Constrained K-means User Clustering and Downlink Beamforming in MIMO-SCMA systems

Sara Norouzi*, Yunlong Cai†, and Benoit Champagne*
*Department of Electrical and Computer Engineering, McGill University, Montreal, Quebec, Canada
†College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China

2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC) | 978-1-7281-7586-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/PIMRC50174.2021.9569415

***Abstract*—In this paper, we study the application of spatial user clustering along with downlink beamforming in multiple-input multiple-output sparse code multiple access (MIMO-SCMA) systems. A user clustering algorithm based on a constrained K-means method is proposed to limit the number of users in each cluster. Subsequently, a two-stage beamforming approach is developed in which a *cluster* beamformer and *user-specific* beamformer obtained from each stage are combined to form the final beamformer for each user. Specifically, in the first stage, the block diagonalization technique is employed to design cluster beamformers so that the inter-cluster interference is removed. In the second stage, an optimization problem is formulated to determine user-specific beamformers for all the users such that the total transmit power is minimized under signal-to-interference-plus-noise ratio (SINR) constraints. The performance of the proposed user clustering and downlink beamforming approaches in MIMO-SCMA systems is evaluated through simulations. The results provide useful insights into the advantages of the proposed scheme in terms of transmit power, and spectral efficiency over benchmark approaches.**

I. Introduction

Through spatial diversity, multiplexing or beamforming gain, the multiple-input multiple-output (MIMO) techniques can offer significant performance improvements in terms of user capacity, spectral efficiency, and peak data rates [1]. Recently, the application of MIMO techniques along with non-orthogonal multiple access (NOMA) has aroused great interest as an enabling technology to meet the exacting demands of fifth generation (5G) and beyond 5G (B5G) wireless networks. In effect, by allowing multiple users to access overlapping time and frequency resources in the same spatial layer, NOMA has the potential to provide higher system throughput and solve the massive connectivity needed for future wireless networks [2].

Sparse code multiple access (SCMA) is a NOMA scheme inspired from the well-known code division multiple access (CDMA) technique. SCMA directly maps groups of bits into a sequence of complex codewords derived from a predefined codebook which can be interpreted as a coding procedure from the binary domain to a multidimensional complex domain [3]. MIMO-SCMA combines MIMO techniques, which increase capacity by transmitting different signals over multiple antennas, and SCMA which improves spectral efficiency and device connectivity by transmitting multiple user signals over the same radio resources.

In [4], a joint sparse graph is constructed for a MIMO-SCMA system model, and the corresponding virtual SCMA codebooks are designed for the detector, wherein the message passing algorithm (MPA) is employed to reconstruct the transmitted data bits. In [5], a joint decoding algorithm is proposed for MIMO-SCMA systems based on space frequency block codes (SFBC), which exhibits lower computational complexity than MPA and yet achieves a similar block error rate (BLER). A novel downlink MIMO mixed-SCMA scheme is proposed in [6], such that the transmitted codewords for each user over different antennas come from different codebooks. The authors in [7] propose near-optimal low-complexity iterative receivers based on a factor graph for a downlink MIMO-SCMA system over frequency selective fading channels. The design of robust radio resource allocation and beamforming approaches for MIMO-SCMA systems under C-RAN is studied in [8], where the aim is to maximize the total sum rate of users subject to a minimum required rate for each slice.

In spite of its potential benefits in terms of spectral efficiency and transmit power, the joint problem of user clustering and beamforming has not received considerable attention in the literature on MIMO-SCMA. Motivated by such consideration, we propose herein novel user clustering and downlink beamforming approaches in a MIMO-SCMA system, assuming that the channel state information (CSI) is known at the base station (BS). We approach the user clustering problem by proposing a so-called *constrained* K-means algorithm, in order to limit the number of users in each cluster. We then conceive a two-stage beamforming approach wherein the beamforming matrices obtained from each stage are multiplied to form a single beamformer. In the first stage, a block diagonalization (BD) technique is adopted to design the cluster beamformers. BD removes the inter-cluster interference, thereby enhancing the quality-of-service (QoS) for intra-cluster users and allowing the system to use a common codebook among users in different clusters. In the second stage, the user-specific beamformers are designed, such that the total transmit power is minimized subject to signal-to-interference-plus-noise ratio (SINR) constraints. Through simulations, it is shown that applying the proposed design to a MIMO-SCMA system can greatly improve spectral efficiency compared to competing ones in the literature.

