Beamspace MIMO-NOMA for Millimeter-Wave Communications Using Lens Antenna Arrays

Bichai Wang[1], Linglong Dai[1], Xiqi Gao[2], and Lajos Hanzo[3]

[1]Tsinghua National Laboratory for Information Science and Technology (TNList), Tsinghua University, Beijing 100084, China
[2]National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China
[3]Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K.
E-mail: wbc15@mails.tsinghua.edu.cn, daill@tsinghua.edu.cn, xqgao@seu.edu.cn, lh@ecs.soton.ac.uk

***Abstract*—The recent concept of beamspace multiple-input multiple-output (MIMO) is capable of significantly reducing the number of radio-frequency (RF) chains required by millimeter-wave (mmWave) massive MIMO systems. However, the fundamental limit of the existing beamspace MIMO is that, the number of supported users cannot be higher than the number of RF chains using the same time-frequency resources. To break this limit, beamspace MIMO is integrated with non-orthogonal multiple access (NOMA) in the proposed MIMO-NOMA system in this paper, where the number of supported users can be higher than the number of RF chains. To reduce the inter-beam interference, a transmit precoding (TPC) scheme based on the principle of zero-forcing (ZF) is designed. Furthermore, a dynamic power allocation scheme is proposed for maximizing the achievable sum rate. Moreover, a low-complexity iterative optimization algorithm is conceived for dynamic power allocation. Simulation results show that the proposed beamspace MIMO-NOMA achieves a higher spectrum and energy efficiency than the existing beamspace MIMO for mmWave communications.**

I. Introduction

Millimeter-wave (mmWave) massive multiple-input multiple-output (MIMO) is able to significantly improve the data rate in future 5G communication systems [1]. However, it is a big challenge to realize mmWave massive MIMO in practice due to high transceiver complexity and energy consumption caused by a large number of radio-frequency (RF) chains [2]. The recent concept of beamspace MIMO can significantly reduce the number of RF chains by using a lens antenna array. Nevertheless, a fundamental limit of beamspace MIMO is that, each RF chain can only support a single user using the same time-frequency resources, hence the maximum number of users that can be supported cannot exceed the number of RF chains [3]–[5].

In this paper, we break this fundamental limit by proposing a spectrum- and energy-efficient mmWave transmission scheme that integrates beamspace MIMO with non-orthogonal multiple access (NOMA), i.e., beamspace MIMO-NOMA[1]. Specifically, NOMA performs superposition coding at the transmitter and successive interference cancellation (SIC) at the receiver to realize non-orthogonal multiplexing [6]–[8]. By integrating NOMA into beamspace MIMO, where the intra-beam superposition coding is performed, more than one user can be simultaneously supported by a single beam, which is more efficient than beamspace MIMO using a single beam to serve a single user. Thus, the number of supported users can be higher than the number of beams. To reduce the inter-beam interference, the equivalent channel vector of each beam is determined to realize the transmit precoding (TPC) based on the principle of zero-forcing (ZF). Furthermore, a dynamic power allocation scheme is proposed for jointly optimizing the power assigned to different users by maximizing the achievable sum rate, and an iterative optimization algorithm is developed for power allocation. Our simulation results show that the proposed beamspace MIMO-NOMA achieves a higher spectrum and energy efficiency than conventional beamspace MIMO for mmWave communications.

The rest of this paper is organized as follows. The system model of beamspace MIMO is introduced in Section II, and Section III discusses the proposed beamspace MIMO-NOMA. Simulation results are provided in Section IV. Finally, conclusions are drawn in Section V.

II. System Model of Beamspace MIMO

We consider a single-cell downlink mmWave communication system, where the base station (BS) is equipped with a lens antenna array having N antennas as well as N_{RF} RF chains, and K single-antenna users are simultaneously served by the BS [4] [5]. By employing a lens antenna array in beamspace MIMO, the spatial channel can be transformed to the beamspace channel [4].

