On Optimal Power Allocation for Downlink Non-Orthogonal Multiple Access Systems

Jianyue Zhu, *Student Member, IEEE*, Jiaheng Wang, *Senior Member, IEEE*, Yongming Huang, *Member, IEEE*, Shiwen He, *Member, IEEE*, Xiaohu You, *Fellow, IEEE*, and Luxi Yang, *Member, IEEE*

Abstract—Non-orthogonal multiple access (NOMA) enables power-domain multiplexing via successive interference cancellation (SIC) and has been viewed as a promising technology for 5G communication. The full benefit of NOMA depends on resource allocation, including power allocation and channel assignment, for all users, which, however, leads to mixed integer programs. In the literature, the optimal power allocation has only been found in some special cases, while the joint optimization of power allocation and channel assignment generally requires exhaustive search. In this paper, we investigate resource allocation in downlink NOMA systems. As the main contribution, we analytically characterize the optimal power allocation with given channel assignment over multiple channels under different performance criteria. Specifically, we consider the maximin fairness, weighted sum rate maximization, sum rate maximization with quality of service (QoS) constraints, and energy efficiency maximization with weights or QoS constraints in NOMA systems. We also take explicitly into account the order constraints on the powers of the users on each channel, which are often ignored in the existing works, and show that they have a significant impact on SIC in NOMA systems. Then, we provide the optimal power allocation for the considered criteria in closed or semi-closed form. We also propose a low-complexity efficient method to jointly optimize channel assignment and power allocation in NOMA systems by incorporating the matching algorithm with the optimal power allocation. Simulation results show that the joint resource optimization using our optimal power allocation yields better performance than the existing schemes.

Index Terms—Non-orthogonal multiple access, power allocation, successive interference cancellation, quality of service, combinatorial optimization, sum rate, fairness, energy efficiency, channel assignment.

I. Introduction

AS A result of the popularity of internet-of-things and cloud-based applications, there is an explosive demand of new services and data traffic for wireless communications. Hence, the fifth generation (5G) communication systems propose higher requirements in data rates, lower latency, and massive connectivity [1]. In order to meet these high demands, some potential technologies, such as massive multiple-input multiple-output (MIMO) [2], millimeter wave [3], small cell [4]–[6] and device to device communication [7], [8] will be introduced into 5G communication systems. Recently, non-orthogonal multiple access (NOMA), which can support overloaded transmission with limited resources and further improve the spectral efficiency [9], arises as a promising technology for 5G communication systems.

The conventional multiple access schemes, which are categorized as orthogonal multiple access technologies, are not sufficient to support a massive connectivity because different users are allocated to orthogonal resources in order to mitigate multiple access interference [10]. On the other side, by using superposition coding at the transmitter with successive interference cancellation (SIC) at the receiver, NOMA allows allocating one (frequency, time, code, or spatial) channel to multiple users at the same time [11], which can lead to better performance in terms of spectral efficiency, fairness, or energy efficiency [12]. Therefore, NOMA has received much attention recently. In [13] and [14], the authors discussed an combination of NOMA with MIMO technologies. NOMA has also been introduced to be used with other technologies, e.g., visible light communication [15] and millimeter wave communication [16].

The basic idea of NOMA is to implement multiple access in the power domain [9]. Hence, the key to achieve the full benefit of NOMA systems is resource allocation, which usually include power allocation and channel assignment. Unfortunately, the joint optimization of power allocation and channel assignment in NOMA systems leads to a mixed integer program [17], which has been proved to be a NP-hard problem in [18]. Hence, finding the jointly optimal resource allocation generally requires exhaustive search [19], which, however, causes prohibitive complexity and is not applicable for practical systems. Therefore, suboptimal but efficient resource optimization methods are more preferred in practice. Such efficient methods are often obtained by optimizing power allocation and channel assignment alternately [17], [20]–[22]. In this paper, we investigate resource allocation with a focus on power allocation for downlink NOMA systems under various criteria.

A. Related Works

In NOMA systems, resource allocation has been studied for different performance measures. In the literature, the sum rate

maximization is the most commonly adopted objective, and there are a number of related works [19], [20], [22], [23]. In [23], the authors investigated the optimal power allocation to maximize the sum rate with QoS constraints only for two users on one channel. In [20], the problem of maximizing the weighted sum rate in a downlink orthogonal frequency division multiplexing access (OFDMA) based NOMA system was studied, where the nonconvex power allocation problem was solved via DC (difference of two convex functions) programming and thus only a suboptimal power allocation solution was provided. In [19], the authors also considered the weighted sum rate maximization and exploited monotonic optimization to develop an optimal joint power allocation and channel assignment policy, which, however, has an exponential complexity and only serves as a system performance benchmark. In [22], the authors introduced a resource allocation method based on waterfilling to improve the total achieved system throughput but there is no guarantee for the optimality of the obtained solution.

Fairness is also an important issue in NOMA systems, where the most common fairness indication is the maximin fairness (MMF). Therefore, a number of works has studied resource allocation for MMF, e.g., [24]–[26]. In [24] and [26], the authors investigated the optimal power allocation based on MMF for users on one channel using statistical channel state information (CSI) and instantaneous CSI, respectively. The proportional fairness scheduling that maximizes the weighted MMF was studied in [25], where the optimal solution was only derived for two users on a single channel.

As energy efficiency (EE) becomes an important performance measure of wireless communication systems, the resource problem that maximizes the EE in NOMA systems has also been considered but only in two works [21], [27]. In [27], the authors developed the optimal power allocation for maximizing the EE with QoS constraints but only for the users on one channel. The joint power allocation and channel assignment for maximizing the EE was considered in [21], whereas only a suboptimal solution was obtained via DC programming.

In summary, so far the optimal power allocation was only found for users on a single channel under particular performance criteria, but unknown in the general case for all users on multiple channels. Furthermore, in almost all existing works, the order constraints on the powers of users were either ignored or not explicitly taken into account.

B. Contributions

In this paper, we investigate resource allocation in downlink NOMA systems with a focus on seeking the optimal power allocation for multiple channels and users under various performance criteria. The contributions in this paper are summarized in the following:

- We consider different criteria that lead to different problem formulations, including the maximin fairness, the weighted sum rate maximization, the sum rate maximization with QoS constraints, the energy efficiency maximization with weights or QoS constraints.

- We take explicitly into account the order constraints on the powers of users on each channel that guarantee the decoding order of SIC on each channel unchanged in NOMA systems.
- Then, we analytically characterize the optimal power allocation and provide closed-form or semi-closed solutions to the formulated power optimization problems.
- It is shown that the power order constraints could result in an equal signal strength, which may cause a failure of SIC or a large error propagation. Thus, we introduce the concept of SIC-stability and identify the conditions that avoid equal power allocation in NOMA systems under different criteria.
- We propose an efficient method to jointly optimize the channel assignment and power allocation by incorporating the matching algorithm with our optimal power allocation and iteratively using them to refine the solution.
- The obtained optimal power allocation can also be used with other channel assignment algorithms and can even reduce the complexity of the exhaustive search for jointly optimal resource allocation.
- Finally, it is shown via simulations that the proposed joint resource optimization method outperforms the existing schemes and achieves near-optimal performance.

The rest of the paper is organized as follows. Section II introduces the NOMA system model and various resource optimization problems under different performance criteria and constraints. In Section III, Section IV, and Section V, we investigate the optimal power allocation for the MMF, sum rate maximization, and EE maximization, respectively. In Section VI, a joint channel assignment and power allocation optimization algorithm is proposed. The performance of the proposed power allocation is evaluated in section VII by simulations and the conclusion is drawn in Section VIII.

II. PROBLEM STATEMENT

A. System Model

Consider a downlink NOMA network wherein a base station (BS) serves N users through M channels. The total bandwidth B is equally divided to M channels so the bandwidth of each channel is $B_c = B/M$. Let $N_m \in \{N_1, N_2, ..., N_M\}$ be the number of users using channel m for $m = 1, 2, \cdots, M$ and $\text{UE}_{n,m}$ denotes user n on channel m for $n = 1, 2, \cdots, N_m$. The signal transmitted by the BS on each channel m can be expressed as

$$x_m = \sum_{n=1}^{N_m} \sqrt{p_{n,m}} s_n$$

where s_n is the symbol of $\text{UE}_{n,m}$ and $p_{n,m}$ is the power allocated to $\text{UE}_{n,m}$. The received signal at $\text{UE}_{n,m}$ is

$$y_{n,m} = \sqrt{p_{n,m}} h_{n,m} s_n + \sum_{i=1,i\neq n}^{N_m} \sqrt{p_{i,m}} h_{n,m} s_i + z_{n,m}$$

where $h_{n,m} = g_{n,m} d_n^{-\alpha}$ is the channel coefficient from the BS to $\text{UE}_{n,m}$, $g_{n,m}$ follows a Rayleigh distribution, d_n is

the distance between the BS and $\text{UE}_{n,m}$, α is the path-loss exponent, and $z_{n,m} \sim \mathcal{CN}(0, \sigma_m^2)$ is the additive white Gaussian noise (AWGN).

According to the principle of NOMA, one channel can be assigned to multiple users, who will use SIC to decode their signals. Specifically, let $\Gamma_{n,m} = \left| h_{n,m} \right|^2 / \sigma_m^2$ be the channel to noise ratio (CNR) of $\text{UE}_{n,m}$. Assume without loss of generality (w.l.o.g.) that the CNRs of the users on channel m are ordered as

$$\Gamma_{1,m} \geq \cdots \geq \Gamma_{n,m} \geq \cdots \geq \Gamma_{N_m, m}$$

i.e., $\text{UE}_{1,m}$ and $\text{UE}_{N_m, m}$ are the strongest and weakest users on channel m, respectively. Then, the NOMA protocol allocates higher powers to the users with lower CNRs [9], [28], leading to $p_{1,m} \leq \cdots \leq p_{n,m} \leq \cdots \leq p_{N_m, m}$. Hence, $\text{UE}_{n,m}$ is able to decode signals of $\text{UE}_{i,m}$ for $i > n$ and remove them from its own signal, but treats the signals from $\text{UE}_{i,m}$ for $i < n$ as interference. Therefore, the signal to interference-plus-noise ratio (SINR) of $\text{UE}_{n,m}$ using SIC is given by

$$\gamma_{n,m} = \frac{p_{n,m} \Gamma_{n,m}}{1 + \sum_{i=1}^{n-1} p_{i,m} \Gamma_{n,m}}.$$

Thus, the data rate of $\text{UE}_{n,m}$ is

$$R_{n,m}(p_{n,m}) = B_c \log \left(1 + \frac{p_{n,m} \Gamma_{n,m}}{1 + \sum_{i=1}^{n-1} p_{i,m} \Gamma_{n,m}} \right).$$

Using SIC at each user's receiver causes additional complexity, which is proportional to the number of users on the same channel. Thus, in practice, each channel is often restricted to be assigned to two users [21], [28], [29], which is also beneficial to reduce the error propagation of SIC. In this paper, we would also like to focus on this typical situation and assume that $N_m = 2$ for $m = 1, 2, \cdots, M$ and $N = 2M$. In this case, suppose w.l.o.g. that the CNRs of $\text{UE}_{1,m}$ and $\text{UE}_{2,m}$ are ordered as $\Gamma_{1,m} \geq \Gamma_{2,m}$. Then, the rates of $\text{UE}_{1,m}$ and $\text{UE}_{2,m}$ on channel m are given respectively by

$$R_{1,m} = B_c \log \left(1 + p_{1,m} \Gamma_{1,m} \right),$$
$$R_{2,m} = B_c \log \left(1 + \frac{p_{2,m} \Gamma_{2,m}}{p_{1,m} \Gamma_{2,m} + 1} \right).$$

B. Problem Formulation

The performance of a NOMA scheme relies on resource allocation, including power allocation and channel assignment, for all users. In this paper, we investigate optimization of resource allocation for NOMA systems. For this purpose, we consider the following performance measures.

