Energy-Efficient Power Allocation for Millimeter Wave Beamspace MIMO-NOMA Systems

PENGLU LIU, **YONG LI**, **WEI CHENG**, **WENJIE ZHANG**, AND **XIANG GAO**

School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710129, China

Corresponding author: Yong Li (ruikel@nwpu.edu.cn)

ABSTRACT Massive multi-input multi-output (MIMO) is envisioned as a key technology for the emerging fifth generation of communication networks (5G). However, considering the energy consumption of the large number of radio frequency (RF) chains, massive MIMO poses a problem to energy efficiency (EE) requirement of 5G. In this paper, we propose an energy-efficient power allocation method for millimeter-wave (mmWave) beamspace MIMO non-orthogonal multiple access (NOMA) systems, where there may be multiple users in each selected beam. First, according to the beam selection (BS) results, we get the precoding matrix through zero-forcing (ZF) beamforming method. Second, we formulate the energy efficiency (EE) maximization optimization problem as a fractional programming. Through sequential convex approximation (SCA) and second-order cone (SOC) transformation, the original optimization problem can be transformed to a convex optimization problem. By using iterative optimization algorithm, we can get the power allocation results. Then, we analyze the convergence of our proposed iterative optimization method and get that the solution in each iteration is a suboptimal solution to the original non-convex optimization problem. Simulation results show that the proposed energy-efficient power allocation scheme has better EE performance comparing with the conventional methods when the transmitted power exceeds the power threshold.

INDEX TERMS Beamspace, NOMA, energy efficiency, power allocation, convex optimization.

I. INTRODUCTION

With the rapid development of mobile Internet and Internet of things (IoT), there will be the prediction of 1000-fold data traffic increase by the year 2020 [1], [2]. Massive multi-input multi-output (MIMO) and non-orthogonal multiple access (NOMA) are two key techniques for the coming fifth generation of communication networks (5G), which work together for satisfying the future large demands of communication service [3]–[6]. Besides the spectral efficiency (SE), massive connectivity for IoT, lower latency and diverse compelling services, energy efficiency (EE) is another key performance indicator (KPI) of 5G, which is more than 100 times the EE of 4G. According to the propagation characteristics of millimeter wave (mmWave), the energy consumption of MIMO-NOMA systems can be reduced by relay [7] and network densification [8], [9]. However, the energy consumption of circuit, which is proportional to the number of radio frequency (RF) chains, can degrade the EE performance

The associate editor coordinating the review of this article and approving it for publication was Amjad Anvari-Moghaddam.

of systems. That's to say, the large-scale antenna array has adverse effect on the EE requirement of 5G, although the large-scale antenna array can offer higher SE by forming directional beams with high gain.

For meeting the requirement of tremendous data increase, in the conventional wireless communication researches, the most widely adopted metrics have been sum rate maximization (SRMax)\spectral efficiency maximization (SEMax) and sum power minimization (SPMin) [31]. In order to deal with the downlink SRMax problem, [10]–[15] exploited the user pairing and power allocation algorithms, which are always NP-hard optimization problems. Focusing on user pairing and power allocation, [16] provided an overview of the resource allocation (RA) algorithms for downlink NOMA in a categorized fashion.

Besides the SE criteria, with consideration of the huge information and communication technology energy consumption, EE has recently drawn significant attention. In [17], under NOMA scenario, the EEMax problem was formulated as a non-convex fractional programming. According to the established feasible range of transmitting power,

an EE-optimal power allocation strategy was proposed. For considering the constraints on minimum user quality of service and the maximum transmitted power constraint, energy-efficient dynamic power allocation in NOMA networks is explored by using the Lyapunov optimization method [18]. Different from energy-efficient power allocation for MIMO-NOMA with multiple users in a cluster [19], Haitham et al. investigated the design of an energy-efficient beamforming technique for downlink transmission in the context of a multiuser multi-input single-output (MISO) NOMA system [20]. Based on sequential convex approximation (SCA) and Dinkelbach's method, the original non-convex fractional programming optimization problem was reformulated, respectively. Two novel algorithms were proposed for solving the downlink beamforming problem for the multiuser MISO-NOMA system. Different from the single-cell EE analyses, [21]–[23] examined the problem of energy-efficient user scheduling and power allocation in NOMA heterogeneous networks (HetNets).

In order to further improve the EE, based on beam selection (BS) method [28], the power consumption of large-scale antenna array can be reduced by introducing discrete lens array (DLA) [24], where the SE performance degeneration is slight comparing with the conventional antenna selection strategies. By classifying all users into two groups, i.e., the interference-users (IUs) and noninterference-users (NIUs), [25] proposed an interference-aware beam selection (IA-BS) method with keeping only one user in one beam. In the conventional beamspace MIMO, the number of supported users cannot be larger than the number of RF chains. Based on NOMA, combing with DLA, Wang et al. proposed a spectrum and energy-efficient beamspace MIMO-NOMA for mm-Wave communications [26]. To maximize the achievable SR, a dynamic power allocation was proposed by solving the joint power optimization problem, which included the intra-beam and inter-beam power allocation. Based on the Sherman-Morrison-Woodbury formula and iterative optimization method [27], the dynamic power allocation problem was solved for mmWave massive MIMO-NOMA with simultaneous wireless information and power transfer. We point out that the main difference between BS MIMO-NOMA method [26] and IA-BS method [25] is the number of users in each selected beam. The former method allows multiple users in each selected beam; the latter just guarantees that there is only one user in each selected beam. Especially, it's more likely that there are multiple users corresponding to one selected beam in the user dense scenarios.

According to the above references, we can find that some papers have discussed the EE performance of mmWave MIMO-NOMA communication systems. Without DLA, reference [20] discussed energy-efficient beamforming design for MISO-NOMA systems. Reference [26] considered the EE of mmWave beamspace MIMO-NOMA. However, the power allocation method was designed with SEMax as metric. Without NOMA and DLA, [29] considered the joint beamforming and antenna selection (JBAS) problem

under conventional OMA schemes. Especially, with precondition of satisfying the users' quality of service (QoS), it makes no sense to purse SRMax blindly.

In this paper, we study on the energy-efficient power allocation method for mmWave beamspace MIMO-NOMA systems. The contributions of this paper can be summarized as follows:

- Based on the BS method for mmWave beamspace MIMO-NOMA [26] and ZF beamforming technique, we formulate the energy-efficient power allocation problem as a non-convex fractional programming, which considers the constraints of all users, i.e., minimum rate constraints, successive interference cancellation (SIC) constraints and maximum power budget. Especially, complex SIC constraints were ignored in [26].
- The original optimization problem is a fractional programming problem. Through SCA and SOC transformation, the EEMax optimization problem can be reformulated as a convex second-order cone programming (SOCP), which is tractable. We choose iterative optimization algorithm to solve the reformulated problem, so get the power allocation results. Then, we analyze the convergence of the proposed algorithm.
- We verify our proposed energy-efficient power allocation method. The simulation results show that the proposed iterative optimization method can converge quickly. Comparing with the conventional power allocation methods, the developed method has higher EE when the power budget exceeds the power threshold.

The rest of this paper is organized as follows. In Section II, we construct the model for the mmWave beamspace MIMO-NOMA systems and formulate the energy-efficient power allocation problem. In order to get the power allocation results, we introduce a series of mathematical transformations to reformulate the original optimization problem in Section III. In Section IV, the convergence and computational complexity analysis is performed. We verify our proposed power allocation method in Section V. Finally, our conclusions are presented in Section VI.

The following notations are used in this paper. The bold lower-case and upper-case letters denote vectors and metrices, respectively. $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^{-1}$, $(\cdot)^\dagger$ represent the transpose, Hermitian transpose, inverse, Moore-Penrose matrix inversion of matrix, respectively. $E\{\cdot\}$ denotes the expectation. $\mathcal{B}$ denotes the selected beams set and $|\mathcal{B}|$ denotes the number of selected beams. $|\mathcal{K}_i|$ and $\mathcal{K}_i$ denote the number of served users and served users set in the i-th selected beam, respectively. $\|\cdot\|_2$ and $|\cdot|$ denote the 2-norm of a vector and absolute value of a complex number, respectively. $\boldsymbol{I}$ denotes the identity matrix.