Specifically, the mathematical function of the lens antenna array is to realize the spatial discrete Fourier transformation with the aid of the $N \times N$ transform matrix $\mathbf{U}$ [5], which contains the array's steering vectors of N directions covering the entire space as follows:

$$\mathbf{U} = \left[\mathbf{a}\left(\bar{\theta}_1\right), \mathbf{a}\left(\bar{\theta}_2\right), \cdots, \mathbf{a}\left(\bar{\theta}_N\right)\right]^H, \tag{1}$$

where $\bar{\theta}_n = \frac{1}{N}\left(n - \frac{N+1}{2}\right)$ for $n = 1, 2, \cdots, N$ are the predefined spatial directions.

[1]Simulation codes are provided to reproduce the results presented in this paper: http://oa.ee.tsinghua.edu.cn/dailinglong/publications/publications.html.

978-1-5090-5935-5/17/$31.00 ©2017 IEEE

Then, the received signal vector $\bar{\mathbf{y}}$ in the downlink can be represented as

$$\bar{\mathbf{y}} = \mathbf{H}^H \mathbf{U}^H \mathbf{W}\mathbf{P}\mathbf{s} + \mathbf{v} = \bar{\mathbf{H}}^H \mathbf{W}\mathbf{P}\mathbf{s} + \mathbf{v}, \tag{2}$$

where $\mathbf{s} = [s_1, s_2, \cdots, s_K]^T$ is the K-element transmitted signal vector for all K users with a normalized power of $\mathrm{E}\left(\mathbf{s}\mathbf{s}^H\right) = \mathbf{I}_K$, $\mathbf{P} = \mathrm{diag}\{\mathbf{p}\}$ includes the transmitted power of all K users with $\mathbf{p} = \left[\sqrt{p_1}, \sqrt{p_2}, \cdots, \sqrt{p_K}\right]$ satisfying $\sum_{k=1}^{K} p_k \leq P$ (the maximum transmitted power at the BS), $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_K]$ is the $(N \times K)$-element TPC matrix with $\|\mathbf{w}_k\|_2 = 1$ for $k = 1, 2, \cdots, K$, and $\mathbf{v}$ is the noise vector obeying the complex Gaussian distribution $\mathcal{CN}\left(0, \sigma^2 \mathbf{I}_K\right)$. $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_K]$ of size $(N \times K)$ is the channel matrix, where $\mathbf{h}_k$ of size $(N \times 1)$ denotes the spatial channel vector between the BS and the kth user.

In this paper, we consider the widely used Saleh-Valenzuela channel model for mmWave communications [3]–[5], hence $\mathbf{h}_k$ can be represented as

$$\mathbf{h}_k = \beta_k^{(0)} \mathbf{a}\left(\theta_k^{(0)}\right) + \sum_{l=1}^{L} \beta_k^{(l)} \mathbf{a}\left(\theta_k^{(l)}\right), \tag{3}$$

where $\beta_k^{(0)} \mathbf{a}(\theta_k^{(0)})$ is the line-of-sight (LoS) component of the kth user with $\beta_k^{(0)}$ being the complex-valued gain and $\mathbf{a}(\theta_k^{(0)})$ being the array's steering vector. Furthermore, $\beta_k^{(l)} \mathbf{a}(\theta_k^{(l)})$ for $1 \leq l \leq L$ is the lth non-line-of-sight (NLoS) component of the kth user, where L is the total number of NLoS components. Finally, the beamspace channel matrix $\bar{\mathbf{H}}$ is defined as

$$\bar{\mathbf{H}} = \mathbf{U}\mathbf{H} = [\mathbf{U}\mathbf{h}_1, \mathbf{U}\mathbf{h}_2, \cdots, \mathbf{U}\mathbf{h}_K] = \left[\bar{\mathbf{h}}_1, \bar{\mathbf{h}}_2, \cdots, \bar{\mathbf{h}}_K\right], \tag{4}$$

where $\bar{\mathbf{h}}_k = \mathbf{U}\mathbf{h}_k$ is the beamspace channel vector between the BS and the kth user, which is actually the Fourier transformation of the spatial channel vector $\mathbf{h}_k$ in (3).

