

Received May 31, 2020, accepted June 12, 2020, date of publication June 16, 2020, date of current version June 26, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3002774

Joint User Clustering and Beamforming in Non-Orthogonal Multiple Access Networks

HA-RYUNG KIM¹, JIASI CHEN², (Member, IEEE), AND JONGWON YOON¹, (Member, IEEE)

¹Department of Computer Science and Engineering, Hanyang University, Ansan 15588, South Korea

²Department of Computer Science and Engineering, University of California, Riverside, CA 92521, USA

Corresponding author: Jongwon Yoon (jongwon@hanyang.ac.kr)

This work was supported by the Hanyang University under Grant HY-2016-N.

ABSTRACT The demand for higher data rate in wireless networks continues to grow, however the wireless spectrum cannot meet ever-increasing demands. Non-orthogonal Multiple Access (NOMA) technology is attracting attentions to overcome the lack of resource spectrum and increase spatial reuse. Traditionally, data is delivered in an orthogonal fashion in the frequency or time domain, whereas NOMA leverages power domain to improve spectral reuse. User clustering is one of the key features that significantly affects NOMA's throughput performance and fairness among the users in the network. In addition to NOMA, beamforming increases spectral efficiency by improving SNR for specific users (and groups) and reducing interference across the clusters. However, combining beamforming with NOMA is a challenging and complex problem because beamforming for improving the reuse of spectral resources is tightly coupled with user clustering in NOMA. In this paper, we introduce a joint user clustering and beamforming scheme that maximizes the network throughput in NOMA with beamforming. Toward this goal, we carefully analyze several features that significantly affect the system throughput and design a heuristic algorithm based on the inferences. Our algorithm increase the signal power in NOMA and effectively reduce both beamforming and NOMA interference. Evaluation results reveal that our solution provide higher utility (sum rate) than other existing schemes and significantly improve fairness among users regardless of the channel conditions.

INDEX TERMS Beamforming, clustering, non-orthogonal multiple access.

I. INTRODUCTION

The rapid growth of hand-held devices and the emergence of various IoT devices generate large amounts of wireless traffic. In addition, the demand for high quality video and real-time streaming is increasing, and the volume of wireless Internet traffic is also increasing. Despite the thirst for high data rate, wireless resources are limited and must be shared by multiple users. To efficiently use scarce resources and to meet the demand for wireless traffic, many researches on resource management have been proposed.

Toward this, Orthogonal Multiple Access (OMA) which orthogonally divides frequency resources into multiple channels and allocates them to each user, is actively used. Unlike OMA, Non-Orthogonal Multiple Access (NOMA) simultaneously encodes two or more signals into the same resource (same time and frequency band) by allocating different

powers to transmitted signals. This characteristic allows NOMA to handle many users at the same time, and allocates resources more efficiently than OMA, which eventually increase data rate. Although NOMA is superior to OMA in terms of data rate and spectral efficiency, however, the performance of NOMA depends on how users are paired with each other, which is called clustering. The larger the difference in user channel gain is, the less interference occurs in NOMA, therefore the higher data rate can be achieved. In order to increase the channel gain difference, standard user clustering algorithms divide users into two groups, cell-center (CC) set and cell-edge (CE) set, and pairs users in each set to form a cluster [1].

In addition to the clustering algorithm, beamforming can be adopted to improve the spectral efficiency in NOMA. To address the problem involving both beamforming vector generation and user clustering, the authors in [2], [3] adopt CC-CE method and decouple the problem; once user clustering is configured, they find the best beamforming in the given

The associate editor coordinating the review of this manuscript and approving it for publication was Ali Afana.

clusters. CC-CE method is an efficient way of minimizing NOMA interference, however it is not able to maximize the benefits of beamforming in NOMA. Even worse, user clustering is not flexible because the CC and CE sets are already fixed and hence there is a lack of variety in pairing users. In particular, carefully designed beamforming in NOMA could improve network performance even though clustering is performed by users in the same category (e.g., CC or CE).

