IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 8, 2019, accepted March 26, 2019, date of current version July 2, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2908673

Spectrum Resource and Power Allocation With Adaptive Proportional Fair User Pairing for NOMA Systems

KEN LONG, PENGYU WANG, WEI LI, AND DEJIAN CHEN
School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Corresponding author: Pengyu Wang (s160101129@stu.cqupt.edu.cn)

This work was supported by the "Research and Development of Low Power Universal Processor Platform for C-RAN" of the Major Project of National Science and Technology "A New Generation of Broadband Wireless Mobile Communications Networks" under Grant 2014ZX03003004-003.

ABSTRACT In this paper, we investigate the spectrum resource and power allocation problem for the tradeoff between maximizing the sum rate and minimum rate requirements of users in non-orthogonal multiple access (NOMA) system. First, we formulate the NOMA techniques, basic principles, and double-objective optimization (DOO) problem. Then, the non-convexity of the DOO problem is converted into a single-objective optimization (SOO) problem by power discretization method. Global optimal search (GOS) algorithm is applied to solve the user-subchannel matching and power allocation problem. Due to its high complexity and unfairness among users, it is only suitable for determining the upper bound of users throughput performance. Finally, yet importantly, a spectrum resource and power allocation algorithm with adaptive proportional fair (APF) user pairing is proposed to convert the original optimization problem into user pairing, sub-channel, and power allocation. The users paired on the sub-channel are determined by the scheduling priority which is based on the equivalent channel gain. The BS dynamically adjusts the forgetting factor in the APF algorithm based on the variance of all the users' scheduling priorities so as to influence the update of users' scheduling weights. The power allocation stage proposes three power allocation schemes to ensure the users' minimum data rate requirements under the condition that effectively guarantees the correct execution of successive interference cancellation (SIC). The simulation results demonstrate that it can not only approach the throughput performance compared with the global optimal search and the classical water-filling (WF) power allocation using matching theory but also can improve the fairness of the users.

INDEX TERMS Non-orthogonal multiple access (NOMA), spectrum resource and power allocation, global optimal search (GOS), adaptive proportional fair (APF) user pairing, fairness.

I. INTRODUCTION

In the coming decades, more and more application scenarios and emerging technologies will place higher demands on 5G networks, such as Artificial Intelligence (AI), Internet of Things (IoT) and Big Data (BD). 5G will develop current mobile and fixed networks into new, integrated and ultra-flexible energy-efficient networks [1]. The traditional orthogonal multi-user access (OMA) can only be used by at most one user in each scheduling period, such as orthogonal frequency division multiple access (OFDMA) for downlink and single-carrier frequency division multiple access (SC-FDMA) for uplink. In order to avoid interference between cells, single-user detection, decoding and a simple receiver design are enabled [2]. However, by its nature, the use of OMA means that scarce bandwidth resources are completely occupied by this user, despite its poor channel conditions. Obviously, this has a negative impact on the throughput and spectral efficiency (SE) of whole system [3].

New multi-user access schemes have emerged and become potential alternatives for OFDMA and SC-FDMA. A promising solution is non-orthogonal multiple access (NOMA) with successive interference cancellation (SIC). Unlike OMA, in order to improve spectral efficiency, NOMA allocates

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

the same spectrum resources to multiple users [4]. In [5], Zeng *et al.* studied energy efficiency (EE) maximization and showed that NOMA can also deliver higher EE than OMA. Through rigorous mathematical analysis of the NOMA and OMA systems with optimal resource allocation strategies, the authors proved that the NOMA is always suprior to any traditional OMA system [6]. In [7], Zeng *et al.* proved that MIMO-NOMA dominates MIMO-OMA in terms of both sum rate and ergodic sum rate. However, since more than one users are multiplexed on the same sub-band, it will inevitably lead to interference among the multiple users. In order to cancellate signal interference among users in NOMA, 5G networks consider designing more advanced interference management technologies [8].

A. EXISTING RESEARCH ON NOMA

Recently, many studies have discussed different aspects of NOMA. For example, the issue of user pairing, power and spectrum resource allocation based on NOMA has attracted widespread attention. In [9], Lei *et al.* analyzed the tractability of the NOMA resource allocation problem using mathematical methods and provided theoretical insights and algorithm solutions for optimizing NOMA power and channel allocation. Some studies are actively finding sub-optimal throughput performance gain. In order to provide a competitive suboptimal solution, the authors combined Lagrangian duality and dynamic programming to optimize multi-user power and sub-channel allocation in the NOMA system [2]. We note that in some studies of NOMA (e.g., [10]–[12]), by setting the users and sub-channels to two groups of participants pursuing their own interests, the authors used a matching algorithm to converge to a pairwise stable match after a limited number of iterations. In [13], Di *et al.* formulated the centralized scheduling and resource allocation problem as equivalent to a multi-dimensional stable roommate matching problem, in which the users and time/frequency resources are considered as disjoint sets of objects to be matched with each other. The particle swarm genetic algorithm allocates power to each sub-channel based on channel state information (CSI) [14], which can achieve better performance than the traditional average power allocation and water-filling algorithm. Considering the majority of resource allocation algorithms divided the entire bandwidth into sub-bands did not fully exploit the potential of NOMA. A concept of vertical pairing is proposed [15], the users can group in pairs and the entire bandwidth is allowed to be occupied, then the Lagrangian duality method is used to solve the problem of power allocation. In [16], Wu *et al.* adopted the vertical decomposition and proposed a layered-algorithm to compute the optimal power allocation for the NOMA downlink relay-transmission. However, such resource allocation algorithm resulted in low spectrum utilization. In [17], Zeng*et al.* illustrated the tradeoff between the sum rate and maximum number of admitted users and proposed an effective user admission scheme. In fact, NOMA resource allocation can be converted to subcarrier allocation (SA) and power allocation (PA) [18], for SA, a greedy algorithm based on user grouping is proposed, for PA, iterative water-filling and specific user rate maximization criteria will be considered together. The results showed that the algorithm had higher spectral efficiency and resisted user overload.

However, the above algorithms did not consider the fairness of users while pursuing the maximum throughput performance gain. In [19], Islam *et al.* introduced the concept of cluster fairness and proposed the divide-and-next-largest-difference-based user pairing algorithm. In [20], for the single-carrier NOMA system, the authors used the proportional fair scheduling algorithm to derive an approximate optimal power allocation solution. In [21], Okamoto proposed an improved proportional fair scheduling scheme that takes into account the fairness of the target frame and the instantaneous fairness of the assigned subbands. However, the above algorithms did not consider the complexity while pursuing fairness. So, the fairness and complexity were weighed [22]–[24]. Genetic algorithm is a powerful heuristic algorithm that can quickly converge to the solution, which can balance the system throughput and user fairness of multi-user NOMA downlink system [22]. In order to avoid unnecessary comparisons of candidate users, in [23], Liu *et al.* addressed preconditions for user pairing. In [24], the authors proposed a low complexity water-filling power allocation algorithm, which is applied to the proportional fair scheduler of the downlink NOMA system. In [25], in order to maintain SC-FDM attributes and reduce the scheduling complexity of non-orthogonal multiplexing users with continuous resource allocation, the authors adopted an enhanced proportional fair scheduling scheme. For the issue of user pairing and resource allocation, the main idea of the problem is not only to find a balance between system performance gain and user fairness, but also the solution of the problem should be low complexity. In [26], Islam *et al.* conducted a comprehensive survey of the most advanced capacity analysis, power allocation strategies, user fairness and user pairing schemes in NOMA and provided solutions to these important issues. However, the above algorithms did not balance throughput, fairness and complexity. In order to satisfy the users' pursuit of high quality of experience (QoE), new solutions must be provided.

B. MOTIVATION AND CONTRIBUTIONS

For the downlink NOMA system, effective user pairing and resource allocation algorithm can significantly affect system throughput performance and user fairness. To date, most studies have transformed the original optimization objective function into a suboptimal step-by-step solution. The NOMA resource allocation problem urgently requires a global optimal solution with a performance upper bound. The first extension contribution of this article is as follows:

- The authors convert the double-objective optimization problem of maximizing the capacity of the NOMA downlink system into a single-objective optimization problem, and then based on the idea of exhaustive search, we propose a global optimal search algorithm with the upper bound of throughput performance.

The highest complexity of global optimal search algorithm compromises the fairness among users while obtaining the optimal system capacity upper bound. Therefore, it poses a challenge to practical applications. So the next contributions of this article are as follows:

- In order to ensure fairness among users, a spectrum resource and power allocation with adaptive proportional fair user pairing algorithm is proposed. In the phase of user pairing and sub-channel allocation, the scheduling priority is determined by the equivalent channel gain, and users with higher priority will be selected for pairing. Then, the forgetting factor is dynamically adjusted in the APF algorithm based on the variance of all the users' scheduling priorities, so as to influence the update of users' scheduling weights.
- For the optimized variables of power allocation, we propose three power allocation schemes. In order to ensure the fairness of the paired users' throughput performance on the sub-channel, the first power allocation scheme directly allocates the total power of the sub-channel with proportional rate constraint. In order to achieve maximum system throughput under conditions that ensure proper detection of the successive interference cancellation, the second power allocation scheme firstly meets the minimum power requirements of the paired users, and then the remaining power on the sub-channel is allocated to the high channel gain user. The third power allocation scheme allocates the remaining power with proportional rate constraint.

C. PAPER ORGANIZATION

The rest of the paper is organized as follows. In Section II, the system model will be described and the optimal objective function will be proposed. In section III a global optimal search algorithm for power discretization is proposed, which has the upper bound of optimal throughput performance. Section IV describes the algorithm of spectrum resource and power allocation with adaptive proportional fair user pairing. Section V carries on the simulations to the above mentioned algorithms. Finally, section VI summarizes the results.

II. FUNDAMENTALS OF DOWNLINK NOMA

In this section, we will discuss the basic concepts and study the user pairing, spectrum resources and power allocation of NOMA based downlink multi-user networks.