It has been shown for mmWave communications that the number of dominant elements of each beamspace channel vector $\bar{\mathbf{h}}_k$ is much smaller than N, since the number of dominant scatters is limited [5]. In this case, the beamspace channel matrix $\bar{\mathbf{H}}$ has a sparse nature [4], which can be exploited for reducing the number of RF chains without obvious performance loss by beam selection. Specifically, according to the sparse beamspace channel matrix, only a small number of beams can be selected for simultaneously serving K users. Then, the received signal vector in (2) can be rewritten as

$$\bar{\mathbf{y}} = \bar{\mathbf{H}}_r^H \mathbf{W}_r \mathbf{P}\mathbf{s} + \mathbf{v}, \tag{5}$$

where $\bar{\mathbf{H}}_r = \bar{\mathbf{H}}(i,:)_{i \in \Gamma}$ of size $(|\Gamma| \times K)$ is the dimension-reduced beamspace channel matrix including the selected beams, and Γ is the index set of selected beams. Furthermore, $\mathbf{W}_r$ of size $(|\Gamma| \times K)$ is the dimension-reduced TPC matrix. Since the row dimension of $\mathbf{W}_r$ is much smaller than N, the number of required RF chains can be significantly reduced, and we have $N_{\mathrm{RF}} = |\Gamma|$ [5].

However, a single beam can only support a single user in existing beamspace MIMO systems [3]–[5], otherwise signal for different users cannot be separated by linear operations, which is a fundamental limit of beamspace MIMO. Note that linear operations have been considered in almost all of massive MIMO systems due to high complexity caused by non-linear operations. To break this limit, in the next section we propose a new transmission scheme called beamspace MIMO-NOMA that integrates NOMA with beamspace MIMO for mmWave communications.

III. Proposed Beamspace MIMO-NOMA

In this section, the basic principle of the proposed beamspace MIMO-NOMA will be introduced at first. Then, the corresponding TPC scheme and the dynamic power allocation scheme will be discussed, separately.

A. The basic principle of beamspace MIMO-NOMA

In the proposed beamspace MIMO-NOMA system, beam selection algorithms, such as the maximum magnitude (MM) selection and signal-to-interference-plus-noise ratio (SINR) maximization based selection [5] for the existing beamspace MIMO, can be used for selecting one beam for each user, where each RF chain corresponds to a single beam. Note that different users may select the same beam, which are termed as "interfering users" in this paper. In contrast to the existing beamspace MIMO systems, where user scheduling is performed to select only one user out of these interfering users [5], they can be simultaneously served within the same beam in the proposed beamspace MIMO-NOMA system. Thus, although the number of selected beams is equal to N_{RF}, the number of users K can be higher than N_{RF}, i.e., $K \geq N_{\mathrm{RF}}$.

Let S_n for $n = 1, 2, \cdots, N_{\mathrm{RF}}$ denote the set of users served by the nth beam with $S_i \cap S_j = \Phi$ for $i \neq j$ and $\sum_{n=1}^{N_{\mathrm{RF}}} |S_n| = K$. The beamspace channel vector with N_{RF} elements between the BS and the mth user in the nth beam after beam selection is denoted by $\mathbf{h}_{m,n}$, and $\mathbf{w}_n$ of size $(N_{\mathrm{RF}} \times 1)$ denotes the uniform TPC vector invoked for users in the nth beam. Without loss of generality, we assume that $\left\|\mathbf{h}_{1,n}^H \mathbf{w}_n\right\|_2 \geq \left\|\mathbf{h}_{2,n}^H \mathbf{w}_n\right\|_2 \geq \cdots \geq \left\|\mathbf{h}_{|S_n|,n}^H \mathbf{w}_n\right\|_2$ for $n = 1, 2, \cdots, N_{\mathrm{RF}}$. Then, in the nth beam using NOMA with SIC, the ith ($i > m$) user's signal is detectable at the mth user, provided that it is detectable at itself [6]. Therefore, the mth user can detect the ith user's signals for $1 \leq m < i \leq |S_n|$, and then remodulate and remove the detected signals from its received signals, in a successive manner [6]. Then, the remaining signal received at the mth user in the nth beam can be written as

$$\begin{aligned} \hat{y}_{m,n} =& \mathbf{h}_{m,n}^H \mathbf{w}_n \sqrt{p_{m,n}} s_{m,n} + \mathbf{h}_{m,n}^H \mathbf{w}_n \sum_{i=1}^{m-1} \sqrt{p_{i,n}} s_{i,n} \\ &+ \mathbf{h}_{m,n}^H \sum_{j \neq n} \sum_{i=1}^{|S_j|} \mathbf{w}_j \sqrt{p_{i,j}} s_{i,j} + v_{m,n}, \end{aligned} \tag{6}$$

where $s_{m,n}$ and $p_{m,n}$ are the transmitted signal and the power transmitted to the mth user in the nth beam, while $v_{m,n}$ is the noise obeying the complex Gaussian distribution $\mathcal{CN}\left(0,\sigma^2\right)$.

