in the second row correspond to the average execution time of one simulation of the whole allocation process.

We can see that the additional complexity driven by NOMA, when applied with a suboptimum waterfilling process, is affordable in comparison to NO_EP and O_WF. However, when optimum power allocation is used, the computational load increases significantly, due to the use of the numerical solver. Therefore, a compromise has to be made between the complexity and the accuracy of the proposed techniques. Such a choice would depend on the requirements of the application in use. However, it should be noted that the allocation algorithms only need to be applied at the BS side, where sufficient hardware and memory resources are supposed to be available to allow efficient and real-time implementations.

When observing the results in the third row of Table 1, together with the comparative results of Fig. 9, we can say that the increase in complexity at the base station allows a better allocation of spectral resources, and therefore a reduction in the average number of subbands allocated to each user. This, in turn, yields a significant decrease in complexity at the user terminal due to the reduced number of executed SIC per user. For instance, the optimum weighted waterfilling-based power allocation requires an average of 0.57 SIC per user, compared to 0.62 or 0.65 SIC with the non-weighted optimum approaches, to 0.68 SIC when suboptimum waterfilling is applied, to 4.54 SIC when an equal power inter-

Table 1 Complexity analysis of the simulated methods

Simulated methods	NO-O-weighted-opt-rate-distance	NO-O-weighted-opt-position	NO-O-opt-alternative1	NO-O-Opt-alternative2	NO-O-WF	NO-WF	NO-O-EP	O-WF
Execution time (ms) at the base station	3858	3857	3570	3850	150	100	50	36
Average number of SIC per user at the terminal user	0.57	5.38	0.65	0.62	0.68	10.80	4.54	0

subband allocation is used, and to 10.80 SIC when suboptimum waterfilling is applied without the dynamic switching to orthogonal signaling.

6 Conclusion

This paper introduces a new strategy for channel and power allocation under a non-orthogonal multiple access scenario. It targets minimizing spectrum usage while satisfying requested data rates by a set of users. Several design issues are thoroughly investigated within the proposed approaches: the choice of user pairing, optimal or suboptimum power allocation, fixed and adaptive intra-subband power allocation, dynamic switching from NOMA to orthogonal signaling, weighting strategies for the optimized sum-rate function, etc. Simulation results show that the proposed framework allows a significant increase in spectral efficiency and in the probability of success, especially when compared to a system purely based on either orthogonal or non-orthogonal signaling. Furthermore, a joint inter and intra subband power allocation obtained by numerically solving an optimized allocation problem yields a substantial gain in performance compared to the suboptimum solutions. In addition, several previously published methods are compared with the proposed techniques within our allocation framework, which proves the relevance of our approach. Moreover, the adoption of appropriate weights in the optimized sum-rate metric shows promising enhancements to either spectral efficiency, cell-edge user throughput, or the necessary number of SIC per user. Such weighting strategies can be used in several practical applications to allow different types of user prioritization. The study conducted here with two users per subband can be adapted to take into account a larger number of cohabiting users in each subband. However, this would incur an important increase in complexity of mobile receivers in downlink transmission because of the necessity to implement a recursive SIC to cancel multiple user interference on allocated subbands [26]. We are currently undergoing further research to include the case of imperfect SIC within our framework and also to study the applicability of the proposed approaches to the context of uplink transmission. Reducing the complexity of the proposed optimal solutions for power allocation is also of a great interest, in order to allow for viable real-time implementations.

References

1. NTT DOCOMO, Requirements, candidate solutions & technology roadmap for LTE Rel-12 onward. 3GPP RWS-120010, 2012.
2. Wunder, G. et al. (2014). 5GNOW: Non-orthogonal, asynchronous waveforms for future mobile applications. *IEEE Communications Magazine* 97–105.
3. Kishiyama, Y., Benjebbour, A., Ishii, H., & Nakamura, T. (2012). Evolution concept and candidate technologies for future steps of LTE-A, IEEE International Conference on Communication Systems.
4. 3GPP TS36.300, Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description.
5. 3GPP TR36.913 (V8.0.0), 3GPP; TSG RAN; Requirements for further advancements for E-UTRA (LTE-Advanced), 2008.
6. 3GPP TR36.814 (V9.0.0), Further advancements for E-UTRA physical layer aspects, 2010.
7. Farah, J., & Marx, F. (2007). Combining strategies for the optimization of resource allocation in a wireless multiuser OFDM system. *AEU International Journal of Electronics and Communications*, 61(10), 665–677.