II. SYSTEM MODEL

In this work, we consider downlink cellular transmission in a MIMO-SCMA system with one BS, as illustrated in Fig. 1. The transmitter is equipped with M antennas, and serves J single-antenna users. The latter are partitioned into K non-overlapping clusters, indexed by $k \in \mathcal{K} \triangleq \{1, \ldots, K\}$ with the kth cluster comprising J_k users, indexed by $j \in \mathcal{J}_k \triangleq \{1, \ldots, J_k\}$. For each user, the BS utilizes the SCMA encoder and beamforming simultaneously.

A. SCMA Encoder

In SCMA, groups of user data bits are directly mapped to sparse N-dimensional codewords. The latter are selected from a codebook specified for each user and then transmitted over N radio resources, e.g., orthogonal frequency division multiple access (OFDMA) subcarriers. Hence, the SCMA encoder for the ith user can be defined as a one-to-one mapping from the set of U-bit tuples to a codebook $\mathcal{X}_i \subset \mathbb{C}^N$, using a function $f_i : \mathbb{B}^U \rightarrow \mathcal{X}_i$, where we define $\mathbb{B} = \{0, 1\}$. The codebook contains N-dimensional codewords, with cardinality $|\mathcal{X}_i| = 2^U$. Specifically, for $\mathbf{b} = [b_1, ..., b_U] \in \mathbb{B}^U$, the corresponding codeword is obtained as,

$$\mathbf{x} = f_i(\mathbf{b}) = [x(1), ..., x(N)] \tag{1}$$

where $\mathbf{x}$ is a sparse vector with $C < N$ non-zero elements.

Each user is assigned C subcarriers such that no two users occupy the same set of subcarriers. Hence, only q users can be supported by SCMA, as given by [9],

$$q = \begin{pmatrix} N \\ C \end{pmatrix} = \frac{N!}{C!(N-C)!}. \tag{2}$$

In this work, we group users into K clusters of size $J_k \leq q$ and remove inter-cluster interference so that the users in different clusters can use common codebooks.

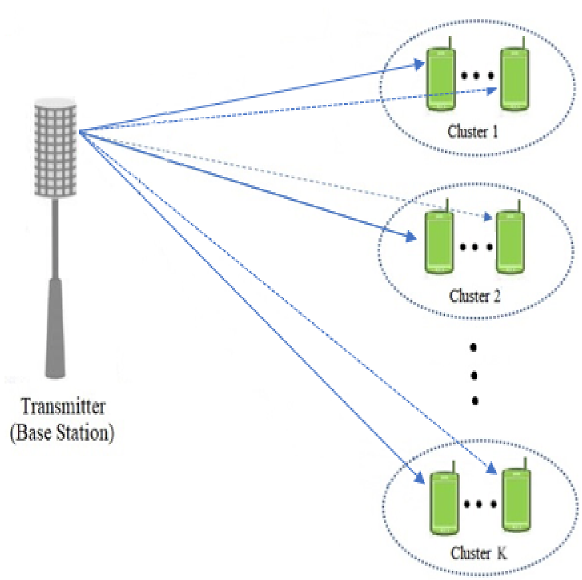


Figure 1: Downlink MIMO-SCMA system model.

The SCMA encoder can be expressed as $f_i(\mathbf{b}) = \mathbf{S}_i g_i(\mathbf{b})$ where $g_i : \mathbb{B}^U \rightarrow \mathbb{C}^C$ is a mapping from the set of U-bit tuples to a C-dimensional constellation point with non-zero elements, while matrix $\mathbf{S}_i$ maps this latter point into an N-dimensional codeword. It is worth noting that $\mathbf{S}_i$ contains $N - C$ all-zero rows and an identity matrix of order C is obtained by removing them. Hence, all the codewords in $\mathcal{X}_i$ contain 0 in the same $N - C$ positions.

The positions (or indices) of the non-zero elements of the binary indicator vector obtained by $\mathbf{f}_i = \text{diag}(\mathbf{S}_i \mathbf{S}_i^T) \in \mathbb{B}^{N \times 1}$ determine the set of subcarriers occupied by user i. In effect, the complete SCMA encoder structure for q users and N subcarriers can be represented by a factor graph, with associated matrix $\mathbf{F} = [\mathbf{f}_1, ..., \mathbf{f}_q] \in \mathbb{B}^{N \times q}$. In this interpretation, subcarrier node n and user node i are connected if and only if the corresponding element of matrix $\mathbf{F}$ is equal to 1, i.e., $[\mathbf{F}]_{n,i} = 1$.