Our goal is to maximize spectral efficiency in NOMA (i.e., improve network capacity) while leveraging beamforming. To achieve this goal, we consider the NOMA and the beamforming features jointly rather than separate considerations as in the above-mentioned works. First, We formulate a problem of maximizing utility (defined in Eq. (7) in Section IV) in NOMA. We propose an algorithm that jointly solve beamforming and user clustering in NOMA. This joint problem is very complex and has a number of features that affect the utility, thus we analyze the features and find inferences for designing an algorithm. Then, based on the design factors and considerations, we propose an efficient but greedy algorithm for user clustering and beamforming in NOMA. Given that the joint problem of user clustering and beamforming in NOMA is hard, there are several studies that use channel difference for clustering and channel coefficient for beamforming in NOMA. However, these factors do not guarantee an effective way to increase spectral efficiency in NOMA. Although several works have shown that these factors are good for clustering or beamforming in NOMA separately, there is no study for these factors to both user clustering and beamforming. We design a joint approach to address both user clustering algorithm and beamforming. Our contributions in this work are multi-fold.

- We analyze various factors affect network performance in NOMA. We leverage beamforming in NOMA to maximize the gains of both, and design a joint solution.
- To the best of our knowledge, we propose the first joint solution for user clustering and beamforming in NOMA and demonstrate superiority in comparison to other approaches.
- Unlike the conventional user clustering in NOMA, we show that clustering users in vicinity is not harmful, and can improve performance when jointly using beamforming.

The rest of this paper is organized as follows. Section II-B describes the basic concepts of NOMA, beamforming, and related works. We motivate the necessity of coupling user clustering and beamforming in Section III. We formulate the problem, and propose a joint user clustering and beamforming algorithm in Section IV. Section V evaluates the performance of JUCAB while comparing to other schemes. Section VI concludes the paper.

II. BACKGROUND AND RELATED WORK

A. NOMA AND BEAMFORMING

NOMA allows users to share the same resources by adapting transmission power, therefore it increases spatial reuse of

spectrum resources. Base station (BS) uses the superposition coding (SC) technique to divide data according to the user's channel, assigning different power to each user. At the receiver side, user leverages successive interference cancellation (SIC) to decode the signal corresponding to a power greater than the assigned power level and then sequentially subtracts the interference. For user clustering, the CC-CE method is widely used in NOMA.

Unlike Omni-directional antenna, beamforming is used to improve signal power for the intended users while mitigating the interference at the unwanted users. The higher the transmit power toward intended direction, the better the SNR (Signal-to-Noise Ratio) for the desired user. It also allows reuse of spectral resources, thus eventually increases spatial reuse. In addition to NOMA that reuses the same resources for multiple user, beamforming with smart algorithm can further improve the reuse gain.

B. RELATED WORK

1) USER CLUSTERING

Ali *et al.* [4] propose two clustering algorithms to solve the problem of power allocation and user clustering. In both algorithms, cluster consists of a set of CC and CE users. The authors in [5] exploit NOMA into mobile edge computing (MEC) in order to obtain multiple gains. One of the contributions of [5] is to reduce the latency of the NOMA-based MEC system. They propose a clustering scheme based on the user channel gain, however it is similar to the clustering methods proposed in [4]. Ma *et al.* [6] distinguish the near and far users from the BS, and then select two users with the least correlation. In the papers mentioned above, clustering is conducted based on CC and CE users for maximizing the NOMA gains, which is a baseline approach for user clustering in NOMA.

Channel correlation is widely adopted for user clustering in NOMA [2], [7], [8]. In [2], clustering algorithm selects two users with high correlation and a large gain-difference between their channels. Lui *et al.* [7] propose user clustering, which pairs two users with large channel correlation taking into account SIC constraints. Ding *et al.* [9] characterize the impact of user pairing on the performance of NOMA. They showed that NOMA with fixed power allocation can offer a large sum rate than conventional OMA.

2) BEAMFORMING

There are several research works on both NOMA and beamforming. The authors in [2] first consider the channel gain difference and the user channel correlation for NOMA, and then separately take into account the beamforming effect. However, NOMA and beamforming are not considered jointly. In [3], users are assigned to clusters based on predefined beamforming vectors, and then clustering is performed sequentially in the channel gain order. In this work, user clustering is performed in a relatively simple manner, whereas beamforming vector is calculated in a more complicated and

complex manner. This work focuses on beamforming gain rather than NOMA. Overall, the authors focus on designing beamforming vector, however user clustering is not taken into account with NOMA. In addition, criteria for determining users for clustering, such as user channel difference and channel gain correlation, are not carefully verified in NOMA and beamforming environments.