Then, according to (6), the SINR at the mth user in the nth beam can be represented as

$$\gamma_{m,n}=\frac{\left\|\mathbf{h}_{m,n}^H\mathbf{w}_n\right\|_2^2 p_{m,n}}{\xi_{m,n}}, \tag{7}$$

where we have

$$\xi_{m,n}=\left\|\mathbf{h}_{m,n}^H\mathbf{w}_n\right\|_2^2\sum_{i=1}^{m-1}p_{i,n}+\sum_{j\neq n}\left\|\mathbf{h}_{m,n}^H\mathbf{w}_j\right\|_2^2\sum_{i=1}^{|S_j|}p_{i,j}+\sigma^2. \tag{8}$$

As a result, the achievable rate of the mth user in the nth beam is

$$R_{m,n}=\log_2\left(1+\gamma_{m,n}\right). \tag{9}$$

Finally, the achievable sum rate of the proposed beamspace MIMO-NOMA system is

$$R_{\text{sum}}=\sum_{n=1}^{N_{\text{RF}}}\sum_{m=1}^{|S_n|}R_{m,n}, \tag{10}$$

which can be improved by carefully designing the precoding $\{\mathbf{w}_n\}_{n=1}^{N_{\text{RF}}}$ and power allocation $\{p_{m,n}\}_{m=1,n=1}^{|S_n|,N_{\text{RF}}}$, which will be introduced in the next two subsections, separately.

B. Precoding Design

To reduce the inter-beam interference, TPC should be carefully designed. In existing beamspace MIMO systems, where only a single user can be served in each beam (i.e., $K\leq N_{\text{RF}}$), the low-complexity linear ZF TPC can be utilized for removing the inter-beam interference [4] [5], which can be simply realized by applying the pseudo-inverse of the beamspace channel matrix to all users. However, in the proposed beamspace MIMO-NOMA system, the number of users is higher than the number of beams, i.e., $K\geq N_{\text{RF}}$, which means that the pseudo-inverse of the beamspace channel matrix of size $(N_{\text{RF}}\times K)$ does not exist. As a result, the conventional ZF precoding cannot be directly used.

To address this problem, an equivalent channel can be derived for each beam for generating the TPC vector. Specifically, since the LoS component is dominant in the mmWave channel and the beamspace channel matrix has a sparse structure representing the directions of different users [5], the beamspace channel vectors of different users in the same beam are highly correlated. Hence, one of the beamspace channel vectors of users multiplexed in the nth beam can be regarded as the equivalent channel vector of the nth beam, and this user will not suffer from substantial inter-beam interference. More particularly, leaning in mind that the first user in each beam should perform SIC to decode all the other users' signals in this beam, we use the beamspace channel vector of the first user in each beam as the equivalent channel vector. Specifically, the equivalent channel matrix of size $(N_{\text{RF}}\times N_{\text{RF}})$ for all N_{RF} beams can be written as

$$\tilde{\mathbf{H}}=\left[\mathbf{h}_{1,1},\mathbf{h}_{1,2},\cdots,\mathbf{h}_{1,N_{\text{RF}}}\right]. \tag{11}$$

Then, the TPC matrix of size $(N_{\text{RF}}\times N_{\text{RF}})$ can be generated by

$$\tilde{\mathbf{W}}=[\tilde{\mathbf{w}}_1,\tilde{\mathbf{w}}_2,\cdots,\tilde{\mathbf{w}}_{N_{\text{RF}}}]=\left(\tilde{\mathbf{H}}\right)^{\dagger}=\tilde{\mathbf{H}}\left(\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}\right)^{-1}. \tag{12}$$