II. SYSTEM MODEL AND PROBLEM FORMULATION
In this section, we introduce the mmWave beamspace MIMO-NOMA model [26]. Then, we formulate the energy-efficient power allocation optimization problem.

A. SYSTEM MODEL

We consider a single-cell downlink mmWave communication system, where the base station has N antennas and N_{RF} RF chains and serves K users. In the conventional MIMO system, each antenna corresponds to one RF chain, which means $N = N_{RF}$. The $K \times 1$ received signal vector $\boldsymbol{y}$ can be described as

$$\boldsymbol{y} = \boldsymbol{H}^H \boldsymbol{W} \boldsymbol{P} \boldsymbol{s} + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{n}$ is a $K \times 1$ additive white Gaussian noise (AWGN) vector, which follows the distribution $\mathcal{CN}\left(0, \sigma^2 \boldsymbol{I}_K\right)$; $\boldsymbol{s} = [s_1, s_2, \cdots, s_K]^T$ denotes the $K \times 1$ transmitted signal of K users and $E\left\{|s_i|^2\right\} = 1, i = 1, \cdots, K$; the $K \times K$ diagonal matrix represents the transmitted power matrix $\boldsymbol{P} = diag\{\boldsymbol{p}\}$, where $\boldsymbol{p}$ can be expressed as a column vector $p = \left[\sqrt{p_1}, \sqrt{p_2}, \cdots, \sqrt{p_K}\right]^T$ and $p_i, i = 1, \cdots, K$ denotes the power allocation for the i-th user; $\boldsymbol{W} = \left[\boldsymbol{w}_1 \; \boldsymbol{w}_2 \; \cdots \; \boldsymbol{w}_K\right]$ denotes the $N \times K$ precoding matrix, where $\boldsymbol{w}_i, i = 1, \cdots, K$ is the $N \times 1$ precoding vector for the i-th user; the $N \times K$ channel state information (CSI) matrix is given by $\boldsymbol{H} = [\boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_K]$, where $\boldsymbol{h}_i, i = 1, \cdots, K$ is the $N \times 1$ CSI vector.

According to the mmWave propagation characteristics [11], the mmWave CSI vector from the BS to the k-th user can be expressed as follows [26]:

$$\boldsymbol{h}_k = \rho_k^0 \boldsymbol{a}\left(\theta_k^0\right) + \sum_{l=1}^{L} \rho_k^l \boldsymbol{a}\left(\theta_k^l\right), \tag{2}$$

where $\rho_k^0 \boldsymbol{a}\left(\theta_k^0\right)$ is the line of sight (LoS) component of $\boldsymbol{h}_k$, in which ρ_k^0 and $\boldsymbol{a}\left(\theta_k^0\right)$ represent the LoS complex gain and spatial direction information, respectively; $\rho_k^l \boldsymbol{a}\left(\theta_k^l\right), i = 1, \cdots, L$ is the non-line of sight (NLoS) component and L denotes the total number of NLoS components, in which ρ_k^l and $\boldsymbol{a}\left(\theta_k^l\right)$ represent the complex gain and spatial direction information for NLoS, respectively. For the uniform linear array (ULA), the array steering vector can be represented as

$$\boldsymbol{a}\left(\theta\right) = \frac{1}{\sqrt{N}}\left[e^{-j2\pi\theta m}\right],$$
$$m \in \left\{q - (N-1)/2, q = 0, 1, \cdots, N-1\right\}. \tag{3}$$

In (3), $\theta = \frac{d}{\lambda} \sin(\varphi)$ is defined as spatial direction, in which $\varphi = \left[-\pi/2, \pi/2\right]$ denotes the physical direction, λ and $d = \lambda/2$ are wavelength and antenna spacing, respectively. In order to moderate the SE degradation when we perform beam selection, by introducing DLA, (1) can be reformulated as

$$\tilde{\boldsymbol{y}} = \boldsymbol{H}^H \boldsymbol{U}^H \boldsymbol{W} \boldsymbol{P} \boldsymbol{s} + \boldsymbol{n} = \tilde{\boldsymbol{H}}^H \boldsymbol{W} \boldsymbol{P} \boldsymbol{s} + \boldsymbol{n}. \tag{4}$$

The DLA acts as an $N \times N$ spatial discrete fourier transform matrix $\boldsymbol{U}$. According to (3), the matrix $\boldsymbol{U}$ can be represented as $\boldsymbol{U} = [\boldsymbol{a}\left(\theta_1\right), \boldsymbol{a}\left(\theta_2\right), \cdots, \boldsymbol{a}\left(\theta_N\right)]$, where $\theta_i, i \in \{1, 2, \cdots, N\}$ are defined uniformly ranging from -1 to 1.

$\tilde{\boldsymbol{H}}$ denotes the beamspace CSI matrix, which can be represented as

$$\tilde{\boldsymbol{H}} = \left[\tilde{\boldsymbol{h}}_1, \tilde{\boldsymbol{h}}_2, \cdots, \tilde{\boldsymbol{h}}_K\right] = [\boldsymbol{U}\boldsymbol{h}_1, \boldsymbol{U}\boldsymbol{h}_2, \cdots, \boldsymbol{U}\boldsymbol{h}_K]. \tag{5}$$

In (5), $\tilde{\boldsymbol{h}}_i$ is the beamspace CSI vector for the i-th user.

B. PROBLEM FORMULATION

In the conventional BS system, there is only one user in one beam. In this part, we focus on cluster based beamspace MIMO-NOMA, which means that more than one user can be simultaneously served in one beam through NOMA. According to [26], we select $|\mathcal{B}|$ beams to serve K users and $|\mathcal{B}| \le K$. The selected beams set and corresponding user set can be represented as $\mathcal{B}$ and $\mathcal{K} \triangleq \{1, 2, \cdots, K\}$, respectively. With the beam selection results, we suppose that the i-th ($i \le |\mathcal{B}|$) selected beam serves $|\mathcal{K}_i|$ users, where $|\mathcal{K}_i|$ and $\mathcal{K}_i$ denote the number of served users and served users set in the i-th selected beam, respectively. $|\mathcal{K}_i|$ meets $\sum_{i=1}^{|\mathcal{B}|} |\mathcal{K}_i| = K$. With the beamspace channel vector of the first user (strong user) in each beam as the equivalent channel vector, the $|\mathcal{B}| \times |\mathcal{B}|$ equivalent channel matrix $\hat{\boldsymbol{H}}$ can be represented as $\hat{\boldsymbol{H}} = \left[\hat{\boldsymbol{h}}_1, \hat{\boldsymbol{h}}_2, \cdots, \hat{\boldsymbol{h}}_{|\mathcal{B}|}\right]$, where $\hat{\boldsymbol{h}}_i, i \in \{1, 2, \cdots, |\mathcal{B}|\}$ denotes the equivalent channel vector of i-th selected beam. Through ZF beamforming technique, we can get the $|\mathcal{B}| \times |\mathcal{B}|$ precoding matrix $\hat{\boldsymbol{W}}$ as

$$\begin{aligned} \hat{\boldsymbol{W}} &= \left[\hat{\boldsymbol{w}}_1, \hat{\boldsymbol{w}}_2, \cdots, \hat{\boldsymbol{w}}_{|\mathcal{B}|}\right] \\ &= \left(\hat{\boldsymbol{H}}^H\right)^\dagger \\ &= \hat{\boldsymbol{H}}\left(\hat{\boldsymbol{H}}^H \hat{\boldsymbol{H}}\right)^{-1}, \end{aligned} \tag{6}$$

where $\hat{\boldsymbol{w}}_i, i \in \{1, 2, \cdots, K\}$ denotes the precoding vector for i-th selected beam, and all the precoding vectors should be normalized. For convenience, we use $\boldsymbol{h}_{i,j}$ to denote the beamspace CSI of j-th user in the i-th selected beam and use $\boldsymbol{w}_i$ instead of $\hat{\boldsymbol{w}}_i$ as the normalized precoding vector. The users are indexed as the descending order of equivalent channel gains as

$$\left|\boldsymbol{h}_{i,1}^H \boldsymbol{w}_i\right|^2 \ge \left|\boldsymbol{h}_{i,2}^H \boldsymbol{w}_i\right|^2 \ge \cdots \ge \left|\boldsymbol{h}_{i,|\mathcal{K}_i|}^H \boldsymbol{w}_i\right|^2. \tag{7}$$

