IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 25, 2018, accepted June 24, 2018, date of publication July 4, 2018, date of current version July 25, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2851609

Spectrum and Energy Efficient Resource Allocation With QoS Requirements for Hybrid MC-NOMA 5G Systems

ZHENGYU SONG[1], QIANG NI[2], (Senior Member, IEEE), AND XIN SUN[1]
[1]School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China
[2]School of Computing and Communications, Lancaster University, Lancaster LA1 4WA, U.K.

Corresponding author: Qiang Ni (q.ni@lancaster.ac.uk)

This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant 2016RC055, in part by the Royal Society Project under Grant IEC\170324, in part by the EPSRC IAA Project under Grant CSA7114, in part by the EPSRC Project under Grant EP/K011693/1, and in part by the EU FP7 CROWN Project under Grant PIRSES-GA-2013-610524.

ABSTRACT In this paper, we investigate the resource allocation problem for achieving spectral efficiency (SE) and energy efficiency (EE) tradeoff with users' minimum rate requirements in hybrid multi-carrier non-orthogonal multiple access (MC-NOMA) systems which incorporate both NOMA and orthogonal multiple access (OMA) modes into one unified framework. All the degrees of freedom involved in resource allocation, including the choice of multiple access (MA) modes, user clustering, subcarrier assignment, and power allocation, are jointly considered. We first formulate the SE-EE tradeoff as a multi-objective optimization (MOO) problem with minimum rate requirement constraints. Then, considering the non-convexity of the MOO problem, it is converted into a single-objective optimization (SOO) problem by utilizing weighted Tchebycheff method. Lagrangian dual decomposition and sequential convex programming are applied to solve the SOO problem. We propose a joint resource allocation algorithm which is applicable to the general case, where an arbitrary number of users can be multiplexed on the same subcarrier. Simulation results demonstrate that users' minimum rate requirements and channel conditions have great impact on the selection of MA modes. The proposed hybrid MC-NOMA mode significantly outperforms MC-NOMA and OMA in terms of SE-EE tradeoff, and the performance gain brought by four or more users sharing the same subcarrier is minimal. Meanwhile, the hybrid MC-NOMA also shows great potential to improve the tradeoff between fairness and system efficiency.

INDEX TERMS MC-NOMA, energy efficiency, spectral efficiency, minimum rate requirement, user clustering, subcarrier assignment, power allocation.

I. INTRODUCTION

Non-orthogonal multiple access (NOMA) has been recognized as one of the promising candidate multiple access (MA) schemes for performance enhancement in the fifth-generation (5G) cellular communications [1]–[4]. Compared to conventional orthogonal multiple access (OMA) such as orthogonal frequency-division multiple access (OFDMA), the major advantages of NOMA include improved spectrum efficiency (SE), massive connectivity and lower transmission latency and signaling cost [5]. In general, NOMA can be divided into two categories, namely code-domain NOMA and power-domain NOMA. In this paper, we focus on the power-domain NOMA. In power-domain NOMA, two or more users are multiplexed in the power domain. Specifically, the base station (BS) transmits a superposition coded signal at the same time, code and frequency, but with different power levels. At the receiver sides, the composite signal of different users is separated by multiuser detection (MUD) algorithms such as successive interference cancellation (SIC) [6]. This renders the resource allocation optimization problem in NOMA more complex and different from the OMA systems.

Recently, related work has emerged to investigate the resource allocation problem in NOMA systems to optimize the system SE. In [7], game theory is applied to allocate power among users in single-carrier NOMA (SC-NOMA) systems to maximize the revenue of BS. Lei *et al.* [8] propose a suboptimal power and channel allocation algorithm combining Lagrangian duality and dynamic programming to

maximize the weighted sum rate in a multi-carrier NOMA (MC-NOMA) system. In [9], for a two-user multiple-input-multiple-output (MIMO) NOMA system, the authors propose two power allocation algorithms to maximize the sum capacity under the total power constraint and the minimum rate requirement of weak users. For uplink NOMA, Mollanoori and Ghaderi [10] propose polynomial time algorithms to solve uplink scheduling problems with fixed received power of users. In [11], user clustering and power allocation algorithms for both uplink and downlink NOMA are proposed, but the user clustering and power allocation are carried out separately, which degrades the performance advantages of NOMA. In [12], by introducing matching theory, a joint resource allocation algorithm is proposed to maximize the weighted sum-rate in MC-NOMA systems, where the subcarrier assignment problem is equivalent to a many-to-many two-sided matching game and the power allocation is performed after the subcarrier assignment is settled.

The aforementioned contributions to NOMA in resource allocation aim at optimizing the system SE, while the energy efficiency (EE) of NOMA systems has not been well investigated. Li *et al.* [13] and Lei *et al.* [14] study the power minimization problem with minimum rate requirement of each user in MC-NOMA systems, but the power minimization does not directly lead to EE optimization. In [15], the EE optimization is studied in fading MIMO-NOMA systems, but the number of users is limited to two. In [16], a power allocation strategy is proposed to maximize the EE subject to a minimum required data rate for each user in SC-NOMA systems. Fang *et al.* [17] optimize the subcarrier assignment and power allocation to maximize the EE for the downlink MC-NOMA networks, while the proposed EE maximization resource allocation algorithms are only valid for the case of two users. Moreover, the optimal EE and SE are not achievable simultaneously, i.e., maximizing EE as in [17] will normally lead to the degradation of SE if EE and SE are not jointly considered in the problem formulation [18]. Considering the spectrum scarcity in wireless communications, it is more advantageous to investigate the SE-EE tradeoff problem compared to the EE maximization only resource allocation in order to provide best levels of SE and EE tradeoff according to different performance preferences [19], [20].

On the other hand, in NOMA systems, rate fairness among users or quality-of-service (QoS) is another key issue that should be guaranteed. This is because if no rate fairness or QoS is imposed on the system, for the users multiplexed on the same subcarrier, all the available transmit power will be allocated to the single user with the best channel quality, which leads to absolute unfairness among users, and NOMA will be reduced to OMA [8]. In [9], the minimum rate requirement is applied to allocate non-trivial data rates to the weaker users to ensure certain level of fairness in MIMO NOMA systems. In [3], [8], and [12], the optimization objectives are all weighted sum-rate in order to guarantee the rate fairness among users to some extent, but the minimum rate requirement for each user cannot be guaranteed. Timotheou and Krikidis [21] investigate the impact of power allocation on the rate fairness of SC-NOMA, whereas only the max-min fairness criterion is considered. To ensure the QoS for each user in NOMA, in this paper, the users' minimum rate requirements are set as constraints in the considered resource allocation problem, and such constraints are also capable of providing various levels of rate fairness among users by tuning the values of minimum rate requirements.

Furthermore, according to the principle of NOMA, it is evident that NOMA encourages multiple users to share the same subcarrier simultaneously. However, for the case where the users' channel conditions are similar, the performance advantages of NOMA over OMA can be diminished [2]. Besides, the implementation of NOMA causes higher decoding complexity and error propagation than that of OMA [22]. This motivates us to investigate a new approach called hybrid NOMA and the corresponding resource allocation problem. Specifically, we investigate hybrid MC-NOMA which incorporates NOMA and OMA into one unified framework, and the users are allowed to choose any one of the MA modes (NOMA or OMA) based on the relationship of all users' channel conditions. The adaptive selection of MA modes introduces an extra degree of freedom in resource allocation, which increases the complexity of the optimization but is beneficial to fully exploiting the joint advantages of both NOMA and OMA. Note that the hybrid NOMA model is also mentioned in [3], where the weighted sum throughput is maximized. Nevertheless, in the system model proposed in [3], the maximum number of users multiplexed on the same subcarrier is limited to two, which restricts the advantage of NOMA. More importantly, as the authors mainly focus on the full-duplex NOMA transmissions, the simulation results do not shed light on the performance gain brought by hybrid NOMA over NOMA and OMA, and how the minimum rate requirement and channel condition of each user impact the selection of MA modes remains unknown.

Motivated by the above observations, in this paper, we focus on a novel joint resource allocation problem, combining the selection of MA modes, user clustering, subcarrier assignment and power allocation together to achieve best levels of SE-EE tradeoff according to different preferences in hybrid MC-NOMA systems. The primary contributions of this paper can be summarized as follows:

1) We propose a hybrid MC-NOMA resource allocation model which incorporates NOMA and OMA to fully exploit the joint advantages of the two kinds of MA modes. In the proposed algorithm, all the degrees of freedom in resource allocation, including the selection of MA modes, user clustering, subcarrier assignment and power allocation, are jointly considered. By performance comparisons, it is found that the hybrid MC-NOMA significantly outperforms MC-NOMA and OMA in terms of SE-EE tradeoff, especially when the number of subcarriers is lower. Moreover, our proposed hybrid MC-NOMA resource allocation scheme shows great potential to improve the tradeoff between user fairness and

system efficiency thanks to the wider capacity region of NOMA compared to OMA.

2) Unlike some previous literature such as [3] and [17] where the maximum number of users sharing the same subcarrier is restricted to 2, we propose a resource allocation algorithm applicable to the general case where arbitrary numbers of users can be multiplexed on the same subcarrier. Moreover, we investigate the SE-EE tradeoff performances when different numbers of users are allowed to share the same subcarrier. It is demonstrated for the first time that the performance gain brought by allowing four or more users to share the same subcarrier is very small, Thus, considering the decoding complexity of receivers and the error propagation problem, our research suggests to set the maximum number of users sharing the same subcarrier as 3 in hybrid MC-NOMA systems. *Such result has not yet been seen in previous literature and can provide deep insights on the parameter design for hybrid MC-NOMA systems*.

3) In our resource allocation problem, the minimum rate requirement for each user is imposed to guarantee the rate fairness and QoS, which is different from [3], [8], and [12]. By accurately counting the number of subcarriers applying NOMA and OMA, it is demonstrated for the first time that the minimum rate requirement has significant impact on the selection of MA modes when NOMA and OMA coexist in the system. In general, with the increase of minimum rate requirement for each user, the ratio of subcarriers applying NOMA mode also rises. A more significant observation is that the users with the best and worst channel conditions are more inclined to transmit in NOMA mode. *To our best knowledge, no similar results have been reported in the existing work from the perspective of optimization*. Besides, by adjusting the minimum rate requirement, the proposed resource allocation algorithm for hybrid MC-NOMA systems can achieve various levels of fairness, but higher fairness results in degraded performance of SE.

The rest of this paper is organized as follows. In Section II, we present the system model for hybrid MC-NOMA systems. In Section III, the resource allocation problem for achieving the SE-EE tradeoff is formulated as a multi-objective optimization problem. In Section IV, the optimization problem is solved by the weighted Tchebycheff method, and a joint resource allocation algorithm is proposed. Simulation results are presented in Section V and Section VI finally concludes this paper.

II. SYSTEM MODEL

Consider a single-cell downlink MC-NOMA system as shown in Fig. 1, where one BS communicates with L users simultaneously through K subcarriers by applying hybrid MA transmission protocols of OMA (e.g., OFDMA) and NOMA. All subcarriers are assumed to undergo quasi-static Rayleigh fading, where the channel coefficients are constants for each transmission block but independent between different blocks.

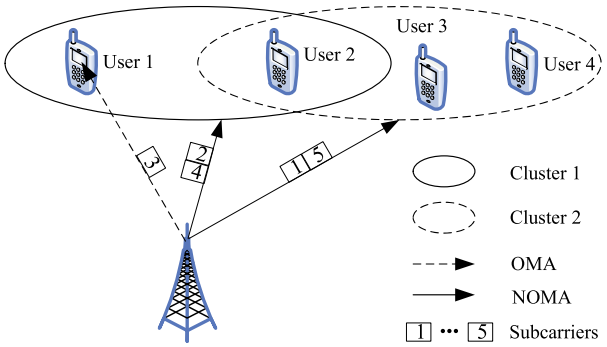

FIGURE 1. System model for the proposed hybrid MC-NOMA.