After normalizing the TPC vectors, the precoding vector for the nth beam $(n=1,2,\cdots,N_{\text{RF}})$ can be written as

$$\mathbf{w}_n=\frac{\tilde{\mathbf{w}}_n}{\|\tilde{\mathbf{w}}_n\|_2}. \tag{13}$$

C. Dynamic Power Allocation

After obtaining the TPC vectors in (13), the power allocation should be optimized. Similar to the existing MIMO-NOMA systems [7] [8], both the inter-beam interference and intra-beam interference should be reduced to improve the achievable sum rate. However, in contrast to the existing MIMO-NOMA systems, where fixed inter-beam power allocation and only two users per beam are usually considered, multiple users (e.g., 1, 2, or 3 users) are allowed in each beam in the proposed beamspace MIMO-NOMA system. Accordingly, the dynamic power allocation is proposed for maximizing the achievable sum rate by solving the joint power optimization problem, which includes both the intra-beam power optimization and the inter-beam power optimization. Specifically, the power allocation problem can be formulated as

$$\begin{aligned}&\max_{\{p_{m,n}\}}\sum_{n=1}^{N_{\text{RF}}}\sum_{m=1}^{|S_n|}R_{m,n},\\ &\text{s.t. } C_1:\ p_{m,n}\geq 0,\quad \forall n,m,\\ &\qquad C_2:\ \sum_{n=1}^{N_{\text{RF}}}\sum_{m=1}^{|S_n|}p_{m,n}\leq P,\\ &\qquad C_3:\ R_{m,n}\geq R_{\min},\quad \forall n,m,\end{aligned} \tag{14}$$

where $R_{m,n}$ is the achievable rate of the mth user in the nth beam, as defined in (9). Furthermore, the constraint C_1 indicates that the power allocated to each user must be positive, C_2 is the transmit power constraint with P being the maximum total power transmitted by the BS, and C_3 is the data rate constraint for each user with $R_{\min}$ being the minimum data rate of each user. By substituting (7)-(9) into the constraint C_3 in (14), we have

$$\begin{aligned}&\left\|\mathbf{h}_{m,n}^H\mathbf{w}_n\right\|_2^2 p_{m,n}-\eta\left\|\mathbf{h}_{m,n}^H\mathbf{w}_n\right\|_2^2\sum_{i=1}^{m-1}p_{i,n}\\ &-\eta\sum_{j\neq n}\left\|\mathbf{h}_{m,n}^H\mathbf{w}_j\right\|_2^2\sum_{i=1}^{|S_j|}p_{i,j}\geq\omega,\end{aligned} \tag{15}$$

where $\eta=2^{R_{\min}}-1$ and $\omega=\eta\sigma^2$. In this way, the non-linear constraint C_3 has been transformed into a linear constraint. Then, the optimization problem (14) can be rewritten as

$$\max_{\{p_{m,n}\}}\sum_{n=1}^{N_{\text{RF}}}\sum_{m=1}^{|S_n|}\log_2\left(1+\gamma_{m,n}\right),\quad \text{s.t. } C_1,\ C_2,\ (15). \tag{16}$$

We can see from (16) that all constraints are linear inequality constraints, while the objective function is non-convex. Therefore, this optimization problem is NP-hard, and it remains an open challenge to obtain the closed-form solution of the optimal power allocation problem (16).

To solve this difficult non-convex problem (16), we propose an iterative optimization algorithm for carrying out the power allocation. Specifically, according to the extension of the Sherman-Morrison-Woodbury formula [9], we have

$$\begin{aligned}&\left(1+\gamma_{m,n}\right)^{-1}\\ =&1-\left\|\mathbf{h}_{m,n}^{H}\mathbf{w}_n\right\|_2^2 p_{m,n}\left(\left\|\mathbf{h}_{m,n}^{H}\mathbf{w}_n\right\|_2^2 p_{m,n}+\xi_{m,n}\right)^{-1},\end{aligned} \tag{17}$$

where $n=1,2,\cdots,N_{\mathrm{RF}}$ and $m=1,2,\cdots,|S_n|$.

