Group Clustering for Block Diagonal Digital Precoder in Multi-user MIMO System

Guanchong Niu, Qi Cao and Man-On Pun‡
School of Science and Engineering
The Chinese University of Hong Kong, Shenzhen
Shenzhen, Guangdong, China, 518172

***Abstract*—Beam division multiple access (BDMA) has recently been proposed for massive multiple-input multiple-output (MIMO) systems by simultaneously transmitting multiple users' data streams via different beams. Meanwhile, the block hybrid precoding has been proposed to reduce the computational complexity. However, previous works mostly rely on a crucial condition that the number of RF chains must be not less than the number of data streams. In this paper, we propose a multipath BDMA based hybrid block precoding system to break the ceiling on minimal required RF chains where the data streams are processed cluster-by-cluster. To overcome the performance degradation arisen from block hybrid precoding scheme, a K-means based heuristic user clustering algorithm is proposed to minimize the inter-cluster interference. Then two digital precoding approaches are investigated to suppress the intra-cluster interference by separately considering the signal-to-interference-and-noise ratio (SINR) and signal-to-leakage-and-noise ratio (SLNR). Simulation results confirm the effectiveness of proposed block clustering precoding scheme compared to conventional hybrid beamforming scheme.**

I. INTRODUCTION

To meet the ever-increasing demand of higher user data rates, it is envisioned that the next-generation cellular systems will be equipped with massive antenna arrays [1]. Capitalizing on the large number of antennas at the base-station (BS), beam division multiple access (BDMA) has recently been proposed to transmit multiple users' data-streams via different beams [2], [3]. In contrast to the more conventional multiple access schemes such as Code Division Multiple Access (CDMA) or Orthogonal Frequency Multiple Division Access (OFDMA) that multiplex users in code, time and frequency domains, BDMA separates users in the beam space by transmitting data to different users in orthogonal beam directions. In [2], BDMA was first proposed to decompose the multiuser multiple-input multiple-output (MU-MIMO) system into multiple single-user MIMO channels by multiplexing multiple users' data onto non-overlapping beams. More recently, joint user scheduling and beam selection for BDMA was formulated under the Lyapunov-drift optimization framework before the optimal user-beam scheduling policy was derived in a closed form [3]. However, the assumption of non-overlapping orthogonal beams is hard to be satisfied which limits the potential of analog-only BDMA beamforming solution. Different from analog-only beamforming system, digital precoding can be implemented to eliminate the inter-user interference, where the baseband requires dedicated RF chain per each antenna [4].

However, the expensive and power-hungry components (e.g. ADC/DAC, filters, mixers and amplifiers) in RF chain constitute an impediment of broad implementation of massive MIMO systems. Thus, hybrid digital and analog beamforming has been developed for massive MIMO transmissions by dividing the procoding process into two steps, namely analog and digital precoding [5], [6]. More specifically, the transmitted signals are first precoded digitally using a smaller number of radio frequency (RF) chains followed by the analog precoding implemented with a much larger number of low-cost phase shifters. As a result, the hybrid analog-digital precoding architecture requires significantly less RF chains as compared to the fully digital precoding in which every available antenna element is supported by one RF chain.

To further reduce the computational complexity and obtaining closed-form solutions in downlink space-division multiple access (SDMA), the notion of *block diagonalization* (BD) was introduced in [7], which is found out helpful in massive MIMO transmissions especially for digital precoder design of hybrid beamforming system. This is because computation and implementation of full zero-forcing precoding matrix in an instant is of great challenge, while separating digital precoders into blocks efficiently mitigates the hard work. With the number of transmit RF chains larger than the total number of receive RF chains (considering multiuser scenario), it can be achieved that, for each block, base station projects all inter-block interference onto the its null spaces [8]. By setting up an optimization problem minimizing the means square error (MSE) between the received signals resulted from full ZF prcoding system and BD precoding system, [4] indicates that utilizing BD precoders can realize an asymptotic performance to full ZF precoding transmissions. Actually, the composition of a block is flexible, *i.e.* it can be a downlink user with multiple antennas or multiple single-antenna users. For instance, in [9], the structure of hybrid BD precoding is also investigated but with each block being a group of single-antenna users. In such a system, the size of each block can even be a optimization variable. Nevertheless, the above hybrid BD massive MIMO systems are all under a crucial condition, that is, the number of RF chains must be no less than the number of supported data-streams.

To circumvent the restriction, [10] uses L RF chains to approximate $K \geq L$ transmitted symbols, in this way, the data-stream to RF chain ratio can be above 1. The success of this scheme relies on that the state-of-art phase shifters and switches can change their state in hundreds of pico seconds, since the approximation needs to be done in each symbol period. However, due to that digital/analog precoder is symbol-dependent, and users have to recover symbols via compressive sensing tools, the overall scheme is of high complexity. In this paper, we also consider the

hybrid BD multiuser massive MIMO system with number of data-streams larger than number of RF chains, and against the current literature, our contributions are threefold:

- It is widely accepted in massive MIMO studies that the number of data-streams is restricted to the number of RF chains. To break the rule, we propose a cluster-by-cluster hybrid precoding process that involves both digital and analog precoders. In our proposed scheme, users are divided into K clusters, their signals are multiplied by corresponding digital precoders block by block, and added up together at the analog precoding stage before simultaneous transmission. Thus, the data-stream per physical RF chain is far above 1 while the data-stream per RF chain usage is still equal to 1.
- Clustering different users together results in different systematic performance. This phenomenon is also known as peer effect. For particular digital and analog precoding strategy, we formulate the user clustering problem as an integer programming problem, which is of NP-hardness. To tackle the problem, we propose a K-means based heuristic user clustering greedy algorithm that could effectively improve system sume-rate capacity.
- We design a whole set of transmission scheme from transmit digital/analog precoding matrix to the receive analog beamforming vectors. With finite number of transmit antennas, the inter-cluster interference inherently exists in our system model, hence we investigate two distinguishing digital precoders respectively coming from equivalent channel inversion and signal-to-leakage-and-noise (SLNR), which show equivalent performance in simulation results.