8. Han, S., Kim, H., Han, Y., Cioffi, J. M., & Leung, V. C. M. (2010). A distributed power allocation scheme for sum-rate maximization on cognitive GMACs. In *The international symposium on personal, indoor and mobile radio communications*, pp. 1854–1858.
9. Mahmoud, H., Yucek, T., & Arslan, H. (2009). OFDM for cognitive radio: Merits and challenges. *IEEE Wireless Communications*, 16(2), 6–15.
10. Rhee, W., & Cioffi, J. (2000). Increase in capacity of multiuser OFDM system using dynamic sub-channel allocation. In *IEEE Vehicular technology conference*, pp. 1085–9.
11. Kim, K., Kim, H., Han, Y., & Kim, S.L. (2004). Iterative and greedy resource allocation in an uplink OFDMA system. In *The international symposium on personal, indoor and mobile radio communications*, pp. 2377–81.
12. Caire, G., & Shamai, S. (2003). On the achievable throughput of a multi-antenna Gaussian broadcast channel. *IEEE Transactions on Information Theory*, 49(7), 1692–1706.
13. Takeda, T., & Higuchi, K. (2011). Enhanced user fairness using non-orthogonal access with SIC in cellular uplink. In *IEEE vehicular technology conference*.
14. Higuchi, K., & Kishiyama, Y. (2012). Non-orthogonal access with successive interference cancellation for future radio access. In *IEEE Vehicular Technology Society, Asia Pacific Wireless communications symposium*.
15. Higuchi, K., & Kishiyama, Y. (2013). Non-orthogonal access with random beamforming and intra-beam SIC for cellular MIMO downlink. In *IEEE Vehicular technology conference*.
16. Otao, N., Kishiyama, Y., & Higuchi, K. (2012). Performance of non-orthogonal access with SIC in cellular downlink using proportional fair-based resource allocation. In *International symposium on wireless communication systems*, pp. 476–480.
17. Saito, Y., Kishiyama, Y., Benjebbour, A., Nakamura, T., Li, A., & Higuchi, K. (2013). Non-orthogonal multiple access (NOMA) for future radio access. In *IEEE Vehicular technology conference*.
18. Umehara, J., Kishiyama, Y., & Higuchi, K. (2012). Enhancing user fairness in non-orthogonal access with successive interference cancellation for cellular downlink. In *International conference on communication systems*.
19. Sharp corporation, Evolving RAN towards Rel-12 and beyond, RWS-120039, 3GPP Workshop on Release 12 Onward Ljubljana, Slovenia (2012).
20. Benjebbour, A., Saito, Y., Kishiyama, Y., Li, A., Harada, A., & Nakamura, T. (2013). Concept and practical considerations of non-orthogonal multiple access (NOMA) for future radio access. *International symposium on intelligent signal processing and communication systems (ISPACS)*.
21. Ding, Z., Yang, Z., Fan, P., & Poor, H. (2014). On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users. *IEEE Signal Processing Letters*, 21(12), 1501–1505.
22. Benjebbour, A., Li, A., Saito, Y., Kishiyama, Y., Harada, A., & Nakamura, T. (2013). System-level performance of downlink NOMA for future LTE enhancements. In *IEEE Globecom*.
23. Schaefferle, J. (2010). Throughput of a wireless cell using superposition based multiple-access with optimized scheduling. In *The international symposium on personal, indoor and mobile radio communications*, pp. 212–21.
24. Kim, B., Lim, S., Kim, H., Suh, S., Kwun, J., Choi, S., Lee, C., Lee, S., & Hong, D. (2013). Non-orthogonal Multiple Access in a Downlink Multiuser Beamforming System, MILCOM.
25. Tse, D., & Viswanath, P. (2005). *Fundamentals of wireless communication*. Cambridge: Cambridge University Press.
26. Tomida, S., & Higuchi, K. (2011). Non-orthogonal Access with SIC in Cellular Downlink for User Fairness Enhancement. In *International symposium on intelligent signal processing and communication systems (ISPACS)*, pp. 1–6.
27. Saito, Y., Benjebbour, A., Kishiyama, Y., & Nakamura, T. (2013). System-level performance evaluation of downlink non-orthogonal multiple access (NOMA). In *The international symposium on personal, indoor and mobile radio communications*, pp. 611–615.
28. Boyd, S., & Vandenberghe, L. (2003). *Convex optimization*. Cambridge: Cambridge University Press.
29. Haykin, S. (2001). *Communication systems* (4th ed.). Colorado: Wiley.
30. Dennis, J. E, Jr, & Schnabel, R. B. (1983). *Numerical methods for unconstrained optimization and nonlinear equations*. Upper Saddle River: Prentice-Hall.
31. Powell, M. J. D. (1984). On the global convergence of trust region algorithms for unconstrained minimization. *Mathematical Programming*, 29(3), 297–303.