Without loss of generality, we first focus on the k-th subcarrier and suppose L_k users are clustered and assigned to transmit on this subcarrier. Here, the number of L_k should satisfy the condition of $1 \leq L_k \leq V$, where V is the maximum number of users allowed to transmit on the k-th subcarrier. If $L_k = 1$, it means that a single user transmits on subcarrier k in OMA mode, while if $L_k > 1$, L_k users transmit on subcarrier k simultaneously by applying NOMA protocol. In such transmission scenario, the user $l \in \{1, \cdots, L_k\}$ may receive interference from other users multiplexed on the same subcarrier. We define Ψ_k as the set of users assigned on subcarrier k and denote $L_k = |\Psi_k|$. Consequently, the received signal for user l on subcarrier k is given by

$$y^k_{\Psi_k}(l) = f^k_l \sum_{i \in \Psi_k} \sqrt{p^k_{\Psi_k}(i)} x^k_{\Psi_k}(i) + n^k_l, \tag{1}$$

where $x^k_{\Psi_k}(i)$ is the transmit signal from BS to user i on subcarrier k; n^k_l is the additive white Gaussian noise (AWGN) at user l on subcarrier k with variance σ^2; f^k_l is the channel coefficient between the BS and user l on subcarrier k; $p^k_{\Psi_k}(i)$ is the transmit power allocated to user i on subcarrier k.

For each subcarrier where NOMA scheme is invoked, it is assumed that the normalized channel gains of L_k users on subcarrier k follow the order as $0 < h^k_1 \leq h^k_2 \leq \cdots \leq h^k_{L_k}$, where $h^k_l = |f^k_l|^2 / \sigma^2, l \in \{1, \cdots, L_k\}$. As a result, SIC can be carried out at the users with stronger channels [23]. If we assume $1 \leq j \leq l < i$, the l-th user can decode the message of the j-th user and treat the message for the i-th user as interference. Specifically, the l-th user decodes the messages of all the first $(l-1)$ users, and then successively subtracts these messages to obtain its own information. Following the above principle, the received signal to interference plus noise ratio (SINR) for the l-th user on subcarrier k is given by

$$\gamma^k_{\Psi_k}(l) = \frac{h^k_l p^k_{\Psi_k}(l)}{h^k_l \sum_{i \in \Psi_k | h^k_i \geq h^k_l, i \neq l} p^k_{\Psi_k}(i) + 1}. \tag{2}$$

Then, the normalized achievable data rate over a unit bandwidth for user l on subcarrier k can be expressed as

$$r^k_{\Psi_k}(l) = \log_2\left(1 + \gamma^k_{\Psi_k}(l)\right). \tag{3}$$

Remark 1: The SINR expression (2) is derived based on the principle of NOMA. However, it is worth noting that if $L_k = 1$, it means a single user has access to subcarrier k orthogonally. In such case, (2) still holds as the first term of the denominator vanishes and thus there is no inter-user interference on subcarrier k, which is just the case of OMA. That is the reason why (2) is applicable to both NOMA and OMA. Hence, our proposed system model is actually a hybrid MC-NOMA model which incorporates NOMA and OMA into one unified framework by the generic SINR expression. The advantage of such a framework lies in that we can perform the optimization and resource allocation algorithm design for both NOMA and OMA by a using unified approach, rather than addressing them separately.

III. PROBLEM FORMULATION

In this section, we provide the problem formulation for the hybrid MC-NOMA resource allocation scheme, where the objective is to achieve the SE-EE tradeoff with minimum rate requirement of each user. All the degrees of freedom in resource allocation are jointly considered and arbitrary numbers of users are allowed to be multiplexed on the same subcarrier.

A. ACHIEVABLE DATA RATE

Suppose at most $V \geq 2$ users are allowed to be clustered and multiplexed on the same subcarrier by applying NOMA protocol. The user cluster for subcarrier k is denoted as $\Psi_k = \{l_1, l_2, \ldots, l_V\}$, where $l_1, l_2, \ldots, l_V \in \{1, 2, \ldots, L\}$. According to (3), if $h^k_{l_1} \geq h^k_{l_2} \geq \ldots \geq h^k_{l_V}$, the normalized achievable data rate of user l_1 is obtained by

$$r^k_{\Psi_k}(l_1) = \log_2\left(1 + p^k_{\Psi_k}(l_1)\, h^k_{l_1}\right), \tag{4}$$

and the normalized achievable data rate of user l_2 is expressed as

$$r^k_{\Psi_k}(l_2) = \log_2\left(1 + \frac{p^k_{\Psi_k}(l_2)\, h^k_{l_2}}{1 + p^k_{\Psi_k}(l_1)\, h^k_{l_2}}\right). \tag{5}$$

Similarly, the normalized achievable data rate of the other users in Ψ_k, i.e., $r^k_{\Psi_k}(l_3), \ldots, r^k_{\Psi_k}(l_V)$, can also be obtained on the basis of (3), which are omitted here for the sake of simplicity.

Accordingly, the total normalized data rate of subcarrier k is

$$R^k_{\mathrm{SC}} = r^k_{\Psi_k}(l_1) + r^k_{\Psi_k}(l_2) + \ldots + r^k_{\Psi_k}(l_V). \tag{6}$$

For the special case of $h^k_{l_1} = h^k_{l_2} = \ldots = h^k_{l_V}$, (4), (5) and (6) still hold. In fact, the channel gains of different users for a given subcarrier cannot be the same due to their mutually independent random nature [24]. In particular, it can be proved that the event of two links having the same channel gain has Lebesgue-measure zero when the fading process has a continuous cumulative distribution function [25]. As a consequence, the special case of $h^k_{l_1} = h^k_{l_2} = \ldots = h^k_{l_V}$ occurs if and only if $l_1 = l_2 = \ldots = l_V$. When $l_1 = l_2 = \ldots = l_V$, it means a single user l_1 (which can be equivalently denoted by l_2, l_3,..., or l_V) is assigned to transmit on subcarrier k, i.e., user l_1 has access to subcarrier k orthogonally. This illustrates again that the proposed system model in this paper is a hybrid MA model which incorporates both NOMA and OMA.

In order to represent NOMA and OMA in a unified framework, for the special case of $h^k_{l_1} = h^k_{l_2} = \ldots = h^k_{l_V}$, we still view user l_1, l_2,..., and l_V as different users from the mathematical perspective. When $h^k_{l_1} = h^k_{l_2} = \ldots = h^k_{l_V}$, no matter what the decoding order is and how the power on subcarrier k is allocated between $p^k_{\Psi_k}(l_1)$, $p^k_{\Psi_k}(l_2)$,..., and $p^k_{\Psi_k}(l_V)$, the total normalized data rate of subcarrier k, R^k_{SC}, remains the same, since l_1, l_2,..., and l_V actually represent the same one user.

Then, to represent the system throughput and power consumption, we introduce a binary variable $x^k_{\Psi_k} \in \{0, 1\}$ and a parameter $y^k_{\Psi_k} \in \{0, 1\}$. $x^k_{\Psi_k}$ is the subcarrier assignment and user clustering indicator. If subcarrier k is assigned to the user cluster Ψ_k, $x^k_{\Psi_k} = 1$; Otherwise, $x^k_{\Psi_k} = 0$. For the parameter $y^k_{\Psi_k}$, we define

$$y^k_{\Psi_k} = \begin{cases} 1, & \text{if } h^k_{l_1} \geq h^k_{l_2} \geq \ldots \geq h^k_{l_V},\ \forall k, \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

Note that $y^k_{\Psi_k}$ is a constant parameter depending on the channel conditions of all users in the cluster Ψ_k, rather than an optimization variable. With the help of $x^k_{\Psi_k}$ and $y^k_{\Psi_k}$, the hybrid MC-NOMA system throughput is the summation of data rates generated by all the subcarriers, which can be expressed as

$$R = \sum_{l_1=1}^{L}\sum_{l_2=1}^{L}\cdots\sum_{l_V=1}^{L}\sum_{k=1}^{K} x^k_{\Psi_k} y^k_{\Psi_k}\left(r^k_{\Psi_k}(l_1) + r^k_{\Psi_k}(l_2) + \ldots + r^k_{\Psi_k}(l_V)\right). \tag{8}$$

We next derive the data rate of any user l. For any user $l \in \{1, 2, \ldots, L\}$, its data rate consists of multiple parts. If user l is the one with the highest channel gain in the user cluster Ψ_k, the generated data rate is given by

$$R_{l,1} = \sum_{l_2=1}^{L}\sum_{l_3=1}^{L}\cdots\sum_{l_V=1}^{L}\sum_{k=1}^{K} x^k_{ll_2l_3\ldots l_V} y^k_{ll_2l_3\ldots l_V} r^k_{ll_2l_3\ldots l_V}(l). \tag{9}$$

For the case where user l is the one with the second highest channel gain in the user cluster Ψ_k, the generated data rate is expressed as

$$R_{l,2} = \sum_{l_1=1}^{L}\sum_{l_3=1}^{L}\cdots\sum_{l_V=1}^{L}\sum_{k=1}^{K} x^k_{l_1ll_3\ldots l_V} y^k_{l_1ll_3\ldots l_V} r^k_{l_1ll_3\ldots l_V}(l). \tag{10}$$

Likewise, we can also derive the other data rate expressions, i.e., $R_{l,3}, R_{l,4}, \ldots$, and $R_{l,V}$, when the channel gain of user l is in the other orders in the user cluster Ψ_k, which take the forms similar to (9) and (10) and are omitted here for brevity.

Therefore, the total data rate of user l is the summation of data rates generated from all the above cases, which is given by

$$\begin{aligned} R_l = & \sum_{l_2=1}^{L}\sum_{l_3=1}^{L}\cdots\sum_{l_V=1}^{L}\sum_{k=1}^{K} x^k_{ll_2l_3\ldots l_V} y^k_{ll_2l_3\ldots l_V} r^k_{ll_2l_3\ldots l_V}(l) \\ & + \sum_{l_1=1}^{L}\sum_{l_3=1}^{L}\cdots\sum_{l_V=1}^{L}\sum_{k=1}^{K} x^k_{l_1ll_3\ldots l_V} y^k_{l_1ll_3\ldots l_V} r^k_{l_1ll_3\ldots l_V}(l) \\ & + \ldots + \sum_{l_1=1}^{L}\sum_{l_2=1}^{L}\cdots\sum_{l_{V-1}=1}^{L} \\ & \times \sum_{k=1}^{K} x^k_{l_1l_2\ldots l_{V-1}l} y^k_{l_1l_2\ldots l_{V-1}l} r^k_{l_1l_2\ldots l_{V-1}l}(l). \end{aligned} \tag{11}$$

B. POWER CONSUMPTION MODEL

With the aid of $x^k_{\Psi_k}$ and $y^k_{\Psi_k}$, the transmit power of BS can be expressed as

$$\begin{aligned} P_{\mathrm{t}} = & \sum_{l_1=1}^{L}\sum_{l_2=1}^{L}\cdots\sum_{l_V=1}^{L}\sum_{k=1}^{K} x^k_{\Psi_k} y^k_{\Psi_k}\left(p^k_{\Psi_k}(l_1) + p^k_{\Psi_k}(l_2)\right. \\ & \left. + \ldots + p^k_{\Psi_k}(l_V)\right). \end{aligned} \tag{12}$$

Then, the total power consumption of BS is generally the summation of the static circuit power and the dynamic amplifier's power [26], which is given by

$$P = P_{\mathrm{C}} + \varepsilon P_{\mathrm{t}}, \tag{13}$$

where P_{C} is the static circuit power of BS and $1/\varepsilon$ represents the amplifier's efficiency.