We can find that the expression (17) has the same form as the MMSE. Specifically, if MMSE detection is used for detecting $s_{m,n}$ from $\hat{y}_{m,n}$ in (6), it can be formulated as

$$c_{m,n}^{o}=\arg\min_{c_{m,n}} e_{m,n}, \tag{18}$$

where

$$e_{m,n}=\mathrm{E}\left\{\left|s_{m,n}-c_{m,n}\hat{y}_{m,n}\right|^2\right\} \tag{19}$$

is the mean square error (MSE), $c_{m,n}$ is the channel equalization coefficient, and $c_{m,n}^{o}$ is the optimal value of $c_{m,n}$ required for minimizing the MSE. Substituting (6) into (19), we have

$$\begin{aligned}e_{m,n}=&\left|1-c_{m,n}\sqrt{p_{m,n}}\mathbf{h}_{m,n}^{H}\mathbf{w}_n\right|^2+\left|c_{m,n}\right|^2\sigma^2\\ &+\left|c_{m,n}\right|^2\left\|\mathbf{h}_{m,n}^{H}\mathbf{w}_n\right\|_2^2\sum_{i=1}^{m-1}p_{i,n}\\ &+\left|c_{m,n}\right|^2\sum_{j\neq n}\left\|\mathbf{h}_{m,n}^{H}\mathbf{w}_j\right\|_2^2\sum_{i=1}^{|S_j|}p_{i,j}.\end{aligned} \tag{20}$$

Then, by solving (18) based on (20), the optimal equalization coefficient $c_{m,n}^{o}$ can be calculated by

$$c_{m,n}^{o}=\left(\sqrt{p_{m,n}}\mathbf{h}_{m,n}^{H}\mathbf{w}_n\right)^{*}\left(p_{m,n}\left\|\mathbf{h}_{m,n}^{H}\mathbf{w}_n\right\|_2^2+\xi_{m,n}\right)^{-1}. \tag{21}$$

Substituting (21) into (20), we obtain the MMSE as

$$e_{m,n}^{o}=1-\left\|\mathbf{h}_{m,n}^{H}\mathbf{w}_n\right\|_2^2 p_{m,n}\left(\left\|\mathbf{h}_{m,n}^{H}\mathbf{w}_n\right\|_2^2 p_{m,n}+\xi_{m,n}\right)^{-1}, \tag{22}$$

which is equal to $\left(1+\gamma_{m,n}\right)^{-1}$ in (17). Then, the achievable rate of the mth user in the nth beam can be written as

$$R_{m,n}=\log_2\left(1+\gamma_{m,n}\right)=\max_{c_{m,n}}\left(-\log_2 e_{m,n}\right). \tag{23}$$

To remove the log function in (23), we introduce the following proposition [9].

Proposition 1: Let $f(a)=-\frac{ab}{\ln 2}+\log_2 a+\frac{1}{\ln 2}$ and a be a positive real number. Then we have $\max\limits_{a>0} f(a)=-\log_2 b$, where the optimal value of a is $a^{o}=\frac{1}{b}$.

Using Proposition 1, (23) can be rewritten as

$$R_{m,n}=\max_{c_{m,n}}\max_{a_{m,n}>0}\left(-\frac{a_{m,n}e_{m,n}}{\ln 2}+\log_2 a_{m,n}+\frac{1}{\ln 2}\right). \tag{24}$$

As a result, the objective function of the optimization problem (16) has been transformed into a quadratic programming function, and (16) can be reformulated as

$$\begin{aligned}&\max_{\{p_{m,n}\}}\sum_{n=1}^{N_{\mathrm{RF}}}\sum_{m=1}^{|S_n|}\max_{c_{m,n}}\max_{a_{m,n}>0}\left(-\frac{a_{m,n}e_{m,n}}{\ln 2}+\log_2 a_{m,n}+\frac{1}{\ln 2}\right),\\ &\text{s.t. } C_1,\ C_2,\ (15).\end{aligned} \tag{25}$$

To solve the reformulated optimization problem (25), we propose to iteratively optimize $\{c_{m,n}\}$, $\{a_{m,n}\}$, and $\{p_{m,n}\}$. Specifically, given the optimal power allocation solution $\{p_{m,n}^{(t-1)}\}$ in the $(t-1)$th iteration, the optimal solution of $\{c_{m,n}^{(t)}\}$ in the tth iteration can be obtained according to (21). The corresponding MMSE $\{e_{m,n}^{o(t)}\}$ can be obtained according to (22), and then the optimal solution of $\{a_{m,n}^{(t)}\}$ in the tth iteration can be obtained by $a_{m,n}^{(t)}=\frac{1}{e_{m,n}^{o(t)}}$.