The received j-th user's signal at the l-th ($l \le j$) user in the i-th selected beam can be represented by

$$\begin{aligned} y_{i,j}^l = \boldsymbol{h}_{i,l}^H \boldsymbol{w}_i \sqrt{p_{i,j}} s_{i,j} + \boldsymbol{h}_{i,l}^H \boldsymbol{w}_i \sum_{n=1}^{j-1} \sqrt{p_{i,n}} s_{i,n} \\ + \boldsymbol{h}_{i,l}^H \sum_{m \ne i} \sum_{n=1}^{|\mathcal{K}_n|} \boldsymbol{w}_m \sqrt{p_{m,n}} s_{m,n} + \upsilon_{i,j}^l. \end{aligned} \tag{8}$$

According to the right side of (8), the first part denotes the desired signal; the second part is the intra-beam interference, the third part denotes the inter-beam interference, and the last part is the noise. The received signal to interference and noise

ratio (SINR) for the l-th user to decode the j-th user in the i-th selected beam can be formulated as

$$r_{i,j}^l = \frac{\left|\boldsymbol{h}_{i,l}^H \boldsymbol{w}_i\right|^2 p_{i,j}}{\left|\boldsymbol{h}_{i,l}^H \boldsymbol{w}_i\right|^2 \sum\limits_{n=1}^{j-1} p_{i,n} + \sum\limits_{m \neq i} \left|\boldsymbol{h}_{i,l}^H \boldsymbol{w}_m\right|^2 \sum\limits_{n=1}^{|\mathcal{K}_m|} p_{m,n} + \sigma^2}. \quad (9)$$

Then, the achievable rate of user j in the i-th beam can be represented as

$$R_{i,j} = \log_2\left(1 + \min\left\{r_{i,j}^j, r_{i,j}^{j-1} \cdots, r_{i,j}^1\right\}\right). \quad (10)$$

In this paper, we suppose the bandwidth is a unit bandwidth, which means SR is equivalent to SE and all K users share the same spectral resource elements.

For formulating the EE maximization optimization, we should firstly analyze the total power consumption of the modeled beamspace MIMO-NOMA communication system. The total power consumption can be denoted as

$$P_{tot} = \frac{1}{\epsilon} P_{tr} + P_{ext}, \quad (11)$$

in which P_{tr} is the transmitted power defined as $P_{tr} = \sum\limits_{i=1}^{|\mathcal{B}|} \sum\limits_{j=1}^{|\mathcal{K}_i|} p_{i,j}$; $0 < \epsilon < 1$ is the energy efficiency of power amplifier; P_{ext} denotes the power consumption of circuits [26], which is given by $P_{ext} = N_{RF} P_{RF} + N_{RF} P_{SW} + P_{BB}$; P_{RF}, P_{SW} and P_{BB} denote the power consumption of each RF chain, switching and baseband, respectively.

According to the above definitions, the EEMax optimization problem can be formulated as follows:

$$OP1: \quad \max_{\{p_{i,j}\}} EE = \frac{\sum\limits_{i=1}^{|\mathcal{B}|} \sum\limits_{j=1}^{|\mathcal{K}_i|} R_{i,j}}{P_{tot}} \quad (12a)$$

$$s.t. \quad R_{i,j} \geq R_{i,j}^{\min}, \quad \forall i \leq |\mathcal{B}|, \, j \leq |\mathcal{K}_i|, \quad (12b)$$

$$p_{i,j} \geq 0, \quad \forall i \leq |\mathcal{B}|, \, j \leq |\mathcal{K}_i|, \quad (12c)$$

$$\sum\limits_{i=1}^{|\mathcal{B}|} \sum\limits_{j=1}^{|\mathcal{K}_i|} p_{i,j} \leq P_{bud}. \quad (12d)$$

In the problem OP1, (12a) denotes the EE of the system, the numerator of (12a) is the SR, the denominator is the total power consumption as (11); (12b) means that each user's achievable rate should meet the minimum rate requirement [1] $R_{i,j}^{\min}$; (12c) insures that the optimal power allocation factors are nonnegative; (12d) guarantees the maximum transmitted power should not exceed the power budget P_{bud}.

---

[1] Under the assumption of a given SIC decoding order, the constraint (12b) guarantees SIC performed successfully. In this paper, the SIC decoding order is the ascending order of equivalent channel gains. Alavi *et al.* [31] and Cui *et al.* [35] also considered about the SIC constraints, where both of them pointed out that the inter-cluster/inter-beam interference also affects the SIC decoding process significantly. While Wang *et al.* [26] simplified the constraints. Especially, Wei *et al.* [36] discussed the relationship among SR, SIC decoding order and target data rate, which in this paper we don't discuss them in detail.

As we know, the optimization problem OP1 is a non-convex fractional programming, which is intractable. In order to transform the original problem to a tractable one, in next section, we will reformulate the above optimization problem by using a series of mathematical transformations, i.e., SCA, SOC transformation, relaxation and iterative optimization algorithm.

III. SOLVING THE ENERGY-EFFICIENT POWER ALLOCATION PROBLEM

Different from the Dinkelbach's method [29], [30], we deal with the objective function by using a slack variable, so the objective function is transformed into a linear objective function. For the non-linear constraints, we deal with them through SCA [30], [32] and SOCP [32], [34]. By solving SPMin problem [31], we get the power allocation results. Then, by using the above results from SPMin as an initial iteration point, we solve the energy-efficient power allocation problem through iterative optimization method.

According to references [20] and [29], by introducing two slack variables α and β, which denotes squared energy efficiency and squared power consumption, respectively. The original problem OP1 can be equivalently reformulated as follows:

$$OP2: \quad \max_{\{p_{i,j}\}} \alpha \quad (13a)$$

$$s.t. \quad \sum\limits_{i=1}^{|\mathcal{B}|} \sum\limits_{j=1}^{|\mathcal{K}_i|} R_{i,j} \geq \sqrt{\alpha\beta}, \quad (13b)$$

$$P_{tot} \leq \sqrt{\beta}, \quad (13c)$$

$$(12b) \, (12c) \, (12d) \,. \quad (13d)$$

In OP2, maximizing $\sqrt{\alpha}$ is equal to maximizing α.

As we know, the constraint (13b) is non-convex. According to (10), (13b) can be rewritten as

$$\sum\limits_{i=1}^{|\mathcal{B}|} \sum\limits_{j=1}^{|\mathcal{K}_i|} \log_2\left(1 + \min\left\{r_{i,j}^j, r_{i,j}^{j-1} \cdots, r_{i,j}^1\right\}\right) \geq \sqrt{\alpha\beta}. \quad (14)$$

For dealing with the non-convexity of (14), by introducing new slack variables $\delta_{i,j}$ and $\eta_{i,j}$, we define two new constraints for the j-th user in the i-th selected beam as follows:

$$\log_2\left(1 + \min\left\{r_{i,j}^j, r_{i,j}^{j-1} \cdots, r_{i,j}^1\right\}\right) \geq \delta_{i,j}, \quad (15)$$

$$1 + \min\left\{r_{i,j}^j, r_{i,j}^{j-1} \cdots, r_{i,j}^1\right\} \geq \eta_{i,j}. \quad (16)$$

Based on (15), the constraint (14) can be rewritten as

$$\sum\limits_{i=1}^{|\mathcal{B}|} \sum\limits_{j=1}^{|\mathcal{K}_i|} \delta_{i,j} \geq \sqrt{\alpha\beta}. \quad (17)$$

The new constraint (17) is still non-convex. By using first-order Taylor series expansion around the point (α^t, β^t), we can approximate $\sqrt{\alpha\beta}$ from the upper bound, which can

be written as

$$
\sqrt{\alpha\beta} \leq \sqrt{\alpha^t \beta^t} + \frac{1}{2}\sqrt{\frac{\beta^t}{\alpha^t}}\left(\alpha - \alpha^t\right) + \frac{1}{2}\sqrt{\frac{\alpha^t}{\beta^t}}\left(\beta - \beta^t\right). \quad (18)
$$

With (18), the non-convex constraint (17) can be transformed into

$$
\sum_{i=1}^{|\mathcal{B}|}\sum_{j=1}^{|\mathcal{K}_i|} \delta_{i,j} \geq \sqrt{\alpha^t \beta^t} + \frac{1}{2}\sqrt{\frac{\beta^t}{\alpha^t}}\left(\alpha - \alpha^t\right) + \frac{1}{2}\sqrt{\frac{\alpha^t}{\beta^t}}\left(\beta - \beta^t\right).
$$
$$(19)$$

Based on (16), we can rewrite the constraint (15) as

$$
\eta_{i,j} \geq 2^{\delta_{i,j}}. \quad (20)
$$

The Hessian Matrix of the right-hand expression can be represented as $\begin{bmatrix} 2^{\delta_{i,j}}(\ln 2)^2 & 0 \\ 0 & 0 \end{bmatrix}$, which means $2^{\delta_{i,j}}$ is a convex function.