A. SYSTEM MODEL

The system consists of a single base station (BS) and the BS sends the signal to a group of mobile users denoted by $N = \{1, \cdots, N\}$. As shown in Fig. 1, the users are assumed to be uniformly deployed in a circular area, all transceivers are equipped with a single antenna. The BS divides the available bandwidth into a set of sub-channels that are smaller than the coherent bandwidth, denoted by $K = \{1, \cdots, K\}$. In this case, the bandwidth of each sub-channel is denoted by $B_s = B/K$, where B is the total system bandwidth. According to the NOMA protocol, one sub-channel can be allocated to multiple users, and one user can receive from the BS through multiple sub-channels. Because more than one users are multiplexed on the same sub-channel, the co-channel interference in the NOMA system will be very strong. Therefore, it is unrealistic for all users in the system to perform NOMA together. It is recommended to adopt a hybrid MA that combines NOMA and traditional MA. In the system, the users are divided into multiple groups, each group implements NOMA, and different groups allocate orthogonal spectrum resource [10]. We assume that the BS fully understands channel state information. Based on the CSI of each sub-channel, the BS allocates different power to the users in the subset of non-orthogonal sub-channels. We consider a block fading channel whose channel gain remains constant for one scheduling period but varies independently of each other. On the sub-channel k, the channel gain coefficient between the multiplexed user n and the BS is denoted by $h_{n,k} = g_{n,k}PL(d_n)$, where $g_{n,k}$ denotes the Rayleigh fading channel gain, d_n is the distance between the user n and the BS, and $PL(\cdot)$ is the path loss function. We assume that M users are multiplexed on the sub-channel k. The signal transmitted by the BS through the sub-channel k can be expressed as [27]

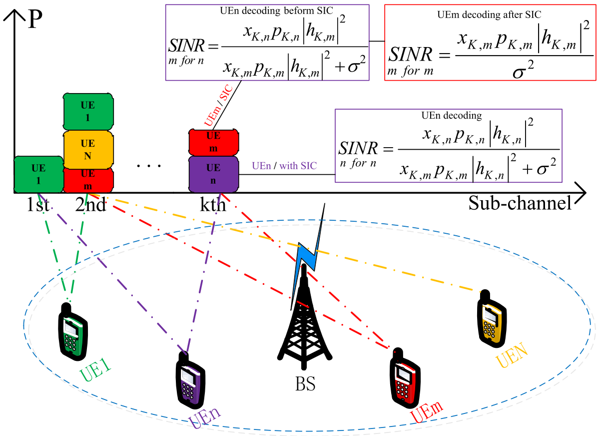


FIGURE 1. System model of the NOMA systems.

$$x_k = \sum_{n \in M} \sqrt{x_{k,n} p_{k,n}} s_n \tag{1}$$

where the binary variable $x_{k,n}$ represents whether sub-channel k is allocated to the user n, where the binary variable $p_{k,n}$ represents the power allocated to the user n on the sub-channel k, and s_n is the modulated symbol, and $E\left[|s_n|^2\right] = 1$. since BS and all users are equipped with a single antenna, the signal received by the user n on the sub-channel k can be formulated as [27]

$$y_{k,n} = h_{k,n} x_k + w_n \tag{2}$$

where $w_n \sim \left(0, \sigma^2\right)$ is the additive white Gaussian noise (AWGN) with zero average, and σ^2 is the noise variance. The equivalent channel gain (ECG) is denoted by $H_{k,n} = \left|h_{k,n}\right|^2/\sigma^2$. At the receiver, successive interference cancellation techniques are used to cancel interference among users, with lower channel gain user treats the signal of the higher channel gain user as a noise, while the user with higher channel gain first reconstructs the signal of the lower channel gain user and then uses the SIC to cancel the signal of the lower channel gain user to obtain its own signal.

B. OPTIMIZATION PROBLEM FORMULATION

Similarly, for the capacity optimization problem of user pairing and resource allocation in NOMA downlink system, we assume that the number of users multiplexed on each sub-channel is M, the throughput achieved by the multiplexing user n on the sub-channel k is expressed as [10]

$$R_{k,n} = \sum_{n\in N} B_k \log_2\left(1 + \frac{x_{k,n}p_{k,n}\left|h_{k,n}\right|^2}{I_{k,n} + \sigma^2}\right) \tag{3}$$

$I_{k,n}$ is the interference for user n introduced by the users within the number of M multiplexed users who have higher channel gain on the sub-channel k

$$I_{k,n} = \sum_{m\in\left\{M \mid \frac{\left|h_{k,m}\right|^2}{\sigma^2} > \frac{\left|h_{k,n}\right|^2}{\sigma^2}\right\}} x_{k,m}p_{k,m}\left|h_{k,n}\right|^2 \tag{4}$$

In order to avoid the form of the OFDMA solution, we assume that the range of M is $M_f \leq M \leq M_u$, where M_f is the lower bound of the multiplexed user on the sub-channel and M_u is the upper bound. The optimization problem is formulated to maximize the total data rate overall system bandwidth

$$p_{k,n} = allocated\ power\ to\ user\ n\ on\ sub-channel\ k$$

$$x_{k,n} = \begin{cases} 1, & if\ sub-channel\ k\ is\ allocated\ to\ user\ n \\ 0, & otherwise \end{cases}$$

$$\max_{x_{k,n},p_{k,n}} \sum_{k\in K}\sum_{n\in N} B_k \log_2\left(1 + \frac{x_{k,n}p_{k,n}\left|h_{k,n}\right|^2}{I_{k,n} + \sigma^2}\right) \tag{5}$$

$$s.t.: \ C1: \sum_{n\in N} x_{k,n}p_{k,n} \leq p_k \quad \forall k \in K \tag{6}$$

$$C2: \sum_{n\in N}\sum_{k\in K} x_{k,n}p_{k,n} \leq P_{total} \tag{7}$$

$$C3: M_f \leq \sum_{n\in N}\sum_{l\in L} x_{k,n} \leq M_u \quad \forall k \in K \tag{8}$$

$$C4: R_n > R_{n,\min} \quad \forall n \in \{N|x_{k,n} = 1\}, k \in K \tag{9}$$

$$C5: x_{k,n} \in \{0, 1\} \quad \forall k \in K \ \ n \in N \tag{10}$$

Due to the power transmitted by the BS and the power allocation for each sub-channel is limited, therefore, the power allocation variable $p_{k,n}$ must satisfy constraints (6), (7). Constraint (8) ensure that each sub-channel can only be allocated at most M_u users. In order to ensure each user's QoS requirement, the constraint (9) indicates that the user implemented data rate must be greater than the user's minimum data rate. Constraint (10) indicates whether a discrete binary variable $x_{k,n}$ is allocated.

Because of the structure of the utility function, the discrete variables $x_{k,n}$, $p_{k,n}$ and the existence of the constraints, the optimization problem has mixed integer nonlinear programming (MINLP) characteristics. It is difficult to obtain the optimal solution for this type of problem. Common solutions include the following categories:

(1) Adopt heuristic algorithm, such as the genetic algorithm mentioned in the literature [14], [22], particle swarm algorithm, etc. This type of algorithm does not have a strict theoretical basis, it is inspired by people's actual life experience and rules of things.

(2) By reducing the degree of freedom in allocation variables, the original problem is decomposed into several sub-problems, such as [18]. The typical application of such method is converted to sub-carrier and power allocation.

(3) Converting the original problem into a convex optimization problem by rationally changing the constraints. For example, literature [9], the convex optimization problem is the easiest to solve for all optimization problems, and its convexity guarantees that the local optimum is the global optimum.

III. GLOBAL OPTIMAL SEARCH(GOS) OF POWER DISCRETIZATION ALGORITHM

In order to achieve maximum system utility, we need to optimize user and sub-channel matching variable $x_{k,n}$, user power allocation variable $p_{k,n}$. Because both the user and sub-channel are discrete values, so the user pairing and user-subchannel matching can be achieved based on exhaustive search idea, but the value of power allocation for the paired users are continuous, so it is impossible to combine all of them based on the idea of exhaustive search. However, in practical systems, the power is typically set in discrete steps. Therefore, by discretizing the total power budget into the number of L uniform power level, the power interval between each level is denoted by $\varsigma = P_{total}/L$. The continuous power value of the discrete power level l is denoted by $p^l = \varsigma * l$, and $l \in \{1, 2, \ldots, L\}$. After the power is discretized, the original binary optimization variable $x_{k,n}$, $p_{k,n}$ are transformed into an optimization variable, denoted by $C^l_{k,n}$. So the original optimization problem can be achieved the optimal utility based on the idea of group search.

A. OPTIMIZATION PROBLEM CONVERSION

Therefore, the utility function and optimization variables after power discretization can be expressed as

$$\max_C U = \sum_{n\in N}\sum_{k\in K}\sum_{l\in L} C^l_{k,n} R^l_{k,n} \tag{11}$$

$$s.t.: \ C1: \sum_{n\in N}\sum_{l\in L} c^l_{k,n} p^l \leq P_k \quad \forall k \in K \tag{12}$$

$$C2: \sum_{k\in K}\sum_{n\in N}\sum_{l\in L} c^l_{k,n} p^l \leq P_{total} \tag{13}$$

$$C3: M_f \le \sum_{n\in N}\sum_{l\in L} c_{k,n}^{l} \le M_u \quad \forall k \in K \tag{14}$$

$$C4: \sum_{l\in L} c_{k,n}^{l} \le 1 \quad \forall k \in K\ n \in N \tag{15}$$

$$C5: R_n > R_{n,\min} \quad \forall n \in \{N|x_{k,n}=1\}, k \in K \tag{16}$$

$$C6:\ C_{k,n}^{l} \in \{0\ 1\} \quad \forall k \in K\ \ n \in N\ \ l \in L \tag{17}$$

Constraint (12) indicates that the power level p^l allocated to the user n multiplexed on the sub-channel k cannot exceed the total power p_k allocated for the sub-channel. Constraint (13) indicates that the sum of power levels allocated by all users in the entire bandwidth cannot exceed the total power P_{total} transmitted by the BS. Constraint (14) indicates that the maximum number of users multiplexed on each sub-channel cannot exceed M_u. Constraint (15) ensures each user can only select one power level on each sub-channel. In order to ensure each user's QoS requirement, the constraint (16) indicates that the user implemented data rate must be greater than the user's minimum data rate. Constraint (17) indicates whether a discrete binary variable $c_{k,n}^{l}$ is allocated. Therefore, the data rate that can be achieved by allocating power level l to user n on the sub-channel k is expressed as

$$R_{k,n}^{l} = B_k \log_2\left(1 + \frac{p^l \left|h_{k,n}\right|^2}{I_{k,n} + \sigma^2}\right) \tag{18}$$

$$I_{k,n} = \sum_{m\in\left\{M \mid \frac{\left|h_{k,m}\right|^2}{\sigma^2} > \frac{\left|h_{k,n}\right|^2}{\sigma^2}\right\}} \sum_{l'\in l} c_{k,m}^{l'} p^{l'} \left|h_{k,n}\right|^2 \tag{19}$$

$I_{k,n}$ is the interference of power discretization for user n introduced by the users within the number of M multiplexed users who have higher channel gain on the sub-channel k.