B. Signal Model

Let $\mathbf{h}_{j,k}(n) \in \mathbb{C}^{1 \times M}$ denote the channel vector from BS to the jth user in the kth cluster over the nth subcarrier. The channel vectors are assumed to be known at the BS. The received signal at the jth user in the kth cluster over the nth subcarrier is given by

$$r_{jk}(n) = \mathbf{h}_{jk}(n)\mathbf{z}(n) + \eta_{jk}(n) \tag{3}$$

where $\mathbf{z}(n) \in \mathbb{C}^{M \times 1}$ is the transmit signal of the BS over the nth subcarrier and $\eta_{jk}(n) \sim \mathcal{CN}(0, \sigma_{jk}^2)$ is an additive noise term.

Let $x_{jk}(n) \in \mathbb{C}$ denote the codeword element intended for the jth user in the kth cluster over the nth subcarrier which can be either 0, or a non-zero element due to the sparsity of the SCMA encoder. If $x_{jk}(n)$ is non-zero, its is normalized to one, i.e., $E\{|x_{jk}(n)|^2\} = 1$ and it is transmitted from the BS by employing the beamforming vector $\mathbf{w}_{jk}(n) \in \mathbb{C}^{M \times 1}$. Hence, we have

$$\mathbf{z}(n) = \sum_{k=1}^{K} \sum_{j \in \mathcal{U}_{n,k}} \mathbf{w}_{jk}(n) x_{jk}(n) \tag{4}$$

where $\mathcal{U}_{n,k}$ denotes the set of users in the kth cluster occupying the nth subcarrier.

Upon substitution of (4) into (3), we can express the received signal of this user as a sum of the desired signal, the interference from the other users in that cluster (intra-cluster interference), the inter-cluster interference and the noise. The SINR of the jth user in the kth cluster over the nth subcarrier with non-zero codeword element is

$$\gamma_{jk}(n) = \frac{|\mathbf{h}_{jk}(n)\mathbf{w}_{jk}(n)|^2}{I_{jk}^{(1)}(n) + I_{jk}^{(2)}(n) + \sigma_{jk}^2} \tag{5}$$

where the first term in the denominator represents the intra-cluster interference and the second term shows the inter-cluster interference, i.e.,

$$I_{jk}^{(1)}(n) = \sum_{j' \neq j, j' \in \mathcal{U}_{n,k}} |\mathbf{h}_{jk}(n)\mathbf{w}_{j'k}(n)|^2 \tag{6}$$

$$I_{jk}^{(2)}(n) = \sum_{k' \neq k} \sum_{j' \in \mathcal{U}_{n,k'}} |\mathbf{h}_{jk}(n)\mathbf{w}_{j'k'}(n)|^2. \quad (7)$$

In this work, we seek to reduce the total transmit power of a MIMO-SCMA system under QoS requirements, while supporting a larger number of users by sharing a common codebook. We conceive efficient algorithms for user clustering and beamforming design with low complexity to obtain the desired solution. Specifically, we propose a user clustering algorithm based on the constrained K-means method in Section III to group users into non-overlapping clusters. Then, the beamformer design is addressed in Section IV by means of a two-stage energy-efficient approach which removes inter-cluster interference and allows codebook sharing among the users in different clusters.

III. User Clustering

Although an exhaustive search for spatial user clustering can achieve maximum capacity, this type of approach entails high computational complexity, which grows exponentially with the number of users. Alternatively, under the assumption of known CSI, we can employ a low-complexity algorithm based on machine learning to group users into clusters. In this section, we first introduce the proposed constrained K-means algorithm for user clustering and then apply the elbow method to determine the number of clusters.

A. Constrained K-means Clustering

K-means is a popular method for clustering data points such that a similarity criterion within clusters is maximized [10]. Given a set of J multi-dimensional data points $\{\mathbf{d}_1, \mathbf{d}_2, ..., \mathbf{d}_J\}$, K-means attempts to group these points into K clusters by finding cluster centers $\{\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_K\}$ such that similarities between the points in the same group are high while similarities between the points in different groups are low. Two key factors in the K-means method are the number of clusters K, which is pre-determined, and the similarity metric [11].

In this work, to account for variations of channel gains due to fading and other propagation effects, the averages of normalized channel vectors over the available subcarriers are treated as the data points in the application of the K-means method, i.e.,

$$\mathbf{d}_j = \frac{1}{N} \sum_{n=1}^{N} \frac{\mathbf{h}_j(n)}{\|\mathbf{h}_j(n)\|_2} \quad (8)$$

where $\mathbf{h}_j(n) \in \mathbb{C}^{1\times M}$ for $j \in \mathcal{J} \triangleq \{1, ..., J\}$ are the known channel vectors of all users prior to clustering. Hence, the subcarrier index is removed in this section and the user clustering algorithm is carried out once for every N subcarriers. In the current MIMO-SCMA application, if users in a cluster have highly correlated channels, a better beamforming performance will be obtained. Indeed, high correlation between the channel vectors of the users in a cluster can provide more degrees of freedom for the inter-cluster interference cancellation (as explained in Section IV). In this work, the Euclidian distance is utilized as a metric to quantify the similarity between a user's data point and the cluster centers.