After obtaining the optimal $\{c_{m,n}^{(t)}\}$ and $\{a_{m,n}^{(t)}\}$ in the tth iteration, the optimal $\{p_{m,n}^{(t)}\}$ in the tth iteration can be obtained by solving the following problem:

$$\min_{\{p_{m,n}^{(t)}\}}\sum_{n=1}^{N_{\mathrm{RF}}}\sum_{m=1}^{|S_n|}a_{m,n}^{(t)}e_{m,n}^{(t)},\qquad \text{s.t. } C_1,\ C_2,\ (15). \tag{26}$$

Then, the convex optimization problem (26) can be solved according to the Karush-Kuhn-Tucker (KKT) [8] conditions.

IV. Simulation Results

In this section, we provide simulation results for verifying the performance of the proposed beamspace MIMO-NOMA. Specifically, we consider a typical downlink mmWave massive MIMO system, where the BS is equipped with $N=256$ antennas and communicates with $K=32$ users. One LoS component and $L=2$ NLoS components are assumed for all users' channels. We consider the channel parameters of user k as follows: 1) $\beta_k^{(0)}\sim\mathcal{CN}(0,1)$, $\beta_k^{(l)}\sim\mathcal{CN}\left(0,10^{-1}\right)$ for $1\leq l\leq L$; 2) $\theta_k^{(0)}$ and $\theta_k^{(l)}$ for $1\leq l\leq L$ obey the uniform distribution within $\left[-\frac{1}{2},\frac{1}{2}\right]$. The signal-to-noise ratio (SNR) is defined as $\frac{E_b}{\sigma^2}$ in this paper.

In the simulations, we consider the following four typical mmWave massive MIMO schemes for comparison: 1) "Fully digital MIMO", where each antenna is connected to one RF chain, i.e., $N_{\mathrm{RF}}=N$; 2) "Beamspace MIMO" [5], where each beam only supports a single user with $N_{\mathrm{RF}}=K$; 3) "MIMO-OMA" [8] with $K\geq N_{\mathrm{RF}}$, where OMA is performed for the interfering users, and users in the same beam are allocated with orthogonal frequency resources; 4) "Proposed beamspace MIMO-NOMA" with $K\geq N_{\mathrm{RF}}$, which integrates NOMA into beamspace MIMO, and different users share the same time-frequency resources. Particularly, ZF precoding is considered in the fully digital MIMO and beamspace MIMO.

In this paper, the spectrum efficiency is defined as the achievable sum rate (10), and the energy-efficiency ε is defined as the ratio between the achievable sum rate R_{sum} and the total power consumption [10], i.e.,

$$\varepsilon=\frac{R_{\mathrm{sum}}}{P+N_{\mathrm{RF}}P_{\mathrm{RF}}+N_{\mathrm{RF}}P_{\mathrm{SW}}+P_{\mathrm{BB}}}\ (\mathrm{bps/Hz/W}), \tag{27}$$

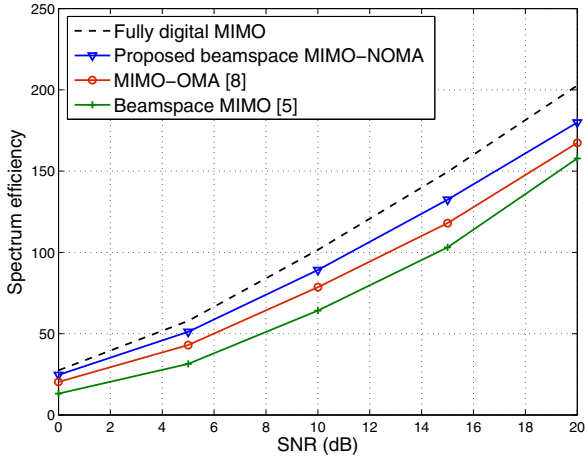

Fig. 1. Spectrum efficiency against SNR.