The structure of this paper is arranged as follows: the massive MIMO system with asynchronous hybrid BD precoding is introduced in Section II; the design of transmit digital/analog precoding matrix and receive analog beamforming vectors are elaborated in Section III; The user clustering algorithm is detailed in Section IV whereas the simulation results are shown in Section V; and finally we conclude the paper in Section VI

Notation: Vectors and matrices are denoted by boldface letters. $\boldsymbol{I}_N$ denotes the identity matrix with size $N \times N$. $\boldsymbol{A}^T$ and $\boldsymbol{A}^H$ denote transpose and conjugate transpose of $\boldsymbol{A}$, respectively. $\boldsymbol{A}^\dagger$ being the pseudo inverse of $\boldsymbol{A}$ while $||\boldsymbol{A}||_0$, $||\boldsymbol{A}||$ and $|\boldsymbol{A}|$ stand for 0 norm, the Frobenius norm and determinant of $\boldsymbol{A}$, respectively. $\boldsymbol{A}(i,j)$ denotes the i row, j column element of $\boldsymbol{A}$; $|\mathcal{I}|$ is the cardinality of the enclosed set $\mathcal{I}$; Finally, $\mathbb{E}[\cdot]$ and $\Re\{\cdot\}$ denote the expectation and real part of a random variable.

II. SYSTEM MODEL

There are N_{tot} users under considered base station, and N_U of them will be selected to serve. We consider a multi-user mmWave MIMO system shown in Fig. 1, in which a transmitter equipped with N_{RF} RF chains and N_T antennas transmits N_U data streams to N_U receivers with N_R receive antennas. Following the same assumption commonly employed in the literature [11], we assume only one data stream is designated to each scheduled receiver. We use $\boldsymbol{s}(n)$ to denote the n-th block of N_U data to be transmitted with $\mathbb{E}\left[\boldsymbol{s}\boldsymbol{s}^H\right] = \frac{1}{N_U}\boldsymbol{I}_{N_U}$. In the sequel, we concentrate on a single block and omit the temporal index n for notational simplicity.

Many papers have proposed block algorithm to reduce the computational complexity. Our basic idea is to group the users into several clusters and then the inter-cluster interference can be minimized by analog precoding with group-scheduling and the intra-cluster interference can be eliminated by digital precoding.

The N_U users are divided into K clusters and each cluster has M_k users. Obviously, we have

$$\sum_{k=1}^{K} M_k = N_U, \quad 0 < M_k \leq N_U \tag{1}$$

And the digital precoder is given by

$$\boldsymbol{F} = \begin{bmatrix} \boldsymbol{\mathcal{F}}_1 & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \boldsymbol{\mathcal{F}}_2 & \vdots & \vdots \\ \boldsymbol{0} & \cdots & \ddots & \boldsymbol{0} \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{\mathcal{F}}_K \end{bmatrix}, \quad \boldsymbol{F}_k \in \mathcal{C}^{N_{RF} \times M_k} \tag{2}$$

Correspondingly, the analog precoder is also divided into K parts

$$\boldsymbol{V} = [\boldsymbol{\mathcal{V}}_1, \boldsymbol{\mathcal{V}}_2, \cdots, \boldsymbol{\mathcal{V}}_K], \quad \boldsymbol{\mathcal{V}}_k \in \mathcal{C}^{N_T \times M_k} \tag{3}$$

Also, the data streams $\boldsymbol{S}$ is divided into K clusters

$$\boldsymbol{s} = \left[\mathbf{s}_1^T, \mathbf{s}_2^T, \cdots, \mathbf{s}_K^T\right]^T, \quad \mathbf{s}_k \in \mathcal{C}^{M_j \times 1} \tag{4}$$

For conventional block hybrid precoding systems, the number of required RF chains is still equal to the number of users although the computational complexity is reduced. Distinct to the conventional systems, each RF chain can serve more than one data stream thus the required RF chains will be significantly reduced in our proposed cluster-by-cluster hybrid precoding system as shown in Fig. 1.

We use $\boldsymbol{f}_{ku}, \boldsymbol{v}_{ku}$ and s_{ku} to represent the digital precoding, analog precoding and data stream for u-th user in k-th cluster.

The resulting precoded signal $\boldsymbol{x}$ of dimension $N_T \times 1$ can be expressed as

$$\boldsymbol{x} = \boldsymbol{V} \cdot \boldsymbol{F} \cdot \boldsymbol{s} = \sum_{k=1}^{K} \boldsymbol{\mathcal{V}}_k \boldsymbol{\mathcal{F}}_k \mathbf{s}_k \tag{5}$$

The precoded signal $\boldsymbol{x}$ is then broadcast to N_U users. The signal received by the u-th user is given by

$$\begin{aligned} \boldsymbol{y}_{ku} &= \underbrace{\boldsymbol{H}_{ku}\boldsymbol{\mathcal{V}}_k\boldsymbol{f}_{ku}s_{ku}}_{\text{Desired Signal}} + \underbrace{\boldsymbol{H}_{ku}\boldsymbol{\mathcal{V}}_k\sum_{\substack{i=1 \\ i\neq u}}^{M_K}\boldsymbol{f}_{ki}s_{ki}}_{\text{Intra-cluster Interference}} \\ &+ \underbrace{\boldsymbol{H}_{ku}\sum_{\substack{j=1 \\ j\neq k}}^{K}\boldsymbol{\mathcal{V}}_j\boldsymbol{\mathcal{F}}_j\boldsymbol{s}_j}_{\text{Inter-cluster Interference}} + \underbrace{\boldsymbol{n}_{ku}}_{\text{Noise}} \end{aligned} \tag{6}$$

where $\boldsymbol{H}_{ku} \in \mathbb{C}^{N_R \times N_T}$ is the MIMO channel matrix between the transmitter and the u-th receiver [6]. Furthermore, $\boldsymbol{n}_u$ is complex additive white Gaussian noise with zero mean and variance equal to σ^2.

