A low complexity user scheduling algorithm aimed for the maximum number of active users in NOMA system

Chengyi Liu | Geng Chen | Song Xing

1 School of Electronic and Information Engineering, Changshu Institute of Technology, Suzhou, China
2 College of Electronic and Information Engineering, Shandong University of Science and Technology, Qingdao, China
3 Department of Information Systems, California State University, Los Angeles, USA

Abstract
User scheduling algorithms in non-orthogonal multiple access (NOMA) have attracted much attention to improve the performance of the communication system. Here, a low complexity user scheduling algorithm aimed for the maximum number of active users in the single NOMA cluster ensuring the individual minimum rate requirement is proposed. Based on the precondition of the maximisation of the number of active users, the maximum sum-rate strategy is further integrated into the algorithm to compose a multi-round user scheduling algorithm. Moreover, a computational-complexity reduction algorithm is also introduced and has proved to be conveniently used in practical operation. Simulation results have shown that, compared with other existing user scheduling algorithms, the proposed user scheduling algorithms can achieve the maximum number of active users while ensuring their individual minimum rate requirements, and significantly improve the performance of the sum-rate among the cases when the maximum number of active users is achieved. Furthermore, the proposed multi-round and computational-complexity reduction user scheduling algorithms have shown better performance than the exhaustive search in terms of the computational complexity.

1 | INTRODUCTION

The non-orthogonal multiple access (NOMA), which superimposes the signals of multiple users over the same spectrum resource via power domain division or code division based on the users’ respective channel gain differences at the base station (BS), has gained significant attention for the fifth generation (5G) and beyond communication systems [1–3]. At each receiver, successive interference cancellation (SIC) technology is applied to retrieve its own signal from the single power-composed signal of all users in the cluster [4]. In NOMA systems, power resource allocation plays an important role and has been studied for performance evaluation and improvement in terms of the sum-rate maximisation, the energy-efficiency maximisation, the outage probabilities minimisation, and the optimal user pairing, some of which are listed below.

1.1 | Existing research on NOMA

1.1.1 | Sum-rate maximisation

Achieving the maximum sum-rate of the NOMA system is the most attractive objective, on which multiple existing literatures focused [5–9]. Specifically, the authors in [5] investigate the optimal power allocation to maximise the network sum-rate with the individual quality of service (QoS) constraints applied in both single-input single-output (SISO) and multiple-input multiple-output (MIMO) scenarios. Weighted sum-rate maximisation and sum-rate maximisation with QoS constraints are proposed in [6] for multiple NOMA clusters scene. To maximise the secrecy sum-rate subject to the users’ QoS requirements, an optimal power allocation policy is given in [7]. A closed form of the power allocation in single cluster to achieve the sum-rate maximisation is obtained in [8], proving that the sum-rate...
maximisation problem of multiple users in one NOMA cluster is convex. For both uplink and downlink NOMAs, a sum-rate maximisation problem in a cell such that the user clustering (i.e. grouping users into multiple clusters) and power allocations in NOMA clusters can be optimised under transmission power constraints, minimum rate requirements of the users, and SIC constraints in [9].

1.1.2 | Energy-efficiency maximisation

Energy efficiency is another important performance evaluation term in communication systems, where the resource allocation to maximise the energy efficiency of the NOMA system is studied in [10–16]. In [10–12], three communication modes such as device-to-device, machine-to-machine, and relay are considered to maximise the energy efficiency in NOMA systems. The authors in [13] investigate the energy-efficient optimisation problem with a dual-connectivity mode by using rate-splitting method to transform the original non-convex problem into a convex one. A joint allocation of power and bandwidth, which is an energy-efficient algorithm in NOMA systems, is proposed with the equivalent difference-of-convex (DC) functions method in [14]. To maximise the energy efficiency, a power allocation strategy subject to meet the individual minimum rate requirement, is proposed by decoupling the optimal problem into two concatenate sub-problems in [15]. The authors in [16] study a low-complexity, energy-efficient maximisation problem by decoupling the power allocation and sub-channel assignment in NOMA systems.

1.1.3 | User scheduling schemes

Furthermore, user scheduling has a significant impact on the performance of NOMA systems. In [17], a clustering and power allocation approach is proposed by splitting the users with high channel gains into different clusters to achieve the sum-rate maximisation. In [9], by classifying the users into two classes, channel gains into different clusters to achieve the sum-rate allocation approach is proposed by splitting the users with high performance of NOMA systems. In [17], a clustering and power allocations method in [14]. To maximise the energy efficiency, a power allocation strategy subject to meet the individual minimum rate requirement, is proposed by decoupling the optimal problem into two concatenate sub-problems in [15]. The authors in [16] study a low-complexity, energy-efficient maximisation problem by decoupling the power allocation and sub-channel assignment in NOMA systems.

1.1.3 | User scheduling schemes

Furthermore, user scheduling has a significant impact on the performance of NOMA systems. In [17], a clustering and power allocation approach is proposed by splitting the users with high channel gains into different clusters to achieve the sum-rate maximisation. In [9], by classifying the users into two classes, a sub-optimal user clustering algorithm with low computational complexity is proposed for downlink and uplink transmission, respectively. Considering additional complexity introduced by SIC at the receiver, each NOMA cluster is restricted to include only two users in some literatures [18–20]. In [18], an optimal user pairing algorithm for multiple NOMA clusters is studied to propose a computational-complexity reduced user pairing scheme to achieve the sum-rate maximisation. A novel low-complexity sub-optimal user scheduling algorithm was proposed to maximise the system energy efficiency with imperfect channel state information (CSI) in [19]. The authors in [20] propose a matching algorithm in the downlink NOMA network to optimise the two users pairing and the power allocation between the weak users located at the edge of the cell and the strong users located close to the BS.

1.2 | Motivation and contributions

According to the aforementioned works, an implicit assumption is made that all users are active at the same time in NOMA systems. However, it is not always true for internet-of-things (IoT) networks connecting a very large number of devices with various data rate requirements or limited resource. On the other hand, more devices are active, more accurate and comprehensive information of the NOMA system can be obtained in IoTs.

However, according to the NOMA principle and the various minimum rate requirements among multiple users in a single NOMA cluster, the user scheduling result will dynamically change, which could be solved by using the exhaustive search but with unapplicable computational complexity. Therefore, in the NOMA system, it is still an open problem on how to choose the active users facing multiple users with various minimum rate requirements with lower computational complexity when the BS cannot support the QoS of all users. And its optimal solution is unknown yet.

Motivated by the introduced works on NOMA systems and the question on all active users assumption, the focus is on developing the practical user scheduling solution for multi-users in single NOMA cluster. The contributions are outlined as follows:

- In the single NOMA cluster, a user scheduling problem is formulated such that the number of active users can be maximised under the constraints of individual minimum rate requirements of the users and the total transmission power.
- Due to the combinatorial nature of the formulated and mixed integer non-linear programming problems, we propose a multi-round user scheduling scheme, which exploits the minimal power requirement values for all combinations of the active users while ensuring their individual minimum rate requirements by dividing the whole power range into multiple intervals to obtain the maximum number of active users.
- Among multiple optimal solutions of the original problem provided by the nature of the integer programming problem, a sum-rate improvement scheme is further integrated into the proposed multi-round user scheduling scheme by specifically choosing the active users.
- According to the given solution attaining the proposed multi-round user scheduling scheme, a practical computational-complexity reduced user scheduling algorithm (CCRUSA) is further derived and proved.