Research on the use of beamforming and user clustering in NOMA is receiving attentions to improve performance and spectral efficiency. The authors in [10] show that the channel differences are important for assigning users in NOMA and the authors in [11] describe that channel coefficients are critical on the performance of beamforming. Choi *et al.* [12] proposed beamforming and user clustering algorithm in MU-MIMO NOMA, however the proposed algorithm solves beamforming and clustering separately. After configuring CC-CE based user clustering, beamforming vector that minimizes the transmit power is found. The authors in [1] proposes a linear beamforming technique that eliminates inter-cluster interference, with a focus on maximizing total spectral efficiency in MU-MIMO NOMA. In addition, a user clustering algorithm using a correlation coefficient is proposed to increase the benefits of beamforming.

3) POWER

In the area of beamforming and power, Islam *et al.* [13] propose D-NLUPA, a resource allocation scheme while focusing on user clustering and power allocation. They present the concept of clustering fairness and exploit the channel gain difference for user clustering. In addition, suggested power allocation method balances system throughput and user fairness. Zeng *et al.* [14] present performance analysis of MIMO-NOMA and propose an optimal user admission scheme for the cluster based on SINR threshold. [15] presents that two-user cluster and beamforming sharing in MIMO-NOMA outperforms MIMO-OMA. This work clearly shows that beamforming improves the performance of MIMO-NOMA. The authors in [16] proposes a cluster beamforming strategy to jointly optimize beamforming vectors and power allocation coefficients for mobile users in MIMO-NOMA clustering. They focus on reducing the total power consumption rather than improving the network capacity and propose an improved coalition game approach to effectively optimize user clustering in MIMO.

4) DIFFERENCES FROM PREVIOUS APPROACHES

Ding *et al.* [17] focused on the problem of minimizing the total power consumption by jointly optimizing the beamforming vector and the power allocation coefficient. They proposed a two-user clustering algorithm in which only two users share one beamforming vector while considering the peer effect. Sayed-Ahmed *et al.* [18] solved the sum rate maximization problem with SIC constraint while assuming two-user clustering. They found the optimal beamforming vector, however two-user clustering cannot maximize the benefits of clustering. They solved the problem while

focusing on beamforming rather than efficient clustering. Our contribution differs from [17], [18] because we proposed a clustering algorithm for multiple users, and identified several factors that effectively perform user clustering. In addition, our algorithm considers the case where the gains due to beamforming compensates for the loss of NOMA gain. The authors in [19] divides users into two groups; cluster user and singleton user (not belong to a cluster). Once classified as a singleton user, it cannot be classified as a cluster user again. When the algorithm reconfigures the clusters, there may be singleton users that can be reclassified as cluster users, but this is not considered. Thus, the flexibility of clustering is significantly lower than that of our work.

III. MOTIVATION

In NOMA, carefully designed user clustering and beamforming can improve the network capacity. We now motivate the necessity of coupling user clustering and beamforming to maximize the spatial reuse gains in NOMA.

A. NEEDS FOR JOINT USER CLUSTERING AND BEAMFORMING

User clustering in NOMA is employed to improve the spectral reuse gain between multiple clients in the cluster so as to effectively maximize the network throughput. Conventional clustering method pairs two clients with the largest channel difference in the cluster and adapts signal interference cancellation while allocating the same spectrum resources to two users in the cluster. In naive scheme, two types of users are identified according to their geographical locations; the set of CC and CE users. CC users are relatively closed to the BS and hence their received signal strength (RSSI) is much higher than that of CE users. Then, one CC user and one CE user are paired and they form the cluster (each cluster has two users with the largest channel difference). The reason for pairing CC and CE users in the cluster is to maximize the spectrum (channel) reuse. One drawback of the above-mentioned approach (called CC-CE) is that CC-CE method only considers clients' locations and does not take into account the beamforming factors.

Ali *et al.* [1] proposed MIMO-NOMA when clustering users who are equipped with multiple antenna arrays. Similar to CC-CE approach, they configure a cluster with one CC and one CE user, however they consider the channel coefficient between CC and CE users for beamforming. Specifically, the pair of CC and CE user that has the highest channel coefficient form a cluster to maximize the benefits of NOMA (in order to easily apply SIC in NOMA). Neither CC-CE nor MIMO-NOMA leverages the benefits of beamforming.