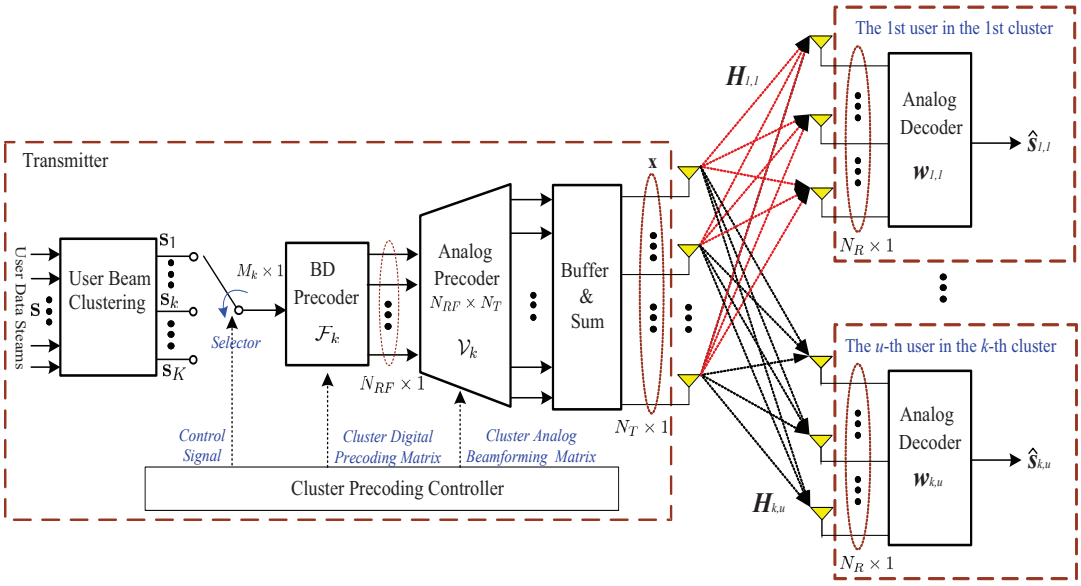

Fig. 1. Block diagram of the hybrid precoding system under consideration

Assuming the receivers are all low-cost terminals that perform analog beamforming only in decoding, the decoded signal by the u-th user in k-th cluster denoted by $\hat{s}_u$ is given as

$$\hat{s}_{ku} = \boldsymbol{w}_{ku}^H \boldsymbol{H}_{ku} \boldsymbol{\mathcal{V}}_k \boldsymbol{f}_{ku} s_{ku} + \boldsymbol{w}_{ku}^H \tilde{\boldsymbol{n}}_{ku}, \tag{7}$$

where $\boldsymbol{w}_{ku}$ of dimension $N_R \times 1$ is the analog beamforming vector employed by the u-th receiver with the power constraint of $|\boldsymbol{w}_u|^2 = 1$ and

$$\tilde{\boldsymbol{n}}_u = \boldsymbol{H}_{ku} \boldsymbol{\mathcal{V}}_k \sum_{\substack{i=1 \\ i \neq u}}^{M_K} \boldsymbol{f}_{ki} s_{ki} + \boldsymbol{H}_{ku} \sum_{\substack{j=1 \\ j \neq k}}^{K} \boldsymbol{\mathcal{V}}_j \boldsymbol{\mathcal{F}}_j \boldsymbol{s}_j + \underbrace{\boldsymbol{w}_{ku}^H \boldsymbol{n}_{ku}}_{\text{Noise}} \tag{8}$$

Note that the first term in Eq. (7) stands for the desired signal while the second term is the sum of its own receiver noise and interference from intra-cluster users and other clusters' users.

A. Channel Model

As shown in [12], the mmWave wireless channel can be well modeled by the Saleh-Valenzuela model. Following the same approach developed in [13], we assume that each scatter only contributes one single propagation path. As a result, the u-th user's channel model can been modeled as:

$$\boldsymbol{H}_u = \sqrt{\frac{N_T N_R}{L_u}} \sum_{l=1}^{L_u} \alpha_{u,l} \cdot \boldsymbol{a}_R(\phi_{u,l}^r, \theta_{u,l}^r) \cdot \boldsymbol{a}_T^H(\phi_{u,l}^t, \theta_{u,l}^t), \tag{9}$$

where L_u is the number of scatters of the u-th user's channel. Furthermore, $\alpha_{u,l}$, $\theta_{u,l}^r/\phi_{u,l}^r$ and $\theta_{u,l}^t/\phi_{u,l}^t$ are the complex path gain, azimuth/elevation angles of arrival(AoA) and azimuth/elevation angles of departure(AoD) of the l-th path of the u-th user, respectively. Finally, $\boldsymbol{a}$ is the array response vector. For an uniform planar array (UPA) of size $P \times Q$ considered in this work, the array response vector $\boldsymbol{a}$ is given by [13]

$$\begin{aligned} \boldsymbol{a}(\phi, \theta) = & \frac{1}{\sqrt{N_T}} \Big[1, e^{jkd(\sin\phi \sin\theta + \cos\theta)}, \cdots, e^{jkd(p \sin\phi \sin\theta + q \cos\theta)}, \\ & \cdots, e^{jkd((P-1)\sin\phi \sin\theta + (Q-1)\cos\theta)}\Big]^T, \end{aligned} \tag{10}$$

where $k = \frac{2\pi}{\lambda}$ is the wavenumber while d is the distance between two adjacent antennas.