The K-means method can be presented as an optimization problem for finding the K best centers such that the sum of squared Euclidean (SSE) distance between the data points and their nearest cluster centers is minimized. Specifically, this optimization problem can be expressed as follows,

$$\min_{\mathbf{C}} \sum_{j} \min_{k \in \mathcal{K}} \|\mathbf{d}_j - \mathbf{c}_k\|_2^2 \quad (9)$$

where $\mathbf{C} \triangleq \{\mathbf{c}_k | k \in \mathcal{K}\}$. By introducing selection variables $\boldsymbol{\iota} \triangleq \{\iota_{j,k} | j \in \mathcal{J}, k \in \mathcal{K}\}$ and using the linear programming duality theory, we can reformulate problem (9) as the following program,

$$\min_{\boldsymbol{\iota}, \mathbf{C}} \sum_{j} \sum_{k} \iota_{j,k} \|\mathbf{d}_j - \mathbf{c}_k\|_2^2 \quad (10a)$$

$$\text{s.t.} \quad \sum_{k} \iota_{j,k} = 1, \quad \forall j \in \mathcal{J} \quad (10b)$$

$$\iota_{j,k} \in \{0, 1\}, \quad \forall j \in \mathcal{J}, k \in \mathcal{K} \quad (10c)$$

where $\iota_{j,k} = 1$ if the jth data point is closest (or belong) to the kth cluster center, and $\iota_{j,k} = 0$ otherwise.

While at most q users can be supported by the SCMA encoder in the current application, the K-means method does not have *a priori* constraint on the number of users in each cluster [12]. We propose adding explicit constraints to problem (10) to avoid solutions with more than q data points in a cluster, i.e.,

$$\min_{\boldsymbol{\iota}, \mathbf{C}} \sum_{j} \sum_{k} \iota_{j,k} \|\mathbf{d}_j - \mathbf{c}_k\|_2^2 \quad (11a)$$

$$\text{s.t.} \quad (10b), (10c) \quad (11b)$$

$$\sum_{j} \iota_{j,k} \leq q, \quad \forall k \in \mathcal{K}. \quad (11c)$$

Problem (11) can be solved iteratively by uncoupling cluster center and selection variables. Specifically, in each iteration, the constrained K-means algorithm alternates between solving a linear program for variable $\boldsymbol{\iota}$ with fixed $\mathbf{c}$ and solving a convex problem for $\mathbf{c}$ with fixed $\boldsymbol{\iota}$. The overall constrained K-means algorithm for solving problem (11) is summarized in Algorithm 1, where the superscript t denotes the iteration index. We can use a mixed integer linear programming solver such as MOSEK [13] for tackling the cluster assignment subproblem.

B. Number of Clusters

The choice of the number of clusters K plays a key role in the performance of K-means clustering [14]. Herein, the elbow method is utilized to determine the number of clusters. In effect, the elbow method runs the clustering algorithm on the dataset and evaluates a clustering criterion for different values of K. The sum of the normalized within-cluster SSE distance (W_K) is generally used as a clustering criterion for applying the elbow method along with K-means. The plot

Algorithm 1: The proposed constrained K-means algorithm for user clustering.

Initialization: Choose K cluster centers $\mathbf{c}^{(0)} = \{\mathbf{c}_1^{(0)}, \mathbf{c}_2^{(0)}, \ldots, \mathbf{c}_K^{(0)}\}$ from the dataset randomly. Set $t = 0$.

Repeat:

1) **Cluster assignment:** Solve the following linear program with fixed $\mathbf{c}^{(t)}$.

$$\boldsymbol{\iota}^{(t)} = \arg\min_{\boldsymbol{\iota}} \sum_j \sum_k \iota_{j,k} \|\mathbf{d}_j - \mathbf{c}_k^{(t)}\|_2^2 \quad \text{s.t. (10b), (10c), (11c).}$$

2) **Cluster update:** Update the cluster centers as,

$$\mathbf{c}_k^{(t+1)} = \frac{\sum_j \iota_{j,k}^{(t)} \mathbf{d}_j}{\sum_j \iota_{j,k}^{(t)}}, \quad \forall k \in \mathcal{K}.$$

3) Set $t \leftarrow t + 1$.