1) *Maximin fairness*: A common criterion is the maximin fairness (MMF), which aims to provide fairness for all users. The corresponding resource allocation problem is given by

$$\max_{m=1,\ldots,M} \min \left\{ R_{1,m}, R_{2,m} \right\}. \tag{1}$$

The similar problems have been studied in [24]–[26], whereas the optimal power allocation was only found

for a few users on a single channel but unknown for all users over multiple channels.

2) *Sum rate:* The most common objective is to maximize the sum rate (SR) of all users. To avoid that the resource on each channel is occupied by one user, weights or QoS constraints are often introduced into SR maximization. In this paper, we consider both the weighted SR maximization:

$$\max \sum_{m=1}^{M} \left(W_{1,m} R_{1,m} + W_{2,m} R_{2,m} \right) \tag{2}$$

where $W_{n,m}$ is the weight of $\text{UE}_{n,m}$, and the SR maximization with QoS constraints:

$$\max \sum_{m=1}^{M} \left(R_{1,m} + R_{2,m} \right)$$
$$\text{s.t. } R_{n,m} \geq R_{n,m}^{\min}, \quad n = 1, 2, \ \forall m \tag{3}$$

where $R_{n,m}^{\min}$ is the QoS threshold of $\text{UE}_{n,m}$. Although these two problems have been studied in a number of works, e.g., [19], [20], [22], [23], the optimal power allocation was only found for two users on one channel [23], while the joint resource optimization is either suboptimal [20], [22] or needs exhaustive search [19].

3) *Energy efficiency*: In this paper, we also consider improving energy efficiency (EE) of the NOMA system, which is defined as the ratio between the sum rate and the power consumption of the whole system. When weights or QoS constraints are introduced, the EE maximization problem can be formulated as

$$\max \frac{\sum_{m=1}^{M} \left(W_{1,m} R_{1,m} + W_{2,m} R_{2,m} \right)}{P_T + \sum_{m=1}^{M} \left(p_{1,m} + p_{2,m} \right)} \tag{4}$$

or

$$\max \frac{\sum_{m=1}^{M} \left(R_{1,m} + R_{2,m} \right)}{P_T + \sum_{m=1}^{M} \left(p_{1,m} + p_{2,m} \right)} \tag{5}$$
$$\text{s.t. } R_{n,m} \geq R_{n,m}^{\min}, \quad n = 1, 2, \ \forall m$$

where P_T is the power consumption of the circuits and SIC on all channels. This problem has only been studied in [27] but only for one channel.

In addition to the above objectives and QoS constraints, one shall also consider power constraints in NOMA systems. The transmit power constraint of the BS is given by

$$\sum_{m=1}^{M} (p_{1,m} + p_{2,m}) \leq P$$

where P is the total power budget of the BS. In NOMA systems, there is an implicit power constraint for the users on each channel m, i.e.,

$$p_{1,m} \leq p_{2,m}, \quad m = 1, \ldots, M$$

which is to guarantee that a higher power is allocated to the user with a lower CNR (i.e.,$\text{UE}_{2,m}$) on channel m so that the decoding order of the SIC is not changed. However, in most existing works, the power order constraints were ignored.

In this paper, we will show that it is important to take such constraints into account explicitly in power allocation for NOMA.

The joint optimization of power allocation and channel assignment in NOMA systems is, unfortunately, a mixed integer problem. Finding the jointly optimal solution requires exhaustive search [17], which results in prohibitive computational complexity. Therefore, in practice, power allocation and channel assignment are often separately and alternatively optimized, i.e., fix one and optimize the other [17], [20], [21], which may lead to, though possibly suboptimal, efficient resource allocation solutions. In this paper, we would also like to use this methodology. Specifically, we first optimize power allocation with given channel assignment, and then optimize channel assignment. The most exciting thing is that, different from all existing works, we are able to find the optimal power allocation for all users over multiple channels for all above considered performance measures. The optimal power allocation is either given in a closed-form expression or can be efficiently obtained via the proposed algorithms. Our results will dramatically simplify the joint resource allocation and improve the system performance.

III. OPTIMAL POWER ALLOCATION FOR MAXIMIN FAIRNESS

The NOMA scheme enables a flexible management of the users' achievable rates and provides an efficient way to enhance user fairness. In this section, we study the optimal power allocation to achieve the maximin fairness (MMF) in the NOMA system. According to (1), with given channel assignment, the MMF problem is equivalent to the following power allocation problem:

$$\mathcal{OP}_1^{\text{MMF}} : \max_{\boldsymbol{p}_1, \boldsymbol{p}_2} \min_{m=1,\ldots,M} \left\{ R_{1,m}(p_{1,m}, p_{2,m}), R_{2,m}(p_{1,m}, p_{2,m}) \right\}$$

$$\text{s.t. } \boldsymbol{0} \leq \boldsymbol{p}_1 \leq \boldsymbol{p}_2, \quad \sum_{m=1}^{M} p_{1,m} + p_{2,m} \leq P$$

where $\boldsymbol{p}_1 = \{p_{1,m}\}_{m=1}^{M}$ and $\boldsymbol{p}_2 = \{p_{2,m}\}_{m=1}^{M}$. However, $\mathcal{OP}_1^{\text{MMF}}$ is a nonconvex problem, as its objective is not concave. Its optimal solution has only been found in the special case $M = 1$ [24]–[26], i.e., a single channel, but unknown in the general case yet.

To address this problem, we first introduce auxiliary variables $\boldsymbol{q} = \{q_m\}_{m=1}^{M}$, where q_m represents the power budget for channel m with $p_{1,m} + p_{2,m} = q_m$. Suppose that the channel power budgets $\{q_m\}_{m=1}^{M}$ are given. Then, $\mathcal{OP}_1^{\text{MMF}}$ is decomposed into a group of subproblems for each channel m:

$$\mathcal{OP}_{2,m}^{\text{MMF}} : \max_{p_{1,m}, p_{2,m}} \min \left\{ R_{1,m}(p_{1,m}, p_{2,m}), R_{2,m}(p_{1,m}, p_{2,m}) \right\}$$

$$\text{s.t. } 0 \leq p_{1,m} \leq p_{2,m}, \quad p_{1,m} + p_{2,m} = q_m.$$

We first solve subproblem $\mathcal{OP}_{2,m}^{\text{MMF}}$ and show that its optimal solution is given in a closed form.

Proposition 1: Suppose that $\Gamma_{1,m} \geq \Gamma_{2,m}$. Then, the optimal solution to $\mathcal{OP}_{2,m}^{\text{MMF}}$ is given by $p_{1,m}^\star = \Lambda_m$ and

$p_{2,m}^\star = q_m - p_{1,m}^\star$, where $\Gamma_{l,m} \triangleq |h_{l,m}|^2 / \sigma_m^2$ and

$$\Lambda_m \triangleq \frac{-(\Gamma_{1,m} + \Gamma_{2,m}) + \sqrt{(\Gamma_{1,m} + \Gamma_{2,m})^2 + 4\Gamma_{1,m}\Gamma_{2,m}^2 q_m}}{2\Gamma_{1,m}\Gamma_{2,m}}.$$

Proof: From the constraint $p_{1,m} + p_{2,m} = q_m$, we have $p_{2,m} = q_m - p_{1,m}$, so $p_{1,m} \leq p_{2,m}$ is equivalent to $p_{1,m} \leq q_m/2$. Substituting $p_{2,m} = q_m - p_{1,m}$ into $R_{1,m}$ and $R_{2,m}$, we obtain

$$R_{1,m}(p_{1,m}) \triangleq B_c \log\left(1 + p_{1,m}\Gamma_{1,m}\right),$$

$$R_{2,m}(p_{1,m}) \triangleq B_c \log\left(\frac{q_m\Gamma_{2,m} + 1}{p_{1,m}\Gamma_{2,m} + 1}\right).$$

If $p_{1,m} \geq \Lambda_m$, then $R_{1,m}(p_{1,m}) \geq R_{2,m}(p_{1,m})$ and the objective of $\mathcal{OP}_{2,m}^{\text{MMF}}$ is $R_{2,m}(p_{1,m})$, which is decreasing in $p_{1,m}$. So the maximizer is the lower bound $p_{1,m} = \Lambda_m$. If $p_{1,m} \leq \Lambda_m$, then $R_{1,m}(p_{1,m}) \leq R_{2,m}(p_{1,m})$ and the objective of $\mathcal{OP}_{2,m}^{\text{MMF}}$ is $R_{1,m}(p_{1,m})$, which is increasing in $p_{1,m}$. So the maximizer is the upper bound $p_{1,m} = \Lambda_m$. Therefore, the optimal point is $p_{1,m}^\star = \Lambda_m$. Finally, it can be verified that $p_{1,m}^\star = \Lambda_m < q_m/2$. ∎

Remark 1: From Proposition 1, we obtain $R_{1,m}(p_{1,m}^\star, p_{2,m}^\star) = R_{2,m}(p_{1,m}^\star, p_{2,m}^\star) = f_m^{\text{MMF}\star}$, where

$$f_m^{\text{MMF}\star}$$
$$\triangleq B_c \log\left(\frac{\Gamma_{2,m} - \Gamma_{1,m} + \sqrt{(\Gamma_{1,m} + \Gamma_{2,m})^2 + 4\Gamma_{1,m}\Gamma_{2,m}^2 q_m}}{2\Gamma_{2,m}}\right),$$
$$(6)$$

i.e., UE$_{1,m}$ and UE$_{2,m}$ achieve the same rate at the optimal point. This indicates that, under the MMF criterion, the NOMA system will provide absolute fairness for two users on one channel.

To elaborate another important insight, we introduce the following definition.

Definition 1: A NOMA system is called *SIC-stable* if the optimal power allocation satisfies $p_{1,m} < p_{2,m}$ on each channel m.

Remark 2: In NOMA systems, SIC is performed according to the order of the CNRs of the users on one channel [9], [28], which is guaranteed by imposing an inverse order of the powers allocated to the users, i.e., $p_{1,m} \leq p_{2,m}$ on channel m. Specifically, UE$_{1,m}$ (the strong user with a higher CNR) first decodes the signal of UE$_{2,m}$ (the weak user with a lower CNR) and then subtracts it from the superposed signal. Therefore, from the SIC perspective, a difference between the signal strengths of UE$_{2,m}$ and UE$_{1,m}$ is necessary [30]. However, even with the power order constraint, the power optimization may lead to $p_{1,m} = p_{2,m}$, i.e., UE$_{1,m}$ and UE$_{2,m}$ have the same signal strength, which is the worst situation for SIC. In this case, SIC may fail or has a large error propagation and thus is unstable. Indeed, the authors in [31] pointed out that the power of the weak user must be strictly larger than that of the strong user, otherwise the users' outage probabilities will always be one. Definition 1 explicitly concretizes such a practical requirement in NOMA systems.

Lemma 1: The NOMA system is SIC-stable for $O\mathcal{P}_1^{MMF}$.