B. Problem Formulation

For notational simplicity, we denote by $\boldsymbol{g}_{ku}^H$ the effective array gain of the u-th user in k-th cluster with

$$\boldsymbol{g}_{ku}^H = \boldsymbol{w}_{ku}^H \boldsymbol{H}_{ku} \boldsymbol{\mathcal{V}}_k. \tag{11}$$

And the effective array gain of u-th user from other clusters is given by

$$\boldsymbol{g}_{ju}^H = \boldsymbol{w}_{ku}^H \boldsymbol{H}_{ku} \boldsymbol{\mathcal{V}}_j. \tag{12}$$

Then, the channel capacity of the u-th user is given by

$$R_{ku} = \log\left(1 + \frac{\frac{P}{N_U} |\boldsymbol{g}_{ku}^H \boldsymbol{f}_{ku}|^2}{\frac{P}{N_U} \sum_{\substack{i=1 \\ i \neq u}}^{M_k} (|\boldsymbol{g}_{ku}^H \boldsymbol{f}_{ki}|^2 + \sum_{\substack{j=1 \\ j \neq k}}^{K} ||\boldsymbol{g}_{ju} \boldsymbol{F}_j||^2) + \sigma^2}\right). \tag{13}$$

where we assume the power of each user is uniformly allocated.

Subsequently, the system average capacity that is a function of $\boldsymbol{V}$ and $\boldsymbol{F}$ can be computed as

$$R_{avg} = \frac{1}{K N_U} \sum_{k=1}^{K} \sum_{u=1}^{N_U} R_{ku}. \tag{14}$$

C. Cluster-by-Cluster(CbC) hybrid precoding

Obviously, the capacity of BD precoding is upper bounded by conventional hybrid precoding although the computational complexity is reduced. In this paper, we proposed a clustering algorithm to reduce the gap between conventional and block hybrid precoding. By introducing an switch matrix $\boldsymbol{T} \in \{0,1\}^{N_U \times N_U}$, the data streams are clustered by distinct beams.

As shown in Eq. (7), the proposed algorithm is also aiming to break the ceiling on the number of steams $N_{RF} \leq N_U$, where we assume the analog and digital precoder can be instantaneously calculated such that the transmitted signal can be processed cluster-by-cluster.

Finally, for the given K clusters, the optimal design of the digital and analog precoding matrices can be formulated as

$$
\begin{aligned}
P_1: \quad & \max_{\boldsymbol{W},\boldsymbol{T},\boldsymbol{V},\boldsymbol{F}} \quad R_{avg}(\boldsymbol{W}, \boldsymbol{VT}, \boldsymbol{F}) \qquad (15)\\
s.t. \quad & C_1: \text{diag}(\boldsymbol{v}_{ku}\boldsymbol{v}_{ku}^H) = \frac{\boldsymbol{I}_{N_T}}{\sqrt{N_T}};\\
& C_2: \text{diag}(\boldsymbol{w}_{ku}\boldsymbol{w}_{ku}^H) = \frac{\boldsymbol{I}_{N_R}}{\sqrt{N_R}};\\
& C_3: ||\boldsymbol{V}\boldsymbol{f}_{ku}||^2 = 1;\\
& C_4: \boldsymbol{F} = \text{diag}(\boldsymbol{\mathcal{F}}_1, \boldsymbol{\mathcal{F}}_2, \cdots, \boldsymbol{\mathcal{F}}_K);\\
& C_5: \boldsymbol{V} = [\boldsymbol{\mathcal{V}}_1, \boldsymbol{\mathcal{V}}_2, \cdots, \boldsymbol{\mathcal{V}}_K];\\
& C_6: \max\{M_k\}_{k=1}^K \leq N_{RF};\\
& C_7: ||\boldsymbol{t}_i||_0 = ||\boldsymbol{t}_j||_0 = 1, \quad [\boldsymbol{T}]_{ij} \in \{0,1\}.
\end{aligned}
$$

where $\boldsymbol{t}_i$ and $\boldsymbol{t}_j$ are the row and column vector of $\boldsymbol{T}$ respectively.

Considering the phase-only constraint of C_1 and C_2, we need to solve P_1 in a two-stage algorithm. The phase-only constraint is very challenging because it makes the problem non-convex and combinatorial. For simplicity, in the first stage, we ignore the couple effect of digital and analog precoding on clustering and only consider analog precoding to minimize the inter-cluster interference by clustering. The second stage is to design the digital precoder to further eliminate the intra-cluster interference cluster-by-cluster.

III. Proposed Block Hybrid Beamforming for RF Chains Reduction

To solve the Problem P_1, We will firstly ignore the constraints on digital precoder and solve the analog precoder. Then two approaches for digital precoding are proposed in this section, namely block zero-forcing (BZF) and block signal-to-leakage-and-noise ratio (SLNR) maximization (BSM). For BZF, the users are clustered like K-means and then the inter-cluster interference can be minimized by analog precoding and the intra-cluster interference can be eliminated by conventional zero-forcing. As to BSM, the analog precoder is calculated same with BC but the digital precoder will be designed by maximizing the SLNR after clustering.