In the context of beamforming in NOMA, user clustering has a bigger role to play, and hence different user clustering could result in different network performance in a given topology. Unlike OMA systems, where distinct spectrum resources have to be assigned to each user, NOMA can improve the spectrum reuse if the user clustering is effectively determined. This requires that the beamforming caters

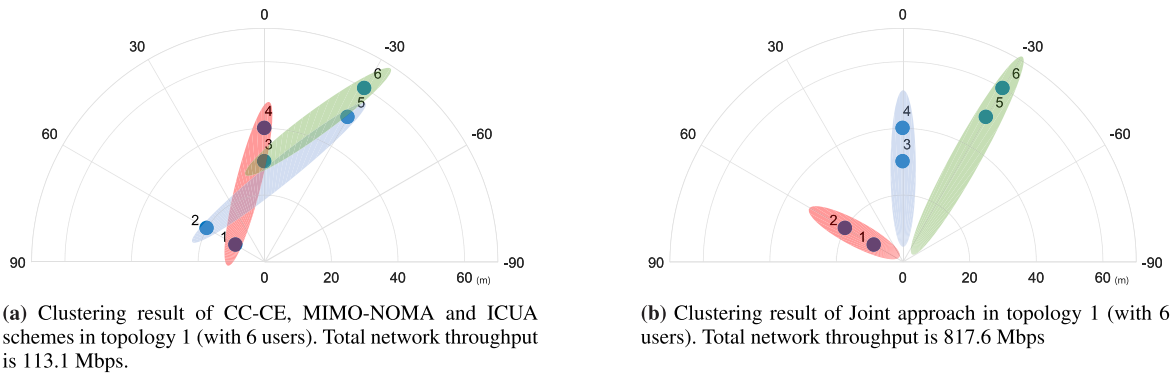
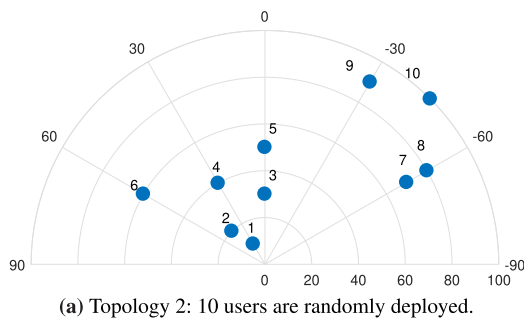


FIGURE 1. Joint approach has different clustering result from CC-CE (MIMO-NOMA and ICUA), and outperforms the network throughput compared to CC-CE.



Scheme	Clustering					Throughput
CC-CE	1, 6	2, 7	3, 8	4, 9	5, 10	181.1 Mbps
MIMO-NOMA	1, 6	2, 8	3, 10	4, 7	5, 9	142.7 Mbps
IUCA	1, 8	2, 7	3, 6	4, 9	5, 10	116.5 Mbps
Joint	3, 10	2, 4	7, 8	6, 9	5, 1	856.9 Mbps

(b) Results of user clustering and corresponding network throughput for each scheme.

FIGURE 2. Joint scheme provides 5.9 times higher throughput than MIMO-NOMA in Topology 2.

effectively to all its users in the network. Since the beamforming impacts the interference and hence performance seen by other users in the network, this naturally results in user clustering being closely coupled with beamforming. This motivate us to carefully design an algorithm for user clustering and beamforming in NOMA.

B. IMPACT OF USER CLUSTERING ON THROUGHPUT

To identify the impact of both user clustering and beamforming in NOMA, we conducted case studies on user clustering. We created two topologies as case studies that have randomly distributed users as depicted in Figures 1a and 2a, where theta and radius are orientation (unit of degree) and distance (unit of meter), respectively. The origin represents the BS and the blue dots are the users. Note that topology1 has 6 users (Figure 1a) and topology 2 has 10 users (Figure 2a).

With given two topologies, we simulated the user clustering results applying the CC-CE method, MIMO-NOMA [1] and ICUA [7] schemes (simulation details are described in Section V-A). Both CC-CE and MIMO-NOMA schemes are used as baseline strategies for performance comparison. The clustering results of CC-CE, MIMO-MONA and ICUA schemes are same as shown in Figures 1a. In case of topology 1, CC-CE scheme forms three clusters based on the locations of the clients (e.g., CC and CE). Thus, we can see that user 1 (CC user) and 4 (CE user) are paired (similarly, user 2 and 5 are paired and user 3 and 6 are paired).