Proof: Given $\Gamma_{1,m} \geq \Gamma_{2,m}$, we have

$$\Lambda_m = \frac{2\Gamma_{2,m}q_m}{\left(\Gamma_{1,m} + \Gamma_{2,m}\right) + \sqrt{\left(\Gamma_{1,m} + \Gamma_{2,m}\right)^2 + 4\Gamma_{1,m}\Gamma_{2,m}^2 q_m}}$$

$$< \frac{\Gamma_{2,m}q_m}{\Gamma_{1,m} + \Gamma_{2,m}} \leq \frac{q_m}{2},$$

which indicates $p_{1,m}^\star < p_{2,m}^\star$ for each m. Therefore, the NOMA system is SIC-stable. ∎

Remark 3: According to Definition 1 and indicated by Lemma 1, the NOMA system is always SIC-stable under the MMF criterion, as in this case the optimal power allocation always satisfies $p_{1,m}^\star < p_{2,m}^\star$, $\forall m$. On the other hand, in the subsequent sections, we will show that a NOMA system is not always SIC-stable under different criteria and constraints.

To obtain the optimal power allocation for all channels, we shall optimize the power budget q_m for each channel m. According to $O\mathcal{P}_1^{MMF}$ and $O\mathcal{P}_{2,m}^{MMF}$, the corresponding power budget optimization problem is given by

$$O\mathcal{P}_3^{MMF} : \quad \max_{\boldsymbol{q}} \ \min_{m=1,\ldots,M} f_m^{MMF\star}(q_m)$$

$$\text{s.t.} \ \sum_{m=1}^{M} q_m \leq P, \quad \boldsymbol{q} \geq \boldsymbol{0}$$

where $f_m^{MMF\star}(q_m)$ is the optimal objective value of $O\mathcal{P}_{2,m}^{MMF}$ and given in (6).

Lemma 2: $f_m^{MMF\star}(q_m)$ *is a concave function.*

Proof: It can be verified that $\partial^2 f_m^{MMF\star}/\partial q_m^2 < 0$ and hence $f_m^{MMF\star}(q_m)$ is concave. ∎

From Lemma 2, $O\mathcal{P}_3^{MMF}$ is actually a convex problem, whose solution can be efficiently found via standard convex optimization tools, e.g., CVX. Nevertheless, we are able to analytically characterize the optimal solution to (8).

Theorem 1: The optimal solution to $O\mathcal{P}_3^{MMF}$ is given by

$$q_m^\star = \frac{\left(Z(\lambda)\Gamma_{2,m} + \Gamma_{1,m}\right)\left(Z(\lambda) - 1\right)}{\Gamma_{1,m}\Gamma_{2,m}}, \quad \forall m \quad (7)$$

where

$$Z(\lambda) \triangleq X + \sqrt{X^2 + \frac{B_c}{2\lambda \sum_{m=1}^{M} 1/\Gamma_{1,m}}},$$

$$X \triangleq \frac{\sum_{m=1}^{M} \left(\Gamma_{2,m} - \Gamma_{1,m}\right)/\left(\Gamma_{1,m}\Gamma_{2,m}\right)}{4\sum_{m=1}^{M} 1/\Gamma_{1,m}}$$

and λ is chosen such that $\sum_{m=1}^{M} q_m^\star = P$.

Proof: We first transform $O\mathcal{P}_3^{MMF}$ into

$$\max_{\boldsymbol{q},t} \ t$$

$$\text{s.t.} \ \boldsymbol{q} \geq \boldsymbol{0}, \ \sum_{m=1}^{M} q_m \leq P, \ f_m^{MMF\star}(q_m) \geq t, \ \forall m \quad (8)$$

where $f_m^{MMF\star}(q_m) \geq t$ is equivalent to $q_m \geq (a^t\Gamma_{2,m} + \Gamma_{1,m})(a^t - 1)/(\Gamma_{1,m}\Gamma_{2,m})$ with $a = 2^{1/B_c}$. Then, the Lagrange of (8) can be written as

$$L = t + \sum_{m=1}^{M} \mu_m \left[q_m - \frac{\left(a^t\Gamma_{2,m} + \Gamma_{1,m}\right)\left(a^t - 1\right)}{\Gamma_{1,m}\Gamma_{2,m}}\right]$$

$$- \lambda\left(\sum_{m=1}^{M} q_m - P\right)$$

where $\{\mu_m\}_{m=1}^{M}$ and λ are the Lagrange multipliers. Since (8) is a convex optimization problem, its optimal solution is characterized by the following Karush-Kuhn-Tucker (KKT) conditions:

$$\frac{\partial L}{\partial q_m} = \mu_m - \lambda = 0, \quad (9)$$

$$\frac{\partial L}{\partial t} = 1 - a^{2t}\sum_{m=1}^{M} \frac{2\mu_m \ln a}{\Gamma_{1,m}} + a^t \sum_{m=1}^{M} \frac{\mu_m \ln a}{\Gamma_{1,m}}$$

$$- a^t \sum_{m=1}^{M} \frac{\mu_m \ln a}{\Gamma_{2,m}} = 0, \quad (10)$$

$$\mu_m\left(q_m - \frac{\left(a^t\Gamma_{2,m} + \Gamma_{1,m}\right)\left(a^t - 1\right)}{\Gamma_{1,m}\Gamma_{2,m}}\right) = 0,$$

$$\lambda\left(\sum_{m=1}^{M} q_m - P\right) = 0. \quad (11)$$

It follows from (9) and (10) that $\mu_m = \lambda \neq 0$. Then, (10) is equivalent to

$$C_1 a^{2t} - C_2 a^t - 1 = 0$$

where $C_1 = 2\lambda \ln a \sum_{m=1}^{M} 1/\Gamma_{1,m}$ and $C_2 = \lambda \ln a \sum_{m=1}^{M} (\Gamma_{2,m} - \Gamma_{1,m})/(\Gamma_{1,m}\Gamma_{2,m})$. By solving this quadratic equation, we obtain

$$a^t = \frac{C_2}{2C_1} + \sqrt{\left(\frac{C_2}{2C_1}\right)^2 + \frac{1}{C_1}}$$

$$= X + \sqrt{X^2 + \frac{B_c}{2\lambda \sum_{m=1}^{M} 1/\Gamma_{1,m}}} = Z(\lambda).$$

Finally, from (11), we have

$$q_m = \frac{\left(a^t\Gamma_{2,m} + \Gamma_{1,m}\right)\left(a^t - 1\right)}{\Gamma_{1,m}\Gamma_{2,m}}$$

$$= \frac{\left(Z(\lambda)\Gamma_{2,m} + \Gamma_{1,m}\right)\left(Z(\lambda) - 1\right)}{\Gamma_{1,m}\Gamma_{2,m}} \geq 0 \quad (12)$$

which completes the proof. ∎

Corollary 1: Under the MMF criterion, the optimal power allocation achieves the absolute fairness for all the users on all channels, i.e., $R_{1,m} = R_{2,m} = r$, $m = 1, \ldots, M$, for some $r \geq 0$.

Proof: Given the optimal q_m in (12), it can be verified that $f_m^{MMF\star}(q_m) = t$ for $m = 1, \ldots, M$. ∎

The optimal power allocation under the MMF criterion is fully characterized by Theorem 1 and Proposition 1. It follows from (7) that $q_m^\star$ is monotonically decreasing in λ, so the optimal λ satisfying $\sum_{m=1}^{M} q_m^\star = P$ can be efficiently found via a simple bisection method.

IV. OPTIMAL POWER ALLOCATION FOR SUM RATE

In this section, we seek the optimal power allocation for maximizing the weighted sum rate (SR) or maximizing the SR with QoS constraints.

A. Weighted SR Maximization (SR1)

According to (2), with given channel assignment, the problem of maximizing the weighted sum rate is equivalent to the following power allocation problem:

$$OP_1^{SR1}: \quad \max_{\boldsymbol{p}_1, \boldsymbol{p}_2} \sum_{m=1}^{M} g(p_{1,m}, p_{2,m})$$
$$\text{s.t. } \boldsymbol{0} \leq \boldsymbol{p}_1 \leq \boldsymbol{p}_2, \quad \sum_{m=1}^{M} (p_{1,m} + p_{2,m}) \leq P$$

where $g(p_{1,m}, p_{2,m}) \triangleq W_{1,m} R_{1,m}(p_{1,m}, p_{2,m}) + W_{2,m} R_{2,m}(p_{1,m}, p_{2,m})$. As the objective of OP_1^{SR1} is not a concave function, OP_1^{SR1} is also a nonconvex problem. Although this problem has been studied in [19], [20], and [22] the solution is either suboptimal or needs exhaustive search.

Introduce auxiliary variables $\boldsymbol{q} = \{q_m\}_{m=1}^{M}$ that represent the power budgets on each channel m with $p_{1,m} + p_{2,m} = q_m$. Then, OP_1^{SR1} is decomposed into a group of subproblems for each channel m:

$$OP_{2,m}^{SR1}: \quad \max_{p_{1,m}, p_{2,m}} g(p_{1,m}, p_{2,m})$$
$$\text{s.t. } 0 \leq p_{1,m} \leq p_{2,m}, \quad p_{1,m} + p_{2,m} = q_m.$$

We first solve the subproblem $OP_{2,m}^{SR1}$ for each channel m. Note that $OP_{2,m}^{SR1}$ is still a nonconvex problem due to the interference between $UE_{1,m}$ and $UE_{2,m}$. Nevertheless, its optimal solution can be characterized in a closed form.

Proposition 2: Suppose that $\Gamma_{1,m} \geq \Gamma_{2,m}$, $1 < W_{2,m}/W_{1,m} < \Gamma_{1,m}/\Gamma_{2,m}$ *and* $q_m > 2\Omega_m$, *with*

$$\Omega_m \triangleq \frac{W_{2,m}\Gamma_{2,m} - W_{1,m}\Gamma_{1,m}}{\Gamma_{1,m}\Gamma_{2,m}(W_{1,m} - W_{2,m})}.$$

Then, the optimal solution to $OP_{2,m}^{SR1}$ *is given by* $p_{1,m}^{\star} = \Omega_m$ *and* $p_{2,m}^{\star} = q_m - p_{1,m}^{\star}$.

Proof: Since $p_{2,m} = q_m - p_{1,m}$, $p_{1,m} \leq p_{2,m}$ is equal to $p_{1,m} \leq q_m/2$, and the objective becomes

$$F(p_{1,m}) \triangleq W_{1,m} B_c \log(1 + p_{1,m}\Gamma_{1,m}) + W_{2,m} B_c \log\left(\frac{q_m\Gamma_{2,m} + 1}{p_{1,m}\Gamma_{2,m} + 1}\right).$$

By setting the derivative of F to zero, we have

$$\frac{dF}{dp_{1,m}} = \frac{W_{1,m} B_c}{1/\Gamma_{1,m} + p_{1,m}} - \frac{W_{2,m} B_c}{1/\Gamma_{2,m} + p_{1,m}} = 0,$$

leading to a unique root $p_{1,m} = \Omega_m$, which satisfies the constraint $p_{1,m} \leq q_m/2$ since $\Omega_m < q_m/2$. Given $\Gamma_{1,m} \geq \Gamma_{2,m}$ and $1 < W_{2,m}/W_{1,m} < \Gamma_{1,m}/\Gamma_{2,m}$, it follows that

$$\frac{\partial^2 F}{\partial p_{1,m}^2} = \frac{B_c W_{2,m}}{(1/\Gamma_{2,m} + \Omega_m)^2} - \frac{B_c W_{1,m}}{(1/\Gamma_{1,m} + \Omega_m)^2}$$
$$= \frac{B_c \Gamma_{1,m}^2 \Gamma_{2,m}^2 (W_{1,m} - W_{2,m})^2}{(\Gamma_{2,m} - \Gamma_{1,m})^2}\left(\frac{1}{W_{2,m}} - \frac{1}{W_{1,m}}\right) < 0,$$

indicating that Ω_m is a maximizer. ∎

Remark 4: In Proposition 2, the conditions $1 < W_{2,m}/W_{1,m} < \Gamma_{1,m}/\Gamma_{2,m}$ and $q_m > 2\Omega_m$ are both to avoid a failure of SIC. Indeed, if $W_{2,m}/W_{1,m} < 1$ or $W_{2,m}/W_{1,m} > \Gamma_{1,m}/\Gamma_{2,m}$, the solution to $OP_{2,m}^{SR1}$ is $p_{1,m}^{\star} = p_{2,m}^{\star} = q_m/2$, i.e., the NOMA system is unstable

according to Definition 1. SIC may also fail on channel m if $q_m \leq 2\Omega_m$, which will lead to $p_{1,m}^{\star} = p_{2,m}^{\star} = q_m/2$ too. Therefore, the NOMA system is SIC-stable on channel m if and only if $1 < W_{2,m}/W_{1,m} < \Gamma_{1,m}/\Gamma_{2,m}$ and $q_m > 2\Omega_m$. For all channels, we have the following result.