1.3 | Paper organisation

The remaining is organised as follows. Section 2 describes the system model and problem formulation. Section 3 proposes a proposed multi-round user scheduling algorithm (MRUSA) and a CCRUSA, followed by the analysis of the computational
2 | SYSTEM MODEL AND PROBLEM FORMULATION

2.1 | System model

Without loss of generality, we assume that there are $M$ single-antenna users denoted by $\mathbb{M} = (m)_m^M$ which intend to be active in a downlink single-NOMA-cluster system. The channel gain between the single-antenna BS and the user $m$ is denoted by $h_m$, and the channel gains are sorted in the ascending order, that is, $b_1 \leq \cdots \leq b_M$. The BS transmits a superimposed signal for the clustered users as

$$ x = \sum_{m=1}^{M} \sqrt{p_m} x_m, $$

(1)

where $x_m$ and $p_m, \forall m$, denote the signal and the allocated power for the user $m$, respectively.

The received signal at the user $m$ can be obtained as

$$ y_m = b_m \sum_{l=1}^{M} \sqrt{p_l} x_l + n_m, \forall m \in \mathbb{M}, $$

(2)

where $n_m$ denotes the additive white noise at the user $m$ with the variance $\sigma^2$.

In downlink NOMA, the user $m$ receives its own signal from the received composite power signals of all users belonging to the same cluster by implementing SIC. The achievable rate of the user $m$ can be given by

$$ R_m = \log_2 \left(1 + \frac{|b_m|^2 p_m}{|b_m|^2 \sum_{l=m+1}^{M} |\beta_l|^2 + \sigma^2}\right), \forall m \in \mathbb{M}. $$

(3)

2.2 | Problem formulation

Our objective is to obtain the maximum number of active users via user scheduling. Thus, it is necessary to define a indicator index vector $\mathbf{\beta} = (\beta_1, ..., \beta_M)^T$ to represent the user activation in the single NOMA cluster, whose element is given by

$$ \beta_m = \begin{cases} 1, \text{ user } m \text{ is active} \\ 0, \text{ otherwise.} \end{cases} $$

(4)

Then, the optimisation problem is formulated as

$$ \begin{align*} & \underset{\mathbf{\beta} \geq 0, \mathbf{\beta} \in \{0,1\}^M}{\max} \sum_{m=1}^{M} \beta_m p_m \\ & \text{s.t.} \sum_{m=1}^{M} \beta_m p_m \leq P_{\text{max}} \\ & \log_2 \left(1 + \frac{|b_m|^2 \beta_m p_m}{|b_m|^2 \sum_{l=m+1}^{M} |\beta_l|^2 |p_l| + \sigma^2}\right) \geq \beta_m R_m, \forall m, \end{align*} $$

(5)

where $R_m$ is the individual minimum rate required for the user $m$, $P_{\text{max}}$ is the maximum transmission power for the single NOMA cluster, and the allocated power vector $\mathbf{p} = (p_1, \ldots, p_M)^T$, respectively.

The maximisation objective function in (5) is to obtain the maximum number of active users, which is the primary motivation of this work. In (5), the first constraint implies that the sum-power allocated to all the active users cannot exceed $P_{\text{max}}$. If $\mathbf{\beta}$ is given, all the active users should obtain the achieved rate no less than their individual minimum rate requirements, which is decided in the second constraint.

3 | SOLUTION OF THE OPTIMISATION PROBLEM

The problem in (5) is a mixed integer non-linear programming problem, in which the optimal solution can be obtained by an exhaustive search. Since the exhaustive search needs to be performed with all combination cases on each user, the optimal solution of user scheduling is easy to obtain only when the number of user equipment is small. However, the computational complexity of exhaustive search is not tolerated when the number of users gets higher.

To reduce the exhaustive search time, in this section, we propose a multi-round user scheduling scheme under general conditions, and then specify a computational-complexity reduced scheme of user scheduling in some practical use. Finally, based on the analysis of the computational complexity of the algorithms, a hybrid user scheduling algorithm is introduced.

3.1 | Multi-round user scheduling algorithm

To obtain the optimal solution of user scheduling problem, a two-stage decision method is proposed. In Stage 1, the feasible points of the user scheduling problem is given. To further achieve the better transmission rate, Stage 2 is then operated.

Stage 1: Decision for the feasible points of the user scheduling problem.

The feasible points of the user scheduling problem are given by the following lemma.

**Lemma 1.** Assuming that all the $C_N^\mathbb{M}$ inactive users combinations, that is, $\pi_1^N, \pi_2^N, \ldots$, and $\pi_{\mathbb{M}}^N$, for only $N$ inactive users case, are sorted in the declining order of $p_{\text{max}}^{\mathbb{M}}$, that is, $P_{\text{max}}^{\pi_1^N} \geq P_{\text{max}}^{\pi_2^N} \geq \cdots \geq P_{\text{max}}^{\pi_{\mathbb{M}}^N}$, the feasible points of the user scheduling problem (5), that is, the optimal candidates of inactive users, are shown in Table 1, where $\pi_N$ is the general expression of the inactive $N$-user set, and $p_{\text{max}}^{\pi_N}$ is given by

$$ p_{\text{max}}^{\pi_N} = \sum_{l=1}^{M} \left( \frac{2^{R_m^{\mathbb{M}}}}{b_m}\right) \left( \prod_{j=1}^{N-1} \left( \frac{2^{R_m^{\mathbb{M}}}}{b_m}\right) \right). $$

(6)

**Proof.** Please refer to Appendix 1. \[\square\]
Remark 1. It should be noticed that the result in Table 1 defaults the condition that \( P_{\text{max}}^{C_{N-1}} > P_{\text{max}}^{1} \). However, the condition \( P_{\text{max}}^{C_{N-1}} < P_{\text{max}}^{1} \) may be satisfied in some special case, in which the BS can remove the user set \( \pi_{N-1}^{C_{N-1}} \) or \( \pi_{N}^{1} \) to maintain the individual minimum rate requirements of the active users. To achieve the objective designed, the BS prefers to remove \( \pi_{N-1}^{C_{N-1}} \) to achieve the maximum number of active users.

Remark 2. In practical, due to the hardware constraints on SIC, the number of accessible users is restricted. Thus, if the number of accessible users in a single NOMA cluster is limited by \( Q = M - N \), the feasible points of the user scheduling problem given in Table 1 can also hold by setting \( P_{\text{max}}^{C_{N-1}} = \infty \), and remove the first \( N \) rows in the table.

Stage 2: Decision for the inactive users.

As seen in Stage 1, there may still have more than one feasible point. Therefore, extra efforts need to be made to achieve better performance. The optimal inactive users can be finally obtained by the following lemma.

Lemma 2. If \( P_{\text{max}}^{k} \leq P_{\text{max}} < P_{\text{max}}^{k+1} \), the final user scheduling solution with maximal number of active users among multiple feasible points of the original problem, for example, \( \pi_{N}^{k}, \ldots, \pi_{N}^{C_{M}} \), can be obtained as

\[
(\pi_{N}^{\ast})^{k} = \arg \max_{q \in \{k, C_{M}\}} R_{\pi_{N}^{q}}^{k},
\]

where

\[
R_{\pi_{N}^{k}}^{k} = \sum_{m=1}^{M-1} \prod_{s=1}^{M-1} R_{m}^{s} \log_{2} \left[ 1 + \frac{P_{\text{max}} |h_{M-s-1}|^{2}}{\sigma^{2} \prod_{m=1}^{M-1} |h_{M-s-1}|^{2} 2^{R_{m}^{s}} - \sum_{k=1}^{M} \prod_{s=1}^{M} |h_{M-s-1}|^{2} 2^{R_{m}^{s}} } \right].
\]

Proof. Please refer to Appendix 2.