B. INTRA-SUB-CHANNEL USER PARING AND POWER ALLOCATION

In this section, we ignore the view of the complexity and aiming to propose an algorithm that approximates the optimal solution, it is used to determine the upper bound of the throughput performance that the system can provide. So, the idea of group search can effectively ensure the optimal performance of the system under the premise of sacrificing complexity. The algorithm which user pairing and resource allocation under global optimal search of power discretization is divided into two steps.

The first step is the user pairing and power allocation on the sub-channel. First, on the sub-channel k, we will pair all the number of N users based on the idea of group search. After pairwise pairing is performed, the paired users are ranked in descending order according to the equivalent channel gain to obtain different paired user pairs $p_{m,n}$. Then, the BS transmits L different total power levels for sub-channel k. For each user in the user pairs $p_{m,n}$, the power level allocated by the BS to them are denoted by $l' = 0:1:l$, $l'' = l'+1:1:l$, and $l'+l'' \le l$. Through implementing iterative processes to maximize the throughput of the currently paired user pair $p_{m,n}$. Then repeat above steps to obtain the maximum throughput of different paired user pairs $p_{m',n'}$ for the current power level l on the sub-channel k and selected the maximum throughput of the user pairs as the maximum utility, denoted by $u_{k,l}$. Finally, the power level l of the sub-channel k is adjusted, and the achieved maximum utility value denoted by $u_{k,l'}$. The specific description of the algorithm is as follows:

Algorithm 1 Intra-Sub-Channel User Paring and Power Allocation

```
Require: the total number of users = $N$; the number of users $M$ on each sub-channel = 2; the power level constraint on each sub-channel = $L$;
Ensure: the maximum utility $u_{k,l}$ for each power level $l \in L$ to user pairs $p_{m,n}$ on the sub-channel $k$;
 1: Initialize $u_{k,l} \leftarrow \varnothing$, for $\forall k \in K, \forall n \in N, \forall l \in L$;
 2: for user $m, n \in N$ and $m \ne n$ do
 3:     (a) get paired user pairs $p_{m,n}$ based on group search
 4:     (b) sort them ECG in descending order
 5: end for
 6: for power level $l$ on the sub-channel $k$ do
 7:     for each paired user pairs $p_{m,n}$ and assume their $ECG$ satisfy $H_{k,m} > H_{k,n}$ do
 8:         for user $m$ allocated power level $l' = 0:1:l$; user $n$ allocated power level $l'' = l'+1:1:l$ do
 9:             if $l' + l'' \le l$ then
10:                 (c) according to (18) calculating their throughput $R_{k,m}^{l'}, R_{k,n}^{l''}$,
11:                 if $R_{k,m}^{l'} \ne 0, R_{k,n}^{l''} \ne 0$ then
12:                     (d) select max $\left\{R_{k,n}^{l''} + R_{k,m}^{l'}\right\}$ among paired user pairs as the maximum utility $u_{k,l}$ achieved by paired users $p_{m,n}$ at this power level $l$ on the sub-channel $k$
13:                 end if
14:             end if
15:         end for
16:     end for
17: end for
18: for power level $l = 1:1:L$ do
19:     repeat steps 6-17, get maximum utility $u_{k,l}$ for each power level $l$ on the sub-channel $k$
20: end for
```

When the power allocation of the paired user is determined, next, the BS will allocate power among sub-channels based on the maximum sum rate.

C. INTER-SUB-CHANNEL POWER ALLOCATION

The second step is the power allocation among sub-channels. For different sub-channels, repeating above steps to get the maximum utility value $u_{k,l}$ which achieved by different paired users at the different power level l. Under the condition that the power level l allocated for all sub-channels less than the total power P_{total} transmitted by the BS, by combinating the maximum throughput utility $u_{k,l}$ based on the idea of group search to achieve the maximum utility $U_{K,N}^{L}$ by all paired users of the entire bandwidth. By determining the combination of $u_{k,l}$, we can be ensured the binary variable $c_{k,n}^{l}$ and the user n allocated power level l on the sub-channel k. The specific description of the algorithm is as follows:

Algorithm 2 Inter-Sub-Channel Power Allocation

```
Require: the maximum utility $u_{k,l}$ for each power level $l \in L$ to user pairs $p_{m,n}$ on the sub-channel $k$; the number of sub-channel $= K$; the power level constrain on each sub-channel $= L$;
Ensure: $c_{k,n}^{l}$ for power level $l$ is allocated to user $n$ on the sub-channel $k$, $U_{K,N}^{L}$ for maximum throughput utility achieved by all paired users across entire bandwidth
Initialize $u_{k,l} \leftarrow \emptyset$, for $\forall k \in K, \forall n \in N, \forall l \in L$;
for sub-channel $k = 1 : K$ do
    perform Algorithm 1, and obtain $u_{k,1}, u_{k,2}, \ldots, u_{k,L}$
end for
for power level $l = 1 : 1 : L$ do
    each sub-channel select one power level $l$ achieved maximum utility $u_{k,l}$
    if the sum of all the $K$ sub-channel's power level $L$ less than the total power level $l$ and $R_n > R_{n,\min}$, $R_m > R_{m,\min}$ then
        exhaustively combinate the maximum utility $u_{k,l}$ achieved at each power level $l$ on each sub-channel $k$
    end if
end for
select maximum $U_{K,N}^{L} = \{u_{1,l} + u_{2,l'} +, \ldots, u_{K,l''}\}$ as the system achieved Maximum utility and via $u_{k,l}$ to determine $c_{k,n}^{l}$
```

To reduce the complexity of SIC, the global optimal search algorithm first describes the case of $M_f = M_u = 2$. As shown in Fig. 2, we describe the implementation of the global optimal search algorithm with $M_u = 3$.

The above algorithm is also applicable to the case where the number of multiplexed users is greater than two users. In the fifth part of this paper, it is verified by simulation that the upper bound of the number of multiplexed users are equal to 3, 4, 5. The global optimal search of power discretization algorithm converges to a maximum utility value $u_{k,l}$ for each power level l on the sub-channel k through multiple iterations. For the case of $M_f = M_u = 2$, its complexity denoted by $O\left((N(N-1)/2)L^3\right)$, the complexity increases with the number of users N, and the power level L, so when the power interval ς is smaller and the number of users N is larger, the complexity is higher, and the fairness between users is not taken into consideration when ensuring the maximum system utility $U_{K,N}^{L}$. Therefore, this algorithm is not practical and is only suitable for evaluating the throughput performance upper bound. Based on this, a low-complexity spectrum resource and power allocation with adaptive proportional fair user pairing algorithm is proposed.

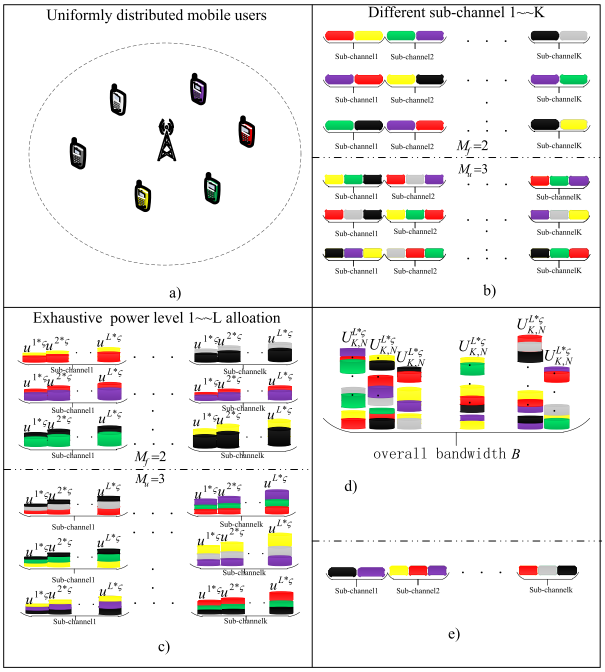


FIGURE 2. Global optimal search for NOMA ($M_u = 3$).

IV. SPECTRUM RESOURCE AND POWER ALLOCATION WITH ADAPTIVE PROPORTIONAL FAIR USER PAIRING ALGORITHM

In this section, the original optimization problem is transformed into the sub-problems of user pairing, sub-channel and power allocation by reducing the degree of freedom in allocation variables. First of all, in order to ensure the user's fairness, an adaptive proportional fair user pairing algorithm based on equivalent channel gain is proposed.

A. ADAPTIVE PROPORTIONAL FAIR USER PAIRING ALGORITHM BASED ON EQUIVALENT CHANNEL GAIN

The existing user pairing algorithms are all based on channel state information which feedback by the users. For a given sub-channel, assuming that the sorted candidate user sets is denoted by $U = \{u_1, u_2, \ldots, u_N\}$ and $H_{k,1} > H_{k,2} > \cdots > H_{K,N}$. In [28], the authors proposed two user pairing algorithms based on channel gain, the one selected the $n-th$ user and the $(N+1-n)-th$ user into a user pair, and another selected the $n-th$ user and the $(N/2-n)-th$ user form a user pair. The authors found that the two user pairing algorithms have the same system detection error rate. However, In [29], the authors pointed out that when the channel gain gap between paired users is maximum, the system can achieve the maximum utility. However, the above user pairing algorithms are based on the principle of user channel gain ranking and did not consider the user's scheduling priority. So proportional fair scheduling policy has been adopted in the majority of papers dealing with NOMA [13], [18], [19]. All above PF scheduling algorithms can guarantee user fairness and throughput performance, but user scheduling priority based on the ratio of average throughput and instantaneous throughput. For the users multiplexed on the same sub-channel, the solution of those algorithms have higher calculation complexity. Because the user's reachable data rate is a monotonically increasing function of its channel quality. In this case, we convert the ratio of user throughput to the ratio of the equivalent channel gain on the sub-channel. Therefore, based on the idea of traditional PF algorithm, an adaptive proportional fair user pairing algorithm based on equivalent channel gain is proposed. For user n on the sub-channel k, its scheduling priority is expressed as

$$\rho_{k,n}(t) = H_{k,n}(t)/\bar{H}_{k,n}(t) \tag{20}$$

In the time slot t, the equivalent channel gain of the user n on the sub-channel k is denoted by $H_{n,k}(t)$, the average equivalent channel gain for the user n on the sub-channel k from the initial time to the time slot $t-1$ denoted by $\bar{H}_{k,n}(t)$. The average channel gain of each user on the sub-channel k is updated in the time slot $t+1$ according to the following formula