Mainly, one CC user (1, 2, and 3) and one CE user (4, 5, and 6) that have the largest channel difference forms a cluster. Recall that MIMO-NOMA is based on the CC-CE scheme for user clustering and MIMO-NOMA considers the channel difference for NOMA interference and the channel coefficient for beamforming distinctively. In this reason, user clustering outcomes are same as that of CC-CE scheme for topology 1. For topology 2, CC-CE and MIMO-NOMA determine the similar user clustering as shown in Figure 2b. Again, we can confirm that user clustering is mainly determined with one CC user (user index 1 to 5) and one CE user (user index 6 to 10) for both schemes.

Unlike above-mentioned two clustering methods, carefully designed clustering algorithm which takes into account the beamforming could reduce the interference, and hence improve the signal strength and eventually spectrum reuse. In a nutshell, we employ joint approach of user clustering and beamforming, therefore user clustering is different from the two schemes as shown in Figures 1b and 2b (details of joint approach are described later in Section IV-C2). In topology 1, Joint scheme configures the clusters focusing on impact of beamforming rather than channel difference (Figure 1b). As we can see, cluster consists of two users with a relatively smaller channel difference (compared to CC-CE method) which significantly reduces beamforming interference. In other words, Joint scheme configures cluster at the expense of NOMA benefit because the benefits of beamforming are larger than those of NOMA.

We also measure the network throughput (aggregating throughput of each user) and summarize the results in Figure 2b. CC-CE, MIMO-NOMA and ICUA have similar network throughput since their clusters are similar, whereas Joint scheme significantly improve the network throughput ($5.9\times$) compared to MIMO-NOMA. Joint scheme considers both the user clustering and beamforming, therefore it can improve the network capacity. We have observed similar results with respect to the topology 1 (CC-CE, MIMO-NOMA and ICUA yield 113.1 Mbps, whereas Joint approach has 817.6Mbps).

In summary, carefully designed user clustering and beamforming is needed in order to maximize the aggregate network throughput.

C. QUANTIFY THE GAINS OF JOINT USER CLUSTERING AND BEAMFORMING

We have observed the throughput gains of joint user clustering and beamforming in NOMA. Now, we analyze and quantify the gains of joint approach in detail.

In beamforming NOMA, there are multiple factors we need to consider for determining the user clusters. The client's location (i.e., chordal distance¹), channel difference and channel coefficient are the most considered factors in general. Thus, we have conducted hundreds of experiments in various topologies to quantify the impact of these factors on network throughput. For each experiment, we measure the aggregated throughput and the corresponding values of three factors; chordal distance, channel difference and channel coefficient. Table 1 summarizes the correlation coefficient between the network throughput and each factor; chordal distance, channel difference and channel coefficient, respectively. We can clearly see that three factors have almost no correlation with network capacity, therefore considering a single factor when forming clusters in NOMA does not result in improving network throughput. Thus, we should consider multi-factors jointly for user clustering in beamforming NOMA.

TABLE 1. There is no correlations between the network throughput and the conventional three factors.

Factors	Correlation coefficient
Chordal distance	0.001
Channel difference	0.331
Channel coefficient	-0.226

Besides the above-mentioned three factors, there are multiple factors we need to consider in beamforming NOMA. We first define three factors that directly affect network capacity; beamforming interference, NOMA interference and signal power as follows (we define the network throughput as well).

- Beamforming interference I_B is defined as the sum of all interference that affect the users in cluster n , which can be considered as inter-cluster interference (definitions of

TABLE 2. Definition of the variable.

Variable	Description
\mathcal{N}	Set of clusters
\mathcal{K}	Set of users
h_k	Channel for user k
w_n	Beamforming vector for cluster n
p_{nk}	Allocated power for user k in cluster n
B	Bandwidth
M	Number of antenna arrays

the other variables are described in Table 2).

$$I_B = \sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{N} \setminus \{n\}} |h_k w_i^T|^2 \sqrt{p_{ik}} \quad (1)$$

- NOMA interference I_N is the sum of interference that introduced by NOMA (reusing the same spectrum to the other user in the same cluster). We can call it as intra-cluster interference.

$$I_N = \sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} \sum_{i=1}^{k-1} |h_k w_n^T|^2 \sqrt{p_{ni}} \quad (2)$$

- The signal power S is the sum of the RSSI for the user.