As is well known, the considered user scheduling problem can be solved by using the exhaustive search, but with an intolerable computational complexity. To alleviate the computational burden, the following analysis is given to propose a user scheduling scheme implemented by Algorithm 1.

The objective of the user scheduling problem considered is to seek the maximum number of active users, such that if a feasible condition can support more active users, other conditions with less active users should not be calculated. For this reason, a multi-round user scheduling scheme is designed to search the optimal solution of user scheduling via the outer and inner multi-round mechanisms given as follows.

1) The outer multi-round mechanism:

It can be seen from Table 1 that the whole range of \( P_{\text{max}} \) is decomposed into \( M + 1 \) feasible conditions to obtain the optimal number of inactive users. However, there are \( 2^{M} \) feasible conditions for the candidates of inactive users. Thus, instead of obtaining the final solutions directly, getting the optimal number of inactive users firstly will have better performance in terms of the computational complexity.

\[
\begin{array}{c|c|c}
\text{Interval of } P_{\text{max}} & \text{The number of inactive user} & \text{Interval of } P_{\text{max}} \\
\hline p_{\text{max}}^{C_{N-1}} & 0 & p_{\text{max}}^{C_{N-1}} < p_{\text{max}} < \infty \\
\hline p_{\text{max}}^{1} & 1 & p_{\text{max}}^{1} \leq p_{\text{max}} < p_{\text{max}}^{C_{N-1}} \\
\hline p_{\text{max}}^{C_{M}} & N & p_{\text{max}}^{C_{M}} \leq p_{\text{max}} < p_{\text{max}}^{1} \\
\hline 0 & M & 0 \leq p_{\text{max}} < p_{\text{max}}^{C_{M}} \\
\end{array}
\]

\[
\pi_{0} \neq \emptyset
\]

\[
\pi_{1}^{2}, \ldots, \pi_{C_{N}}^{2},
\]

\[
\pi_{1}^{1}, \pi_{1}^{2}, \ldots, \pi_{C_{N}}^{1},
\]

\[
\pi_{1}^{k}, \pi_{1}^{k+1}, \ldots, \pi_{C_{N}}^{k},
\]

\[
\pi_{1}^{k}, \pi_{1}^{k+1}, \ldots, \pi_{C_{N}}^{k},
\]

\[
\pi_{1}^{M}
\]

\[
\pi_{1}^{M}
\]

\[
\pi_{1}^{M}
\]

\[
\pi_{1}^{M}
\]

\[
\pi_{1}^{M}
\]

\[
\pi_{1}^{M}
\]

\[
\pi_{1}^{M}
\]

\[
\pi_{1}^{M}
\]

\[
\pi_{1}^{M}
\]

\[
\pi_{1}^{M}
\]

\[
\pi_{1}^{M}
\]
Algorithm 1 The proposed multi-round user scheduling algorithm

\begin{algorithm}
\begin{algorithmic}
\State Input: $h_i, \forall i \in \mathbb{M}$ \; $\hat{R}_i, \forall i \in \mathbb{M}$ \; $P_{\text{max}}$ \;
\State Output: $\{\pi_{N, c}^s\}$ \;
\For{$N = 0; N \leq M$} \; \SetLine
\State Calculate $P_{\text{max}}^{c_{N, k}}, \forall i \in \{1, 2, \ldots, C_{M}^{N}\}$, sorted in declining order of $P_{\text{max}}^{c_{N, k}}$ \; \SetLine
\If{$P_{\text{max}}^{c_{N, k}} < P_{\text{max}}^{c_{N, k+1}}$} \; \SetLine
\State $N^* = N$ \; \SetLine
\For{$k = 1; k \leq C_{M}^{N}$} \; \SetLine
\If{$P_{\text{max}}^{c_{N, k}} < P_{\text{max}}^{c_{N, k+1}}$} \; \SetLine
\State $\{\pi_{N, k}^s\}^* = \text{arg max}\{P_{\text{max}}^{c_{N, k}}, P_{\text{max}}^{c_{N, k+1}}, \ldots, P_{\text{max}}^{c_{N, C_{M}^{N}}}\}$ \; \SetLine
\State go to final; \SetLine
\EndIf \; \SetLine
\EndFor \; \SetLine
\EndIf \; \SetLine
\EndFor \; \SetLine
\EndFor \; \SetLine
\State return $\{\pi_{N, c}^s\}$ \;
\end{algorithmic}
\end{algorithm}

An outer multi-round mechanism is conducted from step 1 to step 12 iteratively in Algorithm 1 to search the minimum number of inactive users.

2) The inner multi-round mechanism:
After obtaining $N^*$, the corresponding feasible condition needs to be further decomposed into $C_{M}^{N^*}$ subintervals seen from Table 1. Thus, an inner multi-round mechanism is conducted from step 5 to step 10 iteratively in Algorithm 1 to obtain the feasible points of the problem (5).

Finally, the user scheduling solution having the maximal number of active users with better sum-rate performance is given by step 7 among multiple feasible points according to (7).

It can be concluded that MRUSA not only obtains the feasible points of the original problem via Stage 1, but also seeks the improved sum-rate by Stage 2. Thus, the obtained user scheduling solution is optimal in terms of active users number with higher transmission rate.

3.2 Computational-complexity reduced user scheduling algorithm

Obviously, when more users participate in the user scheduling scheme, it will also incur greater computational complexity. To further alleviate this computational burden when huge active users present in a single NOMA cluster, a CCRUSA is introduced under some practical conditions, which is shown in the following theorem.

**Theorem 1.** Assuming that $\hat{R}_1 \geq \cdots \geq \hat{R}_M$, if $P_{\text{max}}^{c_{N^*, k}} < P_{\text{max}}^{c_{N^*, k+1}}$, $\forall N \in [1, M]$, the optimal inactive users set is given by

$$\pi_{N^*}^k = (m)_N^{k+1}. \hspace{1cm} (9)$$

**Proof.** Please refer to Appendix 3.

Remark 3. Theorem 1 demonstrates the proposed user scheduling scheme in specific cases when $\hat{R}_1 \geq \cdots \geq \hat{R}_M$, which includes the more practical case when $\hat{R}_1 = \cdots = \hat{R}_M$.

In practice, the proposed computational-complexity reduced user scheduling scheme can be implemented by Algorithm 2.

Algorithm 2 The proposed computational-complexity reduced user scheduling algorithm

\begin{algorithm}
\begin{algorithmic}
\State Input: $h_i, \forall i \in \mathbb{M}$ \; $\hat{R}_i, \forall i \in \mathbb{M}$ \; $P_{\text{max}}$ \;
\State Output: $\{\pi_N\}^*$ \;
\For{$N = 0; N \leq M$} \; \SetLine
\State Calculate $P_{\text{max}}^{c_{N^*, k}}$ according to formula (6) \; \SetLine
\If{$P_{\text{max}}^{c_{N^*, k}} \geq P_{\text{max}}^{c_{N^*, k+1}}$} \; \SetLine
\State $\{\pi_N\}^* = (m)_N^{k+1}$ \; \SetLine
\State go to final; \SetLine
\EndIf \; \SetLine
\EndFor \; \SetLine
\State return $\{\pi_N\}^*$ \;
\end{algorithmic}
\end{algorithm}

3.3 Computational-complexity analysis

The computational complexities of the proposed MRUSA and CCRUSA are analysed with given $h_i, \hat{R}_i, \forall i \in \mathbb{M}$ and $P_{\text{max}}$, which are compared with one of the exhaustive search. To evaluate the complexity, the computing of the number of floating-point operations in the algorithms are collected [21]. For the sake of brevity, the computational complexity is denoted by the notation $\mathcal{F}$ and the flops count of each step in each algorithm is given as follows.