$$\bar{H}_{k,n}(t+1) = \begin{cases} \alpha H_{k,n}(t) + (1-\alpha)\bar{H}_{k,n}(t), & n \in N \\ (1-\alpha)\bar{H}_{k,n}(t), & n \notin N \end{cases} \tag{21}$$

where α called forgetting factor, and denoted by $\alpha = 1/T_c$. T_c is the time window parameter. Taking time slot t_0 as the starting point of an observation interval, we assume that two users m, n multiplexed on the sub-channel k have the same initial average equivalent channel gain, and denoted by $\bar{H}_{k,m}(t_0) = \bar{H}_{k,n}(t_0)$. So, the user's scheduling priority is determined only by the instantaneous equivalent channel gain at the initial scheduling moment. We assume $H_{k,m} > H_{k,n}$, so scheduling priority denoted by $\rho_{k,m} > \rho_{k,n}$, and the user m first scheduled. Next time slot, the user m, n average equivalent channel gain is updated to $\bar{H}_{k,m}(t_0+1) = \alpha H_{k,m}(t_0) + (1-\alpha)\bar{H}_{k,m}(t_0)$ and $\bar{H}_{k,n}(t_0+1) = \alpha H_{k,n}(t_0) + (1-\alpha)\bar{H}_{k,n}(t_0)$. The BS recalculates the priority of the scheduling users in the time slot t_0+1. We assume the time slot t_0+t_g their scheduling priority meet $\rho_{k,n}(t_0+t_g) > \rho_{k,m}(t_0+t_g)$. Therefore, according to the priority calculation formula (20), when we take the bigger T_c, the slower update rate of user's average equivalent channel gain, and denoted by $\bar{H}_{k,n}(t+1) \approx \bar{H}_{k,n}(t)$. Within a long observation interval, it can ensure that the scheduling probabilities of each user are close to each other. However, for users who stay in the system for a short time, the PF algorithm cannot guarantee its fairness. Based on this, in each scheduling slot, the variance of all the users' scheduling priority is first calculated. The variance of scheduling priority $\xi(t)$ for all N users on the sub-channel k in the time slot t is expressed as

$$\xi(t) = N^{-1}\sum_{n\in N}\left[\rho_{k,n}(t) - N^{-1}\sum_{i\in N}\rho_{k,i}(t)\right]^2 \tag{22}$$

The larger $\xi(t)$, the greater priority gap among users. Therefore, for better short-term fairness, we can set larger $\alpha(t)$ to speed up the user's priority approaching speed. When the priority gap between users is smaller, we can set smaller $\alpha(t)$ to ensure users with good channel quality to get more scheduling opportunities and obtain good throughput performance. If $\alpha(t)$ is considered as the amplitude attenuation of the signal, and $\xi(t)$ is equivalent to frequency, we can construct function $\alpha(t) = f[\xi(t)]$. An adaptive proportional fair user pairing algorithm is proposed which dynamically adjusts the forgetting factor α by

$$\alpha(t) = \max\left\{2\alpha_{ref} - 2\alpha_{ref}\left\{1 + [\xi(t)]^{N(t)}\right\}^{-1},\ \varepsilon\right\} \tag{23}$$

where $\varepsilon = 0.002$. Since the forgetting factor α of the traditional proportional fair scheduling algorithm is taken as 0.01, so we dynamically adjust $\alpha(t)$ based on $\alpha_{ref} = 0.01$. The larger the order $N(t)$ and the faster $\xi(t)$ converges to 0, the better the short-term fairness of the system. If $\alpha \to 0$, the scheduling priority of all users will tend to be constant, resulting in the scheduling set tend to be fixed. Therefore, we design $N(t) = \xi(t) + \tau$ as a monotonically increasing function of $\xi(t)$. Where τ is a positive number close to 0, and it ensures $N(t) \neq 0$. When $\xi(t) > 1$, $N(t) > 1$, and with the $\xi(t)$ increasing, $N(t)$ monotonous increase, so it can get bigger $\alpha(t)$, this accelerates the speed in which the user's priority is approaching and can obtain better short-term fairness. When $\xi(t) < 1$, $N(t) < 1$, and as the $\xi(t)$ reduction, the $N(t)$ decreases, so long-term fairness can be guaranteed.

In each scheduling slot, firstly, based on adaptive equivalent channel gain proportional fair scheduling algorithm to calculate the user's scheduling priority on each sub-channel of the entire bandwidth. In order to reduce the complexity of SIC, we assume the case of $M_f = M_u = 2$. Then the BS selects users which have the highest priority on the sub-channel k to schedule. In next scheduling time slot $t+1$ updated the user's average *ECG*. The specific description of the algorithm is as follows:

When the above resource allocation process is completed, the power allocation is performed for the case where two users are multiplexed on the sub-channel k.

B. THREE POWER ALLOCATION SCHEMES

In order to simplify the power allocation, the total power is equally allocated by the BS to all the K sub-channels, and the power on each sub-channel is denoted by $p_k = P_{total}/K$. So the total power of two users which multiplexed on any sub-channel k equal to p_k. The original optimization problem is transformed into

$$\max_{p_{k,n}} \sum_{k\in K}\sum_{n\in N} B_k \log_2\left(1 + \frac{p_{k,n}\left|h_{k,n}\right|^2}{I_{k,n} + \sigma^2}\right) \tag{24}$$

$$s.t.:\ C1: \sum_{n\in N} x_{k,n}p_{k,n} = p_k \quad \forall k \in K \tag{25}$$

$$C4: R_n > R_{n,\min} \quad \forall n \in \{N|x_{k,n} = 1\},\ k \in K \tag{26}$$

Downlink multi-user NOMA power allocation not only guarantees the user's minimum data rate requirements and fairness between users, but also needs to allocate appropriate power for paired users based on the above criteria. The proportional

Algorithm 3 Adaptive Proportional Fair User Pairing Algorithm Based on Equivalent Channel Gain

```
Initialize instantaneous ECG $H_{k,n}(t)$ that the user feedback to the BS; updated average ECG $\bar{H}_{k,n}(t)$ in scheduling time slot $t$
for sub-channel $k = 1 : K$ do
    for user $n = 1 : N$ do
        (a) calculate the scheduling priority $\rho_{k,n}(t)$ according to (20),
        (b) select the two users with the highest scheduling priority to pair.
    end for
    (c) BS count the priority variance $\xi(t)$ of all users according to (22)and determined the order $N(t)$
    (d) determine forgetting factor $\alpha(t)$ according to (23) for each sub-channel $k$
end for
for sub-channel $k = 1 : K$ do
    for user $n = 1 : N$ do
        updat user $n$ average ECG according to (21) in next scheduling time slot $t + 1$
    end for
end for
```

constraint based on the minimum data rate requirement can not only guarantee the fairness of the paired users, but also by analyzing the calculation formulas of the throughput for the high and low channel gain users, it is found that low channel gain user in order to achieve throughput performance similar to that of high channel gain user, the allocated power must be much larger than that of high channel gain user. Moreover, due to the proportional constraint relationship between the data rates of the paired users, the power allocation between the two users remains relatively stable, it can ensure the correct and stable implementation of the SIC. Based on this, the following three power allocation schemes (PAS) are proposed:

1) THE FIRST POWER ALLOCATION SCHEME

The first allocation scheme allocates the total power of the sub-channel p_k with proportional rate constraint. For any sub-channel k, multiplexed users m, n, we assume their equivalent channel gain satisfies $H_{k,m} > H_{k,n}$, then its power allocation with the proportional rate constraint must satisfy the following restrictions

$$\max_{p_{k,n}} \sum_{k\in K}\sum_{n\in N} B_k \log_2\left(1+\frac{p_{k,n}\left|h_{k,n}\right|^2}{I_{k,n}+\sigma^2}\right) \tag{27}$$

$$s.t.:\ C1: \sum_{n\in N} x_{k,n}p_{k,n} = p_k \quad \forall k\in K \tag{28}$$

$$C4: R_{k,m} : R_{k,n} = R_{m,\min} : R_{n,\min} \tag{29}$$

For the optimization problem of the equations (27), (28), and (29), the Lagrangian multiplier is used to solve the problem. we assume all users have the same minimum data rate requirements, the detailed solution is shown in Appendix A. Then, the allocated power for the high channel gain user m and the low channel gain user n are represented respectively as follows

$$p_{k,m} = \frac{R_{\min}\sigma^2\left(\left|h_{k,m}\right|^2-\left|h_{k,n}\right|^2\right)}{2\mu\left|h_{k,m}\right|^2\left|h_{k,n}\right|^2} - \frac{\sigma^2\left(\left|h_{k,m}\right|^2+\left|h_{k,n}\right|^2\right)}{2\left|h_{k,m}\right|^2\left|h_{k,n}\right|^2} \tag{30}$$

$$p_{k,n} = \frac{2p_k\left(\left|h_{k,m}\right|^2+\left|h_{k,n}\right|^2\right)}{2\left|h_{k,m}\right|^2\left|h_{k,n}\right|^2} + \frac{\sigma^2\left(\left|h_{k,m}\right|^2+\left|h_{k,n}\right|^2\right)}{2\left|h_{k,m}\right|^2\left|h_{k,n}\right|^2} - \frac{R_{\min}\sigma^2\left(\left|h_{k,m}\right|^2-\left|h_{k,n}\right|^2\right)}{2\mu\left|h_{k,m}\right|^2\left|h_{k,n}\right|^2} \tag{31}$$

$$\mu = \frac{\sigma^2 R_{\min}\left(\left|h_{k,m}\right|^2-\left|h_{k,n}\right|^2\right)}{\sigma^2\left(\left|h_{k,m}\right|^2+\left|h_{k,n}\right|^2\right)+\psi} \tag{32}$$

$$\psi = \sqrt{\left(\sigma^2\left(\left|h_{k,m}\right|^2+\left|h_{k,n}\right|^2\right)\right)^2+4\left|h_{k,m}\right|^2\left|h_{k,n}\right|^4 p_k\sigma^2} - \sigma^2\left(\left|h_{k,m}\right|^2+\left|h_{k,n}\right|^2\right) \tag{33}$$

The above power allocation scheme does not consider the minimum data rate requirement of the two users multiplexed on sub-channel. So the following algorithm will first meet the minimum power requirements and it is expressed as

$$p_{k,n.\min} = \frac{2^{\frac{R_{n,\min}}{B_k}}-1}{\left|h_{k,n}\right|^2}\left(\sum\nolimits_{m\in\{M|\}}^{p_{k,m.\min}\left|h_{k,n}\right|^2} + \sigma^2\right) \tag{34}$$

When solving the minimum power requirements of the paired users, it is necessary to solve the power requirements of the high channel gain user, and then to solve the power requirements of the low channel gain user. If $\{p_{n.\min} + p_{m.\min}\} < p_k$, then the remaining power is denoted by $p_k^r = p_k - (p_{n.\min} + p_{m.\min})$. The following allocation schemes allocate remaining power to the paired users.