Corollary 2: For OP_1^{SR1}, *the NOMA system is SIC-stable only if* $P > 2\sum_{m=1}^{M} \Omega_m$ *and* $1 < W_{2,m}/W_{1,m} < \Gamma_{1,m}/\Gamma_{2,m}$ *for* $m = 1, \ldots, M$.

Next, we further optimize the power budget q_m for each channel m. To guarantee that the NOMA system is SIC-stable, it is reasonable to assume that $q_m \geq \Theta_m > 2\Omega_m$ and $P \geq \sum_{m=1}^{M} \Theta_m$ for some positive Θ_m. Then, from OP_1^{SR1} and $OP_{2,m}^{SR1}$, the corresponding power budget optimization problem is given by

$$OP_3^{SR1}: \quad \max_{\boldsymbol{q}} \sum_{m=1}^{M} f_m^{SR1\star}(q_m)$$
$$\text{s.t. } \sum_{m=1}^{M} q_m \leq P, \quad q_m \geq \Theta_m, \ \forall m$$

where $f_m^{SR1\star}(q_m)$ is the optimal objective value of $OP_{2,m}^{SR1}$ and given by

$$f_m^{SR1\star}(q_m) = W_{1,m} B_c \log(1 + \Omega_m \Gamma_{1,m}) + W_{2,m} B_c \log\left(\frac{q_m\Gamma_{2,m} + 1}{\Omega_m\Gamma_{2,m} + 1}\right). \quad (13)$$

It is easily seen that $f_m^{SR1\star}(q_m)$ is a concave function, so OP_3^{SR1} is a convex problem, whose solution is provided in the following result.

Theorem 2: The optimal solution to OP_3^{SR1} *is given by*

$$q_m^{\star} = \left[\frac{W_{2,m} B_c}{\lambda} - \frac{1}{\Gamma_{2,m}}\right]_{\Theta_m}^{\infty} \quad (14)$$

where λ *is chosen such that* $\sum_{m=1}^{M} q_m^{\star} = P$.

Proof: The solution of OP_3^{SR1} is given by the well-known waterfilling form. ∎

Consequently, the optimal power allocation for the sum rate maximization with weights in NOMA systems is jointly characterized by Theorem 2 and Proposition 2 under the SIC-stability.

B. SR Maximization With QoS (SR2)

Now, we consider maximizing the SR with QoS constraints. According to (3), in this case the power allocation problem is given by

$$OP_1^{SR2}: \quad \max_{\boldsymbol{p}_1, \boldsymbol{p}_2} \sum_{m=1}^{M} (R_{1,m}(p_{1,m}, p_{2,m}) + R_{2,m}(p_{1,m}, p_{2,m}))$$
$$\text{s.t. } \boldsymbol{0} \leq \boldsymbol{p}_1 \leq \boldsymbol{p}_2, \quad \sum_{m=1}^{M} (p_{1,m} + p_{2,m}) \leq P,$$
$$R_{n,m} \geq R_{n,m}^{\min}, \quad n = 1, 2, \ m = 1, \ldots, M.$$

As a special case of OP_1^{SR2}, [23] studied the power allocation for one channel. Thus, OP_1^{SR2} is still an open problem and its optimal solution is unknown yet.

We use the similar method to address OP_1^{SR2}. By introducing the power budget q_m on each channel m, OP_1^{SR2} decomposes

into the following subproblems for each channel m:

$$OP_{2,m}^{SR2}: \quad \max_{p_{1,m}, p_{2,m}} R_{1,m}(p_{1,m}, p_{2,m}) + R_{2,m}(p_{1,m}, p_{2,m})$$
$$\text{s.t. } 0 \leq p_{1,m} \leq p_{2,m}, \quad p_{1,m} + p_{2,m} = q_m,$$
$$R_{1,m} \geq R_{1,m}^{\min}, \quad R_{2,m} \geq R_{2,m}^{\min}.$$

The optimal solution to $OP_{2,m}^{SR2}$, although it is nonconvex, is provided in the following result.

Proposition 3: Suppose that $\Gamma_{1,m} \geq \Gamma_{2,m}$, $A_{2,m} \geq 2$, and $q_m \geq \Upsilon_m$, with

$$A_{n,m} = 2^{\frac{R_{n,m}^{\min}}{B_c}}, \quad \Upsilon_m \triangleq \frac{A_{2,m}(A_{1,m} - 1)}{\Gamma_{1,m}} + \frac{A_{2,m} - 1}{\Gamma_{2,m}},$$
$$\Xi_m \triangleq \frac{\Gamma_{2,m} q_m - A_{2,m} + 1}{A_{2,m} \Gamma_{2,m}}.$$

Then, the optimal solution to $OP_{2,m}^{SR2}$ is given by $p_{1,m}^\star = \Xi_m$ and $p_{2,m}^\star = q_m - p_{1,m}^\star$.

Proof: Since $p_{2,m} = q_m - p_{1,m}$, $p_{1,m} \leq p_{2,m}$ is equal to $p_{1,m} \leq q_m/2$ and the objective becomes

$$T(p_{1,m}) \triangleq B_c \log\left(1 + p_{1,m}\Gamma_{1,m}\right) + B_c \log\left(\frac{q_m\Gamma_{2,m} + 1}{p_{1,m}\Gamma_{2,m} + 1}\right).$$

Given $\Gamma_{1,m} \geq \Gamma_{2,m}$, we take the derivative of $T(p_{1,m})$ and have

$$\frac{dT}{dp_{1,m}} = \frac{B_c}{1/\Gamma_{1,m} + p_{1,m}} - \frac{B_c}{1/\Gamma_{2,m} + p_{1,m}} \geq 0,$$

implying that $T(p_{1,m})$ is monotonically nondecreasing, so the maximum is achieved at the upper bound of $p_{1,m}$. From $R_{1,m} \geq R_{1,m}^{\min}$ and $R_{2,m} \geq R_{2,m}^{\min}$, we obtain

$$\frac{A_{1,m} - 1}{\Gamma_{1,m}} \leq p_{1,m} \leq \Xi_m$$

which holds if and only if $(A_{1,m} - 1)/\Gamma_{1,m} \leq \Xi_m$, i.e., $q_m \geq \Upsilon_m$. Finally, since $A_{2,m} \geq 2$, $p_{1,m} = \Xi_m < q_m/2$ holds. Thus the optimal solution is $p_{1,m}^\star = \Xi_m$. ∎

Remark 5: Similarly, in Proposition 3, the conditions $A_{2,m} \geq 2$ and $q_m \geq \Upsilon_m$ are to guarantee the SIC-stability. Indeed, if $A_{2,m} < 2$, then $\Xi_m > q_m/2$ and the optimal solution will be $p_{1,m}^\star = p_{2,m}^\star = q_m/2$, which may leads a failure of SIC. At the same time, SIC may also fail on channel m if $q_m < \Upsilon_m$, which will lead to $p_{1,m}^\star = p_{2,m}^\star = q_m/2$ as well. Therefore, the NOMA system is SIC-stable on channel m if and only if $A_{2,m} \geq 2$ and $q_m \geq \Upsilon_m$. For all the channels, we have the following result.

Corollary 3: For OP_1^{SR2}, the NOMA system is SIC-stable only if $P \geq \sum_{m=1}^{M} \Upsilon_m$ and $A_{2,m} \geq 2$ for $m = 1, \ldots, M$.

Remark 6: According to Proposition 3, if the NOMA system is SIC-stable, the optimal solution will be $p_{1,m}^\star = \Xi_m$ and $p_{2,m}^\star = q_m - p_{1,m}^\star$. Hence, we have $R_{2,m}(p_{1,m}^\star, p_{2,m}^\star) = R_{2,m}^{\min}$, implying that the user with a lower CNR (i.e., $UE_{2,m}$) receives the power to meet its QoS requirement exactly, while the remaining power is used to maximize the rate of the user with a higher CNR (i.e., $UE_{1,m}$).

Then, we focus on optimizing the power budget q_m for each channel. Similarly, to guarantee the NOMA system is SIC-stable, we assume that $q_m \geq \Upsilon_m$ and $P \geq \sum_{m=1}^{M} \Upsilon_m$.

According to OP_1^{SR2} and $OP_{2,m}^{SR2}$, the corresponding power budget optimization problem is as follows

$$OP_3^{SR2}: \quad \max_{\boldsymbol{q}} \sum_{m=1}^{M} f_m^{SR2\star}(q_m)$$
$$\text{s.t. } \sum_{m=1}^{M} q_m \leq P, \quad q_m \geq \Upsilon_m, \quad \forall m$$

where $f_m^{SR2\star}(q_m)$ is the optimal objective value of $OP_{2,m}^{SR2}$ and given by

$$f_m^{SR2\star}(q_m) = w(q_m) + R_{2,m}^{\min} \qquad (15)$$

where $w(q_m) = B_c \log \frac{(A_{2,m}\Gamma_{2,m} - A_{2,m}\Gamma_{1,m} + \Gamma_{1,m}\Gamma_{2,m}q_m + \Gamma_{1,m})}{A_{2,m}\Gamma_{2,m}}$. Since $f_m^{SR2\star}(q_m)$ is a concave function, OP_3^{SR2} is a convex problem, whose solution is also given in a waterfilling form.

Theorem 3: The optimal solution to OP_3^{SR2} is given by

$$q_m^\star = \left[\frac{B_c}{\lambda} - \frac{A_{2,m}}{\Gamma_{1,m}} + \frac{A_{2,m}}{\Gamma_{2,m}} - \frac{1}{\Gamma_{2,m}} \right]_{\Upsilon_m}^{\infty}$$

where λ is chosen such that $\sum_{m=1}^{M} q_m^\star = P$.

Proof: The proof is simple and thus omitted. ∎

Therefore, the optimal power allocation for the SR maximization with QoS constraints in NOMA systems is jointly characterized by Proposition 3 and Theorem 3. Note that, unlike the MMF criterion, for the SR maximization with weights or QoS constraints, NOMA systems are not always SIC-stable but have to satisfy some conditions on the weights, power budgets, and QoS thresholds as indicated in this section.

V. OPTIMAL POWER ALLOCATION FOR ENERGY EFFICIENCY

In this section, we investigate the optimal power allocation for maximizing the energy efficiency (EE) of the NOMA systems with weights or QoS constraints.