For a given $N$, $C_{M}^{N}$ combinations of user cases need to be calculated, in which the flops for each combination are expressed by

$$\mathcal{F}_1^N = 10.5(M^2 + N^2) + 14.5(M - N) - 21MN - 1. \hspace{1cm} (10)$$

Further, assuming that the optimal number of inactive users is $N^*, \forall N^* \in [0, M]$. The candidates of inactive users will be $\pi_{N^*}^k$, or ... or $\pi_{N^*}^{C_{M}^{N^*}}$, $\forall k^* \in [1, C_{M}^{N^*}]$. The flops for each of $C_{M}^{N^*} - k^* + 1$ candidate cases in the sum-rate improvement strategy can be given by

$$\mathcal{F}_2^{N^*} = 10.5(M - N^*)^2 + 36.5(M - N^*) - 23. \hspace{1cm} (11)$$

Then, based on the obtained $\mathcal{F}_1^N$ and $\mathcal{F}_2^{N^*}$, the computational-complexity evaluation of the compared algorithms in terms of flops are summarised as follows.

1) Exhaustive search (ES):
In the exhaustive search, all the combinations of user scheduling should be enumerated. Thus, $\sum_{N=0}^{M-1} C_{M}^{N} \mathcal{F}_1^N$ combinations of users cases need to be calculated to obtain the candidates of inactive users, which requires the flops of $\sum_{N=0}^{M-1} C_{M}^{N} \mathcal{F}_1^N$. 


TABLE 2 Flops versus $N$ assuming $M=9$ and $k^* = 1$

| $N$ | ES | MRUSA | CCRUSA |
|-----|----|-------|--------|
| 1   | 15,4798 | 9004 | 1767 |
| 3   | 197,481 | 112,803 | 2846 |
| 5   | 188,195 | 173,051 | 3405 |
| 7   | 156,433 | 156,217 | 3612 |
| 9   | 154,018 | 154,017 | 3635 |

TABLE 3 Flops versus $M$ assuming $N^* = 3$ and $k^* = 1$

| $M$ | ES | MRUSA | CCRUSA |
|-----|----|-------|--------|
| 3   | 441 | 441 | 230 |
| 5   | 4201 | 4081 | 766 |
| 7   | 33,624 | 27,191 | 1638 |
| 9   | 197,481 | 112,803 | 2846 |
| 11  | 1016768 | 344173 | 4390 |

In the sum-rate improvement strategy, additional flops are needed for $C^N_M - k^* + 1$ candidates, which are conducted to obtain the final user scheduling solution.

Therefore, the total flops of the exhaustive search are counted as $\sum_{N=0}^{M-1} C^N_M \sum_{k=0}^{N^*} (C^N_M - k^* + 1)F^2_N$. 

2) Multi-round user scheduling algorithm (MRUSA):

Contrasting with the exhaustive search is the multi-round mechanism to achieve the optimal solution in the proposed MRUSA.

Specifically, according to Algorithm 1, determining the candidates of inactive users needs $\sum_{N=0}^{N^*} C^N_M F^1_N$ flops. The flops count in the sum-rate improvement strategy is equal to that in the exhaustive search.

Therefore, the total flops of the MRUSA are counted as $\sum_{N=0}^{N^*} C^N_M F^1_N + (C^N_M - k^* + 1)F^2_N$.

3) Computational-complexity reduced user scheduling algorithm (CCRUSA):

It can be seen from Algorithm 2 that to obtain the user scheduling solution the proposed algorithm only needs to calculate $P_{\text{max}}^{\text{CCRUSA}}$, $\forall N \in [0, N^*]$, which requires the flops counted as $\sum_{N=0}^{N^*} F^1_N$.

To illustrate the comparison of the computational complexity against $N$ among the algorithms, the numerical results of the flops are listed in Table 2.

It can be seen that the proposed MRUSA can obtain less complexity than the ES when the optimal number of inactive users is not large enough, and the CCRUSA has the lowest complexity.

Furthermore, the numerical results of the flops versus $M$ are listed in Table 3. It shows that the proposed two user scheduling algorithms achieve the significant reduction of the computational complexity than the ES, and when there are more considered users, the computational savings are more significant.

Algorithm 3 The hybrid user scheduling algorithm

| Input: $h_i, v_i \in M; R_i, v_i \in \mathbb{P}$; |
| Output: $(\pi_N^*)$ |
| 1 if $R_1 \geq \cdots \geq R_M$ then |
| 2 \hspace{0.5cm} Do Algorithm 2; |
| 3 else |
| 4 \hspace{0.5cm} Do Algorithm 1; |
| 5 end |
| 6 return $(\pi_N^*)$; |

These observations will be further verified by numerical simulations in Section 4.

Since the same user scheduling solution will be achieved by MRUSA and CCRUSA if both algorithms can work and CCRUSA outperforms MRUSA in terms of computational complexity, CCRUSA has a higher priority than MRUSA to be operated in practice. We further propose a hybrid user scheduling algorithm in Algorithm 3 to describe the tradeoff between the two proposed algorithms.

In the implementation of the proposed schemes, the CSI overhead needs to be discussed. At the beginning of user scheduling, the BS needs to collect the CSI for all users on their sub-channels in the single NOMA cluster. Hence, if more users desire to access the single NOMA cluster, there is more CSI overhead suffered by the system. However, due to additional calculation caused by SIC at each user receiver, the limit of the number of accessible users needs to be considered in each single NOMA cluster[10]. Therefore, the CSI overhead can be controlled by the BS in practice.

4 | SIMULATION AND ANALYSIS

In this section, we evaluate the performance of the proposed user scheduling algorithms in single NOMA cluster network by numerical simulations, where the BS is located in the cell center and the users are randomly distributed in a circular range with a radius of 300 m. The large-scale path loss is $L(d) = 37 + 30\log(d)$, where the unit of $d$ is meter. The noise power is assumed to be $N = -104$ dBm. The individual minimum rate requirement follows an i.i.d. uniform distribution as $R_m \sim U[1, 7]$, for $m = 1, \ldots, M$.

Figure 1 depicts the cumulative distribution function (CDF) of the number of active users using the proposed user scheduling algorithm labelled as Proposed-USA and the existing enhancing sum-rate user scheduling algorithm in [9] labelled as Existing-USA, both of which provide us useful insights on the user scheduling in single NOMA cluster. For the sake of brevity, 10 users are considered in this simulation, which is repeated 10,000 times for the result to get stable statistical performance. It is clearly observed that the Proposed-USA performs better than Existing-USA with more active user numbers, for example, for the $P_{\text{max}} = 30$ dBm case, the numbers of active users in the Existing-USA and Proposed-USA are concentrated on 3.3 and 5.7, respectively. In addition, the number of active users...
the same $N^*$ the performance gap between these two algorithms increases with the number of inactive users $N^*$ candidates sets. The performance improvement in the proposed algorithm results from the operation of Stage 2 in Section 3.1 for further improving the sum-rate of NOMA system. Moreover, at each changing point of the number of inactive users $N^*$, the sum-rate gets higher with $N^*$, which is due to the fact that the proposed scheme is to unload some users at that point so that the remaining active users will be allocated more resources to obtain higher transmission rates.