C. PROBLEM FORMULATION

In this paper, we focus on the resource allocation problem for the SE-EE tradeoff with each user's minimum rate requirement in hybrid MC-NOMA systems, where SE is defined as the normalized system throughput over bandwidth, and EE is defined as the delivered bits per unit energy, i.e., $SE = R$ and $EE = R/P$. All the degrees of freedom in resource allocation, including the selection of MA modes, user clustering, subcarrier assignment and power allocation, are jointly considered. As explained in [27], the tradeoff between SE and EE can be actually equivalent to maximizing the SE and minimizing the total power consumption simultaneously. Hence, we formulate it as a multi-objective optimization (MOO) problem as follows:

$$\min_{\mathbf{x},\mathbf{p}} \; -R(\mathbf{x},\mathbf{p}), \tag{14a}$$

$$\min_{\mathbf{x},\mathbf{p}} \; P(\mathbf{x},\mathbf{p}), \tag{14b}$$

$$\text{s.t. C1}: P_{\mathrm{t}} \le P_{\mathrm{T}}, \tag{14c}$$

$$\text{C2}: p^k_{\Psi_k}(l_1) \ge 0, p^k_{\Psi_k}(l_2) \ge 0, \ldots, p^k_{\Psi_k}(l_V) \ge 0, \quad \forall k, \tag{14d}$$

$$\text{C3}: \sum_{l_1=1}^{L}\sum_{l_2=1}^{L}\cdots\sum_{l_V=1}^{L} x^k_{\Psi_k} \le 1, \quad \forall k, \tag{14e}$$

$$\text{C4}: x^k_{\Psi_k} \in \{0,1\}, \quad \forall k, \tag{14f}$$

$$\text{C5}: R_l \ge R_{\min}, \quad \forall l. \tag{14g}$$

In problem (14), $\mathbf{p} = \left\{p^k_{\Psi_k}(l_1), p^k_{\Psi_k}(l_2), \ldots, p^k_{\Psi_k}(l_V)\right\}$ and $\mathbf{x} = \left\{x^k_{\Psi_k}\right\}$ with $l_1, l_2, \ldots, l_V \in \{1,2,\ldots,L\}$ and $k \in \{1,2,\ldots,K\}$. The constraints C3 and C4 indicate that each subcarrier k can be assigned to only one user cluster $(l_1, l_2, \ldots, l_V)$, while the maximum number of subcarriers that can be occupied by one user is not constrained. C5 is the minimum rate requirement serving as the QoS guarantee for each user.

For the formulated MOO problem (14), as pointed out in [28], after converting it into a scalar problem, the EE maximization corresponds to a specific choice of the weighting parameter. If the largest weighting parameter is imposed on the single objective (14a), the SE is maximized. Hence, although the MOO problem (14) maximizes the SE and minimizes the total power consumption instead of maximizing the SE and EE directly, it can still achieve the SE-EE tradeoff for a specific range of the weighting parameter [27].

IV. JOINT RESOURCE ALLOCATION FOR THE HYBRID MC-NOMA SYSTEMS

In this section, we first solve the MOO problem (14) by utilizing the weighted Tchebycheff method. Then, a joint MA mode selection, user clustering, subcarrier assignment and power allocation algorithm for achieving the SE-EE tradeoff in hybrid MC-NOMA systems is proposed.

A. TRANSFORMATION TO SINGLE OBJECTIVE OPTIMIZATION

For a MOO problem with conflicting objectives, the weighted Tchebycheff method is one of the most general methods to convert it into a single-objective optimization (SOO) problem [29]. Such method is also employed in [30] where it is termed as the utility profile method. In [30], after the conversion of a MOO problem into a SOO problem, it is equivalently reformulated as a max-min problem, and the sequential generalized fractional programming is proposed to operate the max-min problem directly. Since the subcarrier allocation problem is not involved in [30], in this paper, in order to facilitate the subcarrier allocation, the transformed SOO problem will be solved by using the Lagrangian dual method.

To employ the weighted Tchebycheff method, in the first step, we normalize the two conflicting objective functions to ensure a consistent comparison [31] as follows

$$\min_{\mathbf{x},\mathbf{p}} \; \frac{R_{\max} - R(\mathbf{x},\mathbf{p})}{R_{\max}}, \tag{15a}$$

$$\min_{\mathbf{x},\mathbf{p}} \; \frac{P(\mathbf{x},\mathbf{p})}{P_{\max}}, \tag{15b}$$

where $R_{\max}$ is the maximum achievable system throughput and $P_{\max}$ denotes the maximum total power consumption of BS. Mathematically, they are defined as $P_{\max} = P_{\mathrm{C}} + \varepsilon P_{\mathrm{T}}$ and $R_{\max} = \max_{\mathbf{x},\mathbf{p}} R(\mathbf{x},\mathbf{p})\big|_{P_{\mathrm{t}}=P_{\mathrm{T}}}$, respectively.

Then, according to the weighted Tchebycheff method, the MOO problem (14) can be converted into a SOO problem as

$$\min_{\mathbf{x},\mathbf{p}} \max\left\{w\frac{R_{\max}-R(\mathbf{x},\mathbf{p})}{R_{\max}},(1-w)\frac{P(\mathbf{x},\mathbf{p})}{P_{\max}}\right\}, \quad (16a)$$

$$\text{s.t. C1-C5}, \quad (16b)$$

where $w \in [0,1]$ is the weighting parameter, which is used to reflect the importance levels of the two conflicting objectives.

To solve such a *min-max* problem, we introduce an additional auxiliary optimization variable ϕ to make it tractable. Hence, the SOO problem (16) is further rewritten as

$$\min_{\mathbf{x},\mathbf{p},\phi} \phi, \quad (17a)$$

$$\text{s.t. C1-C5}, \quad (17b)$$

$$\text{C6: } w\frac{R_{\max}-R}{R_{\max}}-\phi\le 0, \quad (17c)$$

$$\text{C7: } (1-w)\frac{P}{P_{\max}}-\phi\le 0. \quad (17d)$$

By relaxing the constraints C6 and C7, the partial Lagrangian function is given by

$$\begin{aligned}G(\mathbf{x},\mathbf{p},\phi,\lambda,\mu) &= \phi+\lambda\left(w\frac{R_{\max}-R}{R_{\max}}-\phi\right)\\ &\quad+\mu\left((1-w)\frac{P}{P_{\max}}-\phi\right)\\ &= \lambda w\frac{R_{\max}-R}{R_{\max}}+\mu(1-w)\frac{P}{P_{\max}}\\ &\quad+(1-\lambda-\mu)\phi, \end{aligned} \quad (18)$$

where $\lambda \ge 0$ and $\mu \ge 0$ are Lagrangian multipliers corresponding to the constraints C6 and C7, respectively. Then, the dual function is expressed as

$$H(\lambda,\mu)=\begin{cases}\min\limits_{\mathbf{x},\mathbf{p},\phi} G(\mathbf{x},\mathbf{p},\phi,\lambda,\mu),\\ \text{s.t. C1-C5.}\end{cases} \quad (19)$$

Note that relaxing the constraints C6 and C7 may introduce duality gap due to the non-convex nature of problem (17). In the simulation results, we will illustrate that the performance gap between the Lagrangian dual method and the exhaustive search method is very small.

Observing (18) and (19), for a given set of Lagrangian multipliers (λ,μ), the optimization problem $\min\limits_{\mathbf{x},\mathbf{p},\phi} G(\mathbf{x},\mathbf{p},\phi,\lambda,\mu)$ consists of two parts, which are the joint MA mode selection, user clustering, subcarrier assignment and power allocation optimization problem

$$\min_{\mathbf{x},\mathbf{p}} G_1(\mathbf{x},\mathbf{p},\lambda,\mu)=\lambda w\frac{R_{\max}-R}{R_{\max}}+\mu(1-w)\frac{P}{P_{\max}}, \quad (20a)$$

$$\text{s.t. C1-C5}, \quad (20b)$$

and the adaptive ϕ selection problem

$$\min_{\phi} G_2(\phi,\lambda,\mu)=(1-\lambda-\mu)\phi. \quad (21)$$

B. JOINT MA MODE SELECTION, USER CLUSTERING, SUBCARRIER ASSIGNMENT AND POWER ALLOCATION

In the following, we solve the optimization problem (20), which is divided into an outer and an inner subproblem. In the outer subproblem, joint MA mode selection, user clustering and subcarrier assignment are obtained by the Lagrangian dual decomposition method, while for the inner subproblem, the power allocation problem is solved by employing the sequential convex programming. As the Lagrangian dual decomposition is not guaranteed to be globally solved with a finite number of subcarriers, problem (20) is not sure to obtain the globally optimal solution.

The technological advancement of Lagrangian dual decomposition is that it can decouple the dual objective function into K independent problems. As the decoupled independent problems are unconstrained, they are more manageable than the primal problem. While for the sequential convex programming, the advantage lies in that it can solve the non-convex power allocation problem with low complexity and fulfilling the KKT first-order optimality conditions. More importantly, as we will illustrate in the simulation results, the sequential convex programming actually attains the global optimum.

1) JOINT MA MODE SELECTION, USER CLUSTERING AND SUBCARRIER ASSIGNMENT PROBLEM

Due to the existence of inter-user interference and integer allocation indicators, (20) is a non-convex mixed-integer non-linear optimization problem, which is NP-hard. This motivates us to consider efficient algorithms to find near-optimal solutions in polynomial time. One natural approach is to apply Lagrangian relaxation. Since problem (20) is non-convex, there may exist duality gap between the primal and dual solutions. Nevertheless, in [32] and [33], it is shown that if the optimization problem satisfies the time-sharing condition, the duality gap is zero even if the problem is not convex. We now show the time-sharing condition holds for (20) as $K \to \infty$.

Definition 1 [32]: Let $(\mathbf{x}^1,\mathbf{p}^1)$ and $(\mathbf{x}^2,\mathbf{p}^2)$ be the optimal solutions to the spectrum optimization problem (20) with power and minimum rate constraints $\{P_{\mathrm{T}}^1, R_{\min}^1\}$ and $\{P_{\mathrm{T}}^2, R_{\min}^2\}$, respectively. The optimization problem (20) is said to satisfy the *time-sharing condition* if for any $\{P_{\mathrm{T}}^1, R_{\min}^1\}$ and $\{P_{\mathrm{T}}^2, R_{\min}^2\}$ and for any $\nu \in [0,1]$, there always exists a feasible solution $(\mathbf{x}^3,\mathbf{p}^3)$, such that $P_{\mathrm{t}}^3 \le \nu P_{\mathrm{T}}^1+(1-\nu)P_{\mathrm{T}}^2$, $R_l^3 \ge \nu R_{\min}^1+(1-\nu)R_{\min}^2$, and $G_1(\mathbf{x}^3,\mathbf{p}^3,\lambda,\mu) \ge \nu G_1(\mathbf{x}^1,\mathbf{p}^1,\lambda,\mu)+(1-\nu)G_1(\mathbf{x}^2,\mathbf{p}^2,\lambda,\mu)$, where P_{t}^3 and R_l^3 are the resulting transmit power of BS and the data rate of each user calculated by (12) and (11) with the solution $(\mathbf{x}^3,\mathbf{p}^3)$.

Proposition 1: The *time-sharing condition* holds for problem (20) in the limit $K \to \infty$.