Until: $\mathbf{c}_k^{(t)} = \mathbf{c}_k^{(t-1)}, \forall k \in \mathcal{K}$.

of W_K versus K resembles an arm in which the elbow point (the point where the curve visibly bends) determines the appropriate number of clusters for the dataset.

In a given cluster $\mathcal{C}_k$, the within-cluster SSE distance between the data points is given by,

$$D_k = \frac{1}{2} \sum_{\mathbf{d}_i \in \mathcal{C}_k} \sum_{\mathbf{d}_{i'} \in \mathcal{C}_k} \|\mathbf{d}_i - \mathbf{d}_{i'}\|_2^2. \tag{12}$$

Hence, the sum of the normalized within-cluster SSE distances can be expressed as,

$$W_K = \sum_{k=1}^{K} \frac{1}{|\mathcal{C}_k|} D_k \tag{13}$$

where $|\mathcal{C}_k|$ shows the cardinality of the cluster $\mathcal{C}_k$. It should be noted that although the sum of the normalized within-cluster SSE distance can give a proper measure of the compactness of the clustering, we may encounter cases with more than one elbow point or no elbow point. In such cases, other reliable methods, e.g. [15], can be used to find the best K.

IV. Downlink Beamforming

Our aim is to minimize the total transmit power subject to the SINR constraints, while removing the inter-cluster interference through the design of the beamforming vectors. Hence, the users in different clusters can use a common codebook. In order to reduce the computational complexity, we propose a two-stage energy-efficient beamforming approach such that

$$\mathbf{w}_{jk}(n) = \mathbf{B}_k(n)\mathbf{v}_{jk}(n) \tag{14}$$

where $\mathbf{B}_k(n) \in \mathbb{C}^{M\times a}$ is the kth cluster beamformer obtained in the first stage, which should eliminate the inter-cluster interference, and $\mathbf{v}_{jk}(n) \in \mathbb{C}^{a\times 1}$ is the user-specific beamformer for the jth user in the kth cluster optimized in the second stage. The dimension parameter a will be determined as part of our developments below.

Stage 1: BD beamforming can be adopted in a MIMO-SCMA system to remove the inter-cluster interference. Specifically, the former projects the transmitted signal onto the null-space of the interfering channels. To find the corresponding null-space, let us define,

$$\mathbf{H}_k(n) = [\mathbf{h}_{1k}(n)^T \ldots \mathbf{h}_{Jk}(n)^T] \in \mathbb{C}^{M\times J_k} \tag{15}$$

$$\bar{\mathbf{H}}_k(n) = [\mathbf{H}_1(n) \ldots \mathbf{H}_{k-1}(n)\ \mathbf{H}_{k+1}(n) \ldots \mathbf{H}_K(n)] \tag{16}$$

where $k \in \mathcal{K}$ and $\bar{\mathbf{H}}_k(n) \in \mathbb{C}^{M\times(J-J_k)}$ is the matrix containing all interfering channels for the kth cluster. We seek $\mathbf{B}_k(n)$ orthogonal to the column span of $\bar{\mathbf{H}}_k(n)$, i.e., $\bar{\mathbf{H}}_k(n)^T\mathbf{B}_k(n) = \mathbf{0}$. It is assumed that the total number of antennas M is larger than the total number of users J.

The singular value decomposition (SVD) can be employed to calculate the cluster beamformers. Applying the SVD to $\bar{\mathbf{H}}_k(n)$ yields,

$$\bar{\mathbf{H}}_k(n) = \mathbf{U}_k(n)\mathbf{\Sigma}_k(n)\mathbf{V}_k(n)^H \tag{17}$$

where $\mathbf{U}_k(n) \in \mathbb{C}^{M\times M}$ and $\mathbf{V}_k(n) \in \mathbb{C}^{(J-J_k)\times(J-J_k)}$ are unitary matrices and $\mathbf{\Sigma}_k(n) \in \mathbb{R}^{M\times(J-J_k)}$ is the rectangular diagonal matrix of singular values. Let r denote the rank of matrix $\bar{\mathbf{H}}_k(n)$, which corresponds to the number of non-zero diagonal entries in $\mathbf{\Sigma}_k(n)$. The null-space of the interfering channel matrix $\bar{\mathbf{H}}_k(n)$ is spanned by the left singular vectors (i.e. columns of matrix $\mathbf{U}_k(n)$) associated to the zero singular values of $\bar{\mathbf{H}}_k(n)$. We can express the kth cluster beamformer as,

$$\mathbf{B}_k(n) = [\mathbf{u}_{r+1,k}(n)\ \mathbf{u}_{r+2,k}(n)\ \ldots \mathbf{u}_{M,k}(n)] \tag{18}$$

where $\mathbf{u}_{i,k}(n)$ denotes the ith column of $\mathbf{U}_k(n)$. It can be seen that $a = M - r$.