Figure 3 depicts the achieved individual rate versus the user minimum rate requirement in a four-user NOMA cluster. The sum-power is allocated to users to meet their initial minimum rate requirements, which are set as, for example, $\hat{R}_1 = \hat{R}_2 = \hat{R}_3 = \hat{R}_4 = 3 \text{ b/s/Hz}$ in the simulations. The distances of the four users to the BS are set to 100, 80, 60 and 40 m, respectively. As illustrated in Figure 3(a)–(c), when the variable minimum rate requirement denoted by $\hat{R}$ is less than 3 b/s/Hz, the achieved rate of the user 4 decreases with $\hat{R}$, since more powers are allocated to other users. Moreover, as illustrated in each figure of individual user's rate achievement in Figure 3, when $\hat{R}$ exceeds 3 b/s/Hz, the BS cannot support all the four users, and then user 1, which has the lowest channel gain, becomes inactive first to achieve the improved sum-rate, that is, the rising performance of user 4. Further when $\hat{R}$ for a user is large enough, there is insufficient sum-power to be allocated to that user (user 1, user 2, user 3 and user 4 in (a), (b), (c), and (d), respectively) and hence the proposed user scheduling scheme will offload the corresponding user and re-activate user 1 to sustain the maximum number of active users. It is worth noting that the proposed user scheduling scheme can offload any users to achieve the maximum number of active users, revealing the well-done mechanism of fairness.

Figure 4 depicts the number of active users and sum-rate versus $P_{\text{max}}$ when the number of accessible users in a single NOMA cluster is limited by $Q$. It can be seen from Figure 4(a) that by adding the mechanism for the hardware constraints on SIC proposed in Remark 2, the number of active users can be well restricted no matter how large $P_{\text{max}}$ is. Thus, the error propagation in NOMA can be controlled in the proposed user scheduling algorithm. From Figure 4(b), since more power will be allocated to the user with higher channel gain when $Q$ is small, the sum-rate with lower $Q$ is larger than that with higher one after the restricted mechanism. However, the gap becomes smaller due to more power allocated to the user with higher channel gain will also rise the interference to other users in the same NOMA cluster.

Finally, the computational complexities of the ES, and the proposed MRUSA and CCRUSA are compared in Figure 5. It can be seen that, at any given number of users $M$, CCRUSA requires the lowest computational flops among three algorithms and the gaps of the flops between CCRUSA and others increase significantly with the number of users in the single NOMA cluster. It is also worth noting that MRUSA achieves the significant reduction of the computational complexity than ES when $M$ is large, implying that MRUSA is more applicable than ES in use.
in the multi-user scene. Additionally, the difference of the flops between ES and MRUSA decreases when the optimal number of inactive users increases, which results from the multi-round mechanism in MRUSA. Moreover, since CCRUSA can obtain the final user scheduling solution without performing the sum-rate comparison, the same computational complexity will be obtained with equal $N^\ast$.

5 | CONCLUSIONS
Here, we have investigated the user scheduling algorithm for maximising the number of active users while ensuring their individual minimum rate requirements in a single NOMA cluster. A MRUSA is proposed firstly for common practical environment. To achieve better performance, a CCRUSA is further
introduced under some practical conditions. Simulation results show that the proposed user scheduling algorithms can achieve the maximum number of active users and significantly improve the performance in terms of the network sum-rate when the maximum number of active users is achieved. And, the proposed user scheduling algorithm has the properties of fairness and adapting the hardware constraints on SIC. Moreover, the required number of computational flops of the proposed computational-complexity reduced user scheduling algorithm is much lower than that of the proposed optimal user scheduling algorithm.

**ACKNOWLEDGEMENTS**

This work was supported by the National Natural Science Foundation of China under Grant 61701284, Natural Science Foundation of Jiangsu Province under Grant BK20181032, and the Innovative Research Foundation of Qingdao under Grant No.19-6-2-1-cg.

**REFERENCES**

1. Ding, Z., et al.: Application of non-orthogonal multiple access in LTE and 5G networks. IEEE Commun. Mag. 55(2), 185–191 (2017)
2. Dai, L., et al.: Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends. IEEE Commun. Mag. 53(9), 74–81 (2015)
3. Yang, K., et al.: Non-orthogonal multiple access: Achieving sustainable future radio access. IEEE Commun. Mag. 57(2), 116–121 (2019)
4. Ding, Z., et al.: Cooperative non-orthogonal multiple access in 5G systems. IEEE Commun. Lett. 19(8), 1462–1465 (2015)
5. Wang, C., et al.: Power allocation for a downlink non-orthogonal multiple access system. IEEE Wireless Commun. Lett. 5(5), 532–535 (2016)
6. Zhu, J., et al.: On optimal power allocation for downlink non-orthogonal multiple access systems. IEEE J. Sel. Areas Commun. 35(12), 2744–2757 (2017)
7. Zhang, Y., et al.: Secrecy sum rate maximization in non-orthogonal multiple access. IEEE Commun. Lett. 20(5), 930–933 (2016)
8. Yang, Z., et al.: On the optimality of power allocation for NOMA downlinks with individual QoS constraints. IEEE Commun. Lett. 21(7), 1649–1652 (2017)
9. Ali, M., et al.: Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (NOMA) systems. IEEE Access 4, 6325–6343 (2016)
10. Pei, L., et al.: Energy-efficient D2D communications underlying NOMA-based networks with energy harvesting. IEEE Commun. Lett. 22(5), 914–917 (2018)
11. Yang, Z., et al.: Energy-efficient non-orthogonal multiple access for machine-to-machine communications. IEEE Commun. Lett. 21(4), 817–820 (2017)
12. Liu, Q., et al.: Energy-efficient transmission design in cooperative relaying systems using NOMA. IEEE Commun. Lett. 22(3), 594–597 (2018)
13. Wu, Y., Qian, L.: Energy-efficient NOMA-enabled traffic offloading via dual-connectivity in small-cell networks. IEEE Commun. Lett. 21(7), 1605–1608 (2017)
14. Wang, J., et al.: Energy-efficient joint power and bandwidth allocation for NOMA systems. IEEE Commun. Lett. 22(4), 780–783 (2018)
15. Zhang, Y., et al.: Energy-efficient transmission design in non-orthogonal multiple access. IEEE Trans. Veh. Technol. 66(3), 2852–2857 (2017)
16. Fang, F., et al.: Energy-efficient resource allocation for downlink non-orthogonal multiple access network. IEEE Trans. Commun. 64(9), 3722–3732 (2016)
17. Tsai, Y., Wei, H.: Quality-balanced user clustering schemes for non-orthogonal multiple access systems. IEEE Commun. Lett. 22(1), 113–116 (2018)
18. Zhu, L., et al.: Optimal user pairing for downlink non-orthogonal multiple access (NOMA). IEEE Wireless Commun. Lett. 8(2), 328–331 (2019)
19. Fang, F., et al.: Joint user scheduling and power allocation optimization for energy-efficient NOMA systems with imperfect CSI. IEEE J. Sel. Areas Commun. 35(12), 2874–2885 (2017)
20. Liang, W., et al.: User pairing for downlink nonorthogonal multiple access networks using matching algorithm. IEEE Trans. Commun. 65(12), 5319–5332 (2017)
21. http://ai.stanford.edu/~paskin/slam/javadoc/javaslam/util/Flops.html
A low complexity user scheduling algorithm aimed for the maximum number of active users in NOMA system.