Proof: Please see Appendix A. ■

Since the time-sharing condition holds in the limit $K \to \infty$, problem (20) can be solved by applying Lagrangian

$$\begin{aligned}\min_{\mathbf{x},\mathbf{p}} L_1(\mathbf{x},\mathbf{p},\eta,\boldsymbol{\theta},\lambda,\mu) = &\sum_{l_1=1}^{L}\sum_{l_2=1}^{L}\cdots\sum_{l_V=1}^{L}\sum_{k=1}^{K} x_{\Psi_k}^k y_{\Psi_k}^k \left\{\frac{-\lambda w}{R_{\max}}\left(r_{\Psi_k}^k(l_1)+r_{\Psi_k}^k(l_2)+\ldots+r_{\Psi_k}^k(l_V)\right)\right.\\ &\left.+\left(\frac{\mu(1-w)\varepsilon}{P_{\max}}+\eta\right)\left(p_{\Psi_k}^k(l_1)+p_{\Psi_k}^k(l_2)+\ldots+p_{\Psi_k}^k(l_V)\right)\right\}\\ &-\sum_{l=1}^{L}\sum_{l_2=1}^{L}\sum_{l_3=1}^{L}\cdots\sum_{l_V=1}^{L}\sum_{k=1}^{K} x_{ll_2l_3\ldots l_V}^k y_{ll_2l_3\ldots l_V}^k \theta_l r_{ll_2l_3\ldots l_V}^k(l)\\ &-\sum_{l_1=1}^{L}\sum_{l=1}^{L}\sum_{l_3=1}^{L}\cdots\sum_{l_V=1}^{L}\sum_{k=1}^{K} x_{l_1ll_3\ldots l_V}^k y_{l_1ll_3\ldots l_V}^k \theta_l r_{l_1ll_3\ldots l_V}^k(l)-\ldots\\ &-\sum_{l_1=1}^{L}\sum_{l_2=1}^{L}\cdots\sum_{l_{V-1}=1}^{L}\sum_{l=1}^{L}\sum_{k=1}^{K} x_{l_1l_2\ldots l_{V-1}l}^k y_{l_1l_2\ldots l_{V-1}l}^k \theta_l r_{l_1l_2\ldots l_{V-1}l}^k(l) \qquad (22)\end{aligned}$$

$$\begin{aligned}\min_{\mathbf{x},\mathbf{p}} L_1(\mathbf{x},\mathbf{p},\eta,\boldsymbol{\theta},\lambda,\mu) = &\sum_{l_1=1}^{L}\sum_{l_2=1}^{L}\cdots\sum_{l_V=1}^{L}\sum_{k=1}^{K} x_{\Psi_k}^k y_{\Psi_k}^k \left\{\left(-\frac{\lambda w}{R_{\max}}-\theta_{l_1}\right) r_{\Psi_k}^k(l_1)+\left(-\frac{\lambda w}{R_{\max}}-\theta_{l_2}\right) r_{\Psi_k}^k(l_2)+\ldots\right.\\ &\left.+\left(-\frac{\lambda w}{R_{\max}}-\theta_{l_V}\right) r_{\Psi_k}^k(l_V)+\left(\frac{\mu(1-w)\varepsilon}{P_{\max}}+\eta\right)\left(p_{\Psi_k}^k(l_1)+p_{\Psi_k}^k(l_2)+\ldots+p_{\Psi_k}^k(l_V)\right)\right\}. \qquad (23)\end{aligned}$$

dual method. Note that in practical systems, the number of subcarriers, K, is usually large but not infinite. Hence, although the duality gap is not strictly zero, it is nearly zero. By relaxing the constraints C1 and C5, substituting (8), (11) and (12) into (20) and deleting the constant terms, the Lagrangian function is obtained as (22) at the top of this page.

In the Lagrangian function (22), we note that the term $\sum_{l=1}^{L}\sum_{l_2=1}^{L}\sum_{l_3=1}^{L}\cdots\sum_{l_V=1}^{L}\sum_{k=1}^{K} x_{ll_2l_3\ldots l_V}^k y_{ll_2l_3\ldots l_V}^k \theta_l r_{ll_2l_3\ldots l_V}^k(l)$ actually represents the weighted data rate (Lagrangian multiplier θ_l can be viewed as the weighting parameter) generated by user l when $h_l^k \geq h_{l_2}^k \geq \ldots \geq h_{l_V}^k$. Therefore, in order to unify the notations in (22) and facilitate the dual decomposition method [34], the notation l can be replaced by l_1 without changing the physical significance. Likewise, in the term $\sum_{l_1=1}^{L}\sum_{l=1}^{L}\sum_{l_3=1}^{L}\cdots\sum_{l_V=1}^{L}\sum_{k=1}^{K} x_{l_1ll_3\ldots l_V}^k y_{l_1ll_3\ldots l_V}^k \theta_l r_{l_1ll_3\ldots l_V}^k(l)$, the notation l can be replaced by l_2, and so forth. With these replacements, problem (22) is rewritten as (23) at the top of this page.

The optimization problem (23) can be decomposed into K independent subproblems, and each of the subproblem corresponding to subcarrier k takes the form as

$$\begin{aligned}&\min_{\mathbf{x}^k,\mathbf{p}^k} L_1^k\left(\mathbf{x}^k,\mathbf{p}^k,\eta,\boldsymbol{\theta},\lambda,\mu\right)\\ &\qquad=\sum_{l_1=1}^{L}\sum_{l_2=1}^{L}\cdots\sum_{l_V=1}^{L}\sum_{k=1}^{K} x_{\Psi_k}^k \Theta_{\Psi_k}^k\left(\mathbf{p}^k\right), \qquad (24)\end{aligned}$$

where

$$\mathbf{x}^k=\left\{x_{\Psi_k}^k\right\},\quad \mathbf{p}^k=\left\{p_{\Psi_k}^k(l_1),p_{\Psi_k}^k(l_2),\ldots,p_{\Psi_k}^k(l_V)\right\},$$

and $\Theta_{\Psi_k}^k\left(\mathbf{p}^k\right)$ is defined as

$$\begin{aligned}&\Theta_{\Psi_k}^k\left(\mathbf{p}^k\right)\\ &=y_{\Psi_k}^k\left\{\left(\frac{-\lambda w}{R_{\max}}-\theta_{l_1}\right) r_{\Psi_k}^k(l_1)\right.\\ &\quad+\left(\frac{-\lambda w}{R_{\max}}-\theta_{l_2}\right) r_{\Psi_k}^k(l_2)+\ldots+\left(\frac{-\lambda w}{R_{\max}}-\theta_{l_V}\right) r_{\Psi_k}^k(l_V)\\ &\quad+\left(\frac{\mu(1-w)\varepsilon}{P_{\max}}+\eta\right)\left(p_{\Psi_k}^k(l_1)+p_{\Psi_k}^k(l_2)\right.\\ &\quad\left.\left.+\ldots+p_{\Psi_k}^k(l_V)\right)\right\}. \qquad (25)\end{aligned}$$

According to the constraints C3 and C4 in (14), each subcarrier k is exclusively assigned to only one user cluster $\Psi_k=\{l_1,l_2,\ldots,l_V\}$. If we denote

$$\Theta_{\Psi_k}^{k*}=\min_{\mathbf{p}^k}\Theta_{\Psi_k}^k\left(\mathbf{p}^k\right), \qquad (26)$$

to minimize $L_1^k\left(\mathbf{x}^k,\mathbf{p}^k,\eta,\boldsymbol{\theta},\lambda,\mu\right)$, subcarrier k should be allocated to the user cluster Ψ_k with the minimum value of $\Theta_{\Psi_k}^{k*}$, which yields the optimal user clustering and subcarrier assignment indicator as

$$x_{\Psi_k}^{k*}=\begin{cases}1, & \text{if } (l_1,l_2,\ldots,l_V)=\arg\min\limits_{\Psi_k}\Theta_{\Psi_k}^{k*},\\ 0, & \text{otherwise.}\end{cases} \qquad (27)$$

Remark 2: It is noteworthy that for the obtained $x_{\Psi_k}^{k*}$, if the indexes of the users in Ψ_k are identical, i.e., $l_1=l_2=\ldots=l_V$, it represents that user l_1 (or equivalently denoted by l_2, l_3,..., or l_V) has access to subcarrier k orthogonally; Otherwise, it means that subcarrier k is simultaneously utilized by multiple users $(l_1,l_2,\ldots,l_V)$ in NOMA mode. Hence, $x_{\Psi_k}^{k*}$ is also the indicator of MA mode selection.

2) POWER ALLOCATION PROBLEM

The following problem to be solved is the power allocation optimization problem (26) in order to obtain $\Theta^{k*}_{\Psi_k}$ for any fixed user cluster Ψ_k and subcarrier k, which is the inner subproblem. Due to the existence of inter-user interference, problem (26) is not convex, and therefore obtaining its global optimum is rather difficult. Here, we employ the sequential convex programming [35] to solve the power allocation problem (26), which allows us to develop a low-complexity algorithm guaranteed to converge to a first-order optimal solution and applicable to the general case where arbitrary numbers of users are allowed to share the same subcarrier in NOMA mode.

We observe that for $\forall\gamma$ and $\overline{\gamma} \geq 0$, the following inequality holds [35]:

$$\log_2(1+\gamma) \geq b\log_2\gamma + c, \tag{28}$$

where b and c are defined as

$$b = \frac{\overline{\gamma}}{1+\overline{\gamma}}, \tag{29}$$

$$c = \log_2(1+\overline{\gamma}) - \frac{\overline{\gamma}}{1+\overline{\gamma}}\log_2\overline{\gamma}, \tag{30}$$

respectively. The bound is tight for $\gamma = \overline{\gamma}$.

As a consequence, the upper bound to the objective function in (26) is obtained as

$$\begin{aligned}
&\Theta^k_{\Psi_k}\left(\mathbf{p}^k\right)\\
&\leq y^k_{\Psi_k}\left\{\left(\frac{-\lambda w}{R_{\max}}-\theta_{l_1}\right)\bar{r}^k_{\Psi_k}(l_1)\right.\\
&\quad+\left(\frac{-\lambda w}{R_{\max}}-\theta_{l_2}\right)\bar{r}^k_{\Psi_k}(l_2)+\ldots+\left(\frac{-\lambda w}{R_{\max}}-\theta_{l_V}\right)\bar{r}^k_{\Psi_k}(l_V)\\
&\quad+\left(\frac{\mu(1-w)\varepsilon}{P_{\max}}+\eta\right)\left(p^k_{\Psi_k}(l_1)+p^k_{\Psi_k}(l_2)\right.\\
&\quad\left.\left.+\ldots+p^k_{\Psi_k}(l_V)\right)\right\},
\end{aligned} \tag{31}$$

where

$$\bar{r}^k_{\Psi_k}(l_m) = b^k_{\Psi_k}(l_m)\log_2\left(\gamma^k_{\Psi_k}(l_m)\right) + c^k_{\Psi_k}(l_m), \quad m = 1,\ 2,\ \ldots,\ V. \tag{32}$$

Next, we adopt the transformation $p^k_{\Psi_k}(l_1) = 2^{q^k_{\Psi_k}(l_1)}$, $p^k_{\Psi_k}(l_2) = 2^{q^k_{\Psi_k}(l_2)},\ldots,\ p^k_{\Psi_k}(l_V) = 2^{q^k_{\Psi_k}(l_V)}$ and define $\mathbf{q}^k = \left\{q^k_{\Psi_k}(l_1), q^k_{\Psi_k}(l_2), \ldots, q^k_{\Psi_k}(l_V)\right\}$. Thus, the following optimization problem can be obtained from (31):

$$\begin{aligned}
&\min_{\mathbf{q}^k}\ \bar{\Theta}^k_{\Psi_k}\left(\mathbf{q}^k\right)\\
&= y^k_{\Psi_k}\left\{\left(\frac{-\lambda w}{R_{\max}}-\theta_{l_1}\right)\bar{r}^k_{\Psi_k}(l_1)\right.\\
&\quad+\left(\frac{-\lambda w}{R_{\max}}-\theta_{l_2}\right)\bar{r}^k_{\Psi_k}(l_2)+\ldots+\left(\frac{-\lambda w}{R_{\max}}-\theta_{l_V}\right)\bar{r}^k_{\Psi_k}(l_V)\\
&\quad+\left(\frac{\mu(1-w)\varepsilon}{P_{\max}}+\eta\right)\left(2^{q^k_{\Psi_k}(l_1)}+2^{q^k_{\Psi_k}(l_2)}\right.\\
&\quad\left.\left.+\ldots+2^{q^k_{\Psi_k}(l_V)}\right)\right\}.
\end{aligned} \tag{33}$$

To solve the optimization problem (33), we have the following proposition.