A. Analog Precoding Design

For multi-path channel model as shown in Eq. (9), we need to select the beams with least interference. Then the Problem P1 can be simplified as

$$
\begin{aligned}
P_2: \quad & \max_{\boldsymbol{W},\boldsymbol{T},\boldsymbol{V}} \quad R_{avg}(\boldsymbol{W}, \boldsymbol{VT}) \qquad (16)\\
s.t. \quad & C_1: \text{diag}(\boldsymbol{v}_{ku}\boldsymbol{v}_{ku}^H) = \frac{\boldsymbol{I}_{N_T}}{\sqrt{N_T}};\\
& C_2: \text{diag}(\boldsymbol{w}_{ku}\boldsymbol{w}_{ku}^H) = \frac{\boldsymbol{I}_{N_R}}{\sqrt{N_R}};\\
& C_3: \boldsymbol{V} = [\boldsymbol{\mathcal{V}}_1, \boldsymbol{\mathcal{V}}_2, \cdots, \boldsymbol{\mathcal{V}}_K];\\
& C_4: \max\{M_k\}_{k=1}^K \leq N_{RF};\\
& C_5: ||\boldsymbol{t}_i||_0 = ||\boldsymbol{t}_j||_0 = 1, \quad [\boldsymbol{T}]_{ij} \in \{0,1\}.
\end{aligned}
$$

With the assumption that array response vectors corresponding to distinct beams are asymptotically orthogonal with infinite number of antennas at transmitter

$$
\lim_{N\to+\infty} \boldsymbol{a}_T^H(\phi_{i,l}^t, \theta_{i,l}^t) \cdot \boldsymbol{a}_T(\phi_{j,p}^t, \theta_{j,p}^t) = \delta(i-j)\delta(l-p), \quad (17)
$$

Based on the idea of BDMA, the analog precoder is solved by clustering the transmitted data streams to maximize the sum-rate signal-to-interference ratio (SIR)

$$
\{\boldsymbol{\mathcal{W}}_k^*, \boldsymbol{\mathcal{V}}_k^*, \boldsymbol{T}\}_{k=1}^K = \arg\max \sum_{k=1}^K \frac{||\boldsymbol{w}_{ku}^H \boldsymbol{H}_{ku}(\boldsymbol{T})\boldsymbol{\mathcal{V}}_k||_F^2}{\sum_{j=1,j\neq k}^{N_U} ||\boldsymbol{w}_{ku}^H \boldsymbol{H}_{ku}(\boldsymbol{T})\boldsymbol{\mathcal{V}}_j||_F^2} \qquad (18)
$$

$$
\begin{aligned}
s.t. \quad & \boldsymbol{\mathcal{V}}_k^* \in \{\boldsymbol{a}_T^H(\phi_{u,l}^t, \theta_{u,l}^t)\}_{u\in[1,N_U],l\in[1,L_u]};\\
& \boldsymbol{\mathcal{W}}_k^* \in \{\boldsymbol{a}_R^H(\phi_{u,l}^t, \theta_{u,l}^t)\}_{u\in[1,N_U],l\in[1,L_u]};\\
& \max\{M_k\}_{k=1}^K < N_{RF}
\end{aligned}
$$

where the Assuming that the transmitter has perfect channel state information (CSI), then all AoA and AoD information, *i.e.* $\{\phi_u^t, \theta_u^t, \phi_u^r, \theta_u^r\}$, is perfectly known to the transmitter.

To solve $\boldsymbol{T}$, the intuitive thought is to cluster users such that the inter-cluster interference can be minimized. However, the computational complexity will be large although the optimal solution can be given by exhaustive searching. We propose a greedy clustering algorithm in Algorithm. 1. In this K-means based algorithm, the initial center users are firstly selected and then the left users are clustered by maximizing the sum-rate SIR. From the rows of 20 to 22, the constraint on the number of RF chains is considered for each cluster.

The analog precoder of clustered data streams can be given by $\boldsymbol{\mathcal{V}}_k^* = \boldsymbol{a}_{\mathcal{I}_k}$ for users in k-th cluster. This algorithm is very similar to k-means. The main difference is that we use SIR to replace the Euclidean distance and the initial centers are greedily selected with maximal SIR.

B. Digital Precoder

By analog beamforming procedure, the following deduction can be given

$$
\boldsymbol{\mathcal{W}}_k^* \boldsymbol{H}_{ku} \boldsymbol{\mathcal{V}}_j^* \approx \boldsymbol{0} \quad \text{for} \quad k \neq j \qquad (19)
$$

Then the Eq. (8) can be simplified as

$$
\tilde{\boldsymbol{n}}_{ku} \approx \boldsymbol{g}_{ku} \sum_{\substack{i=1\\ i\neq u}}^{M_k} \boldsymbol{f}_{ki} s_{ki} + \boldsymbol{w}_{ku}\boldsymbol{n}_{ku} \qquad (20)
$$

Algorithm 1 Greedy clustering algorithm for block hybrid beamforming system

Input:

All user index set and path set: $\mathcal{X}$, $\mathcal{L}$
Clustered user and path index set : $\mathcal{I}_k = \emptyset$, $k = 1, 2, \cdots, K$
Number of clusters: $\mathcal{K}$

Procedures:

1. Initial Centers:
Assign a user index with largest channel gain $x^*_{k,l}$ corresponding to $\mathcal{I}_1$, *i.e.* $\mathcal{I}_1 \leftarrow x^*_{k,l}$, $\boldsymbol{\mathcal{I}} \leftarrow \mathcal{I}_1$ and $\mathcal{X} \setminus x^*_{k,l}$,
while $2 \leq k \leq K$ **do**
  for $x_{k,l}$ in $(\mathcal{X}, \mathcal{L})$ **do**
    Calculate the sum-rate SIR $p(x_{k,l})$ for the users in $\boldsymbol{\mathcal{I}}$ by Eq. (18)
  end for
  Find the user index x^*_k with minimum $p(x_k)$ in k cluster
  Update $\mathcal{I}_{k,l} \leftarrow x^*_{k,l}$, $\boldsymbol{\mathcal{I}} \leftarrow \mathcal{I}_k$ and $\mathcal{X} \setminus x^*_{k,\mathcal{L}}$
end while
2. Clustering:
for $x_{k,l}$ in $(\mathcal{X}, \mathcal{L})$ **do**
  for k in $\mathcal{K}$ **do**
    Calculate the sum-rate SIR $p(x_{k,l})$ for the users in $\boldsymbol{\mathcal{I}}$ by Eq. (18)
  end for
  Find the user index x^*_k with max $p(x_{k,l})$
  Update $\mathcal{I}_k \leftarrow x^*_{k,l}$, $\boldsymbol{\mathcal{I}} \leftarrow \mathcal{I}_k$ and $\mathcal{X} \setminus x^*_{k,l}$
  if length($\mathcal{I}_k$)$> N_{RF}$ **then**
    $k \setminus \mathcal{K}$
  end if
end for

Where the inter-cluster interference term is eliminated by Eq. (19).

In general, the number of RF chains is constant but not the number of users. In this section, we will use less RF chains to serve users more than N_{RF} by taking the advantage of user scheduling. For the assumption of Eq. (17), although the infinite antennas can't be practical, the residual interference of difference users can be minimized by Eq. (16). To break the limitation on the minimal required number of RF chains, the above equations show that the data streams can be processed by digital precoder cluster-by-cluster and then the precoded data streams will be transmitted after combining, where the number of required RF chains is reduced to M_k for each cluster.

The digital precoder can be solved by the following optimization problem

$$\begin{aligned} P_3: \quad & \max_{\boldsymbol{F}} \quad \sum_{k=1}^{K} R_{avg}(\boldsymbol{\mathcal{F}}_k) \\ s.t. \quad & C_1: ||\boldsymbol{V}\boldsymbol{f}_{ku}||^2 = 1; \\ & C_2: \boldsymbol{F} = \text{diag}(\boldsymbol{\mathcal{F}}_1, \boldsymbol{\mathcal{F}}_2, \cdots, \boldsymbol{\mathcal{F}}_K); \\ & C_3: \max\{M_k\}_{k=1}^K \leq N_{RF}; \end{aligned} \tag{21}$$

Then two approaches will be introduced to solve the P_3.

1) Blcok Zero Forcing (BZF): For the given K, the digital precoder is assumed to be designed as a block diagonal matrix. Then the block digital precoder can be separately calculated. [13] proposed a zero-forcing approach to solve $\boldsymbol{F}$ by setting

$$\boldsymbol{\mathcal{F}}_{BZF,k} = \boldsymbol{\mathcal{G}}_k^{\dagger} = \boldsymbol{\mathcal{G}}_k^H (\boldsymbol{\mathcal{G}}_k \boldsymbol{\mathcal{G}}_k^H)^{-1}. \tag{22}$$

with $N_{RF} \geq M_k$, where $\boldsymbol{\mathcal{G}}_k = [\boldsymbol{g}_{k1}, \boldsymbol{g}_{k2}, \cdots, \boldsymbol{g}_{kM_k}]$.

To satisfy the power constraint C_3 in P_1, power normalization is performed on each $\boldsymbol{f}_{ku}$ derived from $\boldsymbol{\mathcal{F}}_{BZF,k} = [\boldsymbol{f}_{BZF,k1}, \boldsymbol{f}_{BZF,k2}, \cdots, \boldsymbol{f}_{BZF,kM_k}]$ as

$$\boldsymbol{f}^*_{BZF,k,u} = \frac{\boldsymbol{f}_{BZF,ku}}{||\boldsymbol{V} \cdot \boldsymbol{f}_{BZF,ku}||}. \tag{23}$$

For a special case $K = 1$, this problem can be simply solved by conventional zero-forcing.

Compared to conventional digital precoding, the only difference is that the data streams are processed cluster-by-cluster such that the number of considered users is reduced from N_U to M_k for k-th cluster. Thus, the minimal number of required RF chains is then reduce to $N_{RF} = \max\{M_k\}_{k=1}^K$.

C. Block SLNR Maximization(BSM)

Although the BZF has good performance as shown in the simulation results, it requires to acknowledge all the channel information for each user. The conventional SLNR maximization(SM) is considered as a criterion to reduce co-channel interference(CCI) and noise. For the u-th user, the desired data streams can be expressed as

$$S_{ku} = \frac{P}{N_U} |\boldsymbol{w}_{ku} \boldsymbol{H}_{ku} \boldsymbol{V} \boldsymbol{f}_{SM,ku}|^2 \tag{24}$$

and the power leaked from u-th user to all other users is

$$\begin{aligned} L_{ku} \approx \frac{P}{N_U} (& \sum_{j=1, j\neq k}^{K} \sum_{i=1}^{M_j} |\boldsymbol{w}_{ji} \boldsymbol{H}_{ji} \boldsymbol{V} \boldsymbol{f}_{SM,ku}|^2 \\ & + \sum_{t=1}^{M_k} |\boldsymbol{w}_{kt} \boldsymbol{H}_{kt} \boldsymbol{V} \boldsymbol{f}_{SM,ku}|^2) \end{aligned} \tag{25}$$

By recalling the Eq. (19), the SLNR for u-th user in k-th cluster can be expressed as

$$\text{SLNR}_{BSM,ku} \approx \frac{\gamma |\boldsymbol{w}_{ku} \boldsymbol{H}_{ku} \boldsymbol{\mathcal{V}}_k \boldsymbol{f}_{BSM,ku}|^2}{\gamma \sum_{j=1, j\neq k}^{K} \sum_{i=1}^{M_j} |\boldsymbol{w}_{ji} \boldsymbol{H}_{ji} \boldsymbol{\mathcal{V}}_j \boldsymbol{f}_{BSM,ku}|^2 + 1} \tag{26}$$

where the size of $\boldsymbol{f}_{BSM,ku}$ is $M_k \times 1$ while the the size of $\boldsymbol{f}_{SM,ku}$ is $N_U \times 1$. The γ is the Signal-to-noise ratio (SNR).