How to cite this article: Liu C, Chen G, Xing S. A low complexity user scheduling algorithm aimed for the maximum number of active users in NOMA system. *IET Commun.*, 2021;15:265–278. https://doi.org/10.1049/cmu2.12054

**APPENDIX A**

### A.1 Appendix 1

It is not difficult to see that the problem (5) is non-convex due to the nature of the mixed integer non-linear programming problem. However, by using the enumeration method, the considered problem can be converted into a series of cases. Note that, since the objective function is a fixed maximum number of active users, the obtained feasible points are the optimal solutions for each case.

1. First, we consider a special case of the problem (5) by assuming that $P_{\text{max}}$ is large enough such that all the users are active. As a result, the indicators for all users are equal to 1, that is, $\hat{P}_{\text{max}} = 1, \forall m$. Under this assumption, problem (5) can be transformed to the following form

$$
\begin{align}
\max_{\hat{P}_{\text{max}} \geq 0} & \quad \hat{P}_{\text{max}} \\
\text{s.t.} & \quad \sum_{i=1}^{M} P_{Ji} \leq \hat{P}_{\text{max}}, \\
& \quad \log_{2} \left( 1 + \frac{\left| b_{Ji} \right|^{2} P_{Ji}}{\sum_{l=i+1}^{M} P_{Jl} + \sigma^{2}} \right) \geq \hat{R}_{Ji}, \forall i, \\
\end{align}
(A.1)
$$

where $J_i$ denotes the index of the $i$th active user in $\mathbb{M}$.

Obviously, this case has a fixed maximum number of active users, such that if the two constraints are feasible, the optimal user scheduling solution for this case can be obtained immediately.

Thus, the feasibility condition of this case needs to be found by providing the upper and lower bounds of $P_{\text{max}}$. First, the upper bound of $P_{\text{max}}$ for the problem (12) is obviously $\infty$ due to the given assumption. Then, the lower bound of $P_{\text{max}}$ should merely keep all users active, which can be described by

$$
\begin{align}
\min_{\hat{P}_{\text{max}} \geq 0} & \quad \sum_{i=1}^{M} \hat{P}_{Ji} \\
\text{s.t.} & \quad \log_{2} \left( 1 + \frac{\left| b_{Ji} \right|^{2} \hat{P}_{Ji}}{\sum_{l=i+1}^{M} \hat{P}_{Jl} + \sigma^{2}} \right) \geq \hat{R}_{Ji}, \forall i, \\
\end{align}
(A.2)
$$

Fortunately, the given problem in (13) has been solved in [8], and the lower bound of $P_{\text{max}}$ for the problem (12) can be expressed as

$$
\hat{P}_{\text{max}} = \sum_{i=1}^{M} \frac{\left( 2^{\hat{R}_{Ji}} - 1 \right) \sigma^{2}}{\left| b_{Ji} \right|^{2} \left( \sum_{l=i+1}^{M} \hat{P}_{Jl} + \sigma^{2} \right)^{\frac{1}{2}}},
$$

where $\sigma_{0} = \emptyset$, and the second equation holds due to $C_{M} \cap \mathbb{M} = (J)_{i=1}^{M}$.

It is worth noticing that if $P_{\text{max}} < \hat{P}_{\text{max}}$, there is no feasible point of the case (12). However, the feasible condition $P_{\text{max}} < \infty$ allows case (12) to have only one feasible point of user scheduling, in which the optimal number of inactive users is zero, that is, $N_{e} = 0$, and the optimal candidates of inactive users are obviously $\emptyset$.

2. Next, we need to find the feasibility conditions for $M - 1$ active users. Under this assumption, problem (5) can be transformed to the following form

$$
\begin{align}
\max_{\hat{P}_{\text{max}} \geq 0} & \quad \hat{P}_{\text{max}} \cdot M - 1 \\
\text{s.t.} & \quad \sum_{i=1}^{M-1} \hat{P}_{Ji} \leq \hat{P}_{\text{max}}, \\
& \quad \log_{2} \left( 1 + \frac{\left| b_{Ji} \right|^{2} \hat{P}_{Ji}}{\sum_{l=i+1}^{M-1} \hat{P}_{Jl} + \sigma^{2}} \right) \geq \hat{R}_{Ji}, \forall i, \\
\end{align}
(A.4)
$$

where $i \in \{1, 2, \ldots, M - 1\}$.

From (15), different active users $(J)_{i=1}^{M-1}$ will lead to different feasibility conditions. Thus, $C_{i}^{M}$ inactive users combinations, denoted by $\pi_{1}^{i}, \pi_{2}^{i}, \ldots$, and $\pi_{C_{i}^{M}}^{i}$, will decompose the problem into $C_{i}^{M}$ cases.

For each case, for example, $(J)_{i=1}^{M-1} = M - \pi_{1}^{i}$, the lower bound described by the similar structure as the problem (13) is given by

$$
\begin{align}
\min_{\hat{P}_{\text{max}} \geq 0} & \quad \sum_{i=1}^{M-1} \hat{P}_{Ji} \\
\text{s.t.} & \quad \log_{2} \left( 1 + \frac{\left| b_{Ji} \right|^{2} \hat{P}_{Ji}}{\sum_{l=i+1}^{M-1} \hat{P}_{Jl} + \sigma^{2}} \right) \geq \hat{R}_{Ji}, \forall i, \\
\end{align}
(A.5)
$$

where $i \in \{1, 2, \ldots, M - 1\}$.
Since $\mathcal{C}_M \cap \mathcal{M} = (J_{i+1}^{M-1})$, the lower bound of $P_{\text{max}}$ for this case can be obtained similarly as (14), given by

$$P_{\text{max}}^{k_1} = \sum_{i=1}^{M-1} \left( \frac{2^{k_1} - 1}{|b_{f_i}|^2} \left( \prod_{j=1}^{i-1} 2^{k_j} \right) \right)$$

$$= \sum_{a=1}^M \left( \frac{2^{k_a} - 1}{|b_{f_i}|^2} \left( \prod_{j=1}^{a-1} 2^{k_j} \right) \right),$$

where $k \in \{1, 2, \ldots, C_M \}$. By using the same method, the $C_M$ lower bounds of $P_{\text{max}}$ for the problem (15) can be obtained and labeled as $P_{\text{max}}^{k_1}, P_{\text{max}}^{k_2}, \ldots, P_{\text{max}}^{k_M}$. Without loss of generality, we assume that $P_{\text{max}}^{k_1} \geq P_{\text{max}}^{k_2} \geq \cdots \geq P_{\text{max}}^{k_M}$.

It is worth noticing that if $P_{\text{max}} < P_{\text{max}}^{k_1}$, there is no feasible point for $(J_{i+1}^{M-1}) = \mathcal{M} \setminus \pi^k$ case. Otherwise, $\pi^k$ should be regarded as one of the candidates of inactive users.