Proposition 2: $\overline{\Theta}^k_{\Psi_k}\left(\mathbf{q}^k\right)$ is a convex function of $\mathbf{q}^k$.

Proof: Please see Appendix B. ■

Since $\overline{\Theta}^k_{\Psi_k}\left(\mathbf{q}^k\right)$ is a convex function of $\mathbf{q}^k$, the optimization problem (33) is a standard convex optimization problem. There exist many efficient numerical algorithms such as the interior-point method to obtain the optimal solution. Then, we iteratively update the power allocation vector $\mathbf{p}^k$ by solving (33) to tighten the upper bound in (31) until convergence. In summary, the power allocation algorithm for users sharing the same subcarrier in hybrid MC-NOMA systems is outlined in **Algorithm 1**. Note that Algorithm 1 is a general power allocation algorithm applicable to the general case where arbitrary numbers of users are multiplexed on the same subcarrier in hybrid MC-NOMA systems.

Algorithm 1 Power Allocation for Users Sharing the Same Subcarrier in Hybrid MC-NOMA Systems

1. For any given weighting parameter w and Lagrangian multipliers $\eta, \boldsymbol{\theta}, \lambda, \mu$, initialize the power allocation vector $\mathbf{p}^k$;
2. Calculate $\gamma^k_{\Psi_k}(l_1)$, $\gamma^k_{\Psi_k}(l_2)$,..., and $\gamma^k_{\Psi_k}(l_V)$ according to (2);
3. Set $\overline{\gamma}^k_{\Psi_k}(l_1) = \gamma^k_{\Psi_k}(l_1)$, $\overline{\gamma}^k_{\Psi_k}(l_2) = \gamma^k_{\Psi_k}(l_2)$,..., $\overline{\gamma}^k_{\Psi_k}(l_V) = \gamma^k_{\Psi_k}(l_V)$, and then compute $b^k_{\Psi_k}(l_1)$, $c^k_{\Psi_k}(l_1)$, $b^k_{\Psi_k}(l_2)$, $c^k_{\Psi_k}(l_2)$,..., $b^k_{\Psi_k}(l_V)$, $c^k_{\Psi_k}(l_V)$ according to (29) and (30), respectively;
4. Solve the standard convex optimization problem (33) and obtain the optimal solution $\mathbf{q}^{k*}$;
5. Update the power allocation vector as $\mathbf{p}^k = 2^{\mathbf{q}^{k*}}$;
6. Repeat 2 to 5 until convergence.

Proposition 3: Algorithm 1 monotonically decreases the value of $\Theta^k_{\Psi_k}\left(\mathbf{p}^k\right)$ at each iteration and finally converges. The convergent solution satisfies the KKT optimality conditions of the optimization problem (26).

Proof: Please see Appendix C. ■

Remark 3: Simulation results will show that the sequential convex programming in Algorithm 1 actually attains the global optimum, as it achieves the same solution as the one obtained by the exhaustive search method.

C. ADAPTIVE OPTIMIZATION OF ϕ

The optimization problem regarding ϕ is

$$\min_{\phi}\ G_2(\phi,\lambda,\mu) = (1-\lambda-\mu)\phi. \tag{34}$$

Observing (17c) and (17d), it is known that $\phi \geq w\frac{R_{\max}-R}{R_{\max}}$ and $\phi \geq (1-w)\frac{P}{P_{\max}}$. Furthermore, since $\frac{R_{\max}-R}{R_{\max}}$, $\frac{P}{P_{\max}}$ and w are all in the interval of [0,1], we have $w\frac{R_{\max}-R}{R_{\max}} \in [0,1]$ and $(1-w)\frac{P}{P_{\max}} \in [0,1]$. Thus, it is evident that

$$\max\left\{w\frac{R_{\max}-R}{R_{\max}}, (1-w)\frac{P}{P_{\max}}\right\} \leq \phi \leq 1. \tag{35}$$

Then, from (34), the optimal solution of ϕ denoted by ϕ^* can be readily obtained as

$$\phi^* = \begin{cases} 1, & \text{if } \lambda + \mu > 1, \\ \max\left\{ w\dfrac{R_{\max} - R}{R_{\max}}, (1-w)\dfrac{P}{P_{\max}} \right\}, & \text{if } \lambda + \mu \leq 1. \end{cases} \tag{36}$$

D. LAGRANGIAN MULTIPLIERS UPDATE

To obtain the optimal solution by dual method, the dual multipliers are updated by the subgradient method as

$$\eta(t+1) = [\eta(t) + \upsilon(t)(P_{\mathrm{t}} - P_{\mathrm{T}})]^+, \tag{37}$$

$$\theta_l(t+1) = [\theta_l(t) + \chi(t)(R_{\min} - R_l)]^+, \quad \forall l, \tag{38}$$

$$\lambda(t+1) = \left[\lambda(t) + \alpha(t)\left(w\frac{R_{\max} - R}{R_{\max}} - \phi\right)\right]^+, \tag{39}$$

$$\mu(t+1) = \left[\mu(t) + \beta(t)\left((1-w)\frac{P}{P_{\max}} - \phi\right)\right]^+, \tag{40}$$

where $[x]^+ \triangleq \max\{0, x\}$, t is the iteration index and $\upsilon(t)$, $\chi(t)$, $\alpha(t)$ and $\beta(t)$ are sufficiently small positive step sizes. Typical step size rules are constant and square summable but not summable [32]. In the paper, we choose the latter and set $\upsilon(t) = \chi(t) = \alpha(t) = \beta(t) = \frac{0.1}{t}$.

E. THE PROPOSED JOINT RESOURCE ALLOCATION ALGORITHM

With the optimization results obtained in this section, the proposed joint resource allocation algorithm for achieving SE-EE tradeoff in hybrid MC-NOMA systems, including MA mode selection, user clustering, subcarrier assignment and power allocation, is summarized in **Algorithm 2**. Note that Algorithm 2 is a general resource allocation algorithm applicable to the cases of arbitrary numbers of users can be multiplexed on the same subcarrier.

Algorithm 2 MA Mode Selection, User Clustering, Subcarrier Assignment and Power Allocation for Hybrid MC-NOMA systems ($V \geq 2$)

1. Initialize η, $\boldsymbol{\theta}$, λ and μ.
2. **for** $k = 1 : K$
3. **for** $l_1 = 1 : L$
4. **for** $l_2 = 1 : L$
5. ……
6. **for** $l_V = 1 : L$
7. **if** $h^k_{l_1} \geq h^k_{l_2} \geq \ldots \geq h^k_{l_V}$
8. Solve the optimization problem (26) by Algorithm 1, and thus obtain $\Theta^{k*}_{\Psi_k}$ and the optimal power allocation vector $\mathbf{p}^k$;
9. **end if**
10. **end for**
11. ……
12. **end for**
13. **end for**
14. Determine the MA mode selection, user clustering and subcarrier assignment indicator $x^{k*}_{\Psi_k}$ by (27);
15. **end for**
16. Optimize ϕ by (36);
17. Update η, $\boldsymbol{\theta}$, λ and μ according to (37)-(40);
18. Repeat 2-17 until convergence of the Lagrangian dual method.

F. CONVERGENCE AND COMPLEXITY ANALYSIS

As presented, the proposed resource allocation Algorithm 2 is divided into an outer and an inner problem. The convergence of the inner power allocation problem (i.e., Algorithm 1) is proved by Proposition 3. Since the inner problem actually attains the global optimum, the subgradient update of dual multipliers for the outer problem is guaranteed to converge as long as the step size is chosen to be sufficiently small [32]. In Section V, we will further illustrate the convergence evolution of Algorithm 2 by simulations.

Next, we analyze the complexity of the proposed resource allocation Algorithm 2 and compare it with other approaches. For the exhaustive search method, if at most V users can be clustered and multiplexed on the same subcarrier, the matching times of user clustering for each subcarrier are $\Lambda = C_L^1 + C_L^2 + \ldots + C_L^V$, and therefore, the total complexity for user clustering and subcarrier assignment is $\mathcal{O}\left(\Lambda^K\right)$, which is in exponential complexity. In addition, for the monotonic optimization proposed in [3], although its complexity is much lower than the exhaustive search method, it is still exponential in the number of variables. While for our proposed Algorithm 2, the total complexity of user clustering and subcarrier assignment is $\mathcal{O}\left(KL^V\right)$. As we will illustrate in the simulations, compared to $V = 3$, the performance gain brought by the cases of four or more users sharing the same subcarrier ($V \geq 4$) is very small. Accordingly, in order to keep the receiver complexity comparatively low and restrict the error propagation, it is suggested to set the maximum number of users multiplexed on the same subcarrier as $V = 3$ in hybrid MC-NOMA systems. Hence, our proposed resource allocation algorithm is in polynomial complexity and helpful to the practical implementations.

V. SIMULATION RESULTS

In this section, we evaluate the performance of the proposed MA mode selection, user clustering, subcarrier assignment and power allocation algorithm for hybrid MC-NOMA systems. In the simulations, we consider a single cell with circular coverage, where the BS is in the center of the cell and the users are placed within the coverage area following the uniform distribution. The detailed simulation parameters are given in Table 1.

A. CONVERGENCE AND OPTIMALITY OF THE PROPOSED ALGORITHMS

For the proposed power allocation Algorithm 1, the convergence evolution is shown in Fig. 2 with one random fading

TABLE 1. Simulation parameters.

Parameters	Default value
Cell radius	1000 m
P_{T}	30 dBm
P_{C}	1 W
$1/\varepsilon$	38%
Noise power spectrum density	-174 dBm/Hz
Path-loss exponent	3.8
Standard deviation of shadowing	8 dB
Small-scale fading	Rayleigh fading
K	64
L	4/8

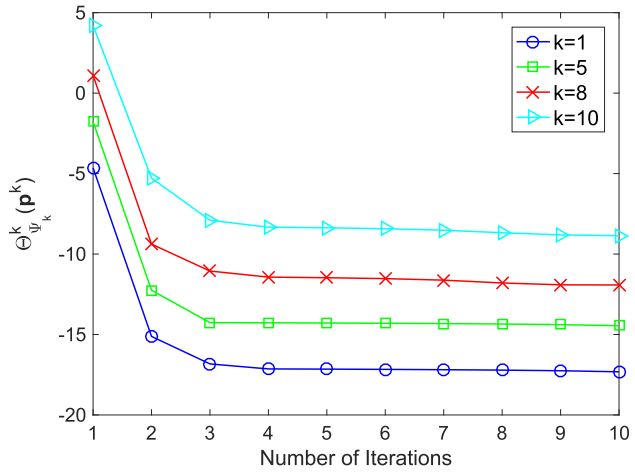

FIGURE 2. Convergence of the power allocation Algorithm 1.

sample of four specific subcarriers, i.e., $k = 1, 5, 8, 10$ with $\Psi_1 = \{1, 3, 6\}$, $\Psi_5 = \{2, 7, 8\}$, $\Psi_8 = \{2, 5, 6\}$ and $\Psi_{10} = \{3, 4, 8\}$. Apparently, just as we proved in Proposition 3, in Fig. 2, the value of $\Theta^k_{\Psi_k}(\mathbf{p}^k)$ is decreased at each iteration and Algorithm 1 converges within five iterations. As for the optimality of Algorithm 1, it has been proved in Proposition 3 that the convergent solution satisfies the KKT conditions of the power allocation problem (26). Furthermore, we discretize the transmit power into uniform steps and set 0.01 W as the power step size to perform the exhaustive search for problem (26). As shown in Table 2, the sequential convex programming (SCP) in Algorithm 1 can actually attain global optimality, as it achieves almost the same value of $\Theta^k_{\Psi_k}(\mathbf{p}^k)$ obtained by the exhaustive search method with a tolerance of around 1%, which is in accordance with the results in [36]. Since Algorithm 1 is the inner algorithm for Algorithm 2, the global optimality of Algorithm 1 is a necessary condition for the convergence of Algorithm 2.