Marie-Rita Hojeij received her B.E. degree in computer and communication engineering (major in telecommunications) from the Holy Spirit University of Kaslik, Lebanon, in 2013. She is currently pursuing the Ph.D. degree in the field of information and communications technologies at Institut Mines-Telecom/Telecom Bretagne, France. Her thesis work deals with 5G transmission systems based on Non-Orthogonal Multiple Access techniques. Her current research interests include resource allocation techniques, Non-Orthogonal Multiple Access systems, Orthogonal Frequency Division Multiplexing systems, and signal processing.



Joumana Farah received the B.E. degree in Electrical Engineering from the Lebanese University, in 1998, the M.E. degree in Signal, Image, and Speech processing, in 1999, and the PhD in third generation mobile communication systems, in 2002, from the Polytechnic Institute of Grenoble, France. Since January 2010, she holds an Accreditation to Supervise Research (HDR) from University Pierre and Marie Curie (Paris VI), France. She is currently a full-time professor at the faculty of engineering of the Lebanese University. She has been the recipient of several research grants from the Lebanese National Council for Scientific Research and the Franco-Lebanese CEDRE program. She has four registered patents and software and has co-authored a research book and more than seventy papers in international journals and conferences. She was the General Chair of the 19th International Conference on Telecommunications (ICT 2012), and serves as TPC member and reviewer for several journals (IEEE Journal on Selected Areas in Communications, IEEE Communications Letters,

Signal Processing: Image Communication, Digital Signal Processing, Annals of Telecommunications, etc.) and conferences (IEEE VTC, IEEE Globecom, IEEE ICECS, EUSIPCO, ICT, etc). Her current research interests include channel coding techniques, distributed video coding, multi-carrier systems, cooperative and wireless sensor networks, and resource allocation techniques.



Charbel Abdel Nour obtained his computer and communications engineering degree in 2002 from the Lebanese University, his Masters degree in digital communications from the University of Valenciennes, France, in 2003 and his PhD in digital communications from Telecom Bretagne, France in 2008. Since June 2007, he has been working as a post-doctoral fellow at the Electronics Department of Telecom Bretagne. His interests concern the radio mobile communications systems, broadcasting systems, coded modulations, error correcting codes, MIMO and iterative receivers. He is involved in several research projects related to broadcasting and satellite communications. Since 2007, he is active in the Digital Video Broadcasting DVB consortium where he had important contributions. Starting November 2011, Charbel holds an associate professor position at the Electronics Department of Telecom Bretagne.



Catherine Douillard received the engineering degree in telecommunications from the Ecole Nationale Supérieure des Télécommunications de Bretagne, France, in 1988, the Ph.D. degree in electrical engineering from the University of Western Brittany, Brest, France, in 1992, and the accreditation to supervise research from the University of Southern Brittany, Lorient, France, in 2004. She is currently a Professor in the Electronics Department of Telecom Bretagne where she is in charge of the Algorithm-Silicon Interaction research team. Her main research interests are turbo codes and iterative decoding, iterative detection, the efficient combination of high spectral efficiency modulation and turbo coding schemes, diversity techniques and turbo processing for multi-carrier, multi-antenna and multiple access transmission systems. In 2009, she received the SEE/IEEE Glavieux Award for her contribution to standards and related industrial impact.