2) THE SECOND POWER ALLOCATION SCHEME

According to the capacity maximization criterion, the second power allocation scheme allocates remaining power p_k^r for the paired users on the sub-channel k. In order to ensure maximum throughput on the sub-channel, the power should be allocated as much as possible to the user with higher channel gain that are depicted in Appendix B. However, in order to ensure the correctly implement of the SIC, the transmit power of any user with low channel gain must be greater than the transmit power of all users with relatively strong channel gain. Therefore, for the case of two users multiplexed on the same channel, we propose a power allocation scheme that ensures the correct detection of the SIC, and is expressed as

$$p_{k,n}H_{k,m} - p_{k,m}H_{k,m} \geq p_{sic} \tag{35}$$

$$p_{k,m} + p_{k,n} = p_k \tag{36}$$

$$p_{k,m,\max} \leq \frac{p_k - p_{sic}/H_{k,m}}{2} \tag{37}$$

where the p_{sic} is the minimum power difference needed to distinguish between the decoded signal and the rest of undecoded signal. The maximum power $p_{k,m,\max}$ is the high channel gain user to ensure the correct implementation of the SIC that can be obtained. Therefore, when allocating power, the relationship between the remaining power p_k^r and $\{p_{k,m,\max} - p_{k,m,\min}\}$ should be firstly determined, if $p_k^r \leq p_{k,m,\max} - p_{k,m,\min}$, then $p_{k,m} = p_{k,m,\min} + p_k^r$, else $p_k^r > p_{k,m,\max} - p_{k,m,\min}$, then $p_{k,m} = p_{k,m,\max}$. $p_{k,n} = p_k - p_{k,m}$.

3) THE THIRD POWER ALLOCATION SCHEME

The third allocation scheme allocate the remaining power p_k^r of the sub-channel k with proportional rate constraint. So the original optimization problem is converted into

$$\max_{p_{k,n}^r} \sum_{k\in K}\sum_{n\in N} B_k \log_2\left(1+\frac{p_{k,n}^r|h_{k,n}|^2}{I_{k,n}+\sigma^2}\right) \tag{38}$$

$$s.t.:\ C1: p_{k,m}^r + p_{k,n}^r = p_k^r \tag{39}$$

$$C4: R_{k,m}^r : R_{k,m}^r = R_{m,\min} : R_{n,\min} \tag{40}$$

where $p_{k,m}^r$, $p_{k,n}^r$ are the remaining power allocated for the users with high channel gain and low channel gain, respectively. For optimization problems consisting of equations (38), (39) and (40), we use the Lagrangian multiplier to solve. Then, the remaining power allocated for the paired user are expressed as

$$p_{k,m}^r = \frac{R_{\min}\sigma^2\left(|h_{k,m}|^2 - |h_{k,n}|^2\right)}{2\mu|h_{k,m}|^2|h_{k,n}|^2} - \frac{\sigma^2\left(|h_{k,m}|^2 + |h_{k,n}|^2\right)}{2|h_{k,m}|^2|h_{k,n}|^2} \tag{41}$$

$$p_{k,n}^r = \frac{2p_k^r|h_{k,m}|^2|h_{k,n}|^2}{2|h_{k,m}|^2|h_{k,n}|^2} + \frac{\sigma^2\left(|h_{k,m}|^2 + |h_{k,n}|^2\right)}{2|h_{k,m}|^2|h_{k,n}|^2} - \frac{R_{\min}\sigma^2\left(|h_{k,m}|^2 - |h_{k,n}|^2\right)}{2\mu^r|h_{k,m}|^2|h_{k,n}|^2} \tag{42}$$

$$\mu^r = \frac{\sigma^2 R_{\min}\left(|h_{k,m}|^2 - |h_{k,n}|^2\right)}{\sigma^2\left(|h_{k,m}|^2 + |h_{k,n}|^2\right) + \psi^r} \tag{43}$$

$$\psi^r = \sqrt{\left(\sigma^2\left(|h_{k,m}|^2 + |h_{k,n}|^2\right)\right)^2 + 4|h_{k,m}|^2|h_{k,n}|^4 p_k^r\sigma^2} - \sigma^2\left(|h_{k,m}|^2 + |h_{k,n}|^2\right) \tag{44}$$

So, the total power allocated for the users m, n on the sub-channel k are expressed as

$$p_{k,m} = p_{k,m}^r + p_{k,m,\min} \tag{45}$$

$$p_{k,n} = p_{k,n}^r + p_{k,n,\min} \tag{46}$$

V. SIMULATION RESULTS

A. SIMULATION SCENARIO SETUP

In this section, we evaluate the performance of the proposed global optimal search algorithm, spectrum resource and power allocation with adaptive proportional fair user pairing algorithm, and compare their performance with the most classical matching theory with water-filling power allocation algorithm and OFDMA scheme. To guarantee the number of scheduled users, we use the proportional fair scheduling algorithm to assign the sub-channels to users. Firstly, the complexity of all algorithms is evaluated, then the simulation analysis mainly evaluates two system-level performance parameters, system throughput performance and users fairness. For simulation, we assumed that all users are uniformly distributed in a circular area with a radius of 200 m. The channel state information is an ideal condition. The main simulation parameters are shown in Table 1.

TABLE 1. Simulation Parameter.

Parameter	Value
Cell radius	200 m
Total bandwidth (B)	5 M
Total power (P_{total})	30 dBm
Number of power levels (L)	10
Minimum power unit (ς)	0.1 W
Parameter M	2
Parameter τ	0.1
Parameter α_{ref}	0.01
Parameter ε	0.002
Path loss	COST-231-HATA
Fading	Rayleigh flat fading
Noise power spectral density	-173 dBm/Hz
Minimum data rate requirement	1 Mbps
Path loss exponent (ν)	3

B. COMPLEXITY ANALYSIS

In this section, we will analyze the complexity of all NOMA schemes, including global optimal search, spectrum resource and power allocation with adaptive proportional fair user pairing and the most classical matching theory and water-filling power allocation. The global optimal search requires the exhaustive search. In the intra-sub-channel user paring and power allocation stage, we assume that the upper limit of the multiplexed user on each sub-channel is M_u. So the number of candidate user sets is $C_N^{M_u} = \frac{N!}{M_u!N-M_u!}$. For the power allocation of the each paired user, we exhaustively consider all the power level L and allocated power to each user in a dynamic iterative manner. Then the most profitable matched pair is selected according to maximizing utility value $u_{k,l}$. For the worst case, the candidate power sets is L^{M_u+1}. In inter-sub-channel power allocation stage, by combining the utility value $u_{k,l}$ implemented on all the sub-channels and then selected the most profitable according to the objective function as discussed in Eq. (11) of Section III. The candidate utility value sets is $C_{L*K}^K = \frac{(L*K)!}{K!(L*K-K)}$. So, for the worst case, in order to achieve maximum utility $U_{K,N}^L$, the complexity of the global optimal search can be expressed as $O\left(\frac{N!}{M_u!N-M_u} * L^{M_u+1} * K\right)$. For the matching theory and water-filling power allocation scheme, its complexity comes from two stages, the sorting phase and the matching phase, and in a few existing works, e.g., [10], [12] proved that the

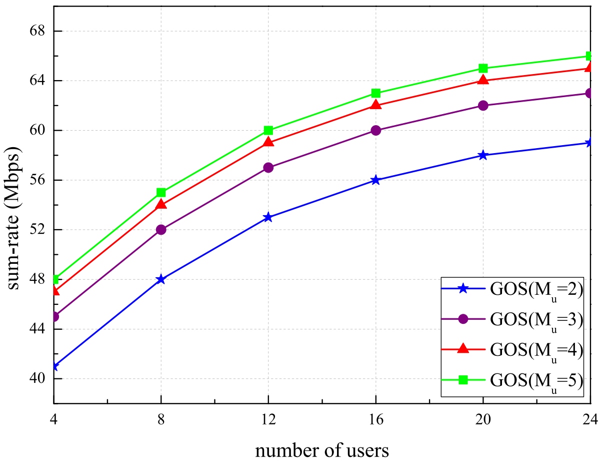


FIGURE 3. GOS performance for different parameter M_u vs. number of users.

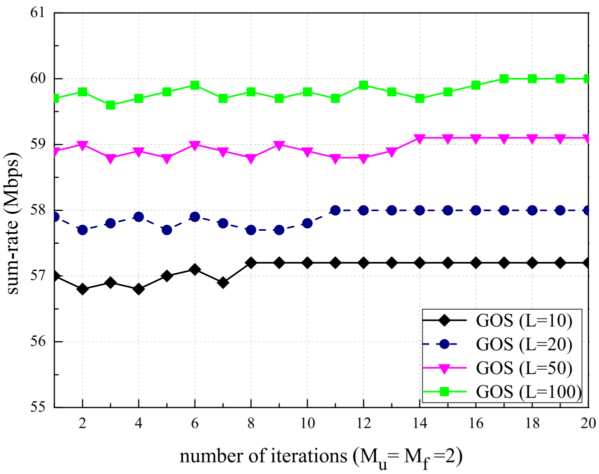


FIGURE 4. GOS performance for different parameter L vs. number of users.

complexity can be expressed as $O(NK^2)$. For the spectrum resource and power allocation with adaptive proportional fair user pairing algorithm, its complexity comes from three stages. First, the complexity of calculating scheduling priority of all users on each sub-channel and allocating power for the paired users with higher priority is $O(KN)$. Then, the BS counted the priority variance $\xi(t)$ of all users and determined the order $N(t)$ to update forgetting factor $\alpha(t)$ on each sub-channel k, the complexity is $O(KN)$. Finally, in next time slot, the complexity of recalculating the average *ECG* of all the users on each sub-channel is $O(KN)$. The above complexity is the summation relationship, so the total complexity can be expressed $O(KN)$. Therefore, the spectrum resource and power allocation with adaptive proportional fair user pairing has the lowest complexity.