Without considering about the other constraints, (20) is a non-linear convex constraint, which has the exponential function $2^{\delta_{i,j}}$. About the non-linear programming, the available non-linear solver such as Fmincon, which is a MATLAB optimization toolbox, can be used to solve non-linear problem. The constraint (20) satisfies the constraint types, which was defined in the CVX guide (Section 4.4) [34], but it's a non-linear constraint. Reference [29] proposed that the constraint (20) can be approximated by a system of SOC constraints with a given accuracy. Fortunately, the new version of CVX tool can approximate the exponential function dynamically. In this paper, we just ignore the non-linearity of the constraint (20) and let CVX solver deal with it. Especially, we should pay attention to the difference between Bit and Nat when we approximate the exponential function before using CVX toolbox. That's because the approximating transformation in [29] is based on Nat[2].

According to (9), the constraint (16) can be reformulated as

$$
\frac{\left|\boldsymbol{h}_{i,l}^H \boldsymbol{w}_i\right|^2 p_{i,j}}{\left|\boldsymbol{h}_{i,l}^H \boldsymbol{w}_i\right|^2 \sum_{n=1}^{j-1} p_{i,n} + \sum_{m \neq i}\left|\boldsymbol{h}_{i,l}^H \boldsymbol{w}_m\right|^2 \sum_{n=1}^{|\mathcal{K}_m|} p_{m,n} + \sigma^2} \geq \eta_{i,j} - 1,
$$
$$(21)$$

in which i, j and l meet $\forall i \leq |\mathcal{B}|$, $j \leq |\mathcal{K}_i|$, $l \leq j$, respectively. We point out that i, j and l satisfy the above constraints in the following content of section III if we don't state them again. The constraint (21) is still non-convex, by introducing a new auxiliary variable $\vartheta_{i,j} \in \mathbb{R}^1_+$ ($\forall i \leq |\mathcal{B}|$, $j \leq |\mathcal{K}_i|$), then the constraint can be reformulated into two new constraints as follows:

$$
\left|\boldsymbol{h}_{i,l}^H \boldsymbol{w}_i\right|^2 p_{i,j} \geq \left(\eta_{i,j} - 1\right)\vartheta_{i,j}, \quad (22a)
$$

[2]Bit and Nat are two different units of information.

$$
\left|\boldsymbol{h}_{i,l}^H \boldsymbol{w}_i\right|^2 \sum_{n=1}^{j-1} p_{i,n} + \sum_{m \neq i}\left|\boldsymbol{h}_{i,l}^H \boldsymbol{w}_m\right|^2 \sum_{n=1}^{|\mathcal{K}_m|} p_{m,n} + \sigma^2 \leq \vartheta_{i,j}.
$$
$$(22b)$$

The constraint (22a) is a quadratic non-convex constraint. With first-order Taylor series expansion around the point $\left(\eta_{i,l}^t, \vartheta_{i,l}^t\right)$, (22a) can be rewritten as

$$
\left|\boldsymbol{h}_{i,l}^H \boldsymbol{w}_i\right|^2 p_{i,j} \geq \left(\eta_{i,j}^t - 1\right)\vartheta_{i,j}^t + \vartheta_{i,j}^t\left(\eta_{i,j} - \eta_{i,j}^t\right)
$$
$$
+ \left(\eta_{i,j}^t - 1\right)\left(\vartheta_{i,j} - \vartheta_{i,j}^t\right). \quad (23)
$$

The constraint (22b) is a linear convex constraint about variables $p_{i,j}$ and $\vartheta_{i,j}$.

We can certify that the right-hand expression of (13c) is a concave function, so the constraint (13c) is convex. By introducing auxiliary variable $\tilde{\beta}$, we can transform the original constraint (13c) into new constraints as follows:

$$
\sqrt{\beta} \geq \tilde{\beta}, \quad (24a)
$$
$$
\tilde{\beta} \geq P_{tot}. \quad (24b)
$$

According to SOCP, the constraint (24a) can be reformulated as

$$
\frac{\beta + 1}{2} \geq \left\|\left[\frac{\beta - 1}{2}, \tilde{\beta}\right]^T\right\|_2. \quad (25)
$$

The constraint (25) is equal to (24a).

According to the minimum rate requirement $R_{i,j}^{\min}$ of i-th user in the j-th selected beam, the minimum SINR requirement $r_{i,j}^{\min}$ can be represented as $r_{i,j}^{\min} = 2^{R_{i,j}^{\min}} - 1$. On the basis of (9), the constraint (12b) can be rewritten as

$$
\frac{\left|\boldsymbol{h}_{i,l}^H \boldsymbol{w}_i\right|^2 p_{i,j}}{\left|\boldsymbol{h}_{i,l}^H \boldsymbol{w}_i\right|^2 \sum_{n=1}^{j-1} p_{i,n} + \sum_{m \neq i}\left|\boldsymbol{h}_{i,l}^H \boldsymbol{w}_m\right|^2 \sum_{n=1}^{|\mathcal{K}_m|} p_{m,n} + \sigma^2} \geq r_{i,j}^{\min}.
$$
$$(26)$$

The above inequation can be reformulated as

$$
\left|\boldsymbol{h}_{i,l}^H \boldsymbol{w}_i\right|^2 p_{i,j} - r_{i,j}^{\min}\left(P_{Intra} + P_{Inter}\right) \geq r_{i,j}^{\min}\sigma^2, \quad (27)
$$

where P_{Intra} denotes the power of intra-beam interference $\left|\boldsymbol{h}_{i,l}^H \boldsymbol{w}_i\right|^2 \sum_{n=1}^{j-1} p_{i,n}$, the power of inter-beam interference $\sum_{m \neq i}\left|\boldsymbol{h}_{i,l}^H \boldsymbol{w}_m\right|^2 \sum_{n=1}^{|\mathcal{K}_m|} p_{m,n}$ is denoted by P_{Inter}. Based on the above mathematical transformations, the original optimization problem OP1 can be reformulated as follows:

$$
OP3: \quad \max_{\{p_{i,j}\}} \alpha \quad (28a)
$$
$$
s.t. \quad (19)\ (20)\ (23)\ (22b)\ (25)
$$
$$
(24b)\ (27)\ (12c)\ (12d) \quad (28b)
$$

Especially, we point out that the constraint (13c) is convex and non-linear. By using SOC, we transform it into two new

constraints (24b) and (25). That's because with considering about the constraint (20), the solver for SOCP is faster than the average time of non-linear solver [29].

In order to solve the problem OP3, we choose iterative method. The initial iterative point can be got by solving a SPMin problem, which can be formulated as follows:

$$OP4: \quad \min_{\{p_{i,j}\}} \sum_{i=1}^{|\mathcal{B}|} \sum_{j=1}^{|\mathcal{K}_i|} p_{i,j} \tag{29a}$$

$$s.t. \quad R_{i,j} \geq R_{i,j}^{\min}, \tag{29b}$$

$$p_{i,j} \geq 0, \tag{29c}$$

$$\sum_{i=1}^{N_{RF}} \sum_{j=1}^{|S_i|} p_{i,j} \leq P_{bud}. \tag{29d}$$

By solving OP4, we can get the initial power allocation results defined as a $K \times 1$ vector p^0. We denote the initial point by $(\alpha^0, \beta^0, \eta^0, \vartheta^0)$. And β^0 can be calculated from (13c); based on (13b) and β^0, α^0 can be got; $\vartheta_{i,j}^0$ can be achieved by (22b); $\eta_{i,j}^0$ can be got through (22a).