A. EE Maximization With Weights (EE1)

According to (4), with given channel assignment, the problem of maximizing the EE with weights is equivalent to the following power allocation problem:

$$OP_1^{EE1}: \quad \max_{\boldsymbol{p_1}, \boldsymbol{p_2}} \frac{\sum_{m=1}^{M} g(p_{1,m}, p_{2,m})}{P_T + \sum_{m=1}^{M} (p_{1,m} + p_{2,m})}$$
$$\text{s.t. } \boldsymbol{0} \leq \boldsymbol{p_1} \leq \boldsymbol{p_2}, \quad \sum_{m=1}^{M} (p_{1,m} + p_{2,m}) \leq P$$

where $g(p_{1,m}, p_{2,m}) \triangleq W_{1,m} R_{1,m}(p_{1,m}, p_{2,m}) + W_{2,m} R_{2,m}(p_{1,m}, p_{2,m})$. The difficulties in solving OP_1^{EE1} lie in its nonconvex and fractional objective. In the literature, only [21], [27] investigated this problem, whereas [27] only found the optimal solution in the special case $M = 1$, i.e., a single channel, and [21] obtained a suboptimal power allocation solution. In the following, we will show that this problem can also be optimally solved.

We use the similar trick to address this problem, i.e., introducing the auxiliary variables $\{q_m\}_{m=1}^{M}$ with $p_{1,m} + p_{2,m} = q_m$

for each channel m. Then, $O\mathcal{P}_1^{\text{EE1}}$ is decomposed into the following subproblems for each channel m:

$$O\mathcal{P}_{2,m}^{\text{EE1}} : \quad \max_{p_{1,m}, p_{2,m}} \frac{g(p_{1,m}, p_{2,m})}{P_T + \sum_{k=1}^{M} q_k}$$
$$\text{s.t. } 0 \leq p_{1,m} \leq p_{2,m}, \quad p_{1,m} + p_{2,m} = q_m$$

whose optimal solution is provided in the following.

Proposition 4: Suppose that $\Gamma_{1,m} \geq \Gamma_{2,m}$, $1 < W_{2,m}/W_{1,m} < \Gamma_{1,m}/\Gamma_{2,m}$ and $q_m > 2\Omega_m$, with

$$\Omega_m \triangleq \frac{W_{2,m}\Gamma_{2,m} - W_{1,m}\Gamma_{1,m}}{\Gamma_{1,m}\Gamma_{2,m}\left(W_{1,m} - W_{2,m}\right)}.$$

Then the optimal solution to $O\mathcal{P}_{2,m}^{\text{EE1}}$ is same with $O\mathcal{P}_{2,m}^{\text{SR1}}$, i.e., $p_{1,m}^\star = \Omega_m$ and $p_{2,m}^\star = q_m - p_{1,m}^\star$.

Remark 7: It is not difficult to see that with given channel power budgets $\{q_m\}_{m=1}^{M}$, $O\mathcal{P}_{2,m}^{\text{EE1}}$ is actually equivalent to $O\mathcal{P}_{2,m}^{\text{SR1}}$, so they have the same optimal solution. Therefore, we obtain the same SIC-stability conditions for $O\mathcal{P}_{2,m}^{\text{EE1}}$: the NOMA system is SIC-stable on channel m if and only if $q_m > 2\Omega_m$ and $1 < W_{2,m}/W_{1,m} < \Gamma_{1,m}/\Gamma_{2,m}$, and is SIC-stable on all channels only if $P > 2\sum_{m=1}^{M} \Omega_m$ and $1 < W_{2,m}/W_{1,m} < \Gamma_{1,m}/\Gamma_{2,m}$ for $m = 1, \ldots, M$.

Then, we concentrate on searching the optimal power budget q_m for each channel. Similarly, to guarantee the NOMA system is SIC-stable, it is assumed that $q_m \geq \Theta_m > 2\Omega_m$ and $P \geq \sum_{m=1}^{M} \Theta_m$ for some positive Θ_m. According to Proposition 4, $O\mathcal{P}_1^{\text{EE1}}$, and $O\mathcal{P}_{2,m}^{\text{EE1}}$, the power budget optimization problem is formulated as

$$O\mathcal{P}_3^{\text{EE1}} : \quad \max_{\boldsymbol{q}} \eta(\boldsymbol{q}) \triangleq \frac{\sum_{m=1}^{M} f_m^{\text{SR1}\star}(q_m)}{P_T + \sum_{m=1}^{M} q_m}$$
$$\text{s.t. } \sum_{m=1}^{M} q_m \leq P, \quad q_m \geq \Theta_m, \; \forall m$$

where $f_m^{\text{SR1}\star}(q_m)$ is the optimal value of $O\mathcal{P}_{2,m}^{\text{SR1}}$ and given in (13). Although $f_m^{\text{SR1}\star}(q_m)$ is a concave function, $O\mathcal{P}_3^{\text{EE1}}$ is nonconvex due to the fraction form. To solve it, we introduce the following objective function:

$$H(\boldsymbol{q}, \alpha) \triangleq \sum_{m=1}^{M} f_m^{\text{SR1}\star}(q_m) - \alpha\left(P_T + \sum_{m=1}^{M} q_m\right)$$
$$= \sum_{m=1}^{M} \left(\tilde{R}_{1,m} + W_{2,m}B_c \log\left(\frac{q_m\Gamma_{2,m} + 1}{\Omega_m\Gamma_{2,m} + 1}\right)\right)$$
$$- \alpha\left(P_T + \sum_{m=1}^{M} q_m\right)$$

where $\tilde{R}_{1,m} \triangleq W_{1,m}B_c \log\left(1 + \Omega_m\Gamma_{1,m}\right)$ and α is a positive parameter. Then, we consider the following convex problem with given α:

$$O\mathcal{P}_4^{\text{EE1}} : \quad \max_{\boldsymbol{q}} H(\boldsymbol{q}, \alpha)$$
$$\text{s.t. } \sum_{m=1}^{M} q_m \leq P, \quad q_m \geq \Theta_m, \; \forall m.$$

The relation between $O\mathcal{P}_3^{\text{EE1}}$ and $O\mathcal{P}_4^{\text{EE1}}$ is given by the following lemma.

Lemma 3 ([32, pp. 493-494]): Let $H^\star(\alpha)$ be the optimal objective value of $O\mathcal{P}_4^{\text{EE1}}$ and $\boldsymbol{q}^\star(\alpha)$ be the optimal solution of $O\mathcal{P}_4^{\text{EE1}}$. Then, $\boldsymbol{q}^\star(\alpha)$ is the optimal solution to $O\mathcal{P}_3^{\text{EE1}}$ if and only if $H^\star(\alpha) = 0$.

Lemma 3 indicates the optimal solution to $O\mathcal{P}_3^{\text{EE1}}$ can be found by solving $O\mathcal{P}_4^{\text{EE1}}$ parameterized by α and then updating α until $H^\star(\alpha) = 0$. For this purpose, we first solve $O\mathcal{P}_4^{\text{EE1}}$ with given α, whose solution is provided in the following result.

Theorem 4: The optimal solution to $O\mathcal{P}_4^{\text{EE1}}$ is

$$q_m^\star = \left[\frac{W_{2,m}B_c}{\alpha + \lambda} - \frac{1}{\Gamma_{2,m}}\right]_{\Theta_m}^{\infty} \tag{16}$$

where λ is chosen such that $\sum_{m=1}^{M} q_m^\star = P$.

Proof: The solution is obtained by exploiting the KKT conditions of $O\mathcal{P}_4^{\text{EE1}}$. ∎

After the optimal solution to $O\mathcal{P}_4^{\text{EE1}}$ is obtained, we shall find an α such that $H^\star(\alpha) = 0$. This can be achieved by Algorithm 1, which is guaranteed to converge to the desirable α [32]. Thereby, the optimal power allocation for the EE maximization with weights in NOMA systems is provided by Algorithm 1, Proposition 4 and Theorem 4.

Algorithm 1 Channel Power Budget Optimization for EE

1: **Initialization**: set $\alpha_{ini} = 0$, $H_{ini}^\star = \infty$ and precision $\delta > 0$.
2: **While** $|H^\star(\alpha)| > \delta$ **do**
3:     Find the optimal $\boldsymbol{q}^\star$ according to Theorem 4;
4:     Calculate $H^\star(\alpha)$;
5:     Update $\alpha = \eta(\boldsymbol{q}^\star)$;
6: **Return** α and $\boldsymbol{q}^\star$.

B. EE Maximization With QoS (EE2)

In this subsection, we consider maximizing the EE with QoS constraints. According to (5), in this case the power allocation problem is given by

$$O\mathcal{P}_1^{\text{EE2}} : \quad \max_{\boldsymbol{p}_1, \boldsymbol{p}_2} \frac{\sum_{m=1}^{M}\left(R_{1,m}\left(p_{1,m}, p_{2,m}\right) + R_{2,m}\left(p_{1,m}, p_{2,m}\right)\right)}{P_T + \sum_{m=1}^{M}\left(p_{1,m} + p_{2,m}\right)}$$
$$\text{s.t. } \boldsymbol{0} \leq \boldsymbol{p}_1 \leq \boldsymbol{p}_2, \quad \sum_{m=1}^{M}\left(p_{1,m} + p_{2,m}\right) \leq P,$$
$$R_{n,m} \geq R_{n,m}^{\min}, \quad n = 1, 2, \; m = 1, \ldots, M.$$

Similarly, $O\mathcal{P}_1^{\text{EE2}}$ is a nonconvex fractional optimization problem. In the literature, the EE maximization with QoS constraints has only been studied in [27], but the optimal solution was only found for one channel, i.e., $M = 1$. Therefore, $O\mathcal{P}_1^{\text{EE2}}$ is an open problem and its solution in the general case is still unknown.

To solve $O\mathcal{P}_1^{\text{EE2}}$, we also adopt $\{q_m\}_{m=1}^{M}$ with $p_{1,m} + p_{2,m} = q_m$ and decompose $O\mathcal{P}_1^{\text{EE2}}$ into a group of subproblems for each channel m:

$$O\mathcal{P}_{2,m}^{\text{EE2}} : \quad \max_{p_{1,m}, p_{2,m}} \frac{R_{1,m}\left(p_{1,m}, p_{2,m}\right) + R_{2,m}\left(p_{1,m}, p_{2,m}\right)}{P_T + \sum_{k=1}^{M} q_k}$$
$$\text{s.t. } 0 \leq p_{1,m} \leq p_{2,m}, \quad p_{1,m} + p_{2,m} = q_m,$$
$$R_{1,m} \geq R_{1,m}^{\min}, \quad R_{2,m} \geq R_{2,m}^{\min}$$

whose solution is given in the following closed form.

Proposition 5: Suppose that $\Gamma_{1,m} \geq \Gamma_{2,m}$, $A_{2,m} \geq 2$, and $q_m \geq \Upsilon_m$, with

$$A_{l,m} = 2^{\frac{R_{l,m}^{\min}}{B_c}}, \quad \Upsilon_m \triangleq \frac{A_{2,m}(A_{1,m} - 1)}{\Gamma_{1,m}} + \frac{A_{2,m} - 1}{\Gamma_{2,m}},$$

$$\Xi_m \triangleq \frac{\Gamma_{2,m} q_m - A_{2,m} + 1}{A_{2,m} \Gamma_{2,m}}.$$

Then, the optimal solution to $O\mathcal{P}_{2,m}^{EE2}$ *is* $p_{1,m}^\star = \Xi_m$ *and* $p_{2,m}^\star = q_m - p_{1,m}^\star$.