Define $\tilde{\boldsymbol{g}}_{ku}$ is the collection of all effective array gain matrix that excludes $\boldsymbol{w}_{ku} \boldsymbol{H}_{ku} \boldsymbol{\mathcal{V}}_k$. Thus the SLNR of each user can be represented as

$$\text{SLNR}_{ku} \approx \frac{\boldsymbol{f}^H_{BSM,ku} \boldsymbol{g}^H_{ku} \boldsymbol{g}_{ku} \boldsymbol{f}_{BSM,ku}}{\boldsymbol{f}^H_{BSM,ku} (\frac{1}{\gamma} \boldsymbol{I} + \bar{\boldsymbol{g}}^H_{ku} \bar{\boldsymbol{g}}_{ku}) \boldsymbol{f}_{BSM,ku}} \tag{27}$$

The digital precoder for each user can be solved by [14]

$$\boldsymbol{f}_{BSM,ku} = \lambda_{max} \left(\left(\frac{1}{\gamma} \boldsymbol{I} + \bar{\boldsymbol{g}}^H_{ku} \bar{\boldsymbol{g}}_{ku} \right)^{-1} \boldsymbol{g}^H_{ku} \boldsymbol{g}_{ku} \right) \tag{28}$$

The equivalence of BZF and BSM will be shown in simulation results.

IV. SIMULATION RESULTS

In this section, we use computer simulation to compare the performance of sum-rate capacity for RF chains-reduction algorithm. Unless specified otherwise, we consider a transmitter equipped with an 12×12 UPA (*i.e.* $N_T = 144$) and $N_U = 16$ users each equipped with a 8×8 UPA (*i.e.* $N_R = 64$). The channels are multi-path with the azimuth AoAs/AoDs being uniformly distributed over $[0, 2\pi]$ and the elevation AoAs/AoDs being uniformly distributed in $[-\pi/2, \pi/2]$, respectively. For each computer experiment, we compute the average over 500 realizations.

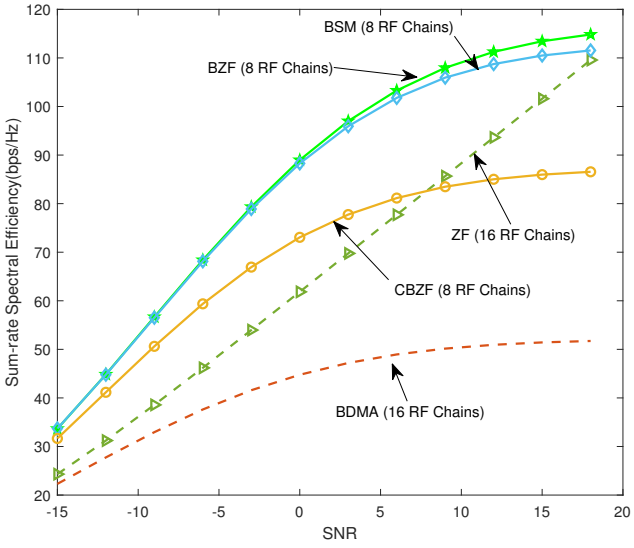

Fig. 2. Sum-rate capacity comparison with different algorithm.

We firstly compare the two proposed algorithm for RF chains reduction with conventional algorithm. As shown in Fig. 2, the dash line "ZF" is the conventional zero-forcing precoding system where 2×8 RF chains are required to serve 16 users. The line "Single user" is fully digital precoding system implemented by SVD. The BDMA is the analog-only precoding system. For the sake of fairness, the sum-rate capacities of dash lines are divided by 2. The solid lines represent the proposed RF chains reduction hybrid beamforming systems where only 8 RF chains are used to serve 16 users. The performance conventional block zero-forcing (CBZF) can be significantly increased by the proposed block zero-forcing(BZF) and block SLNR maximization (BSM) with clustering. The performances of CBZF and BSM are very similar.

In Fig. 3, we can see that the sum-rate capacity will increase as more RF chains added. The upper bound is the zero-forcing precoding system with 16 RF chains for 16 users. The conventional ZF is the hybrid precoding system with 8 RF chains for 8 users.

The Fig. 5 shows that the BZF and BSM are lower bounded by the BDMA and upper bounded by conventional zero-forcing.

Finally, the performance for different systems with increasing number of antennas is shown in Fig. 4. By varying the number of antennas, we investigate the sum-rate capacity improvement.

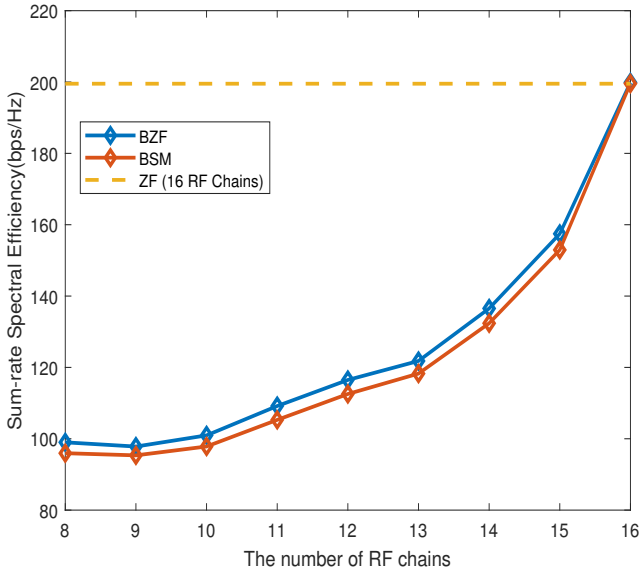

Fig. 3. Different number of RF chains.