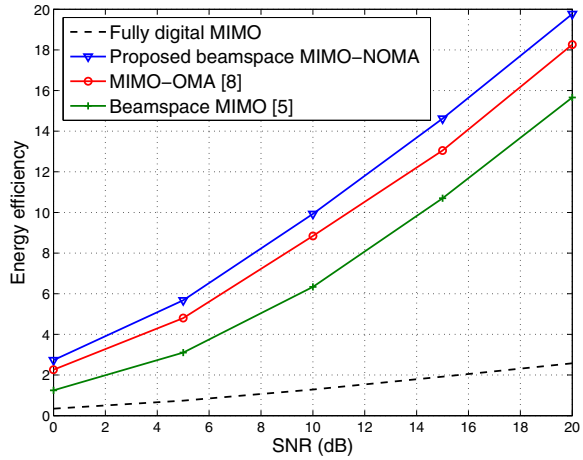

Fig. 2. Energy efficiency against SNR.

where P is the maximum transmitted power, P_{RF} is the power consumed by each RF chain, P_{SW} is the power consumption of the beam switch, and P_{BB} is the baseband power consumption. Specifically, we adopt the typical values of $P = 32$ mW, $P_{\mathrm{RF}} = 300$ mW, $P_{\mathrm{SW}} = 5$ mW, and $P_{\mathrm{BB}} = 200$ mW [10].

Fig. 1 shows the spectrum efficiency against SNR, where the number of iterations required for solving the power allocation optimization problem is set to 10 (We have verified through simulations that the spectrum efficiency tends to a constant after 10 iterations). We find that the proposed beamspace MIMO-NOMA achieves a higher spectrum efficiency than beamspace MIMO [5] as well as MIMO-OMA [8]. Quantitatively, the proposed beamspace MIMO-NOMA has about 3 dB SNR gain over beamspace MIMO, which benefits from the employment of NOMA to serve multiple users in each beam. Additionally, the proposed beamspace MIMO-NOMA also outperforms MIMO-OMA, since NOMA can achieve a higher spectrum efficiency than that of OMA [6].

Fig. 2 shows the energy efficiency against SNR. We find that the proposed beamspace MIMO-NOMA achieves a higher energy efficiency than the other three schemes. Particularly, the proposed beamspace MIMO-NOMA has about 4 dB SNR gain over beamspace MIMO, which benefits from the employment of NOMA to serve multiple users in each beam. Additionally, the proposed beamspace MIMO-NOMA achieves a higher energy efficiency than the fully digital MIMO, where the number of RF chains is equal to the number of BS antennas. By contrast, the number of RF chains is lower than the number of antennas in the proposed beamspace MIMO-NOMA.

V. Conclusions

In this paper, we have proposed a new mmWave transmission scheme called beamspace MIMO-NOMA to integrate beamspace MIMO and NOMA in order to break the fundamental limit of beamspace MIMO, where more than one user can be served in each beam using the same time-frequency resources. To mitigate the inter-beam interference, the equivalent channel vector was determined for each beam for ZF-based TPC. Furthermore, we proposed to jointly optimize the power allocation of all users by maximizing the achievable sum rate, and an iterative optimization algorithm has been developed for power allocation. Our simulation results have shown that the proposed beamspace MIMO-NOMA achieves better performance than beamspace MIMO in terms of spectrum and energy efficiency.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant Nos. 1320106003 and 61571270), and the Royal Academy of Engineering under the UK China Industry Academia Partnership Programme Scheme (Grant No. UK-CIAPP\49).

References

[1] S. Mumtaz, J. Rodriquez, and L. Dai, *MmWave Massive MIMO: A Paradigm for 5G*, Academic Press, Elsevier, 2016.

[2] R. W. Heath, N. Gonzlez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436-453, Apr. 2016.

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

[4] Y. Zeng and R. Zhang, "Millimeter wave MIMO with lens antenna array: A new path division multiplexing paradigm," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1557-1571, Apr. 2016.

[5] 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.

[6] 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.

[7] 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.

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

[9] Q. Zhang, Q. Li, and J. Qin, "Robust beamforming for non-orthogonal multiple access systems in MISO channels," to appear in *IEEE Trans. Veh. Technol.*, 2017.

[10] X. Gao, L. Dai, S. Han, C.-L. I, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998-1009, Apr. 2016.