Next, we display the convergence behaviour of the proposed resource allocation Algorithm 2. In Fig. 3, the convergence of the outer Lagrangian dual method (Steps 2-17) is demonstrated, where one of the Lagrangian multipliers $\boldsymbol{\theta} = (\theta_l)$ is taken as an example. Note that this figure is also plotted based on one random channel sample. It is observed that the outer loop of Algorithm 2 converges fast.

TABLE 2. The values of $\Theta^k_{\Psi_k}(\mathbf{p}^k)$ for sequential convex programming (SCP) and exhaustive search method.

Subcarrier Indexes	SCP	Exhaustive Search
$k = 1$	-17.318	-17.481
$k = 5$	-14.456	-14.593
$k = 8$	-11.923	-12.052
$k = 10$	-8.863	-8.945

Since the outer Lagrangian dual method and the inner Algorithm 1 are both converged, the proposed resource allocation Algorithm 2 is guaranteed to converge.

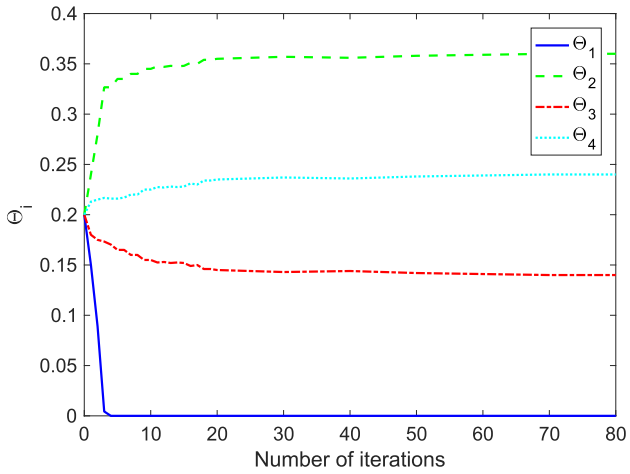

FIGURE 3. Convergence of the outer Lagrangian dual method for Algorithm 2.

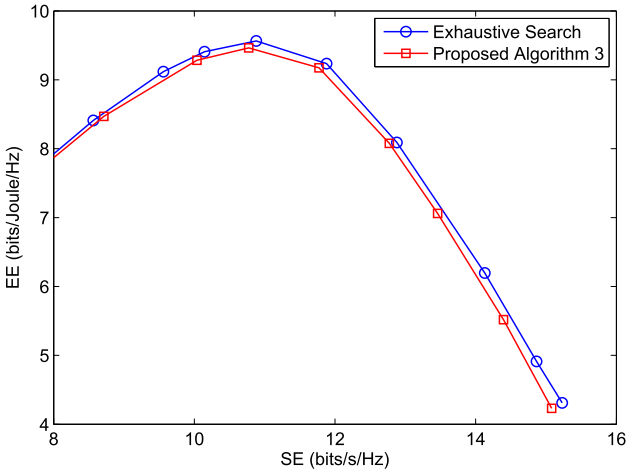

FIGURE 4. SE-EE tradeoff performance comparison between the exhaustive search method and the proposed Algorithm 3.

In Fig. 4, we compare the performances of exhaustive search method and the proposed resource allocation Algorithm 2 for hybrid MC-NOMA systems. Due to the high computational complexity of exhaustive search method, a small-scale problem with 3 users and 6 subcarriers is simulated, and at most 3 users are allowed to share the same

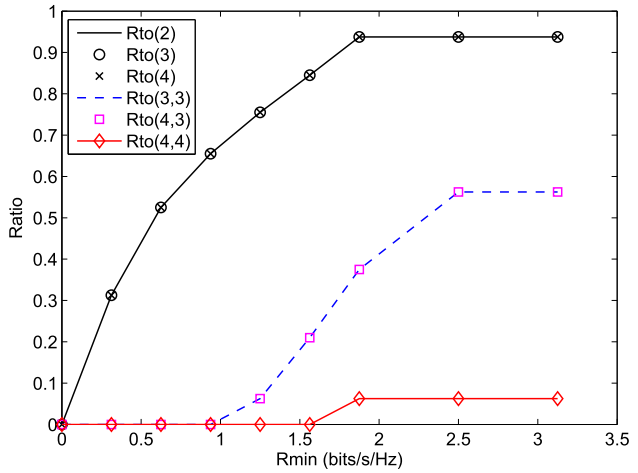

FIGURE 5. Ratio of subcarriers applying NOMA mode *vs.* $R_{\min}$.

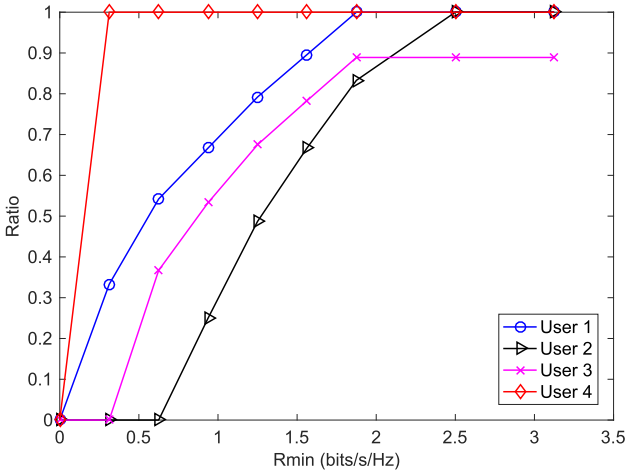

FIGURE 6. Ratio of subcarriers applying NOMA mode for each user *vs.* $R_{\min}$.

subcarrier in NOMA mode. As shown in Fig. 4, the SE-EE tradeoff performance gap between the two methods is very small, which mainly stems from the duality gap. As the duality gap approaches zero when the number of subcarriers goes to infinity, it is predicted that the performance gap will further diminish with the increase of the number of subcarriers.

B. THE RATIO OF SUBCARRIERS APPLYING NOMA MODE

Next, we investigate the impact of $R_{\min}$ on the ratio of subcarriers applying NOMA mode. In Fig. 5, it includes two kinds of legends. The legend Rto(V) represents the ratio of subcarriers applying NOMA mode when the maximum number of users allowed to share the same subcarrier is V, while the legend Rto(V, S) denotes the ratio of subcarriers multiplexed by S users in NOMA mode with the parameter V. As shown in Fig. 5, when $R_{\min} = 0$, no subcarrier is applied in NOMA mode. This is because in this case, the constraint C5 can be satisfied naturally. Then, all the users choose OMA mode since OMA provides the largest summation of data rates if no fairness or QoS requirement is imposed on the system, which is consistent with our analysis in Section IV. For OMA mode, each subcarrier is allocated to a single user with the highest channel gain and the power allocation among subcarriers is conducted by the classical water-filling algorithm. With the increase of $R_{\min}$, the ratio of subcarriers applying NOMA mode also rises. As NOMA is capable of providing a wider capacity region, when $R_{\min}$ cannot be satisfied naturally, the data rates of weak users can be increased by applying NOMA mode with less reduction of strong users' data rates compared to OMA mode. As observed in Fig. 5, when $R_{\min} \geq 1.85$ bits/s/Hz, more than 90% of the subcarriers are accessed by users in NOMA mode.

Besides, when $V > 2$, the ratio of subcarriers shared by over two users, i.e., Rto(3,3), Rto(4,3) and Rto(4,4), also rises with the increase of $R_{\min}$. As allowing more users to share the same subcarrier provides more degrees of freedom in user clustering and subcarrier assignment, some subcarriers may be shared by more than two users to fully exploit the potential performance gain of NOMA. However, even if $R_{\min}$ is large enough, the ratio of subcarriers multiplexed by four users (i.e., Rto(4,4)) is still very low, which implies that allowing four or more users to share the same subcarrier in NOMA mode may not obtain extra gain. This is mainly because more users sharing one subcarrier will also introduce larger inter-user interference, which reduces the performance advantage of NOMA to some extent.

In the following, we demonstrate the ratio of subcarriers applying NOMA mode from the perspective of each user, where a four-user case is taken as an example. The indexes of the four users in Fig. 6 follow the same order with their distances to the BS, i.e., user 1 and user 4 are closest to and furthest from the BS, respectively, while users 2 and 3 are located between user 1 and user 4. As can be seen from Fig. 6, with the increase of $R_{\min}$, the ratio of subcarriers applying NOMA mode for each user also rises, consistent with the trend in Fig. 5. A more significant observation is that all the subcarriers allocated to user 4 are applied in NOMA mode except $R_{\min} = 0$, and the ratio of user 1's subcarriers applying NOMA mode is the second highest. This observation indicates that the users with the best and worst channel conditions are more inclined to transmit in NOMA mode. Interestingly, in our simulations, it is found that user 1 and user 4 are most likely to be clustered in one group to perform NOMA, which implies it is more preferable for NOMA to cluster users whose channel conditions are more distinctive.

C. COMPARISONS OF SE-EE TRADEOFF PERFORMANCES FOR DIFFERENT MA MODES

Fig. 7 shows the SE-EE tradeoff performances of MC-OMA, MC-NOMA and the proposed hybrid MC-NOMA systems by tuning the weighting parameter w. Note that the performance of MC-NOMA is obtained in Algorithm 2 by excluding the case that two users with $l_1 = l_2$ are assigned on the same subcarrier k. As illustrated, with the increase of SE,

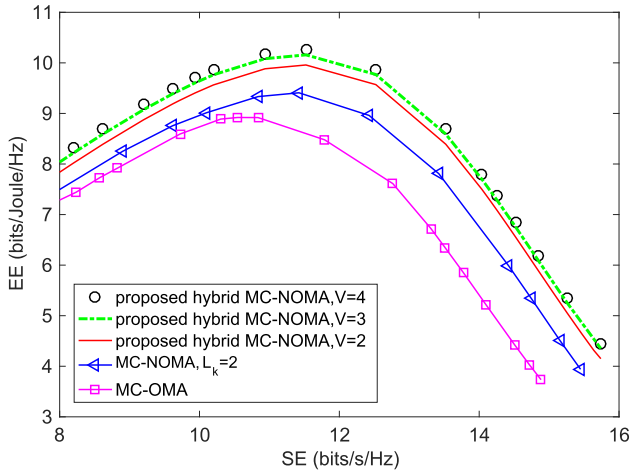

FIGURE 7. Energy efficiency *vs.* spectral efficiency with different MA modes.

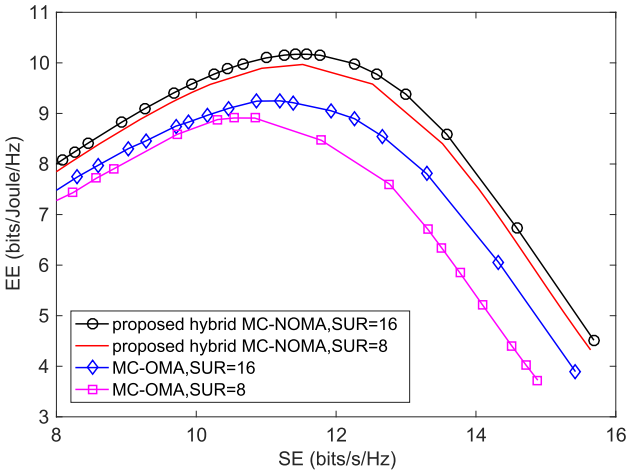

FIGURE 8. Energy efficiency *vs.* spectral efficiency with different SURs and MA modes.

the optimal achievable EE gradually rises to the maximum point, and then decreases to a lower level. By tuning w, it is flexible to perform the tradeoff between SE and EE for different performance preferences. Moreover, it is observed that hybrid MC-NOMA significantly outperforms MC-OMA and MC-NOMA, which indicates that our proposed hybrid MC-NOMA resource allocation algorithm is able to fully exploit the joint advantages of both OMA and NOMA. Meanwhile, although NOMA introduces inter-user interference, the weak users are enabled to transmit with strong users simultaneously on the same subcarrier. Consequently, as illustrated in Fig. 7, for the same level of SE, the achievable EE gain of MC-NOMA over MC-OMA is approximately 15%-25%, which implies that the increase of data rate brought by the power domain multiplexing is larger than the negative impact of inter-user interference.