After getting the initial point, we use iterative method to solve the problem OP2. We supposed that the initial point of $(t+1)$-th iteration is defined as $(\alpha^t, \beta^t, \eta^t, \vartheta^t)$. With the above initial point, we can get p^{t+1} and $(\alpha^t, \beta^t, \eta^t, \vartheta^t)$ by solving OP3. The iteration procedure continues until achieving the pre-defined accuracy ε or maximum number of iterations $T_{\max}$. The accuracy [3] ε is defined as the difference between two sequential optimal values (the square root of α). Especially, without the pre-defined accuracy, the algorithm will continue doing iteration with just few changes of optimal value (EE value). Above all, we summarize the procedure of the proposed iterative optimization method in **Algorithm 1**.

Algorithm 1 Proposed Iterative Optimization Algorithm

Input: $(\alpha^0, \beta^0, \eta^0, \vartheta^0)$
Output: The optimal power allocation result p^*.
1 **Initialization:** Set $t = 0$, by solving OP4, get the initial power allocation vector p^t, then generate the initial point $(\alpha^t, \beta^t, \eta^t, \vartheta^t)$;
2 **while** $\varepsilon \leq 0.01 \times 10^{-3}$ *or* $t \leq T_{max}$ **do**
3 $\quad$ $t = t + 1$;
4 $\quad$ Solve OP3, get the power allocation vector p^{t+1} and the next initial point $(\alpha^{t+1}, \beta^{t+1}, \eta^{t+1}, \vartheta^{t+1})$;
5 **end**

---

[3]The accuracy is set differently according to the unit of power. Based on OP2, the optimal EE value of each iteration is calculated by taking square root of α. If the unit of power is Watt (W), the accuracy ε is formulated as $\varepsilon = \sqrt{\alpha^{t+1}} - \sqrt{\alpha^t}$ and the termination criterion is expressed as $\varepsilon \leq 0.01$. Otherwise, if the unit is Milliwatt (mW), we can rewrite the termination criterion as $\varepsilon \leq 0.01 \times 10^{-3}$. That's because, comparing with Watt, the optimal EE of t-th $\sqrt{\alpha^t}$ should be multiplied by 10^3 under mW scenario, which ensures the unit of EE is bit/J/Hz. In this paper, we choose mW, so the adopted termination criterion is $\varepsilon \leq 0.01 \times 10^{-3}$.

IV. CONVERGENCE AND COMPUTATIONAL COMPLEXITY ANALYSIS

A. CONVERGENCE ANALYSIS

The reference [29] analyzed the convergence of its proposed SCA-based beamformer design for EEMax, where iterative optimization algorithm was introduced to solve the original optimization problem. About the optimality, [20] used the Dinkelbach's method as the benchmark to verify the optimality of the proposed algorithm; [29] proved that the solutions of iterative algorithm satisfy the Karush-Kuhn-Tucker conditions via simulations.

Before we analyze the convergence of the proposed iterative algorithm, we point out that the SCA transformation is a sequential parametric convex approximation (SPCA) type method [38]. Given that $\sqrt{\alpha\beta}$ in (18) and $(\eta_{i,j} - 1)\vartheta_{i,j}$ in (22a) are two nonconvex functions, the SCA transformation of this paper is a convex upper approximation[4]. That's because the first Taylor series expansion result is less or equal to the original nonconvex function.

For ensuing analysis, we use $\{\alpha^t\}$ to denote the set of optimal value of each iteration via the proposed algorithm; $g(\alpha, \beta)$ and $G(\alpha, \beta)$ denote the left part and the right part of (18), respectively; $SR_t = \sum_{i=1}^{|\mathcal{B}|} \sum_{j=1}^{|\mathcal{K}_i|} \delta_{i,j}^t$ denotes the sum rate of the t-th iteration, where $\delta_{i,j}^t$ is the optimal value of slack variable $\delta_{i,j}$ in the t-th iteration. According to OP2 and OP3, we can reformulate a new optimization problem, where the nonconvex constraints (17) and (22a) have not been transformed via SCA method. The new optimization OP5 is presented as follows:

$$OP5: \quad \min_{\{p_{i,j}\}} \quad \alpha \tag{30a}$$

$$s.t. \quad (17)\ (20)\ (22a)\ (22b)\ (25)$$
$$\quad (24b)\ (27)\ (12c)\ (12d). \tag{30b}$$

Based on the analyses in Section III, we point out that OP5 is equivalent to OP1.

According to the Lemma 2.2 and Corollary 2.3 in [38], we can get the conclusion that the sequence $\{\alpha^t\}$ is convergent. Next, we proof the above conclusion.

Proof

For $t \geq 0$, in the t-th iteration, the constraint (19) can be represented as

$$g(\alpha, \beta) \leq G(\alpha, \beta, \alpha^{t-1}, \beta^{t-1}) \leq SR_t. \tag{31}$$

We get α^t and β^t via solving the convex optimization problem OP3. By substituting α^t and β^t into (31), we can get the following result

$$g(\alpha^t, \beta^t) \leq G(\alpha^t, \beta^t, \alpha^{t-1}, \beta^{t-1}) \leq SR_t. \tag{32}$$

---

[4]We perform SCA transformation by using first order Taylor series expansion, which is a convex upper approximation of original nonconvex function in this paper. If the original function is concave, the SCA is different, where the original concave function is approximated from the lower bound [37].

TABLE 1. Optimal value (maximum EE) of each iteration.

System	EE of each iteration (bit/J/Hz)										
	1	2	3	4	5	6	7	8	9	10	11
BS-NOMA EEMax	1.4605	7.2685	18.5910	21.7543	22.7069	22.8309	22.8852	22.9111	22.9306	22.9429	22.9527

In the $(t+1)$-th iteration, (19) can be described as

$$g\left(\alpha, \beta\right) \leq G\left(\alpha, \beta, \alpha^t, \beta^t\right) \leq SR_{t+1}. \quad (33)$$

As the number of iterations increases, the SR meets $SR_t \leq SR_{t+1}$, which means that SR is a nondecreasing function. According to (32) and (33), we can get

$$g\left(\alpha^t, \beta^t\right) \leq SR_{t+1}, \quad (34)$$

which means that $\left(\alpha^t, \beta^t\right)$ is a feasible point of the optimization problem corresponding to the $(t+1)$-th iteration. Suppose that the maximum sum rate corresponding to the original non-convex optimization problem OP5 is SR^*, according to (34), we can get

$$g\left(\alpha^t, \beta^t\right) \leq SR_{t+1} \leq SR^*, \quad (35)$$

which means $\left(\alpha^t, \beta^t\right)$ is a feasible point of non-convex optimization problem OP5. By solving OP3, we can get that the next optimal value α^{t+1} meets $\alpha^{t+1} \geq \alpha^t$, which indicates that the optimal value is nondecreasing. Considering the bounded power constraint, our proposed energy-efficient power allocation design is convergent.

According to (32) and (35), we supposed that the iteration terminates in the n-th iteration, (32) can be reformulated as

$$g\left(\alpha^n, \beta^n\right) \leq G\left(\alpha^n, \beta^n, \alpha^{n-1}, \beta^{n-1}\right) \leq SR_n \leq SR^*, \quad (36)$$

where the equality can be established when $\left(\alpha^n, \beta^n\right) = \left(\alpha^{n-1}, \beta^{n-1}\right)$. With the above analyses, we can analyze the convergence of (23) in the same way.

About the optimality, according to the Proposition 3.2 in [38], the proposition clarified that if the convergent point generated by the SPCA method is a regular point, then the convergent point is a KKT point of original nonconvex problem. In this paper, it's hard to prove the strong convexity of the objective function and whether the convergent point α^n is a regular point or not, which are beyond the scope of this paper. However, according to (36), we can get that the solution to OP3 in each iteration is a suboptimal solution to the original non-convex optimization problem OP5.

B. COMPUTATIONAL COMPLEXITY ANALYSIS

The CVX toolbox uses interior-point algorithm for solving SOCP programming. According to [37] and [39], we can get that the computational complexity of interior algorithm for SOCP programming is based on the number of variables, constraints and constraint dimensions.