Stage 2: To guarantee the QoS among users, our aim is to determine the user-specific beamformers such that the total transmit power is minimized subject to SINR constraints. This optimization problem can be formulated as

$$\min_{\mathbf{v}_{j,k}(n)} \sum_n \sum_k \sum_j \|\mathbf{w}_{jk}(n)\|^2 \tag{19a}$$

$$\text{s.t.}\quad \mathbf{w}_{jk}(n) = \mathbf{B}_k(n)\mathbf{v}_{jk}(n), \quad \forall\, j,k,n \tag{19b}$$

$$\gamma_{jk}(n) \geq \gamma_{\text{min}}, \quad \forall n \in \mathcal{N}_{jk} \tag{19c}$$

where γ_{min} is the minimum required SINR for the user over the subcarriers and $\mathcal{N}_{jk}$ denotes the set of subcarriers occupied by the jth user in the kth cluster.

As mentioned before, using cluster beamformer $\mathbf{B}_k(n)$ obtained from BD can remove inter-cluster interference. Hence, the SINR of the jth user in the kth cluster over the nth subcarrier can be expressed as,

$$\gamma_{jk}(n) = \frac{|\mathbf{h}_{jk}(n)\mathbf{w}_{jk}(n)|^2}{\sum_{j'\neq j}^{J_k} |\mathbf{h}_{jk}(n)\mathbf{w}_{j'k}(n)|^2 + \sigma_{jk}^2} \tag{20}$$

where the inter-cluster interference term is removed. Then, problem (19) can be equivalently rewritten as

$$\min_{\mathbf{v}_{jk}(n)} \quad \sum_n \sum_k \sum_j \|\mathbf{w}_{jk}(n)\|^2 \tag{21a}$$

$$\text{s.t. } \mathbf{w}_{jk}(n) = \mathbf{B}_k(n)\mathbf{v}_{jk}(n) \tag{21b}$$

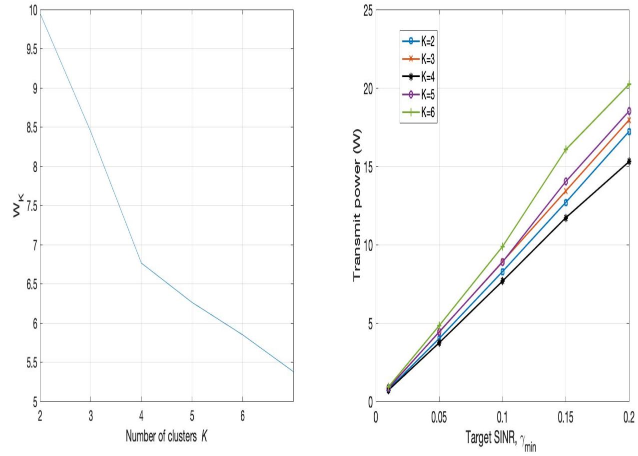
$$\sqrt{\sum_{j'\neq j} |\mathbf{h}_{jk}(n)\mathbf{w}_{j'k}(n)|^2 + \sigma_{jk}^2} \leq \frac{\mathbf{h}_{jk}(n)\mathbf{w}_{jk}(n)}{\sqrt{\gamma_{\min}}} \tag{21c}$$

where in (21c), we have restricted $\mathbf{h}_{jk}(n)\mathbf{w}_{jk}(n)$ to be positive, which incurs no loss of optimality since we can always phase-rotate the vector $\mathbf{w}_{jk}(n)$ such that $\mathbf{h}_{jk}(n)\mathbf{w}_{jk}(n)$ is positive real without affecting the cost function or the constraints. Problem (21) is convex and can be solved via any general-purpose solver using interior-point methods [16].

V. Numerical results

In this section, numerical experiments are carried out to illustrate the performance of the proposed user clustering and beamforming approaches in a MIMO-SCMA system. For this purpose, we consider a single cell with one BS, which is equipped with M antennas for downlink transmission. In our simulations, we model the channel coefficients between the transmit antennas of the BS and the receive antenna of each user as the product of a path loss and complex Rayleigh fading coefficient, with zero-mean and unit variance. The path loss is given by $L = \gamma(d_0/d)^\eta$, where d is the distance between the transmit and receive antennas, $d_0 = 1$m is the reference distance, $\gamma = 0.01$ is a constant, and $\eta = 2$ is the path loss exponent. The users are distributed uniformly over a circular cell with radius 200 m. Throughout the experiments, it is assumed that the noise variance is the same for all users, which is normalized to $\sigma_{jk}^2 = -147$ dBm.