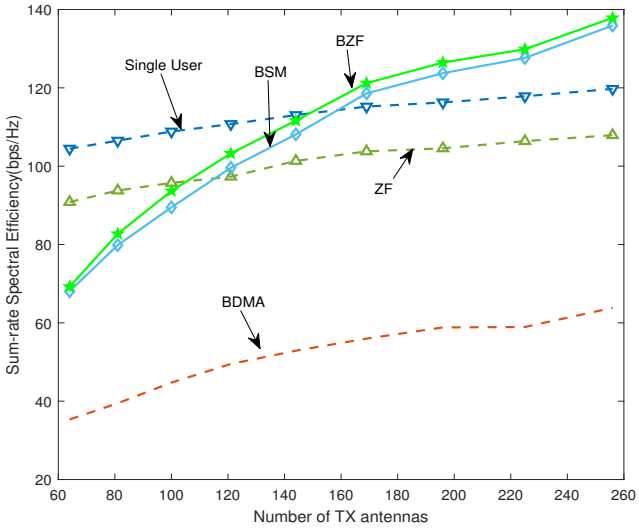

Fig. 4. The sum-rate capacity for different number of TX antennas.

The capacity of BZF and BSM will significantly increase as the more antennas. The reason is that the residual interference from Eq. (19) will be reduced for more antennas.

V. CONCLUSION

In this work, we have developed block clustering precoding scheme for mmWave massive MIMO systems by jointly performing hybrid analog-digital precoding and user-beam clustering. First, we have modeled the block hybrid precoder design to reduce the required RF chains by processing the data streams cluster-by-cluster and the analog precoder can be solved by greedily maximizing the sum-SIR. Furthermore, two approaches

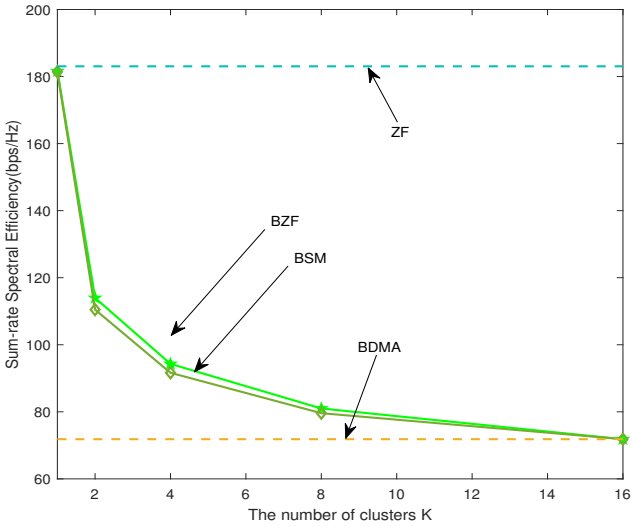

Fig. 5. The sum-rate capacity for different number of TX antennas.

are proposed, namely BZF and BSM, to obtain the digital precoder. Simulation results have confirmed that the proposed block clustering precoding scheme can achieve better sum-rate capacity with less RF chains compared to the conventional hybrid precoding scheme. The two approaches BZF and BSM can be considered to have the equivalent performance.

REFERENCES

[1] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, 2014.

[2] C. Sun, X. Gao, S. Jin, M. Matthaiou, Z. Ding, and C. Xiao, "Beam Division Multiple Access Transmission for massive MIMO communications," *IEEE Transactions on Communications*, vol. 63, no. 6, pp. 2170–2184, June 2015.

[3] Z. Jiang, S. Chen, S. Zhou, and Z. Niu, "Joint user scheduling and beam selection optimization for beam-based massive MIMO downlinks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 4, pp. 2190–2204, April 2018.

[4] T. E. Bogale and L. B. Le, "Beamforming for multiuser massive MIMO systems: Digital versus hybrid analog-digital," *arXiv preprint arXiv:1407.0446*, 2014.

[5] S. Han, I. Chih-Lin, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 186–194, 2015.

[6] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, 2014.

[7] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser mimo channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 461–471, 2004.

[8] W. Ni and X. Dong, "Hybrid block diagonalization for massive multiuser MIMO systems," *IEEE transactions on communications*, vol. 64, no. 1, pp. 201–211, 2016.

[9] A. Liu and V. Lau, "Phase only RF precoding for massive MIMO systems with limited RF chains," *IEEE Transactions on Signal Processing*, vol. 62, no. 17, pp. 4505–4515, 2014.

[10] N. Garcia, H. Wymeersch, and E. G. Larsson, "MIMO with more users than RF chains," *arXiv preprint arXiv:1709.05200*, 2017.

[11] A. Alkhateeb, G. Leus, and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Transactions on Wireless Communications*, vol. 14, no. 11, pp. 6481–6494, 2015.

[12] T. S. Rappaport, R. W. Heath Jr, R. C. Daniels, and J. N. Murdock, *Millimeter wave wireless communications*. Pearson Education, 2014.

[13] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, 2014.

[14] J. Wang, S. Jin, X. Gao, K.-K. Wong, and E. Au, "Statistical eigenmode-based SDMA for two-user downlink," *IEEE Transactions on Signal Processing*, vol. 60, no. 10, pp. 5371–5383, 2012.