According to the optimization problem OP3, the total number of constraints is $4\left(1+K\right) + q_c$, where q_c is a constant related to the number of constraints which are from the relaxation of the exponential constraints [20].

Then, the total number of iterations to decrease the duality gap to a constant fraction of itself is bounded above by $\mathcal{O}\left(\log\left(\frac{1}{\varepsilon}\right)\sqrt{4\left(1+K\right)+q_c}\right)$, where ε is the required accuracy of the iterative algorithm. The amount of work per iteration is $\mathcal{O}\left(\left(4\left(1+K\right)+q_c\right)^2\left(5\left(1+K\right)+q_c\right)\right)$, where $4\left(1+K\right)+q_c$ and $5\left(1+K\right)+q_c$ are the total number of variables and constraint dimensions, respectively.

V. SIMULATION RESULTS

In this section, the performances of the proposed energy-efficient power allocation method for mmWave beamspace MIMO-NOMA scheme are evaluated through numerical simulations. In this paper, we just consider a single-cell downlink communication system, where the base station has $N = 256$ antennas and the antenna spacing is half of wavelength. We define $\epsilon = 1$. The number of selected RF chains is variable according to the different transmission strategies (schemes), which are based on different beam selection methods. For example, if we adopt full digital (FD) transmission strategy, the number of selected RF chains is $N_{RF} = N = 256$; if we choose BS strategy, the number is $N_{RF} \leq K$, which has been discussed in the above section. We suppose that the base station has perfect beamspace CSI of all users. The parameters of the proposed iterative algorithm are identical to **Algorithm 1**.

A. CONVERGENCE OF PROPOSED POWER ALLOCATION METHOD

In order to evaluate the convergence of the proposed iterative algorithm, we set $K = 20$, transmitted power[5] $P_{tr} = 25dBm$ and $SNR = 10dB$. Specially, in order to show the convergence property, we just set the termination criterion as $T_{max} = 30$. The optimal values of each iteration are showed in TABLE 1, where we just display the results of preceding eleven iterations. Fig. 1 demonstrates the convergence of our proposed algorithm. According to TABLE 1, with pre-defined accuracy $\varepsilon \leq 0.01 \times 10^{-3}$, we can find that the iteration will terminate in the 11-th iteration.

B. SE AND EE AGAINST TRANSMITTED POWER

We study the performance of our proposed energy-efficient power allocation method against different transmitted power P_{tr}. For comparing, we also study the power allocation method with maximizing SR as the criterion and analyze the fixed power allocation under FD system and BS-OMA

---

[5]In the following simulations, we substitute transmitted power P_{tr} for power budget P_{bud}. For maximizing the sum rate, we know that FD, BS-OMA and BS-NOMA SRMax allocate all the transmitted power. However, the proposed power allocation method just consumes a portion of transmitted power when P_{tr} is higher than a power threshold, which is defined as an abscissa value of inflection point.
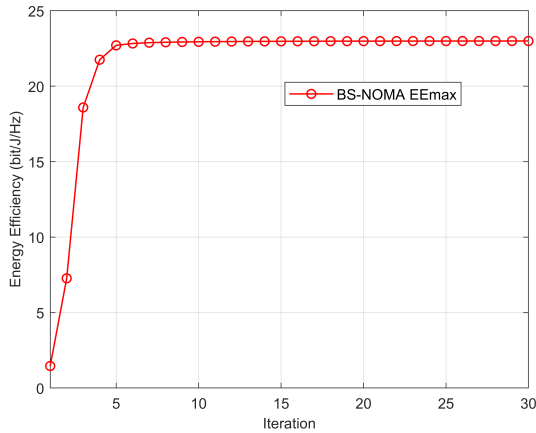
FIGURE 1. The convergence of the proposed iterative optimization algorithm, where $K = 20$, $P_{tr} = 25dBm$, $SNR = 10dB$.
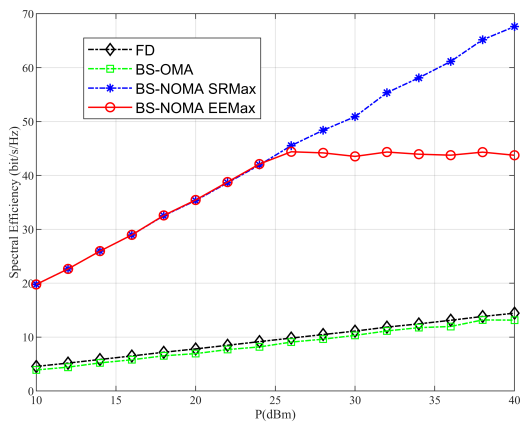

FIGURE 2. The SE against transmitted power, where $K = 5$, $SNR = 10dB$.

FIGURE 3. The EE against transmitted power, where $K = 5$, $SNR = 10dB$.

FIGURE 4. The SE against transmitted power, where $K = 20$, $SNR = 10dB$.

FIGURE 5. The EE against transmitted power, where $K = 20$, $SNR = 10dB$.

system. In this part, we set $K_1 = 5$ and $K_2 = 20$. The simulations are performed under $SNR_1 = 10dB$.

From Fig. 2, we find that when the transmitted power is high enough, the SEs of the proposed energy-efficient power allocation schemes (BS-NOMA EEMax) remain stable, where "high enough" means that the transmitted power is bigger than the power threshold (the abscissa value of inflection point of asterisked curve in Fig. 3). The asterisked curve denotes the EE of BS-NOMA SRMax.

However, the SEs of the other three schemes augment with the increase of transmitted power. Especially, the BS-NOMA EEMax and BS-NOMA SRMax have the nearly same SE performance when the transmitted power is lower, which means that P_{tr} is smaller than the power threshold. Fig. 3 depicts that the EE of the proposed scheme remains stable when the transmitted power is high enough. The BS-NOMA SRMax increases and then goes down when the transmitted is enough high.

From Fig. 5, we can find that the inflection point of asterisked curve shifts to the right side, which means that the power threshold augments. That's to say, as the number of users increases, more transmitted power is required to reach the maximum EE point. Comparing with Fig 2,
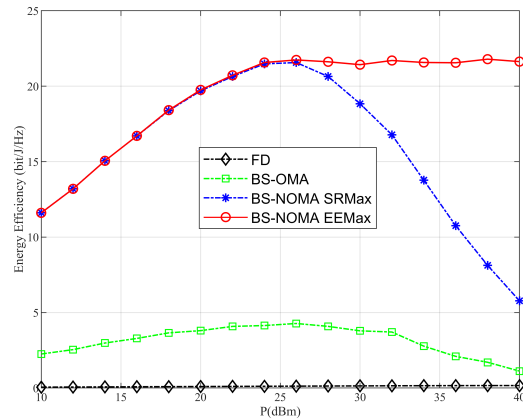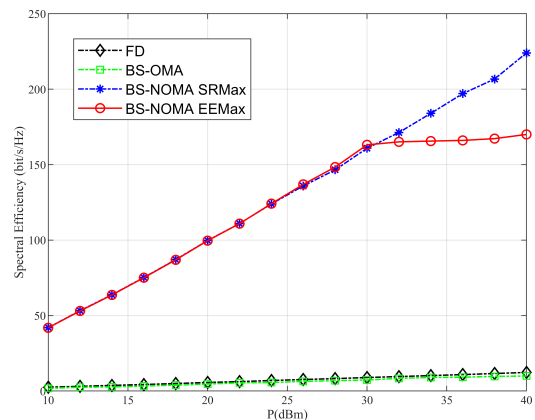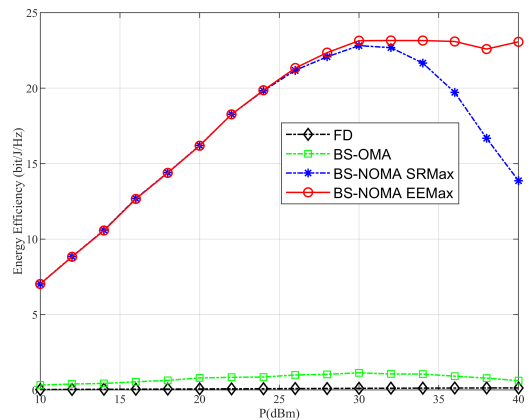
the Fig. 4 depicts that the SE value of inflection point of BS-NOMA EEMax is much bigger than that in Fig. 2. However, according to Fig. 3 and Fig. 5, we can find that the EE gap between maximum EE of BS-NOMA EEMax in Fig. 3 and that in Fig. 5 is small. That's because as the number of user increases, more RF chains are required to be activated. The power consumption of the circuits can't be overlooked, which is the main reason of carrying out beam

TABLE 2. Power allocation results (mW), where $K = 20$, $SNR = 10dB$, $P_{tr} = 40dBm$.