Thus, if $P_{\text{max}}^{k_1} \leq P_{\text{max}}^{k}$, the candidates of inactive users should include $\pi^k, \pi^{k+1}, \ldots, \pi^{C_M}$ since $P_{\text{max}} \geq P_{\text{max}}^{k} \geq P_{\text{max}}^{k+1} \geq \cdots \geq P_{\text{max}}^{C_M}$ holds, and if $P_{\text{max}} < P_{\text{max}}^{k+1}$, the candidates of inactive users cannot include $\pi^{k-1}, \pi^{k-2}, \ldots, \pi^1$ since $P_{\text{max}} < P_{\text{max}}^{k-1} \leq P_{\text{max}}^{k-2} \leq \cdots \leq P_{\text{max}}^{k}$ holds.

According to the above analysis, the feasible condition of $P_{\text{max}}^{k_1} = P_{\text{max}} < P_{\text{max}}^{k+1}$, $\forall k \in \{1, \ldots, C_M \}$, allows case (15) to have $C_M - k + 1$ feasible points of inactive users, that is, $N^* = 1$, the optimal candidates of inactive users are $\pi^1, \pi^{k+1}, \ldots, \pi^{C_M}$, and $P_{\text{max}}^{k_1} \Delta = P_{\text{max}}^{k}$ is the upper bound of $P_{\text{max}}$ for problem (15).

Further, to obtain the general form of the feasible condition, we consider $N$ inactive users case, where $N \in \{1, 2, \ldots, M\}$, and thus the problem (5) can be transformed as

$$\max_{p_0} \quad M - N$$

subject to

$$\sum_{i=1}^{M-N} p_{f_i} \leq P_{\text{max}}$$

$$\log_2 \left( 1 + \frac{|b_{f_i}|^2 p_{f_i}}{\sum_{i=1}^{M-N} |b_{f_i}|^2 + \sigma^2} \right) \geq R_{f_i}, \forall i,$$

where $i \in \{1, 2, \ldots, M-N\}$.

The feasible condition can be obtained similarly as problem (15), and is thus omitted.

It can be concluded that the feasible condition $P_{\text{max}}^{k_1} \leq P_{\text{max}} < P_{\text{max}}^{k+1}$, $\forall k \in \{1, \ldots, C_N \}$, allows case (18) to have $C_N - k + 1$ feasible points of inactive users, in which the optimal number of inactive users is $N$, that is, $N^* = N$, the optimal candidates of inactive users are $\pi^k, \pi^{k+1}, \ldots, \pi^{C_N}$, and $P_{\text{max}}^{k_1} \Delta = P_{\text{max}}^{k}$ is the upper bound of $P_{\text{max}}$ for problem (18).

Now, the whole range of $P_{\text{max}}$ has been decomposed into $2^M$ feasible conditions, and for each condition the feasible points, that is, the optimal candidates of inactive users, of problem (5) are listed in Table 1.

The proof of the lemma is now complete.

### Appendix 2

From Lemma 1, the feasible condition $P_{\text{max}}^{k_1} \leq P_{\text{max}} < P_{\text{max}}^{k+1}$ will lead to multiple feasible points of the original problem, for example, $\pi^k, \pi^{k+1}, \ldots, \pi^{C_N}$.

According to formula (16) in [8], the maximum sum-rate has been obtained when the active users are given. Without loss of generality, we assume that in $\mathcal{M}$ the $n$ users with the highest channel gains are inactive, the user $(M - n)$ is active, and other active users are randomly placed.

Thus, for $\forall n \in \{0, 1, \ldots, N\}$ and $\forall i \in \{k, 2, \ldots, C_M\}$, the optimal sum-rate can be given as

$$R_{\text{opt}}^{\pi^k} = \sum_{i=1}^{M-N-1} R_{f_i} + \log_2 \left[ 1 + \frac{P_{\text{max}} |b_{M-n}|^2}{\sigma^2 \prod_{i=1}^{M-N-1} |b_{f_i}|^2} \right] \Delta \sum_{i=1}^{M-N-1} \xi_{f_i},$$

where $\xi_{f_i} = \frac{|b_{M-n}|^2 (2^{k_i} - 1)}{|b_{f_i}|^2 \prod_{i=1}^{M-N-1} |b_{f_i}|^2}$.

Since $J_{M-N} = M - n$, $(J_{i+1})^{M-N-1} = (J_{i+1})^{M-N-1} \cap (J_{i+1})^{M-N-1}$, thus, $R_{\text{opt}}^{\pi^k}$ can be rewritten as

$$R_{\text{opt}}^{\pi^k} = \sum_{i=1}^{M-N-1} R_{f_i} + \log_2 \left[ 1 + \frac{P_{\text{max}} |b_{M-n}|^2}{\sigma^2 \prod_{i=1}^{M-N-1} |b_{f_i}|^2} \right] \Delta \sum_{i=1}^{M-N-1} \xi_{f_i}.$$
To pursue the better performance in terms of transmission rate, the feasible point having the maximum sum-rate should be chosen as the finally optimal inactive users being scheduled formulated by (7).

The proof of the lemma is now complete.

### A.3 Appendix 3

To demonstrate Theorem 1, we first prove \( P_{\max}^{(m)} = \min \{ P_{\max}^{(m)} \} \).

Without loss of generality the \( N \) inactive users \( \pi_N \) can be denoted by \( (L_{m})_{m=1}^{N} \) which follows a general relationship as

\[
1 \leq L_{1} < \cdots < L_{K} \leq N < L_{K+1} < \cdots < L_{N} \leq M. \tag{A.11}
\]

1. For \( K = N, \pi_N = (m)_{m=1}^{N} \), leads to \( P_{\max}^{(m)} = P_{\max}^{m} \);

2. Next, we show that \( P_{\max}^{(m)} < P_{\max}^{m}, \forall K \in [0, N - 1] \), as follows. Start with the defined variable as

\[
\Delta P = P_{\max}^{m} - P_{\max}^{(m)} = \Delta P_{(1)} \tag{A.12}
\]

Then, by substituting (6) into (A.12) and manipulating it, it yields \( \Delta P \) (see (24)), where \( \chi_{i} = \frac{(\Delta P_{(1)})}{|b_{L_{(K+1)}}^{2}\phi_{i}^{2}} - \sum_{j=1}^{N} \frac{X_{i}}{\prod_{j=1}^{N} 2^{R_{j}}} \).

It can be observed that the denominator items \(|b_{i}|^{2}\) in (24) are sorted in ascending order of the index \( i \). Since \( \bar{R}_{i} \geq \bar{R}_{i+1} \geq 0, \forall i \in \{1, \ldots, M - 1\} \), the first three items in (24) are non-negative. Furthermore, since \( \phi_{i} = \pi_{N} \cap (m)_{m=1}^{N} \), thus \( \phi_{i} \subset (m)_{m=1}^{N} \), \( \forall i \in [N + 1, L_{K+1} - 1] \), and we can obtain that

\[
\frac{X_{i}}{\prod_{j=1}^{N} 2^{R_{j}}} \geq 0 \text{ which means any component of the fourth item is also non-negative. Additionally, } 1/|b_{i}|^{2} \geq 1/|b_{i+1}|^{2}, \forall i \in [1, M - 1] \text{ since } 0 < |b_{i}|^{2} \leq |b_{i+1}|^{2} \text{. Therefore, } \Delta P_{(1)} \text{, a reduced version of } \Delta P \text{ (see (25)) is constructed, with } \Delta P \leq \Delta P_{(1)} \text{.}
\]