Remark 8: It is not surprising that the solutions of $O\mathcal{P}_{2,m}^{EE2}$ and $O\mathcal{P}_{2,m}^{SR2}$ coincide, since $O\mathcal{P}_{2,m}^{EE2}$ is equivalent to $O\mathcal{P}_{2,m}^{SR2}$ with given $\{q_m\}_{m=1}^M$. Therefore, the same SIC-stability conditions hold for $O\mathcal{P}_{2,m}^{EE2}$ and $O\mathcal{P}_1^{EE2}$: the NOMA system is SIC-stable on channel m if and only if $q_m \geq \Upsilon_m$ and $A_{2,m} \geq 2$, and is SIC-stable on all channels only if $P \geq \sum_{m=1}^M \Upsilon_m$ and $A_{2,m} \geq 2$ for $m = 1, \ldots, M$.

Next, we optimize the channel power budget q_m for each channel. First, we assume that $q_m \geq \Upsilon_m$ and $P \geq \sum_{m=1}^M \Upsilon_m$ to guarantee the SIC-stability. Then, according to Proposition 5, $O\mathcal{P}_1^{EE2}$, and $O\mathcal{P}_{2,m}^{EE2}$, the power budget optimization problem is given by

$$O\mathcal{P}_3^{EE2}: \quad \max_{\boldsymbol{q}} \ \eta(\boldsymbol{q}) \triangleq \frac{\sum_{m=1}^M f_m^{SR2\star}(q_m)}{P_T + \sum_{m=1}^M q_m}$$

$$\text{s.t.} \ \sum_{m=1}^M q_m \leq P, \quad q_m \geq \Upsilon_m, \ \forall m$$

where $f_m^{SR2\star}(q_m)$ is the optimal value of $O\mathcal{P}_{2,m}^{SR2}$ and given in (15). To solve $O\mathcal{P}_3^{EE2}$, we also introduce a parameterized objective function

$$Q(\boldsymbol{q}, \alpha) \triangleq \sum_{m=1}^M f_m^{SR2\star}(q_m) - \alpha \left(P_T + \sum_{m=1}^M q_m \right)$$

$$= \sum_{m=1}^M \left(w(q_m) + R_{2,m}^{\min} \right) - \alpha \left(P_T + \sum_{m=1}^M q_m \right)$$

where $w(q_m) = B_c \log \frac{(A_{2,m}\Gamma_{2,m} - A_{2,m}\Gamma_{1,m} + \Gamma_{1,m}\Gamma_{2,m}q_m + \Gamma_{1,m})}{A_{2,m}\Gamma_{2,m}}$, α is a positive parameter, and formulate the following problem with given α:

$$O\mathcal{P}_4^{EE2}: \quad \max_{\boldsymbol{q}} \ Q(\boldsymbol{q}, \alpha)$$

$$\text{s.t.} \ \sum_{m=1}^M q_m \leq P, \quad q_m \geq \Upsilon_m, \ \forall m.$$

Then, according to Lemma 3, the optimal solution to $O\mathcal{P}_3^{EE2}$ can be found by solving $O\mathcal{P}_4^{EE2}$ for a given α and then updating α until the optimal objective value of $O\mathcal{P}_4^{EE2}$, denoted by $Q^\star(\alpha)$, satisfies $Q^\star(\alpha) = 0$. Therefore, we first solve $O\mathcal{P}_4^{EE2}$, which is a convex problem since $Q(\boldsymbol{q}, \alpha)$ is concave in $\boldsymbol{q}$. In particular, the optimal solution to $O\mathcal{P}_4^{EE2}$ is provided as follows.

Theorem 5: The optimal solution to $O\mathcal{P}_4^{EE2}$ is

$$q_m^\star = \left[\frac{W_{1,m} B_c}{\lambda + \alpha} - \frac{A_{2,m}}{\Gamma_{1,m}} + \frac{A_{2,m}}{\Gamma_{2,m}} - \frac{1}{\Gamma_{2,m}} \right]_{\Upsilon_m}^\infty$$

where λ is chosen such that $\sum_{m=1}^M q_m^\star = P$.

Proof: The solution is obtained by exploiting the KKT conditions of $O\mathcal{P}_4^{EE2}$. ∎

After obtaining the optimal solution to $O\mathcal{P}_4^{GEE2}$, we shall find an α such that $Q^\star(\alpha) = 0$. This can also be achieved by Algorithm 1, where Theorem 4 and $H^\star(\alpha)$ are replaced by Theorem 5 and $Q^\star(\alpha)$, respectively. Consequently, the optimal power allocation for the EE maximization with QoS constraints in NOMA systems is obtained by using Theorem 5, Proposition 5, and Algorithm 1.

VI. CHANNEL ASSIGNMENT

In the previous sections, we have found the optimal power allocation with given channel assignment under various performance criteria for NOMA systems. Specifically, we first achieve the optimal power allocation of the users on each channel, which can be expressed as functions of the power budget of each channel. Then, we further optimized the power budgets of all channels and thus obtained the optimal multichannel power allocation, which can be characterized in a closed or semi-closed form. In this section, we consider the joint optimization of power allocation and channel assignment in NOMA systems. Unfortunately, such a joint optimization problem has been shown to be NP-hard [18]. Hence, finding the jointly optimal solution generally requires exhaustive search [17], which results in prohibitive computational complexity. Therefore, in practice suboptimal but efficient joint optimization methods are more preferred [17], [20]–[22].

Enlightened by the optimal multichannel power allocation obtained above, in this paper we propose a low-complexity method to jointly optimize the power allocation and channel assignment in NOMA systems. Specifically, we incorporate the dynamic matching algorithm [33], which is an efficient method to deal with assignment problems, with our optimal power allocation, and iteratively exploit them to refine the solution.

To describe the dynamic matching between the users and the channels, we consider channel assignment as a two-sided matching problem between the set of N users and the set of M channels, where $N = 2M$ since each channel is shared by two users. Denote channel m by C_m. We say UE_n and C_m are matched with each other if UE_n is assigned on C_m. Moreover, denote $PF(UE_n)$ for $n = 1, 2, \cdots N$ and $PF(C_m)$ for $m = 1, 2, \cdots M$ to be the preference lists of the users and channels, respectively. We say UE_n prefers C_i to C_j if UE_n has a higher channel gain on C_i than on C_j, and it can be expressed as

$$C_i(n) \succ C_j(n).$$

In addition, we say that C_m prefers user set ς_l to user set ς_k (where ς_l and ς_k are the subsets of $\{1, 2, \cdots N\}$) if the users in set ς_l can provide better performance than the users in set ς_k on C_m. This preference is expressed as

$$O_m(\varsigma_l) > O_m(\varsigma_k), \quad \varsigma_l, \varsigma_k \subset \{UE_1, UE_2 \cdots UE_{N_m}\}$$

where $O_m(\varsigma_l)$ denotes the performance measure of user set ς_l on C_l, which could be the maximin fairness, sum rate, or energy efficiency introduced in Section II-B. Now, we are

ready to formulate the channel assignment optimization as a two-side matching problem according to matching theory [34]. The following definition formally introduces a matching in the NOMA system.

Definition 2: Consider the users and channels as two disjoint sets. A two-to-one matching Φ is a mapping from all the subsets of users N into the channels set M, which satisfies the following properties for $UE_n \in N$ and $C_m \in M$

(a) $\Phi(UE_n) \in M$;

(b) $\Phi^{-1}(C_m) \subseteq N$;

(c) $|\Phi(UE_n)| = 1$, $|\Phi^{-1}(C_m)| = 2$;

(d) $C_m \in \Phi(UE_n) \iff UE_n \in \Phi^{-1}(C_m)$.

In Definition 2, property (a) states that each user matches with one channel, property (b) indicates that each channel can be matched with a subset of users, property (c) states that each channel can only be assigned to two users, and property (d) means that UE_n and C_m are matched with each other. Consequently, the channel assignment problem is to identify a matching between the users and the channels. However, the globally optimal matching that maximizes the aggregate performance of all users is hard to find and usually requires exhaustive search. Instead, in practice, people are more interested in seeking a so-called stable matching, which can be efficiently found by the deferred acceptance (DA) procedure [35].

Definition 3: Given a matching Φ such that $UE_n \notin \Phi^{-1}(C_m)$ and $C_m \notin \Phi(UE_n)$. If $O_m(S_{\text{new}}) > O_m(\Phi^{-1}(C_m))$ where $S_{\text{new}} \subseteq \{UE_n\} \cup S$ and $S = \Phi^{-1}(C_m)$, then S_{new} is the preferred user set for C_m and (UE_n, C_m) is a preferred pair.

According to the DA procedure, each user sends a matching request to its most preferred channel according to its preference list, while this preferred channel has the right to accept or reject the user according to the performance that all users can achieve on this channel. Thus, the DA procedure is to find preferred pairs for each user and each channel, which is formally described in Algorithm 2. In Algorithm 2, the first step is to initialize the preference lists of channels and users according to the CNRs. Meantime, $S_{\text{Match}}(m)$ and $S_{\text{UnMatch}}(m)$ for $m = 1, 2, \cdots M$ are respectively initialized to record the allocated users on C_m and unallocated users. The next step is the matching procedure, where at each round each user sends a matching request according to its preferred list $PF(UE_n)$ for $n = 1, 2, \cdots N$. Then, the channel accepts the user directly if the number of the users on this channel is less than two, otherwise only the user that can improve the performance will be accepted. This matching process will terminate when there is no user left to be matched.

Now, we are able to jointly optimize channel assignment and power allocation by using Algorithm 2 and the optimal power allocation obtained in this paper, which is described in Algorithm 3. In the initialization, the BS allocates equal power budgets to all channels. In the next step, we obtain the channel assignment using Algorithm 2, then update the optimal power allocation for each user and power budget for each channel, and so on.

Remark 9: It is worth pointing out that the optimal power allocation provided in this paper can be jointly used not only with the DA matching algorithm (i.e., Algorithm 2) but

---

Algorithm 2 Channel Assignment via Matching

1: **Initialize**:

    1) $S_{\text{Match}}(m)$ is the matched list to record users matched on C_m, $m = \{1, 2 \cdots M\}$.

    2) S_{UnMatch} is the set of unmatched users.

    3) Obtain preference lists $PF(UE_n)$, $n = \{1, 2, \cdots N\}$ and $PF(C_m)$, $m = \{1, 2 \cdots M\}$ according to CNRs.

2: **while** $\{S_{\text{UnMatch}}\}$ is not empty

3:   **for** $n = 1$ to N **do**

    Each user sends matching request to its most preferred channel m^* according to its preference list $PF(UE_n)$.

4:     **if** $|S_{\text{Match}}(m^*)| < 2$ **then**

      Channel m^* adds UE_n to $S_{\text{Match}}(m^*)$ and removes UE_n from $\{S_{\text{UnMatch}}\}$.

     end if

5:     **if** $|S_{\text{Match}}(m^*)| = 2$ **then**

      1) Identify the power allocation for every two users in $S_{\varsigma L}, S_{\varsigma L} \subset \{S_{\text{Match}}(m^*), n\}$ according to the corresponding proposition with q_{m^*}.

      2) Channel m^* selects a set of 2 users $S_{\varsigma L}$ satisfying the objective functions $O_{m^*}(\varsigma_l) > O_{m^*}(\varsigma_k)$, $\varsigma_l, \varsigma_k \subset \{S_{\text{Match}}(m^*), n\}$.