Figs. 2a and 2b illustrate the determination of the optimal number of clusters K and evaluate its impact on the performance of the proposed scheme. In Fig. 2a, we plot W_K which serves as clustering criterion in the elbow method described in Secition III.B. It can be seen W_K decreases when K increases and the elbow point can be found at $K = 4$. To gain further insight into the impact of the number of clusters, we investigate the transmit power performance versus target SINR γ_{min} in Fig. 2b, where the number of clusters increases from 2 to 6. It is observed that the total transmit power increases monotonically as γ_{min} increases. Moreover, the best performance is achieved when the number of clusters is $K = 4$. On the one hand, for $K < 4$, an increase in the number of users in a cluster results in larger intra-cluster interference which results in higher transmit power. On the other hand, increasing the number of clusters intensifies inter-cluster interference which increases power consumption in the first stage beamforming for interference cancellation.



(a) Elbow method (b) Transmit power vs. $\gamma_{\min}$

Figure 2: Impact of the number of clusters

Fig. 3 compares the achievable sum rate versus the total transmit power among different transmission schemes. We consider orthogonal multiple access (OMA) and power domain non-orthogonal multiple access (PD-NOMA) as benchmarks with similar parameters, except for the number of multiplexed signals over each subcarrier. Specifically, in OMA, each user is assigned only one subcarrier such that no interference occurs with other user signals. Hence, the maximum number of users in each cluster is equal to the number of subcarriers, i.e., $q = N$. In PD-NOMA, all users have access to all the subcarriers and no constraint is applied to the maximum

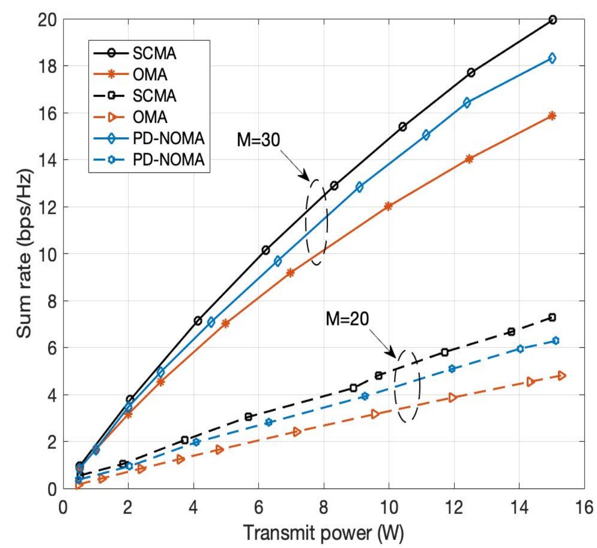


Figure 3: Sum rate vs. transmit power of different schemes

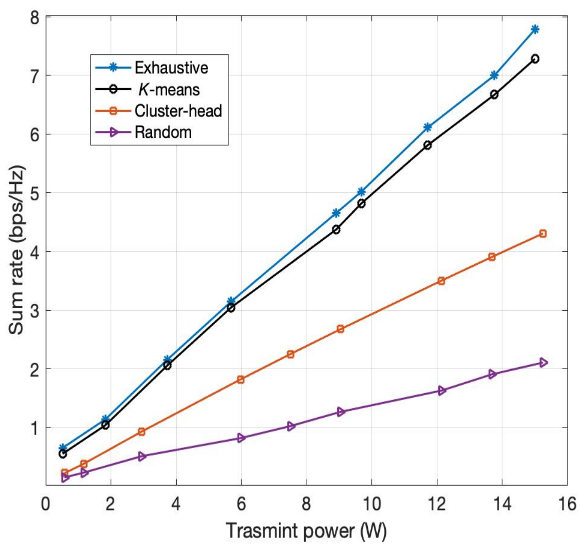


Figure 4: Sum rate vs. transmit power

number of users in a cluster. Two different number of antennas are considered for this purpose, i.e. $M = 20$ and 30. It is observed that in both cases, the proposed SCMA scheme outperforms OMA and PD-NOMA in terms of sum rate and the performance gap gets larger as the transmit power increases. For larger M, better beamforming results are expected as more degrees of freedom will be left for the inter-cluster interference cancellation. We can observe that when sum rate is 4 bps/Hz, the total transmission power decreases by 7 W as the number of antennas increases from 20 to 30 which is a consequence of narrower beamforming.