Systems	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	User 9	User 10
BS-NOMA SRMax	494.123	345.505	616.247	770.529	480.056	351.115	748.712	588.414	359.369	600.788
BS-NOMA EEMax	55.153	2.702	67.513	56.429	60.055	61.678	61.699	61.863	62.052	62.013
Systems	User 11	User 12	User 13	User 14	User 15	User 16	User 17	User 18	User 19	User 20
BS-NOMA SRMax	355.538	598.450	350.948	383.682	464.413	535.259	728.652	381.091	343.906	503.195
BS-NOMA EEMax	62.061	61.992	62.070	62.061	61.952	61.953	61.994	61.987	61.534	61.241
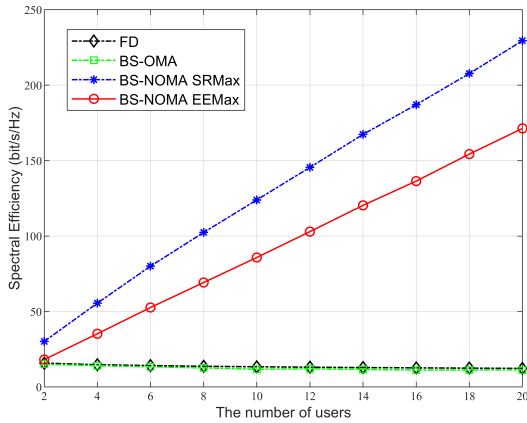
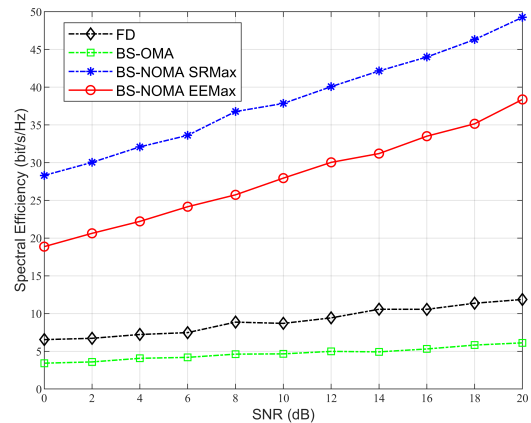**FIGURE 6.** The SE against number of users, where $P_{tr} = 40dBm$, $SNR = 10dB$.
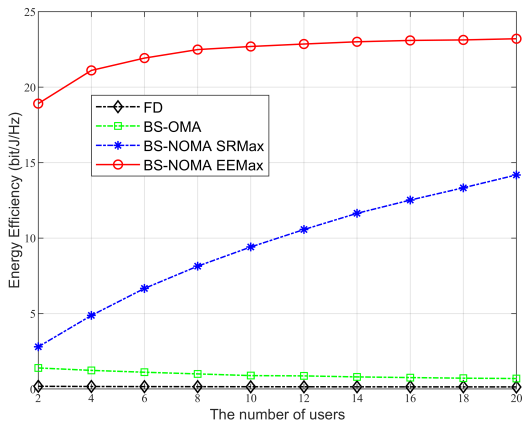

FIGURE 7. The EE against number of users, where $P_{tr} = 40dBm$, $SNR = 10dB$.

selection/antenna selection. The power exhausted by RF circuits can degrade the EE performance of communication system.

C. SE AND EE AGAINST NUMBER OF USERS

In this section, we evaluate the performance of the proposed energy-efficient power allocation scheme under different number of users. Fig. 6 depicts that the SEs of FD and BS-OMA remain almost unchanged, while the other two schemes augment as the number of users increases. Fig. 7 indicates that the proposed power allocation scheme has best EE performance comparing with the other three schemes. As shown in Fig. 5, we find that the power threshold

FIGURE 8. The SE under dense user scenario, where $K = 20$, $P_{tr} = 35dBm$, $SNR = 10dB$.

is $P_{tr} = 30dBm$. Therefore, the proposed scheme has better EE performance when the transmitted power is $P_{tr} = 40dBm$.

D. POWER ALLOCATION RESULTS UNDER DIFFERENT SCHEMES

We study on the power allocation results of BS-NOMA SRMax, BS-NOMA EEMax. TABLE 2 depicts that the BS-NOMA SRMax consumes the total transmitted power to maximize the SR, while BS-NOMA EEMax just uses a portion of the transmitted power. That's because the transmitted power exceeds the power threshold, where the power threshold is $30dBm$ as depicted in Fig. 5 and the transmitted power in TABLE 2 is $40dBm$. In the BS-NOMA schemes (BS-NOMA SRMax and BS-NOMA EEMax), User 2 and User 3 are in the same beamspace and there exists power gap between the above two users.

E. SIMULATION UNDER DENSE USER SCENARIO

In this simulation, we consider a dense user scenario, where there is a high probability that several users may be in one selected beam. Therefore, different from the previous CSI setup, we suppose that several users have the same spatial direction information of LoS path but with different path gain, and NLoS components of each user can modeled according to the corresponding LoS path. The constructed dense user scenario is suitable for large-scale conference room, sports center and so on, where multiple users may locate in the same beam.

From Fig. 8 and Fig. 9, we can find that under dense user scenario, the BS-NOMA EEMax scheme has better
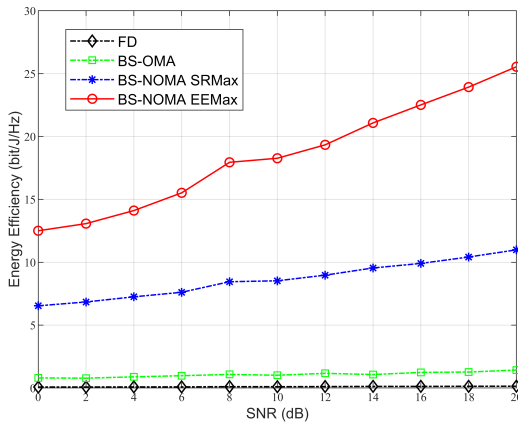
FIGURE 9. The EE under dense user scenario, where $K = 20$, $P_{tr} = 35dBm$, $SNR = 10dB$.

EE performance comparing with the BS-NOMA SRMax, although the latter scheme outperforms the former scheme in terms of SE. The transmitted power is $35dBm$, which exceeds the power threshold corresponding to 6 users. Hence, in terms of EE, the BS-NOMA EEMax performs better than the other three schemes.

VI. CONCLUSION

In this paper, we focus on the energy-efficient power allocation for mmWave beamspace MIMO-NOMA communication systems. According to the BS results, we get the precoding matrix trough ZF beamforming technique. The EEMax problem is formulated as a fractional programming problem. By using SCA and SOC transformation, the original fractional optimization problem is transformed into a convex optimization problem, which is solved through iterative optimization algorithm. Then, we analyze the convergence of the proposed iterative optimization algorithm. Simulation results show that the proposed power allocation scheme has same EE performance with the BS-NOMA SRMax when the transmitted power is lower than the power threshold; the proposed scheme has better EE performance when the transmitted power exceeds the threshold.

REFERENCES

[1] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2294–2323, 3rd Quart.,2018.

[2] L. Zhang, M. Xiao, G. Wu, M. Alam, Y.-C. Liang, and S. Li, "A survey of advanced techniques for spectrum sharing in 5G networks," *IEEE Wireless Commun.*, vol. 24, no. 5, pp. 44–51, Oct. 2017.

[3] 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.

[4] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Nonorthogonal multiple access for 5G and beyond," *Proc. IEEE*, vol. 105, no. 12, pp. 2347–2381, Dec. 2017.

[5] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, C.-L. I, and H. V. Poor, "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 185–191, Feb. 2017.