In the same way, the first two items in \( \Delta P_{(1)} \) are checked to be non-negative, which leads to \( \Delta P_{(2)} \), a reduced version of \( \Delta P_{(1)} \), that is, \( \Delta P_{(1)} \geq \Delta P_{(2)} \), in which the expression of \( \Delta P_{(2)} \) is omitted due to space limitation.
After $L_N - L_{(K+1)} + 1$ same operations, we can obtain the following expression as

$$\Delta P_{(L_N - L_{(K+1)} + 1)} \geq \Delta P_{(L_N - L_{(K+1)} + 1)} \geq 0,$$  \hspace{1cm} (A.15)

where

$$\Delta P_{(L_N - L_{(K+1)} + 1)} = \frac{\sigma^2}{|b_i|^2 \prod_{j \neq i}^{L_N} \frac{x_j}{\prod_{j \neq i}^{N} x_j^{2^j}}} + \sum_{i=1}^{M} \left( \sum_{j=1}^{L_N} \frac{1}{\prod_{j \neq i}^{N} x_j^{2^j}} \prod_{j \neq i}^{N} \frac{x_j}{\prod_{j \neq i}^{N} x_j^{2^j}} \right) \text{ and (a)}$$

holds because

$$\prod_{j \neq i}^{N} x_j^{2^j} > 0, \forall i.$$

Therefore, $P_{\text{max}}^{\text{(a)}} < P_{\text{max}}^{\text{(a)}}, \forall K \in [0, N - 1]$, is proved.

We are now in position to prove $R_{\text{a}}^{\text{(a)}} = \max \{ R_{\text{a}}^{K_N} \}$. Without loss of generality, we assume that in $\mathbb{M}$ the $n$ users with the highest channel gains are inactive, the user $(M - n)$ is active, and other active users are randomly placed.

1) For $K = N$, $\pi_N = (m)_{n=1}^{N}$ leads to $R_{\text{a}}^{\text{(a)}} = R_{\text{a}}^{K_N}$, $\forall \eta$.

2) $R_{\text{a}}^{\text{(a)}} \geq R_{\text{a}}^{K_N}$, $\forall K \in [0, N - 1]$, $\forall \eta \in [0, N]$, which is approved as below.

Let us define a variable as

$$\Xi = 2^{R_{\text{a}}^{K_N} - R_{\text{a}}^{K_{(a)}}}. \hspace{1cm} (A.16)$$

By substituting (8) into (27) and manipulating it, it yields $\Xi$ (see (28)).

Due to the fact that $P_{\text{max}}^{\text{(a)}} \geq P_{\text{max}}^{\text{(a)}}$, we can obtain that $\Xi \geq \Xi^{(1)}$, where $\Xi^{(1)}$ is shown in (29), $L_{(a)} = 0$ and $\epsilon_{i} = \pi_{N - n}^{N - n} \cap (m)_{n=1}^{N}$.

It can be seen that the first three items in $\Xi^{(1)}$ are non-negative, resulting in $\Xi^{(2)}$, a reduced version of of $\Xi^{(1)}$, that is, $\Xi^{(1)} \geq \Xi^{(2)}$, where $\Xi^{(2)}$ is shown in (30).
\[
\Xi^{(2)} = \frac{|b_{M-n}|^2}{|b_{L_{(K+1)}}|} \left( \prod_{i=1}^{L_{(K+1)}-1} 2^{\frac{h_i}{2}} - \prod_{i=1}^{L_{(K+1)}} 2^{\frac{h_i}{2}} \right) \\
+ \sum_{i=K+1}^{N-n} \left[ \prod_{j=1}^{L_{(i+1)}/(K+1)} \left( \frac{|b_{M-n}|^2 X_i}{\sigma^2 \prod_{j \in \mathcal{E}_j} 2^{R_j}} - \frac{|b_{M-n}|^2 X_M}{\sigma^2 \prod_{j=1}^{N} 2^{R_j}} \right) - \frac{|b_{M-n}|^2 X_{L_{(i+1)}}}{\sigma^2 \prod_{j=1}^{N} 2^{R_j}} \right] \\
- \sum_{i=M-n+1}^{M-1} \left( \frac{|b_{M-n}|^2 X_i - \frac{|b_{M-n}|^2 X_M}{\sigma^2 \prod_{j=1}^{N} 2^{R_j}} + \frac{\prod_{i=1}^{M-n-1} 2^{h_i} - \prod_{i=1}^{M-n} 2^{h_i}}{\prod_{i=1}^{N} 2^{R_i}} \right) \\
\] (A.19)

Since \( \prod_{j=1}^{N} 2^{R_j} \geq 2^{h_{L_{(K+1)}}} \prod_{j \in \mathcal{E}_j} 2^{R_j}, \forall K \in [0, N - 1] \), and
\[
\prod_{i=1}^{L_{(K+1)}} 2^{h_i} \left( \frac{\prod_{j=1}^{N} 2^{R_j}}{2^{R_{L_{(K+1)}}} \prod_{j \in \mathcal{E}_j} 2^{R_j}} - 1 \right), \text{ the first item in } \Xi^{(2)}
\]
is also non-negative. Thus, \( \Xi^{(3)} \), a reduced version of \( \Xi^{(2)} \), is constructed as \( \Xi^{(2)} \geq \Xi^{(3)} \). The expression of \( \Xi^{(3)} \) is omitted due to space limitation.

Consequently, after \( N - n - K \) same operations, we can obtain the following expression as
\[
\Xi^{(N-n-K+1)} \geq \Xi^{(N-n-K+1)}, \quad \text{(A.20)}
\]
where
\[
\Xi^{(N-n-K+1)} = \frac{\prod_{i=1}^{M-n-1} 2^{h_i} - \prod_{i=1}^{M-n} 2^{h_i}}{\prod_{i=1}^{N} 2^{h_i}} - \sum_{i=M-n+1}^{M-1} \left( \frac{|b_{M-n}|^2 X_i - \frac{|b_{M-n}|^2 X_M}{\sigma^2 \prod_{j=1}^{N} 2^{R_j}} + \frac{\prod_{i=1}^{M-n-1} 2^{h_i} - \prod_{i=1}^{M-n} 2^{h_i}}{\prod_{i=1}^{N} 2^{R_i}} \right).
\]

Since \(-1/|h_i|^2 \leq -1/|b_{i+1}|^2, \forall i \in [M-n, M-2] \),
\[
\Xi^{(N-n-K+1)} \geq \Xi^{(N-n-K+2)} \geq 0, \quad \text{(A.21)}
\]
where \( \Xi^{(N-n-K+2)} = \frac{|b_{M-n+1}|^2 - \frac{|b_{M-n}|^2 X_M}{\sigma^2 \prod_{i=1}^{N} 2^{h_i}} \text{ and (b) holds because } |b_{M-n+1}|^2 \geq |b_{M-n}|^2, \forall n. \)

Therefore, \( R_n^{(n,n)} \geq R_n^{(n,N)}, \forall K \in [0, N - 1], \forall n \in [0, N] \) is proved.

Hence, the candidates of inactive users always include the user set \( \{m\}_{n=1}^{N}, \forall N \in [1, M] \), due to \( R_{\text{max}}^{(n,N)} = \min \{ R_n^{(n,N)} \} \) and it will be selected as the final inactive users due to \( R_n^{(n)} = \max \{ R_n^{(n,N)} \} \).

The proof of the theorem is now complete.