In Fig. 7, we also demonstrate the SE-EE tradeoff performances of hybrid MC-NOMA when the maximum numbers of users sharing the same subcarrier are 3 and 4, i.e., $V = 3$ and $V = 4$. It is observed that the performance gain brought by four users over three users sharing the same subcarrier is very small. This is consistent with the result in Fig. 5, where the ratio of subcarriers multiplexed by four users is very low. Thus, considering the decoding complexity of receivers and the error propagation problem, it is suggested to set the maximum number of users sharing the same subcarrier as 3 in hybrid MC-NOMA systems.

D. IMPACT OF THE NUMBERS OF SUBCARRIERS AND USERS ON THE SE-EE TRADEOFF PERFORMANCE

In Fig. 8, we investigate the impact of SUR, i.e., the ratio between the numbers of subcarriers and users, on the SE-EE tradeoff performances. It can be seen that when the SUR increases from 8 to 16, the SE-EE tradeoff performances of MC-OMA and hybrid MC-NOMA are both improved. This is because the system benefits from larger diversity gain brought by more subcarriers in resource allocation.

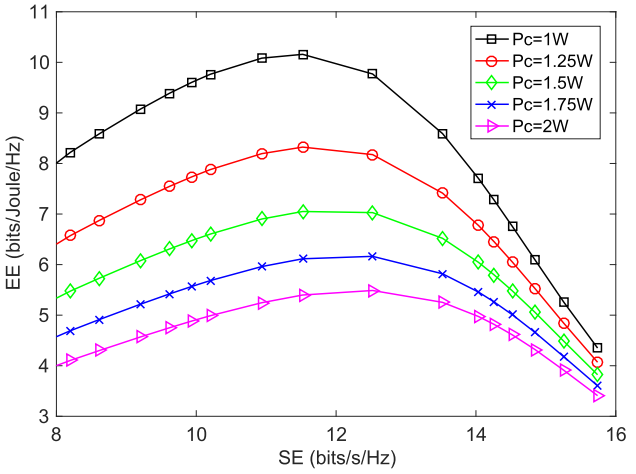

FIGURE 9. Energy efficiency *vs.* spectral efficiency for the proposed hybrid MC-NOMA with different circuit power.

Regarding the individual performance of the two MA modes, the SE-EE tradeoff performance of hybrid MC-NOMA is improved marginally, while for the MC-OMA mode, it is improved substantially, and accordingly, the performance gap between MC-OMA and hybrid MC-NOMA diminishes when SUR goes up. Such an observation demonstrates that the hybrid MC-NOMA is more effective than MC-OMA when the ratio of the numbers of subcarriers to users is lower.

E. IMPACT OF THE CIRCUIT POWER ON THE SE-EE TRADEOFF PERFORMANCE

The SE-EE tradeoff performances for the proposed hybrid MC-NOMA systems with different circuit power are displayed in Fig. 9, where the maximum number of users allowed to be multiplexed on the same subcarrier is 3. Apparently, as the circuit power has no contributions to the achievable SE, EE is always reduced by increasing the circuit power. Therefore, it can be seen from Fig. 9 that increasing

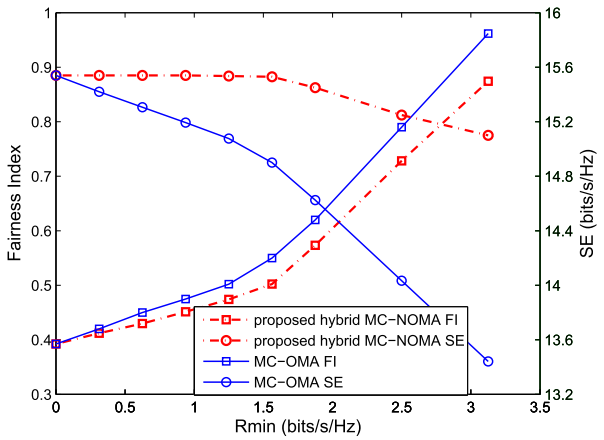

FIGURE 10. Fairness index and spectral efficiency *vs.* $R_{\min}$.

circuit power always degrades the performance of SE-EE tradeoff. Besides, the maximum achievable EE with larger circuit power corresponds to a larger SE.

F. FAIRNESS AND EFFICIENCY TRADEOFF FOR DIFFERENT MA MODES

Finally, Fig. 10 shows the rate fairness and SE of MC-OMA and hybrid MC-NOMA by varying the values of $R_{\min}$, where the rate fairness is measured by Jain's fairness index [37], which can be calculated by $\left(\sum_{l=1}^{L} R_l\right)^2 \Big/ \left(L \sum_{l=1}^{L} R_l^2\right)$. The value range of Jain's fairness index is $\left[1/L,\ 1\right]$ with the maximum achieved by identical users' rates. As the disparity of data rates increases, the fairness index gradually decreases to $1/L$. As can be seen from Fig. 10, with the increase of $R_{\min}$, the rate fairness of MC-OMA and hybrid MC-NOMA both rises. By contrast, since fairness and efficiency are generally conflicting objectives, the SE of MC-OMA and hybrid MC-NOMA both decline with the increase of $R_{\min}$. However, the SE of MC-OMA decreases much more significantly than the hybrid MC-NOMA mode. This verifies that our proposed hybrid MC-NOMA resource allocation scheme has great potential to improve the tradeoff between user fairness and system efficiency thanks to the wider capacity region of NOMA compared to OMA.

VI. CONCLUSIONS

In this paper, we have studied the joint resource allocation for achieving SE-EE tradeoff in hybrid MC-NOMA systems. The SE-EE tradeoff was formulated as a MOO problem with the minimum rate requirement serving as QoS guarantee for each user. The hybrid MC-NOMA scheme incorporates both NOMA and OMA into one unified framework, and all the degrees of freedom in resource allocation, including MA mode selection, user clustering, subcarrier assignment and power allocation, were jointly considered. We have proposed a joint resource allocation algorithm for hybrid MC-NOMA systems which is applicable to the general case where an arbitrary number of users can be multiplexed on the same subcarrier.

In the simulations, we have found that users' minimum rate requirements show significant impact on the selection of MA modes when NOMA and OMA coexist in the system. The hybrid MC-NOMA mode significantly outperforms both NOMA and OMA in terms of SE-EE tradeoff, and it also shows great potential to improve the tradeoff between user fairness and system efficiency. Moreover, since the performance gain brought by four or more users sharing the same subcarrier is very small, it is suggested to set the maximum number of users sharing the same subcarrier as 3 in hybrid MC-NOMA systems to reduce the decoding complexity of receivers and restrict the error propagation while guaranteeing the SE-EE tradeoff performance.

APPENDIX A
PROOF OF PROPOSITION 1

To prove the time-sharing condition, we first consider the continuous version of optimization problem (20), which is given by

$$\min_{\mathbf{p}(f)} G_1\left(\mathbf{p}\left(f\right), \lambda, \mu\right) = \lambda w \frac{R_{\max} - R\left(f\right)}{R_{\max}} + \mu\left(1-w\right)\frac{P\left(f\right)}{P_{\max}}, \tag{41a}$$

$$\text{s.t. } P_{\mathrm{t}}\left(f\right) \le P_{\mathrm{T}}, \tag{41b}$$

$$p_m\left(f\right) \ge 0,\, p_n\left(f\right) \ge 0,\, \forall m, n, \tag{41c}$$

$$R_l\left(f\right) \ge R_{\min},\ \forall l, \tag{41d}$$

where

$$R\left(f\right) = \sum_{m=1}^{L}\sum_{n=1}^{L}\int_{f_L}^{f_H} y_{mn}\left(f\right)\left(\log_2\left(1+p_m\left(f\right)h_m\left(f\right)\right) + \log_2\left(1+\frac{p_n\left(f\right)h_n\left(f\right)}{1+p_m\left(f\right)h_n\left(f\right)}\right)\right)df, \tag{42}$$

$$P_{\mathrm{t}}\left(f\right) = \sum_{m=1}^{L}\sum_{n=1}^{L}\int_{f_L}^{f_H} y_{mn}\left(f\right)\left(p_m\left(f\right)+p_n\left(f\right)\right)df, \tag{43}$$

and

$$R_l\left(f\right) = \sum_{n=1}^{L}\int_{f_L}^{f_H} y_{ln}\left(f\right)\log_2\left(1+p_l\left(f\right)h_l\left(f\right)\right)df + \sum_{m=1}^{L}\int_{f_L}^{f_H} y_{ml}\left(f\right)\log_2\left(1+\frac{p_l\left(f\right)h_l\left(f\right)}{1+p_m\left(f\right)h_l\left(f\right)}\right)df. \tag{44}$$

Note that $[f_L, f_H]$ is a colsed bounded interval corresponding to the system bandwidth.

We first assume the channel gains $h_m\left(f\right)$ and $h_n\left(f\right)$ are constant functions of f for $\forall m, n$. Let $\mathbf{p}^1\left(f\right)$ and $\mathbf{p}^2\left(f\right)$ be the optimal solutions to the spectrum optimization problem (41) with power and minimum rate constraints $\left\{P_{\mathrm{T}}^1, R_{\min}^1\right\}$ and $\left\{P_{\mathrm{T}}^2, R_{\min}^2\right\}$. With the constant channel gain assumption, since the same KKT condition must be satisfied, the optimal solutions $\mathbf{p}^1\left(f\right)$ and $\mathbf{p}^2\left(f\right)$ must also be constant over f. Next, we need to construct a new power allocation vector $\mathbf{p}^3\left(f\right)$ such that it satisfies the constraints

$\left\{vP_{\mathrm{T}}^{1}+(1-v)P_{\mathrm{T}}^{2},\ vR_{\min}^{1}+(1-v)R_{\min}^{2}\right\}$, and the objective function achieves equal to or larger than $vG_1\left(\mathbf{x^1},\mathbf{p^1},\lambda,\mu\right)+(1-v)G_1\left(\mathbf{x^2},\mathbf{p^2},\lambda,\mu\right)$ for $v\in[0,1]$. Such a $\mathbf{p}^3(f)$ can be constructed by subdividing the frequency band into two, v proportion of which has $\mathbf{p}^3(f)=\mathbf{p}^1(f)$, and $(1-v)$ proportion of which has $\mathbf{p}^3(f)=\mathbf{p}^2(f)$. Apparently, with the constructed $\mathbf{p}^3(f)$, the constraints $P_{\mathrm{t}}^3(f)\le vP_{\mathrm{T}}^1+(1-v)P_{\mathrm{T}}^2$ and $R_l^3(f)\ge vR_{\min}^1+(1-v)R_{\min}^2$ are satisfied, and $G_1\left(\mathbf{p^3}(f),\lambda,\mu\right)=vG_1\left(\mathbf{p^1}(f),\lambda,\mu\right)+(1-v)G_1\left(\mathbf{p^2}(f),\lambda,\mu\right)$. Thus, for the case of constant channel gains, the time-sharing condition can be implemented with frequency division.