C. NUMERICAL RESULTS

1) PERFORMANCE IN THROUGHPUT AND UPPER BOUND

The initial stage of the simulations, we analyze the GOS that provides the optimal performance upper bound. First, the impact of the parameter M_u with the different number of multiplexed users 2, 3, 4, 5 and the parameter L with the different power levels 10, 20, 50, 100 is evaluated, respectively in Fig. 3. from Fig. 4. We find out that the total sum-rate increase with the number of users increase. Increasing M_u leads to more total throughput for NOMA schemes, because more users are allowed to share the same sub-channel. But the sum-rate growth becomes slower as M_u increase. As shown in Fig. 4, as the power level gradually becomes larger, a greater granularity of power discretization is provided, the number of iterations that are aggregated into the optimal solution increased. Therefore, the solution quality can be improved, so it can provide a higher system performance. This is because, in order to provide an upper bound, the objective function $U_{K,N}^L$ is over-computed. From the Fig. 4, a majority of the iterations is part of the tailing-off effect. The utility $U_{K,N}^L$ both approach the achievable values with 15 iteration or fewer. Note that each iteration has polynomial-time complexity.

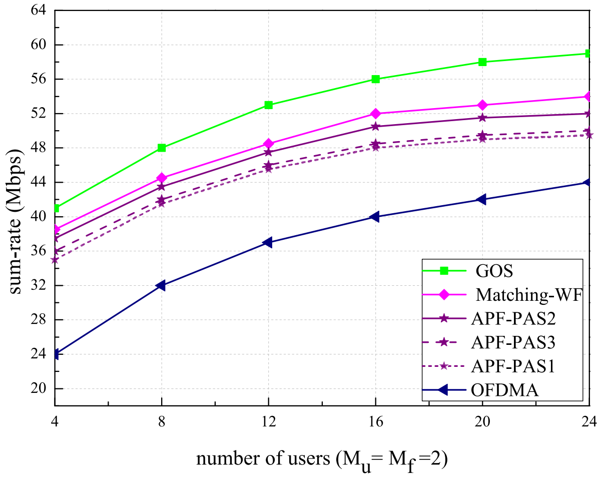


FIGURE 5. Performance for different algorithms vs. number of users.

The next part of our simulation study the characteristics of users throughput performance for GOS, matching theory and water-filling power allocation, spectrum resource and power allocation with adaptive proportional fair user pairing and OFDMA scheme of proportional fair scheduling algorithm, respectively in Fig. 5. For the GOS, in the phase of intra-sub-channel user pairing and power allocation, the global optimal search performs group search for paired users, then dynamically adjust the power to converge to a maximum utility value $u_{k,l}$ for each power level l. In the phase of inter-sub-channel power allocation, under the constraint of total power, we exhaustively combinate the $u_{k,l}$ to get the maximum utility value $U_{K,N}^L$. The matching theory performs bilateral matching through the sub-channels' and the users' preference list, and the users on the sub-channel use water-filling for power allocation. Both of above algorithms have higher throughput performance gains. For spectrum resource and power allocation with adaptive proportional fair user pairing, the

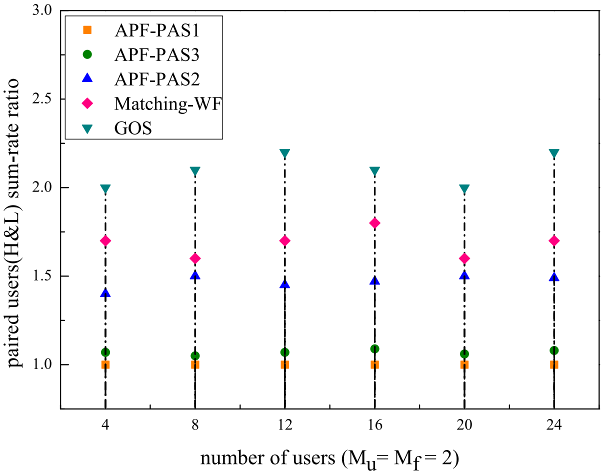


FIGURE 6. Performance ratio for different algorithms vs. number of users.

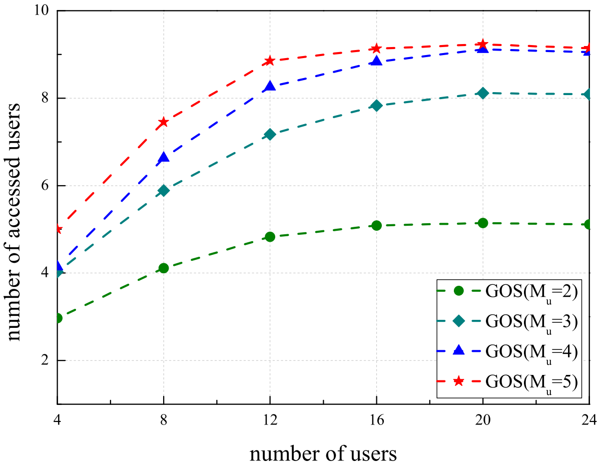


FIGURE 7. GOS number of accessed users for different parameter M vs. number of the users.

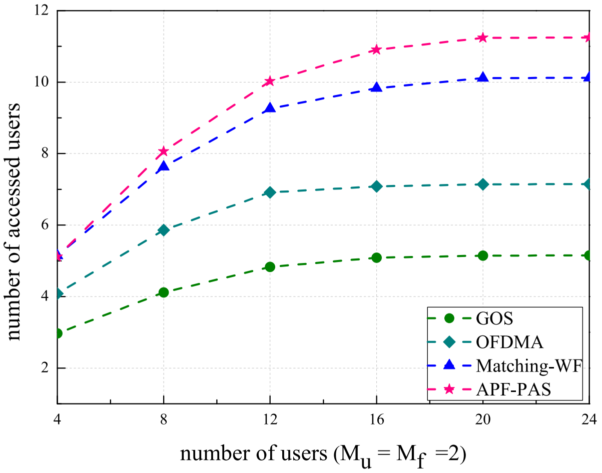


FIGURE 8. Number of accessed users for different algorithms vs. number of the users.

matching process of users and sub-channels is simplified by the determined user pairing relationship. The power allocation is based on the above three schemes, respectively. Under the constraint that the SIC is correctly executed, the scheme 2 allocates the remaining power to the user with high channel gain as much as possible. As shown in Fig. 5, its system throughput performance is superior to scheme 1 and scheme 3 with proportional rate constraint. Scheme 1 directly allocates the total power of the sub-channel with proportional rate constraint, but the scheme 3 first satisfies the minimum data rate, and then allocates the remaining power with proportional rate constraint. we found that scheme 3 performs slightly better than scheme 1. OFDMA has the lowest performance gain since each sub-channel can only be assigned to one user.

From Fig. 6, the performance of cell-edge and cell-center users is evaluated. We add up all the paired users throughout, respectively, and calculating the performance ratio for high(cell-center) and low(cell-edge) ECG users. The GOS and matching theory with water-filling power allocation algorithms achieved throughput performance between the users of the high- and low-channel gain users are quite different. Because, the high ECG users usually occupied more than one sub-channel, but the paired low ECG users always different, and the water-filling power allocation more willing to allocate power to high ECG users. However, the APF user pairing algorithm improved the fairness for all the users sharing the sub-channels. The PAS. 2 allocates all the remaining power to high ECG user, and result a larger sum-rate ratio of paired users. The PAS. 1 allocates the total power of the each sub-channel with the proportional rate constraint, so the paired users sum-rate ratio is 1:1, and PAS. 3 allocates the total remaining power of each sub-channel with the proportional rate constraint, its sum-rate ratio also close to 1:1. Therefore, the PAS. 3 effectively guarantees the fairness and QoS of the paired users without losing more throughput performance.

2) PERFORMANCE IN FAIRNESS

The next part of our performance firstly shows the number of accessed users vs. the number of users. The GOS is firstly evaluated in Fig. 7, since more users have the opportunity to access the sub-channels, the number of accessed users is higher as the parameter M_u increased. But with the parameter M_u increased, the gain of the number of scheduled users is gradually reduced. So could not increase the number of scheduled users by increasing the number of multiplexed users without restrictions. Moreover, for the GOS, when the sub-channel power is large, the best channel gain user is usually selected for pairing, and when the sub-channel power is small, the total power of the sub-channel is generally allocated to the high-channel gain user. So it results in very few users accessed the sub-channel. It impairs the right of cell-edge users to access the spectrum resources. So the GOS is unrealistic, it is only suitable for determining the system performance upper bound.

From Fig. 8, when the number of users is smaller than the number of sub-channels, all the users have accessed to the

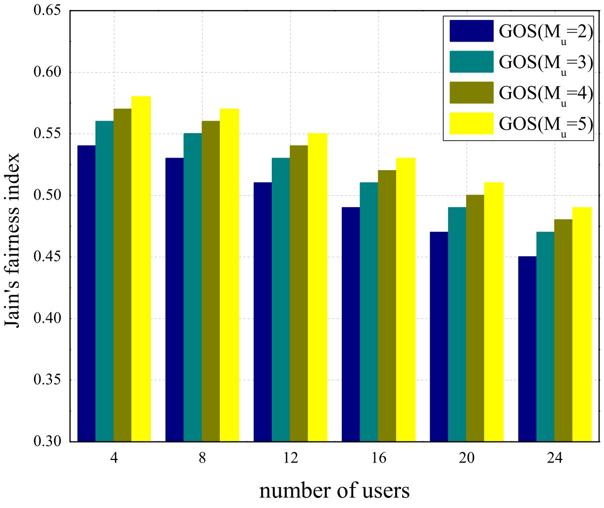


FIGURE 9. Fairness for different parameter M_u vs. number of users.