In Fig. 4, we compare the sum rate versus transmit power among different clustering algorithms. The cluster-head approach proposed in [17], which selects the K users with the highest channel gains as the cluster centers, is used as a benchmark for user clustering. The cluster assignment is then used to group users into clusters. We also consider the performance of the MIMO-SCMA system using exhaustive and random search. As it can be seen from Fig. 4, the proposed constrained K-means clustering algorithm exhibits better performance in terms of sum rate and can partition users more efficiently compared to cluster-head approach and random search. While the performance of the K-means is close to that of exhaustive search, the former entails less computational complexity.

VI. Conclusion

In this work, we considered the joint design of user clustering and downlink beamforming approach in a MIMO-SCMA system, assuming that CSI is available at the BS. The constrained K-means algorithm was proposed and applied to spatially partition users into non-overlapping clusters based on the correlation between channel vectors. The cluster and user-specific beamformers were designed in a two-stage beamforming approach to remove the inter-cluster interference and reduce total transmit power, respectively. In the first stage, the BD beamforming approach was employed to design the cluster beamformers. In the second stage, the design of user-specific beamformers was determined by minimizing the total transmit power under the SINR constraints. Simulation results showed that applying the proposed design to MIMO-SCMA systems can effectively decrease total transmit power and improve spectral efficiency compared to the benchmark approaches.

References

[1] Y. Cai, Z. Qin, F. Cui, G. Y. Li, and J. A. McCann. "Modulation and multiple access for 5G networks," *IEEE Commun. Surveys and Tutorials*, vol. 20, no. 1, pp. 629-646, 2018.

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

[3] M. Taherzadeh, H. Nikopour, A. Bayesteh, and H. Baligh. "SCMA codebook design," *IEEE Vehi. Technol. Conf. (VTC-Fall)*, pp. 1-5, Vancouver, Canada, Sept. 2014.

[4] Y. Du, B. Dong, Z. Chen, P. Gao, and J. Fang. "Joint sparse graph-detector design for downlink MIMO-SCMA systems," *IEEE Wireless Commun. Lett.*, vol. 6, no. 1, pp. 14-17, Nov. 2016.

[5] Z. Wu, C. Zhang, X. Shen, and H. Jiao. "Low complexity uplink SFBC-based MIMO-SCMA joint decoding algorithm," *IEEE Int. Conf. on Computer and Commun. (ICCC)*, pp. 968-972, Chengdu, China, Dec. 2017.

[6] C. Yan, N. Zhang, and G. Kang. "Downlink multiple input multiple output mixed sparse code multiple access for 5G system," *IEEE Access*, vol. 6, pp. 20837-20847, Apr. 2018.

[7] W. Yuan, N. Wu, Q. Guo, Y. Li, C. Xing, and J. Kuang. "Iterative receivers for downlink MIMO-SCMA: Message passing and distributed cooperative detection," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 3444-3458, June 2018.

[8] M. Moltafet, S. Parsaeefard, M. R. Javan, and N. Mokari. "Robust radio resource allocation in MISO-SCMA assisted C-RAN in 5G networks," *IEEE Trans. Vehi. Technol.*, vol. 68, no. 6, pp. 5758 - 5768, Apr. 2019.

[9] M. Vameghestahbanati, I. Marsland, R. H. Gohary, and H. Yanikomeroglu. "Multidimensional constellations for uplink SCMA systems—A comparative study," *IEEE Commun. Surveys & Tutorials*, vol. 21, no. 3, pp. 2169 - 2194, April 2019.

[10] T. Niknam, E. T. Fard, N. Pourjafarian, and A. Rousta. "An efficient hybrid algorithm based on modified imperialist competitive algorithm and K-means for data clustering," *Eng. Applications of Artificial Intelligence*, vol. 24, no. 2, pp. 306-317, Mar. 2011.

[11] A. K. Jain. "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651-666, Jun. 2010.

[12] P. S. Bradley, K. P. Bennett, and A. Demiriz. " Constrained K-means clustering," *Technical Report Microsoft Research*, Redmond, WA, USA, May 2000.

[13] E. D. Andersen, and K. D. Andersen. " *The MOSEK interior point optimizer for linear programming: an implementation of the homogeneous algorithm,*" Springer, Boston, MA, 2000.

[14] M. Inaba, N. Katoh, and H. Imai. "Applications of weighted Voronoi diagrams and randomization to variance-based K-clustering: (extended abstract)," *Proc. of the Tenth Annual Symposium on Computational Geometry*, NY, USA, 1994.

[15] P. J. Rousseeuw. "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, no. 1, pp. 53-65, Jan. 1987.

[16] M. Grant and S. Boyd. "CVX: Matlab software for disciplined convex programming, Version 2.1," Mar. 2014.

[17] 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, Dec. 2016.