      3) Channel m^* sets $S_{\text{Match}}(m^*) = \varsigma_l$, and reject other users. Remove the allocated users from $\{S_{\text{UnMatch}}\}$, add the unallocated user to $\{S_{\text{UnMatch}}\}$.

      4) The rejected user remove channel m^* from their preference lists.

     end if

   end for

 end while

Algorithm 3 Joint Channel Assignment and Power Allocation Optimization

1: **Initialize**: $q_m = \frac{P}{M}$ for $m = 1, \ldots, M$;

2: **Repeat**

3:     Obtain channel assignment $\{S_{\text{Match}}(m)\}_{m=1}^{M}$ using Algorithm 2;

4:     Compute $p_1^\star$, $p_2^\star$, $q^\star$ according to the results in this paper;

5: **Until** the prescribed iteration number is reached;

6: **Output**: $\{S_{\text{Match}}(m)\}_{m=1}^{M}$, $p_1^\star$, $p_2^\star$.

---

also with any other assignment algorithms. One just needs to replace Algorithm 2 by the desirable assignment algorithm in Algorithm 3. Furthermore, our results can also reduce the complexity of exhaustive search. Indeed, in [19], the exhaustive search was performed in the joint continuous-discrete dimension of powers and channels. Now, given the optimal power allocation of all users over multiple channels for fixed channel assignment, one can focus on searching the optimal channel assignment by, e.g., checking all possible user-channel matchings, which is a pure combinatorial problem but not a mixed one anymore. In fact, we will show in the next section that the performance of the proposed low-complexity joint resource optimization method is quite close to that of the globally optimal solution found by exhaustive search.
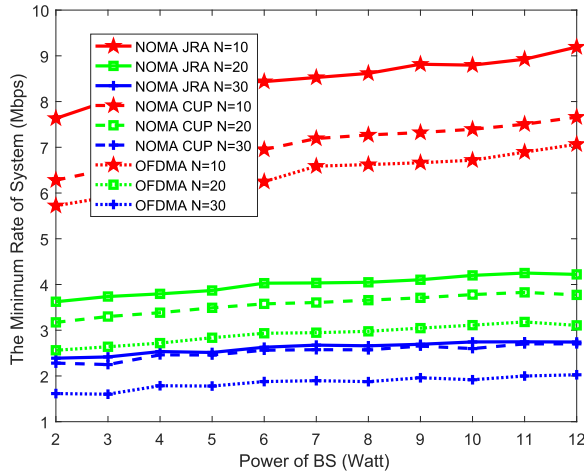
Fig. 1.   Minimum user rate for different number of users versus BS power.

Fig. 2.   Sum rate versus BS power.

Fig. 3.   Spectral efficiency versus the number of users.

VII. Numerical Results

In this section, we evaluate the performance of the optimal power allocation and the proposed joint resource optimization method via numerical simulations. In simulations, the base station is located in the cell center and the users are randomly distributed in a circular range with a radius of 300m. The minimum distance between users is set to be 30m, and the minimum distance between users and BS is 40m. Each channel coefficient follows an i.i.d. Gaussian distribution as $g_m \sim \mathcal{CN}(0,1)$ for $m = 1, \ldots, M$ and the path loss exponent is $\alpha = 2$. The total power budget of the BS is $P = 41$dBm and the circuit power consumption is $P_T = 30$dBm. The noise power is $\sigma_m^2 = BN_0/M$, where the bandwidth is $B = 5$MHz and the noise power spectral density is $N_0 = -174$dBm. We set the user weights to be $W_{1,m} = 0.9$ and $W_{2,m} = 1.1$ for $\forall m$ and the QoS thresholds to be $R_{n,m}^{\min} = 2$ bps/Hz for $n = 1, 2, \forall m$. We compare the proposed joint resource allocation (JRA) method that uses the optimal power allocation and matching algorithm, with OFDMA where each user occupies a bandwidth B/N and the power is optimized in a waterfilling manner, the DC method used in in [20] and [21] where the power allocation was optimized via DC programming, the conventional user pairing (CUP) method used in [36] where the same channel is assigned to the users with a significant channel gain difference, and the exhaustive search.

Fig. 1 depicts the minimum user rates of the NOMA system using the proposed JRA method under the MMF criterion (NOMA JRA), the CUP method with our optimal power allocation for MMF (NOMA CUP), and the OFDMA system for different total power budgets and user numbers. It is clearly seen that NOMA is better than OFDMA in terms of user fairness and the performance gap between NOMA and OFDMA becomes larger as the number of users increases. Meantime, the proposed JRA method outperforms the CUP method.

In Fig. 2, we display the sum rates achieved by various schemes. SR1 JRA and SR2 JRA denote the proposed JRA method for maximizing the weighted SR and the SR with QoS constraints, while SR1 DC represents the DC method for maximizing the weighted SR. Meanwhile, SR1
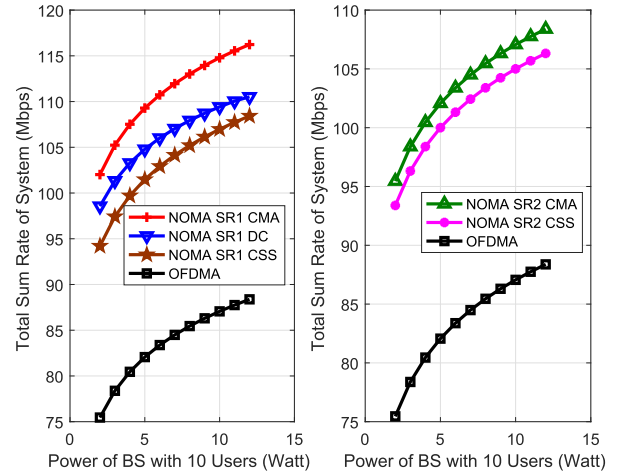
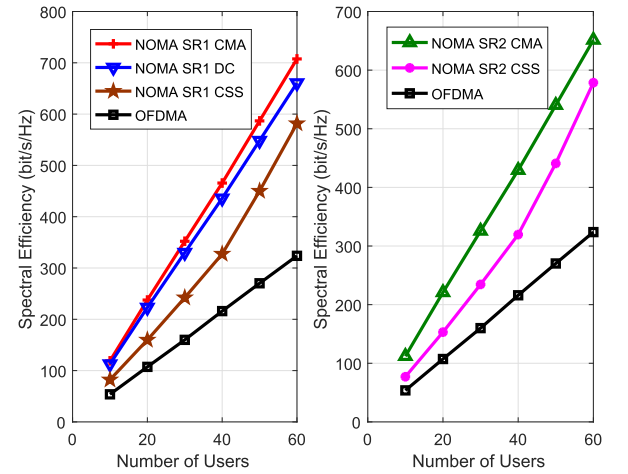CUP and SR2 CUP use the CUP method with our optimal power allocation under the criteria of maximizing the weighted SR and the SR with QoS constraints, respectively. The number of users is 10 in this scenario. As expected, all NOMA schemes (SR1 JRA, SR2 JRA , SR1 CUP, SR2 CUP and SR1 DC) outperform OFDMA. Moreover, it is also observed that SR1 JRA outperforms SR1 DC. This is because the proposed resource allocation uses the optimal power allocation while the DC method leads to a suboptimal power allocation. In addition, both SR1 JRA and SR2 JRA achieve better performance than SR1 CUP and SR2 CUP, which implies the proposed channel assignment method is essential to the performance of the NOMA system.

Fig. 3 shows the spectral efficiency versus the number of users. One can observe the similar phenomenon as in Fig. 2. where the spectral efficiency of NOMA outperforms that of OFDMA and the proposed JRA method leads to higher spectral efficiency than the DC method and the CUP method.

Fig. 4 and Fig. 5 display the EE versus the power budget of BS and the number of users, respectively. The proposed methods for maximizing EE with weights or QoS constraints are denoted by EE1 JRA and EE2 JRA, respectively, while EE1 DC denotes the EE maximization using DC programming. EE1 CUP and EE2 CUP represent the method of CUP with
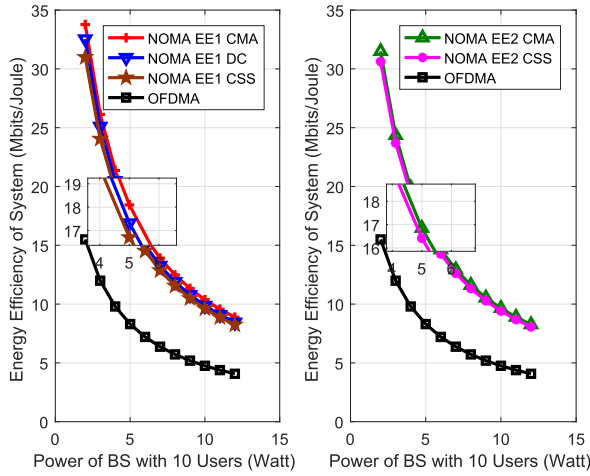
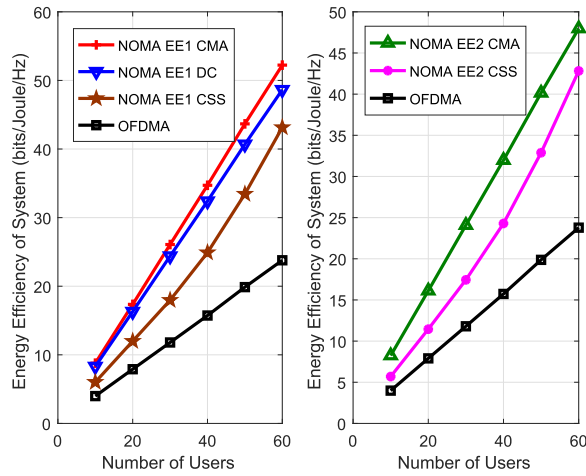Fig. 4.   Energy efficiency versus BS power.

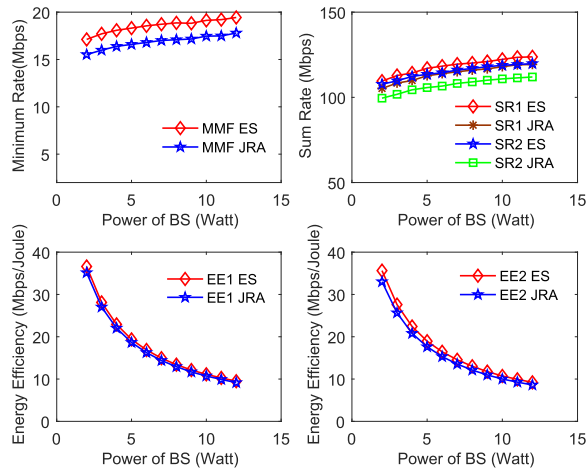Fig. 5.   Energy efficiency versus the number of users.

Fig. 6.   Comparison with the exhaustive search (ES).

our optimal power solutions to maximize EE with weights or QoS constraints, respectively. From Figs. 4 and 5, one can see that the EE of NOMA is significantly higher than that of OFDMA. Meanwhile, EE1 JRA achieves better performance than EE1 DC as a result of using the optimal power allocation. EE1 JRA and EE2 JRA are respectively better than EE1 CUP and EE2 CUP because the channel assignment is optimized by the proposed joint optimization method.