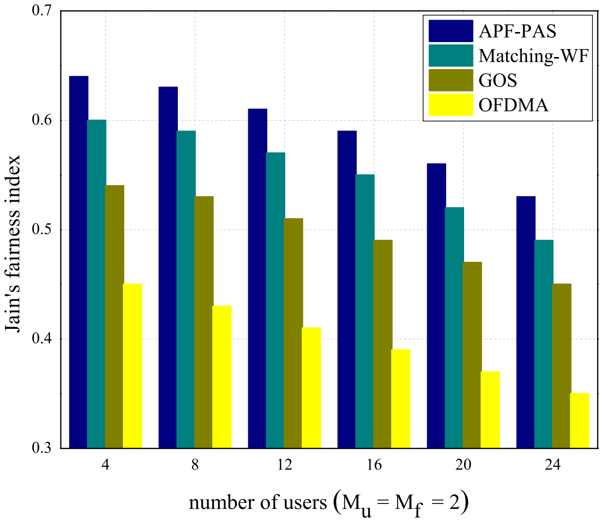


FIGURE 10. Fairness for different algorithms vs. number of users.

spectrum resources in both OFDMA and NOMA schemes. For the matching and adaptive proportional fair user pairing algorithms, as the number of users increases, the number of scheduled users tends to a fixed value, and the number of scheduling users of the APF user pairing is higher than the matching theory. But all of them still much larger than the OFDMA scheme. Because, only one user can be allocated to one sub-channel of the OFDMA scheme, there can only be at most K users accessed, this results in a rapid decrease in the percentage of accessed users.

As our finally part of results, we evaluate the fairness of all the NOMA schemes and OFDMA scheme. The fairness of the system is evaluated based on the Jain's fairness index [30].

$$Jain's\ fairness\ index = \frac{\left(\sum_{n=1}^{N} R_n\right)^2}{N \sum_{n=1}^{N} R_n^2} \tag{47}$$

From Fig. 9, the fairness of GOS increases as the parameter M_u increased. Because a large M_u provides more flexibility in sub-channel allocation among the users. Due to the saturation effect, the influence of constraint increases as the number of multiplexed users per sub-channel is increased. But, note that in Fig. 3, the improvement of throughput is marginal as the parameter M_u increased. Since, it is appropriate to choose a moderate parameter M_u for the NOMA system.

From Fig. 10 and Fig. 11, the fairness of all the NOMA schemes are better than that of OFDMA scheme. The fairness index of the adaptive proportional fair user pairing algorithm is obtained with the average of three power allocation schemes. It dynamically adjusts the forgetting factor α to ensure the fairness of long-term and short-term users. But the GOS and matching theory sacrificing the interests of vulnerable groups at the same time as maximizing benefits. In fact, in NOMA, users having a low ECG are given the possibility of being paired with other users (high ECG) on certain sub-channel and allocated higher power level than that of the users close to the center of the cell. On the contrary, when PF scheduling is used for the OFDMA, only one user is scheduled on each sub-channel, therefore, the user of the cell-edge is deprived of accessing a large number of sub-channels. From this perspective, we can see that NOMA is more fair than OFDMA, because, it compensates for the impact of distance on users' ECG by providing appropriate power levels. However, the fairness of all schemes decreases as the number of users increases and the fairness of all schemes increases as the number of sub-channel increases in Fig. 10 and Fig. 11.

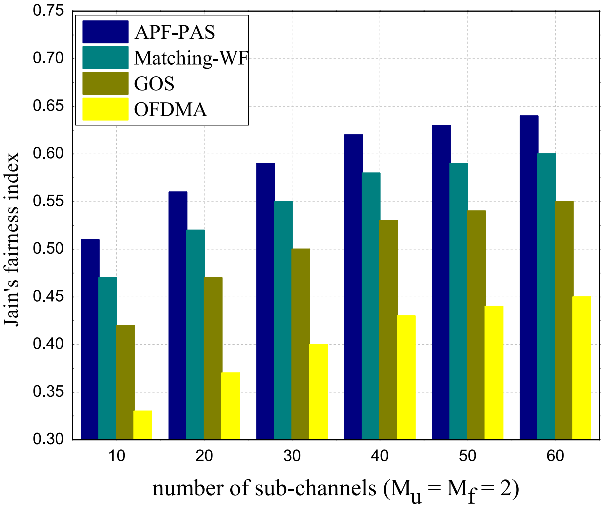


FIGURE 11. Fairness for different algorithms vs. number of sub-channels.

VI. CONCLUSION

This paper studied the spectrum resource and power allocation problem in the downlink NOMA systems. An extension of the work is the consideration of upper bound. In this case, one solution is to perform a global optimal search. Numerical results demonstrated that the proposed algorithm can ensure the best throughput performance by the different power discretization granularities. But, because the highest complexity and the worst fairness that is considered unrealistic. So selecting the moderate power discretization parameters and the number of multiplexed users are necessary to reduce complexity. In this case, a low complexity spectrum resource and power allocation with adaptive proportional fair user pairing algorithm is proposed, the APF user pairing can achieve both the fairness and high system sum-rate. Additionally, by exploring different power allocation schemes with proportional rate constraint, we found that it can not only ensure the correct execution of SIC, but also the throughput performance achieved close to the upper bound. The development of optimization algorithm for this problem is subject to further study.

APPENDIX A

Taking into account the objective function and constraints, using the Lagrangian dual decomposition approach to solve the formulated problem in (27) to (29) and (38) to (40). First of all describe the problem of first power allocation scheme solving process, the third power allocation scheme is similar to the first, therefore, only the results are given.

$$\begin{aligned} \gamma_{k,m} &= \frac{p_{k,m}\left|h_{k,m}\right|^2}{\sigma^2} \\ \gamma_{k,n} &= \frac{p_{k,n}\left|h_{k,n}\right|^2}{p_{k,m}\left|h_{k,n}\right|^2+\sigma^2} \end{aligned} \tag{48}$$

$$\begin{aligned} F &= B_k\log_2\left(\left(1+\gamma_{k,m}\right)\left(1+\gamma_{k,n}\right)\right) \\ &\quad -\mu\left(\frac{B_k\log_2\left(1+\gamma_{k,m}\right)}{R_{\min}}-\frac{B_k\log_2\left(1+\gamma_{k,n}\right)}{R_{\min}}\right) \\ &\quad -\lambda\ \left(p_{k,m}+p_{k,n}-p_k\right) \end{aligned} \tag{49}$$

where μ and λ represent the Lagrange multipliers. Differentiating against $p_{k,n}$, $p_{k,m}$, μ and λ respectively, we obtain

$$\frac{dF}{dp_{k,n}} = -\lambda+\frac{\gamma_{k,n}\mu B_k}{R_{\min}p_{k,n}\left(1+\gamma_{k,n}\right)\ln 2}+\frac{\gamma_{k,n}B_k}{p_{k,n}\left(1+\gamma_{k,n}\right)\ln 2} \tag{50}$$

$$\begin{aligned} \frac{dF}{dp_{k,m}} &= \frac{B_k\left(\frac{\gamma_{k,m}\left(1+\gamma_{k,n}\right)}{p_{k,m}}-\frac{\left(\gamma_{k,n}\right)^2\left(1+\gamma_{k,m}\right)}{p_{k,n}}\right)}{\left(1+\gamma_{k,n}\right)\left(1+\gamma_{k,m}\right)\ln 2}-\lambda \\ &\quad -\mu\left(\frac{B_k\gamma_{k,m}}{R_{\min}p_{k,m}\left(1+\gamma_{k,m}\right)\ln 2}\right. \\ &\quad \left.+\frac{B_k\left(\gamma_{k,n}\right)^2}{R_{\min}p_{k,n}\left(1+\gamma_{k,n}\right)\ln 2}\right) \end{aligned} \tag{51}$$

$$\frac{dF}{d\lambda} = p_k-\left(p_{k,m}+p_{k,n}\right) \tag{52}$$

$$\frac{dF}{d\mu} = \frac{B_k\log_2\left(1+\gamma_{k,m}\right)}{R_{\min}}-\frac{B_k\log_2\left(1+\gamma_{k,n}\right)}{R_{\min}} \tag{53}$$

Setting each of these equations to zero and solving (49) for the Lagrange variable λ which can be used to solve (50) for $p_{k,m}$

$$p_{k,m} = \frac{R_{\min}\sigma^2\left(\left|h_{k,m}\right|^2-\left|h_{k,n}\right|^2\right)}{2\mu\left|h_{k,m}\right|^2\left|h_{k,n}\right|^2}-\frac{\sigma^2\left(\left|h_{k,m}\right|^2+\left|h_{k,n}\right|^2\right)}{2\left|h_{k,m}\right|^2\left|h_{k,n}\right|^2} \tag{54}$$

Next, solving for $p_{k,n}$ by using (51), we assume that the total transmit power of all sub-channel are equal, so which we can obtain that

$$\begin{aligned} p_{k,n} &= \frac{2p_k\left|h_{k,m}\right|^2\left|h_{k,n}\right|^2}{2\left|h_{k,m}\right|^2\left|h_{k,n}\right|^2}+\frac{\sigma^2\left(\left|h_{k,m}\right|^2+\left|h_{k,n}\right|^2\right)}{2\left|h_{k,m}\right|^2\left|h_{k,n}\right|^2} \\ &\quad -\frac{R_{\min}\sigma^2\left(\left|h_{k,m}\right|^2-\left|h_{k,n}\right|^2\right)}{2\mu\left|h_{k,m}\right|^2\left|h_{k,n}\right|^2} \end{aligned} \tag{55}$$

Since (53), (54) still have the Lagrange multiplier μ so it must be solve in order to allocate the transmission power for user m, n, Using the (52) for the Lagrange variable μ we obtain

$$\mu = \frac{\sigma^2R_{\min}\left(\left|h_{k,m}\right|^2-\left|h_{k,n}\right|^2\right)}{\sigma^2\left(\left|h_{k,m}\right|^2+\left|h_{k,n}\right|^2\right)+\psi} \tag{56}$$

where

$$\begin{aligned} \psi &= \sqrt{\left(\sigma^2\left(\left|h_{k,m}\right|^2+\left|h_{k,n}\right|^2\right)\right)^2+4\left|h_{k,m}\right|^2\left|h_{k,n}\right|^4p_k\sigma^2} \\ &\quad -\sigma^2\left(\left|h_{k,m}\right|^2+\left|h_{k,n}\right|^2\right) \end{aligned} \tag{57}$$

As for oblective function (38) and constrations (39) and (40), We use the same solution idea.