[6] Y. Liu, H. Xing, C. Pan, A. Nallanathan, M. Elkashlan, and L. Hanzo, "Multiple-antenna-assisted non-orthogonal multiple access," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 17–23, Apr. 2018.

[7] X. Chen, R. Jia, and D. W. K. Ng, "The application of relay to massive non-orthogonal multiple access," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5168–5180, Nov. 2018.

[8] C. Pan, M. Elkashlan, J. Wang, J. Yuan, and L. Hanzo, "User-centric C-RAN architecture for ultra-dense 5G networks: Challenges and methodologies," *IEEE Commun. Mag.*, vol. 56, no. 6, pp. 14–20, Jun. 2018.

[9] Y. Li, Y. Zhang, K. Luo, T. Jiang, Z. Li, and W. Peng, "Ultra-dense HetNets meet big data: Green frameworks, techniques, and approaches," *IEEE Commun. Mag.*, vol. 56, no. 6, pp. 56–63, Jun. 2018.

[10] M. F. Hanif, Z. Ding, T. Ratnarajah, and G. K. Karagiannidis, "A minorization-maximization method for optimizing sum rate in the downlink of non-orthogonal multiple access systems," *IEEE Trans. Signal Process.*, vol. 64, no. 1, pp. 76–88, Jan. 2015.

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

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

[13] S. Ali, E. Hossain, and D. I. Kim, "Non-orthogonal multiple access (NOMA) for downlink multiuser MIMO systems: User clustering, beamforming, and power allocation," *IEEE Access*, vol. 5, pp. 565–577, 2016.

[14] J. Cui, Z. Ding, P. Fan, and N. Al-Dhahir, "Unsupervised machine learning-based user clustering in millimeter-wave-NOMA systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7425–7440, Nov. 2018.

[15] X. Shao, C. Yang, D. Chen, N. Zhao, and F. R. Yu, "Dynamic IoT device clustering and energy management with hybrid NOMA systems," *IEEE Trans. Ind. Informat.*, vol. 14, no. 10, pp. 4622–4630, Oct. 2018.

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

[17] 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.

[18] H. Zhang, B. Wang, C. Jiang, K. Long, A. Nallanathan, V. C. M. Leung, and H. V. Poor, "Energy efficient dynamic resource optimization in NOMA system," *IEEE Trans. Wireless Commun.*, vol. 17, no. 9, pp. 5671–5682, Sep. 2018.

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

[20] H. M. Al-Obiedollah, K. Cumanan, J. Thiyagalingam, A. G. Burr, Z. Ding, and O. A. Dobre, "Energy efficient beamforming design for miso non-orthogonal multiple access systems," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4117–4131, Jun. 2019.

[21] L. D. Nguyen, H. D. Tuan, T. Q. Duong, O. A. Dobre, and H. V. Poor, "Downlink beamforming for energy-efficient heterogeneous networks with massive MIMO and small cells," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 3386–3400, May 2018.

[22] I. B. Sofi and A. Gupta, "A survey on energy efficient 5G green network with a planned multi-tier architecture," *J. Netw. Comput. Appl.*, vol. 118, pp. 1–28, Sep. 2018.

[23] H. Zhang, F. Fang, J. Cheng, K. Long, W. Wang, and V. C. M. Leung, "Energy-efficient resource allocation in NOMA heterogeneous networks," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 48–53, Apr. 2018.

[24] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Trans. Antennas Propag.*, vol. 61, no. 7, pp. 3814–3827, Jul. 2013.

[25] X. Gao, L. Dai, Z. Chen, Z. Wang, and Z. Zhang, "Near-optimal beam selection for beamspace mmWave massive MIMO systems," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 1054–1057, May 2016.

[26] B. Wang, L. Dai, Z. Wang, N. Ge, and S. Zhou, "Spectrum and energy-efficient beamspace MIMO-NOMA for millimeter-wave communications using lens antenna array," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2370–2382, Oct. 2017.

[27] L. Dai, B. Wang, M. Peng, and S. Chen, "Hybrid precoding-based millimeter-wave massive MIMO-NOMA with simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 1, pp. 131–141, Jan. 2019.

[28] R. Pal, A. K. Chaitanya, and K. V. Srinivas, "Low-complexity beam selection algorithms for millimeter wave beamspace MIMO systems," *IEEE Commun. Lett.*, vol. 23, no. 4, pp. 768–771, Apr. 2019.

[29] O. Tervo, L.-N. Tran, and M. Juntti, "Optimal energy-efficient transmit beamforming for multi-user MISO downlink," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5574–5588, Oct. 2015.

[30] X. Sun, N. Yang, S. Yan, Z. Ding, D. W. K. Ng, C. Shen, and Z. Zhong, "Joint beamforming and power allocation in downlink NOMA multiuser MIMO networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5367–5381, Aug. 2018.

[31] F. Alavi, K. Cumanan, Z. Ding, and A. G. Burr, "Beamforming techniques for nonorthogonal multiple access in 5G cellular networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9474–9487, Oct. 2018.

[32] Z.-Q. Luo and W. Yu, "An introduction to convex optimization for communications and signal processing," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1426–1438, Aug. 2006.

[33] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[34] M. Grant and S. Boyd. (Dec. 2018). *The CVX Users' Guide. Version 2.1*. [Online]. Available: http://web.cvxr.com/cvx/doc/

[35] J. Cui, Y. Liu, Z. Ding, P. Fan, and A. Nallanathan, "Optimal user scheduling and power allocation for millimeter wave NOMA systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1502–1517, Mar. 2017.

[36] Z. Wei, D. W. K. Ng, J. Yuan, and H. M. Wang, "Optimal resource allocation for power-efficient MC-NOMA with imperfect channel state information," *IEEE Trans. Commun.*, vol. 65, no. 9, pp. 3944–3961, Sep. 2017.

[37] N. Zhao, W. Wang, J. Wang, Y. Chen, Y. Lin, Z. Ding, and N. C. Beaulieu, "Joint beamforming and jamming optimization for secure transmission in MISO-NOMA networks," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 2294–2305, Mar. 2019.

[38] A. Beck, A. Ben-Tal, and L. Tetruashvili, "A sequential parametric convex approximation method with applications to nonconvex truss topology design problems," *J. Global Optim.*, vol. 47, no. 1, pp. 29–51, May 2010.

[39] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra Appl.*, vol. 284, nos. 1–3, pp. 193–228, Nov. 1998.

YONG LI received the B.S. degree in avionics engineering and the M.S. and Ph.D. degrees in circuits and systems from Northwestern Polytechnical University, Xi'an, China, in 1983, 1988, and 2005, respectively, where he joined the School of Electronic Information, in 1993, and was promoted to a Professor, in 2002. His research interests include digital signal processing and radar signal processing.

WEI CHENG received the B.S. degree in electronic and information engineering and the M.S. and Ph.D. degrees in communication and information system from Northwestern Polytechnical University, Xi'an, China, in 2003, 2006, and 2011, respectively. From April 2011 to April 2013, he worked in the Postdoctoral Research Station of the School of Electronic Information, Northwestern Polytechnical University, where he has been a Lecturer, since 2013, and was promoted to an Associate Professor, in 2015. His research interests include wireless sensor networks and ad hoc networks, and radar signal processing.

WENJIE ZHANG received the B.S. degree in electronic and information engineering from Yangtze University, Jinzhou, China, in 2011, and the M.S. degree in communication and information system from Guizhou University, Guiyang, China, in 2015. He is currently pursuing the Ph.D. degree in electronic science and technology with Northwestern Polytechnical University, Xi'an, China. His research interests include wireless communications, information theory, massive MIMO, Millimeter wave communication, and signal processing for communication.

PENGLU LIU received the M.S. degree from the Xi'an University of Posts and Telecommunications, China, in 2017. He is currently pursuing the Ph.D. degree with Northwestern Polytechnical University. His current research interests include resource allocation of wireless communication networks, millimeter wave communication, and mobile edge computing.

XIANG GAO received the B.S. degree in communication engineering from the North University of China, Taiyuan, China, in 2017. He is currently pursuing the M.S. degree in communication and information system with Northwestern Polytechnical University, Xi'an, China. His research interests include physical layer security, signal processing, convex optimization, and its applications.

• • •