Finally, in Fig. 6, all of our JRA methods are compared to the exhaustive search (ES). Due to the high complexity of ES, we set the number of users $N = 6$ and the power budget of the BS ranges from 2W to 12W. From Fig. 6, the performance achieved the proposed methods is very close to the globally optimal value and the maximum gap is less than 5%. Therefore, the proposed joint channel assignment and power allocation method is able to achieve near-optimal performance with low complexity.

VIII. CONCLUSION

In this paper, we have studied the power allocation in downlink NOMA systems to maximize the MMF, sum rate, and EE with weights or QoS constraints. The optimal power allocation has been characterized in closed or semi-closed forms for all considered performance criteria. We have explicitly considered the power order constraints in power allocation problems and introduced the concept of the SIC-stability to avoid an equal power allocation on each channel in NOMA systems. We have proposed an efficient method to jointly optimize the channel assignment and power allocation in NOMA systems by exploiting the matching algorithm along with the optimal power allocation. The simulation results have shown that the proposed joint resource optimization method achieve near-optimal performance.

REFERENCES

[1] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[2] V. Jungnickel *et al.*, "The role of small cells, coordinated multipoint, and massive MIMO in 5G," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 44–51, May 2014.

[3] S. He, Y. Huang, J. Wang, L. Yang, and W. Hong, "Joint antenna selection and energy-efficient beamforming design," *IEEE Signal Process. Lett.*, vol. 23, no. 9, pp. 1165–1169, Sep. 2016.

[4] J. G. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. C. Reed, "Femtocells: Past, present, and future," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 497–508, Apr. 2012.

[5] J. Wang, W. Guan, Y. Huang, R. Schober, and X. You, "Distributed optimization of hierarchical small cell networks: A GNEP framework," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 2, pp. 249–264, Feb. 2017.

[6] H. Wang, J. Wang, and Z. Ding, "Distributed power control in a two-tier heterogeneous network," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6509–6523, Dec. 2015.

[7] J. Wang, D. Zhu, C. Zhao, J. C. F. Li, and M. Lei, "Resource sharing of underlaying device-to-device and uplink cellular communications," *IEEE Commun. Lett.*, vol. 17, no. 6, pp. 1148–1151, Jun. 2013.

[8] J. Wang, Q. Tang, C. Yang, R. Schober, and J. Li, "Security enhancement via device-to-device communication in cellular networks," *IEEE Signal Process. Lett.*, vol. 23, no. 11, pp. 1622–1626, Nov. 2016.

[9] Z. Wei, J. Yuan, D. W. K. Ng, M. Elkashlan, and Z. Ding. (2016). "A survey of downlink non-orthogonal multiple access for 5G wireless communication networks." [Online]. Available: https://arxiv.org/abs/1609.01856

[10] P. Wang, J. Xiao, and L. P, "Comparison of orthogonal and non-orthogonal approaches to future wireless cellular systems," *IEEE Veh. Technol. Mag.*, vol. 1, no. 3, pp. 4–11, Sep. 2006.

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

[12] 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.

[13] Z. Ding, R. Schober, and H. V. Poor, "A general MIMO framework for NOMA downlink and uplink transmission based on signal alignment," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4438–4454, Jun. 2016.

[14] Z. Ding, F. Adachi, and H. V. Poor, "The application of MIMO to non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 537–552, Jan. 2016.

[15] X. Zhang, Q. Gao, C. Gong, and Z. Xu, "User grouping and power allocation for NOMA visible light communication multi-cell networks," *IEEE Commun. Lett.*, vol. 21, no. 4, pp. 777–780, Apr. 2016.

[16] Z. Ding, P. Fan, and H. V. Poor, "Random beamforming in millimeter-wave NOMA networks," *IEEE Access*, vol. 5, pp. 7667–7681, Feb. 2017.

[17] S. Zhang, B. Di, L. Song, and Y. Li, "Radio resource allocation for non-orthogonal multiple access (NOMA) relay network using matching game," in *Proc. Int. Conf. Commun.*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[18] L. Lei, D. Yuan, C. K. Ho, and S. Sun, "Joint optimization of power and channel allocation with non-orthogonal multiple access for 5G cellular systems," in *Proc. Proc. IEEE Global Telecommun. Conf.*, Dec. 2015, pp. 1–6.

[19] 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.

[20] P. Parida and S. S. Das, "Power allocation in OFDM based NOMA systems: A DC programming approach," in *Proc. IEEE Globecom Workshops*, Austin, TX, USA, Dec. 2014, pp. 1026–1031.

[21] 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.

[22] M.-R. Hojeij, J. Farah, C. A. Nour, and C. Douillard, "Resource allocation in downlink non-orthogonal multiple access (NOMA) for future radio access," in *Proc. IEEE Veh. Technol. Conf.*, Dresden, Germany, May 2015, pp. 1–6.

[23] C.-L. Wang, J.-Y. Chen, and Y.-J. Chen, "Power allocation for a downlink non-orthogonal multiple access system," *IEEE Wireless Commun. Lett.*, vol. 5, no. 5, pp. 532–535, Oct. 2016.

[24] J. Cui, Z. Ding, and P. Fan, "A novel power allocation scheme under outage constraints in NOMA systems," *IEEE Signal Process. Lett.*, vol. 23, no. 9, pp. 1226–1230, Sep. 2016.

[25] J. Choi, "Power allocation for max-sum rate and max-min rate proportional fairness in NOMA," *IEEE Commun. Lett.*, vol. 20, no. 10, pp. 2055–2058, Oct. 2016.

[26] 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.

[27] 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.

[28] Z. Ding, M. Peng, and H. V. Poor, "Cooperative non-orthogonal multiple access in 5G systems," *IEEE Commun. Lett.*, vol. 19, no. 8, pp. 1462–1465, Aug. 2015.

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

[30] 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, Aug. 2016.

[31] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1501–1505, Dec. 2014.

[32] W. Dinkelbach, "On nonlinear fractional programming," *Manage. Sci.*, vol. 13, no. 7, pp. 492–498, Mar. 1967. [Online]. Available: http://www.jstor.org/stable/2627691

[33] G. Hua, J. Goh, and V. L. L. Thing, "A dynamic matching algorithm for audio timestamp identification using the ENF criterion," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 7, pp. 1045–1055, Jul. 2014.

[34] L. Gao, L. Duan, and J. Huang, "Two-sided matching based cooperative spectrum sharing," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 538–551, Feb. 2017.

[35] Y. Gao, Z. Qin, Z. Feng, Q. Zhang, O. Holland, and M. Dohler, "Scalable and reliable IoT enabled by dynamic spectrum management for M2M in LTE-A," *IEEE Internet Things J.*, vol. 3, no. 6, pp. 1135–1145, Dec. 2016.

[36] H. Zhang, D.-K. Zhang, W.-X. Meng, and C. Li, "User pairing algorithm with SIC in non-orthogonal multiple access system," in *Proc. Int. Conf. Commun.*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

Jianyue Zhu (S'17) received the B.S. degree from Nanjing Agricultural University, Nanjing, China, in 2015. She is currently pursuing the Ph.D. degree in information and communication engineering with the School of Information Science and Engineering, Southeast University, Nanjing. Her current research interests include non-orthogonal multiple access and optimization theory.

Jiaheng Wang (M'10–SM'14) received the B.E. and M.S. degrees from the Southeast University, Nanjing, China, in 2001 and 2006, respectively, and the Ph.D. degree in electronic and computer engineering from The Hong Kong University of Science and Technology, Hong Kong, in 2010. From 2010 to 2011, he was with the Signal Processing Laboratory, KTH Royal Institute of Technology, Stockholm, Sweden. He also held visiting positions with the Friedrich Alexander University Erlangen-Nürnberg, Nürnberg, Germany, and also with the University of Macau, Macau. He is currently a Full Professor with the National Mobile Communications Research Laboratory, Southeast University. He has authored over 80 articles on international journals and conferences. His research interests are mainly on optimization in communication systems and wireless networks. He serves as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS.. He was a recipient of the Humboldt Fellowship for Experienced Researchers, and was the recipient of the Best Paper Award in WCSP 2014.

Yongming Huang (M'10) received the B.S. and M.S. degrees from Nanjing University, Nanjing, China, in 2000 and 2003, respectively, and the Ph.D. degree in electrical engineering from Southeast University, Nanjing, China, in 2007. Since 2007, he has been a Faculty Member with the School of Information Science and Engineering, Southeast University, where he is currently a Full Professor. From 2008 to 2009, he visited the Signal Processing Laboratory, School of Electrical Engineering, Royal Institute of Technology, Stockholm, Sweden. He has authored over 200 peer-reviewed papers, hold over 40 invention patents, and submitted over ten technical contributions to the IEEE standards. His current research interests include MIMO wireless communications, cooperative wireless communications, and millimeter wave wireless communications. Since 2012, he has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, *EURASIP Journal on Advances in Signal Processing*, and *EURASIP Journal on Wireless Communications and Networking*.
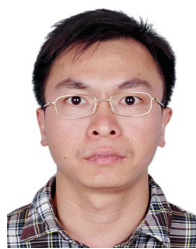
Shiwen He (M'14) received the M.S. degree from the Chengdu University of Technology, Chengdu, China, and the Ph.D. degree in information and communication engineering from Southeast University, Nanjing, China, in 2009 and 2013, respectively. From 2013 to 2015, he was a Post-Doctoral Researcher with the State Key Laboratory of Millimeter Waves, Department of Radio Engineering, Southeast University. Since 2015, he has been with the School of Information Science and Engineering, Southeast University. He has authored or co-authored over 80 technical publications, holds over 16 granted invention patents, and submitted over 20 technical proposals to the IEEE standards. His current research interests include multiuser MIMO communication, cooperative communications, energy efficient communications, millimeter wave communication, and optimization theory. From 2015, he serves as a Technical Sub-Editor for the IEEE 802.11aj.

Xiaohu You (F'11) was born in 1962. He received the master's degree and Ph.D. degrees in electrical engineering from Southeast University, Nanjing, China, in 1985 and 1988, respectively. Since 1990, he has been with National Mobile Communications Research Laboratory, Southeast University, where he held the rank of Director and Professor. His research interests include mobile communication systems, signal processing and its applications. He has contributed over 100 IEEE journal papers and two books in the areas of adaptive signal processing, neural networks and their applications to communication systems. From 1999 to 2002, he was the Principal Expert of the C3G Project, responsible for organizing China's 3G Mobile Communications Research and Development Activities. From 2001 to 2006, he was the Principal Expert of the China National 863 Beyond 3G FuTURE Project. Since 2013, he has been the Principal Investigator of China National 863 5G Project. He was the General Chairs of the IEEE WCNC 2013 and the IEEE VTC 2016. He is currently the Secretary General with the FuTURE Forum, VICE Chair of the China IMT-2020 Promotion Group, the Vice Chair of China National Mega Project on New Generation Mobile Network. He was a recipient of the National First Class Invention Prize in 2011, and he was selected as the IEEE Fellow in same year due to his contributions to development of mobile communications in China.

Luxi Yang (M'96) received the M.S. and Ph.D. degrees in electrical engineering from the Southeast University, Nanjing, China, in 1990 and 1993, respectively. Since 1993, he has been with the Department of Radio Engineering, Southeast University, where he is currently a Full Professor of information systems and communications, and the Director of Digital Signal Processing Division. He has authored or co-authored two published books and over 100 journal papers, and holds 30 patents. His current research interests include signal processing for wireless communications, MIMO communications, cooperative relaying systems, and statistical signal processing. He received the first and second class prizes of science and technology progress awards of the state education ministry of China, in 1998, 2002, and 2014. He is currently a member of Signal Processing Committee of Chinese Institute of Electronics.