APPENDIX B

For the second power allocation scheme for user m, n on the sub-channel k, we give the proof in this section. The objective function under the second scheme can be expressed as

$$\begin{aligned} F &= B_k\log_2\left(\left(1+\gamma_{k,m}\right)\left(1+\gamma_{k,n}\right)\right) \\ &= B_s\log_2\left(\left(1+\frac{p_{k,m}\left|h_{k,m}\right|^2}{\sigma^2}\right)\left(1+\frac{p_{k,n}\left|h_{k,n}\right|^2}{p_{k,m}\left|h_{k,n}\right|^2+\sigma^2}\right)\right) \end{aligned} \tag{58}$$

where F is the total throughput. Differentiating against $p_{k,m}$, $p_{k,n}$, respectively, we obtain

$$\begin{aligned} \frac{dF}{dp_{k,m}} &= \frac{\left|h_{k,m}\right|^2}{\left(p_{k,m}\left|h_{k,m}\right|^2+\sigma^2\right)\ln 2}-\frac{\left|h_{k,n}\right|^2}{\left(p_{k,m}\left|h_{k,n}\right|^2+\sigma^2\right)\ln 2} \\ &\quad +\frac{\left|h_{k,n}\right|^2}{\left(\left(p_{k,m}\left|h_{k,n}\right|^2+\sigma^2\right)+p_{k,n}\left|h_{k,n}\right|^2\right)\ln 2} \end{aligned} \tag{59}$$

$$\frac{dF}{dp_{k,n}} = \frac{\left|h_{k,n}\right|^2}{\left(\left(p_{k,m}\left|h_{k,n}\right|^2+\sigma^2\right)+p_{k,n}\left|h_{k,n}\right|^2\right)\ln 2} \tag{60}$$

It is obvious that

$$\frac{\left|h_{k,m}\right|^2}{\left(p_{k,m}\left|h_{k,m}\right|^2+\sigma^2\right)\ln 2} \div \frac{\left|h_{k,n}\right|^2}{\left(p_{k,m}\left|h_{k,n}\right|^2+\sigma^2\right)\ln 2} = \frac{\left|h_{k,m}\right|^2\left|h_{k,n}\right|^2 p_{k,m}+\left|h_{k,m}\right|^2\sigma^2}{\left|h_{k,m}\right|^2\left|h_{k,n}\right|^2 p_{k,m}+\left|h_{k,n}\right|^2\sigma^2} \quad (61)$$

For the constraint that

$$\left|h_{k,m}\right|^2 > \left|h_{k,n}\right|^2 \quad (62)$$

Thus we can get

$$\frac{dF}{dp_{k,m}} > \frac{dF}{dp_{k,n}} \quad (63)$$

According to (62), for the same power increment, the user with higher channel gain has the a greater contribution to the total throughput. Thus, after ensuring minimum data rate requirements for the user m, n on the sub-channel, the BS should allocate the remaining power to the high channel gain user as much as possible.

REFERENCES

[1] S. V. D. Meer, J. Keeney, and L. Fallon, "5G networks must be autonomic," in *Proc. NOMS.*, Apr. 2018, pp. 1–5.

[2] L. Lei, D. Yuan, C. K. Ho, and S. Sun, "Joint optimization of power and channel allocation with non-orthogonal multiple access for 5g cellular systems," in *Proc. IEEE GLOBECOM* Dec. 2014, pp. 1–6.

[3] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181–2195, Oct. 2017.

[4] Y. Wu, K. Ni, C. Zhang, L. P. Qian, and D. H. K. Tsang, "NOMA-assisted multi-access mobile edge computing: A joint optimization of computation offloading and time allocation," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 12244–12258, Dec. 2018.

[5] M. Zeng, A. Yadav, O. A. Dobre, and H. V. Poor, "Energy-efficient power allocation for MIMO-NOMA with multiple users in a cluster," *IEEE Access*, vol. 6, pp. 5170–5181, 2018.

[6] Z. Chen, Z. Ding, X. Dai, and R. Zhang, "An optimization perspective of the superiority of NOMA compared to conventional OMA," *IEEE Trans. Sig. Proc.*, vol. 65, no. 19, pp. 5191–5202, Oct. 2017.

[7] M. Zeng, A. Yadav, O. A. Dobre, G. I. Tsiropoulos, and H. V. Poor, "On the sum rate of MIMO-NOMA and MIMO-OMA systems," *IEEE Wireless Commun. Lett.*, vol. 6, no. 4, pp. 534–537, Aug. 2017.

[8] Q. Li, H. Niu, A. T. Papathanassiou, and G. Wu, "5G network capacity: Key elements and technologies," *IEEE Veh. Technol. Mag.*, vol. 9, no. 1, pp. 71–78, Mar. 2014.

[9] 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 Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8580–8594, Dec. 2016.

[10] B. Di, S. Bayat, L. Song, and Y. Li, "Radio resource allocation for downlink non-orthogonal multiple access (NOMA) networks using matching theory," in *Proc. IEEE GLOBECOM*, Dec. 2015, pp. 1–6.

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

[12] W. Liang, Z. Ding, Y. Li, and L. Song, "User pairing for downlink non-orthogonal multiple access networks using matching algorithm," *IEEE Trans. Commun.*, vol. 65, no. 12, pp. 5319–5332, Dec. 2017.

[13] B. Di, L. Song, Y. Li, and G. Y. Li, "Non-orthogonal multiple access for high-reliable and low-latency V2X communications in 5G systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2383–2397, Oct. 2017.

[14] Z. Sheng, X. Su, and X. Zhang, "A novel power allocation method for non-orthogonal multiple access in cellular uplink network," in *Proc. Int. Conf. Intell. Environ.*, Aug. 2017, pp. 157–159.

[15] Z. Q. Al-Abbasi and D. K. C. So, "Resource allocation in non-orthogonal and hybrid multiple access system with proportional rate constraint," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6309–6320, Oct. 2017.

[16] Y. Wu, L. P. Qian, H. Mao, X. Yang, H. Zhou, and X. Shen, "Optimal power allocation and scheduling for non-orthogonal multiple access relay-assisted networks," *IEEE Trans. Mobile Comput.*, vol. 17, no. 11, pp. 2591–2606, Nov. 2018.

[17] M. Zeng, A. Yadav, O. A. Dobre, G. I. Tsiropoulos, and H. V. Poor, "Capacity comparison between MIMO-NOMA and MIMO-OMA with multiple users in a cluster," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2413–2424, Oct. 2017.

[18] W. Cai, C. Chen, L. Bai, Y. Jin, and J. Choi, "Subcarrier and power allocation scheme for downlink OFDM-NOMA systems," *IET Signal Process.*, vol. 11, no. 1, pp. 51–58, Feb. 2017.

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

[20] F. Liu and M. Petrova, "Proportional fair scheduling for downlink single-carrier NOMA systems," in *Proc. IEEE GLOBECOM*, Dec. 2017, pp. 1–7.

[21] E. Okamoto, "An improved proportional fair scheduling in downlink non-orthogonal multiple access system," in *Proc. IEEE 82nd VTC-Fall*, Sep. 2015, pp. 1–5.

[22] Ö. F. Gemici, F. Kara, I. Hokelek, G. K. Kurt, and H. A. Çirpan, "Resource allocation for NOMA downlink systems: Genetic algorithm approach," in *Proc. IEEE TSP*, Jul. 2017, pp. 114–118.

[23] F. Liu, P. Mähönen, and M. Petrova, "Proportional fairness-based user pairing and power allocation for non-orthogonal multiple access," in *Proc. IEEE 26th PIMRC*, Aug. 2015, pp. 1127–1131.

[24] M.-R. Hojeij, C. A. Nour, J. Farah, and C. Douillard, "Waterfilling-based proportional fairness scheduler for downlink non-orthogonal multiple access," *IEEE Wireless Commun. Lett.*, vol. 6, no. 2, pp. 230–233, Apr. 2017.

[25] X. Chen, A. Benjebbour, A. Li, and A. Harada, "Multi-user proportional fair scheduling for uplink non-orthogonal multiple access (NOMA)," in *Proc. IEEE 79th VTC-Spring*, May 2014, pp. 1–5.

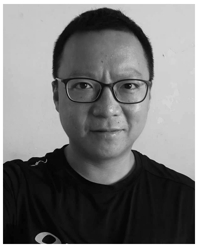
[26] S. M. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Commun. Survey Tuts.*, vol. 19, no. 2, pp. 721–742, 2nd Quart. 2016.

[27] F. Fang, H. Zhang, J. Cheng, S. Roy, and V. C. M. Leung, "Joint user scheduling and power allocation optimization for energy-efficient NOMA systems with imperfect CSI," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2874–2885, Dec. 2017.

[28] J. Guo, X. Wang, J. Yang, J. Zheng, and B. Zhao, "User pairing and power allocation for downlink non-orthogonal multiple access," in *Proc. IEEE Globecom Workshops*, Dec. 2016, pp. 1–6.

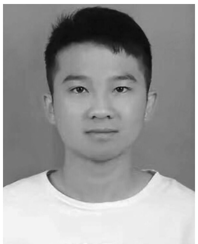
[29] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6010–6023, Aug. 2016.

[30] A. V. Babu and L. Jacob, "Fairness analysis of IEEE 802.11 Multirate wireless LANs," *IEEE Trans. Veh. Technol.*, vol. 56, no. 5, pp. 3073–3088, Sep. 2007.



KEN LONG received the M.S. and Ph.D. degrees in signal and information processing from the Beijing University of Posts and Telecommunications, Beijing, China, in 2004 and 2009, respectively. Since 2009, he has been the Major Investigator of the Mobile Communications Key Laboratory, Chongqing University of Posts and Telecommunications (CQUPT), where he has been a Research Supervisor. He has published some papers on these topics, and holds one patents on wireless technology. His current research interests include wireless communications, next generation mobile networks, and signal processing.

PENGYU WANG is currently pursuing the M.S. degree in information and communication engineering with the Chongqing University of Posts and Telecommunications, Chongqing, China (CQUPT). Since 2016, he has been studying in the Wireless Communication Technology Innovation Laboratory, Broad Band Equipment Mobilization Center. His research interests include wireless communications, next generation mobile networks, and software defined radio.



WEI LI is currently pursuing the M.S degree in information and communication engineering from the Chongqing University of Posts and Telecommunications (CQUPT), Chongqing. Since 2017, he has been studying with the Wireless Communication Technology Innovation Laboratory, Broadband Equipment Mobilization Center. His research interest includes the key technologies for next generation wireless communication systems, especially the signal processing and resource allocation in communication networks.

DEJIAN CHEN received the B.S. degree from the Shaanxi University of Science and Technology, Xi'an, China, in 2016. He is currently pursuing the master's degree in electronic and communication engineering with the Chongqing University of Posts and Telecommunications, Chongqing, China. Since 2016, he has been studying in the Wireless Communication Technology Innovation Laboratory, Broad Band Equipment Mobilization Center. His research interests include the NOMA technology in 5G and software defined radio.

•••