Next, for the general case with continuous channel gains, we can divide the total frequency into a set of infinitesimal frequency bands. By continuity, the channel gains within each band approaches a constant. Then, the above argument for the case of constant channel gains can be applied in each infinitesimal band. Therefore, the time-sharing condition is satisfied for the entire frequency. Furthermore, for the discretized version of the optimization problem (20), it can be straightforward concluded that the time-sharing condition holds as the number of subcarriers $K\to\infty$. ■

APPENDIX B
PROOF OF PROPOSITION 2

Substituting (2) into (32) and rearranging $\bar{r}_{\Psi_k}^k(l_m)$, $m=1,2,\ldots,V$, we have

$$\begin{aligned}\bar{r}_{\Psi_k}^k(l_m)&=b_{\Psi_k}^k(l_m)\log_2\left(\frac{2^{q_{\Psi_k}^k(l_m)}h_{l_m}^k}{1+\sum_{j=1}^{m-1}2^{q_{\Psi_k}^k(l_j)}h_{l_m}^k}\right)+c_{\Psi_k}^k(l_m)\\&=b_{\Psi_k}^k(l_m)\left[q_{\Psi_k}^k(l_m)-\log_2\left(1+\sum_{j=1}^{m-1}2^{q_{\Psi_k}^k(l_j)}h_{l_m}^k\right)\right]\\&\quad+b_{\Psi_k}^k(l_m)\log_2\left(h_{l_m}^k\right)+c_{\Psi_k}^k(l_m).\end{aligned}\tag{45}$$

Apparently, $\bar{r}_{\Psi_k}^k(l_m)$ is a concave function of $\mathbf{q}^k$ because of the convexity of the log-sum-exp function [38]. Observing (33), one can find that $\overline{\Theta}_{\Psi_k}^k\left(\mathbf{q}^k\right)$ is actually the summation of a set of convex terms of $\mathbf{q}^k$. Hence, it is straightforward to conclude that $\overline{\Theta}_{\Psi_k}^k\left(\mathbf{q}^k\right)$ is a convex function of $\mathbf{q}^k$. ■

APPENDIX C
PROOF OF PROPOSITION 3

Denote $\mathbf{q}_t^k$ as the optimal solution to $\overline{\Theta}_{\Psi_k}^k\left(\mathbf{q}^k\right)$ after the t-th iteration of Algorithm 1 and set $\mathbf{p}_t^k=2^{\mathbf{q}_t^k}$. Then, we have the following inequalities

$$\Theta_{\Psi_k,t}^k\left(\mathbf{p}_t^k\right)\overset{(a)}{=}\overline{\Theta}_{\Psi_k,t}^k\left(\mathbf{q}_t^k\right)\overset{(b)}{\ge}\overline{\Theta}_{\Psi_k,t}^k\left(\mathbf{q}_{t+1}^k\right)\overset{(c)}{\ge}\Theta_{\Psi_k,t}^k\left(\mathbf{p}_{t+1}^k\right),\tag{46}$$

where the equality (a) holds because $b_{\Psi_k}^k(l_m)$ and $c_{\Psi_k}^k(l_m)$ are computed at $\overline{\gamma}_{\Psi_k}^k(l_m)=\gamma_{\Psi_k}^k(l_m)$ with $m=1,2,\ldots,V$ such that the bound is tight; the inequality (b) holds because $\mathbf{q}_{t+1}$ is the global optimum solution to (33); the inequality (c) is due to the fact that $\overline{\Theta}_{\Psi_k,t}^k\left(\mathbf{q}_{t+1}^k\right)$ is an upper-bound of $\Theta_{\Psi_k,t}^k\left(\mathbf{p}_{t+1}^k\right)$. As a result of (46), the objective function $\Theta_{\Psi_k}^k\left(\mathbf{p}^k\right)$ decreases after each iteration. Since $\Theta_{\Psi_k}^k\left(\mathbf{p}^k\right)$ is lower-bounded, the proposed Algorithm 1 must converge.

If denoting $\mathbf{p}^{k*}$ as the power allocation solution at convergence of Algorithm 1, $\mathbf{p}^{k*}$ must satisfy the KKT conditions of (33). Problem (26) and (33) actually have different objective functions of $\Theta_{\Psi_k}^k\left(\mathbf{p}^k\right)$ and $\overline{\Theta}_{\Psi_k}^k\left(\mathbf{q}^k\right)$, respectively. However, once Algorithm 1 converges, we have $\overline{r}_{\Psi_k}^k(l_m)=r_{\Psi_k}^k(l_m)$, $m=1,2,\ldots,V$, and thus $\Theta_{\Psi_k}^k\left(\mathbf{p}^k\right)=\overline{\Theta}_{\Psi_k}^k\left(\mathbf{q}^k\right)$. Therefore, $\mathbf{p}^{k*}$ must also satisfy the KKT conditions of (26). ■

REFERENCES

[1] 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. Surveys Tuts.*, vol. 19, no. 2, pp. 721–742, 2nd Quart., 2017.

[2] Z. Ding *et al.*, "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 185–191, Feb. 2017.

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

[4] Y. Liu, Z. Qin, M. Elkashlan, Y. Gao, and L. Hanzo, "Enhancing the physical layer security of non-orthogonal multiple access in large-scale networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1656–1672, Mar. 2017.

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

[6] Y. Liu, Z. Ding, M. Elkashlan, and H. V. Poor, "Cooperative non-orthogonal multiple access with simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 938–953, Apr. 2016.

[7] C. Li, Q. Zhang, Q. Li, and J. Qin, "Price-based power allocation for non-orthogonal multiple access systems," *IEEE Wireless Commun. Lett.*, vol. 5, no. 6, pp. 664–667, Dec. 2016.

[8] 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.

[9] Q. Sun, S. Han, C.-L. I, and Z. Pan, "On the ergodic capacity of MIMO NOMA systems," *IEEE Wireless Commun. Lett.*, vol. 4, no. 4, pp. 405–408, Aug. 2015.

[10] M. Mollanoori and M. Ghaderi, "Uplink scheduling in wireless networks with successive interference cancellation," *IEEE Trans. Mobile Comput.*, vol. 13, no. 5, pp. 1132–1144, May 2014.

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

[12] 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.

[13] X. Li, C. Li, and Y. Jin, "Dynamic resource allocation for transmit power minimization in OFDM-based NOMA systems," *IEEE Commun. Lett.*, vol. 20, no. 12, pp. 2558–2561, Dec. 2016.

[14] L. Lei, D. Yuan, and P. Värbrand, "On power minimization for non-orthogonal multiple access (NOMA)," *IEEE Commun. Lett.*, vol. 20, no. 12, pp. 2458–2461, Dec. 2016.

[15] Q. Sun, S. Han, C.-L. I, and Z. Pan, "Energy efficiency optimization for fading MIMO non-orthogonal multiple access systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 2668–2673.

[16] Y. Zhang, H.-M. Wang, T.-X. Zheng, and Q. Yang, "Energy-efficient transmission design in non-orthogonal multiple access," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2852–2857, Mar. 2017.
[17] F. Fang, H. Zhang, J. Cheng, and V. C. M. Leung, "Energy-efficient resource allocation for downlink non-orthogonal multiple access network," *IEEE Trans. Commun.*, vol. 64, no. 9, pp. 3722–3732, Sep. 2016.
[18] Z. Song, Q. Ni, K. Navaie, S. Hou, S. Wu, and X. Sun, "On the spectral-energy efficiency and rate fairness tradeoff in relay-aided cooperative OFDMA systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6342–6355, Sep. 2016.
[19] J. Zhang, L. Dai, Z. He, S. Jin, and X. Li, "Performance analysis of mixed-ADC massive MIMO systems over Rician fading channels," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1327–1338, Jun. 2017.
[20] Z. Song, Q. Ni, K. Navaie, S. Hou, and S. Wu, "Energy- and spectral efficiency tradeoff with α-fairness in downlink OFDMA systems," *IEEE Commun. Lett.*, vol. 19, no. 7, pp. 1265–1268, Jul. 2015.
[21] S. Timotheou and I. Krikidis, "Fairness for non-orthogonal multiple access in 5G systems," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1647–1651, Oct. 2015.
[22] L. Song, Y. Li, Z. Ding, and H. V. Poor, "Resource management in non-orthogonal multiple access networks for 5G and beyond," *IEEE Netw.*, vol. 31, no. 4, pp. 8–14, Jul./Aug. 2017.
[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.
[24] Y. Li *et al.*, "Energy-efficient subcarrier assignment and power allocation in OFDMA systems with max-min fairness guarantees," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3183–3195, Sep. 2015.
[25] X. Wang and G. Giannakis, "Resource allocation for wireless multiuser OFDM networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4359–4372, Jul. 2011.
[26] J. Tang, D. K. C. So, E. Alsusa, and K. A. Hamdi, "Resource efficiency: A new paradigm on energy efficiency and spectral efficiency tradeoff," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4656–4669, Aug. 2014.
[27] O. Amin, E. Bedeer, M. H. Ahmed, and O. A. Dobre, "Energy efficiency-spectral efficiency tradeoff: A multiobjective optimization approach," *IEEE Trans. Veh. Technol.*, vol. 65, no. 4, pp. 1975–1981, Apr. 2016.
[28] A. Zappone and E. Jorswieck, "Energy efficiency in wireless networks via fractional programming theory," *Found. Trends Commun. Inf. Theory*, vol. 11, nos. 3–4, pp. 185–396, 2015.
[29] M. R. Mili, L. Musavian, and D. W. K. Ng, "Rate-power-interference optimization in underlay OFDMA CRNs with imperfect CSI," *IEEE Commun. Lett.*, vol. 21, no. 7, pp. 1657–1660, Jul. 2017.
[30] O. Aydin, E. A. Jorswieck, D. Aziz, and A. Zappone, "Energy-spectral efficiency tradeoffs in 5G multi-operator networks with heterogeneous constraints," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5869–5881, Sep. 2017.
[31] R. T. Marler and J. S. Arora, "Survey of multi-objective optimization methods for engineering," *Struct. Multidiscipl. Optim.*, vol. 26, no. 6, pp. 369–395, Apr. 2004.
[32] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1310–1322, Jul. 2006.
[33] Z.-Q. Luo and S. Zhang, "Dynamic spectrum management: Complexity and duality," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 57–73, Feb. 2008.
[34] D. P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1439–1451, Aug. 2006.
[35] A. Zappone, L. Sanguinetti, G. Bacci, E. Jorswieck, and M. Debbah, "Energy-efficient power control: A look at 5G wireless technologies," *IEEE Trans. Signal Process.*, vol. 64, no. 7, pp. 1668–1683, Apr. 2016.
[36] A. Zappone, E. Björnson, L. Sanguinetti, and E. Jorswieck, "Globally optimal energy-efficient power control and receiver design in wireless networks," *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2844–2859, Jun. 2017.
[37] H. Shi, R. V. Prasad, E. Onur, and I. G. M. M. Niemegeers, "Fairness in wireless networks: Issues, measures and challenges," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 5–24, 1st Quart., 2014.
[38] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

ZHENGYU SONG received the B.Sc. and M.Sc. degrees in information and communication engineering from Beijing Jiaotong University, Beijing, China, in 2008, and the Ph.D. degree in information and communication engineering from the Beijing Institute of Technology, Beijing, in 2016. He is currently with the School of Electronic and Information Engineering, Beijing Jiaotong University. His research interests include radio resource management in multicarrier systems, non-orthogonal multiple access, heterogeneous networks, and mobile edge computing.

QIANG NI (M'04–SM'08) received the B.Sc., M.Sc., and Ph.D. degrees in engineering from the Huazhong University of Science and Technology, China, in 1993, 1996, and 1999, respectively. He is currently a Professor and the Head of the Communication Systems Group, School of Computing and Communications, Lancaster University, Lancaster, U.K. His research interests include the area of future generation communications and networking, including green communications and networking, cognitive radio network systems, heterogeneous networks, 5G, SDN, cloud networks, energy harvesting, wireless information and power transfer, IoTs, cyber physical systems, machine learning, big data analytics, and vehicular networks. He has authored or co-authored over 200 papers in these areas.

XIN SUN received the Ph.D. degree in electromagnetic measurement technology and instrument from the Harbin Institute of Technology, Harbin, China, in 1998. She is currently a Professor with the School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, China. Her research interests include professional mobile communications, wireless personal communications, and green networking.