$$S = \sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} |h_k w_n^T|^2 \sqrt{p_{nk}} \quad (3)$$

- The network throughput T_{net} is the sum of throughput received by all users in the network, where I_{nk} is the power of interference plus noise for user k in cluster n .

$$T_{net} = \sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} B \log_2 \left(1 + \frac{|h_k w_n^T|^2 \sqrt{p_{nk}}}{I_{nk}} \right) \quad (4)$$

Then we analyze the experiment results in previous subsection III-B (topology 1 and 2) to quantify the impact of above-defined three factors. Table 3 summarizes the impact of each factor and its network throughput. In topology 1, CC-CE, MIMO-NOMA and ICUA configure the cluster with one CC user and one CE user. Their algorithm are simple and fast, but various clusterings are not considered, and hence the beamforming gain is not maximized. In contrast, Joint scheme significantly reduces beamforming interference when forming clusters in a different manner. It has similar NOMA interference and slightly higher signal power compared to other two schemes, therefore it can provide $7.2\times$ higher network throughput. Beamforming interference has much larger impact than NOMA interference and signal power in topology 1. Joint scheme is able to leverage the effect of minimizing beamforming interference.

In densely deployed network (topology 2), ICUA significantly reduce NOMA interference compared to Joint scheme because it considers channel difference for user clustering. However, ICUA has high beamforming interference and low signal power. CC-CE, MIMO-NOMA and ICUA have one CC and one CE user in a cluster, whereas Joint scheme forms

¹The distance between two points located on a curve.

TABLE 3. In topology 1, beamforming interference has significant impact on network throughput compared to NOMA interference. Higher signal power and reduced beamforming interference for Joint scheme in topology 2 increase network capacity.

Scheme	Topology 1				Topology 2			
	Beamforming Interference	NOMA Interference	Signal Power	Network Throughput	Beamforming Interference	NOMA Interference	Signal Power	Network Throughput
CC-CE	-49.88 dBm	-53.75 dBm	-48.98 dBm	113.1 Mbps	-58.85 dBm	-71.87 dBm	-67.10 dBm	181.1 Mbps
MIMO-NOMA	-49.88 dBm	-53.75 dBm	-48.98 dBm	113.1 Mbps	-61.03 dBm	-64.83 dBm	-60.06 dBm	142.7 Mbps
IUCA	-49.88 dBm	-53.75 dBm	-48.98 dBm	113.1 Mbps	-59.87 dBm	-95.74 dBm	-76.27 dBm	116.5 Mbps
Joint	-121.37 dBm	-54.17 dBm	-45.97 dBm	817.6 Mbps	-62.13 dBm	-59.79 dBm	-54.94 dBm	846.9 Mbps

a cluster with the same types (CC or CE) of users. In this sense, Joint scheme has higher NOMA interference compared to others, however it has higher signal power and smaller beamforming interference. Although NOMA interference is high, enhanced signal power and reduced beamforming interference lead to higher network capacity (4.7×). This in turn motivates the need to jointly address user clustering with beamforming, whereby user clustering can be effectively used to maximize the spatial reuse potential of beamforming in NOMA.

IV. JOINT USER CLUSTERING AND BEAMFORMING (JUCAB)

In this section, we formulate the problem and present the joint algorithm for both user clustering and beamforming in NOMA.

A. PROBLEM FORMULATION

We assume a downlink cellular network where K users with a single antenna are associated with a base station (BS) equipped with M array-antenna for beamforming. We consider N number of clusters in the network. \mathcal{N} and \mathcal{K} are the set of clusters and the set of users, respectively. The transmitted signal intended to users in cluster n is defined as $\mathbf{x}_n = \sum_{k=1}^K p_{nk} \cdot s_k$, where p_{nk} is the allocated power for user k in cluster n , and s_k is the desired signal for user k . We denote $\mathbf{w}_n \in \mathbb{C}^{M \times 1}$ as the beamforming vector for cluster n and describe beamforming vector composed at the BS as $\mathbf{w} = [w_1^T w_2^T \dots w_N^T]^T$ (herein $[\cdot]^T$ indicates the transpose of matrix). Then, the transmitted signal $\mathbf{x} \in \mathbb{C}^{M \times N}$ at the BS can be expressed as $\mathbf{x} = \mathbf{w}[x_1 x_2 \dots x_N]^T$.

The received signal \mathbf{y}_k for user k can be described as $\mathbf{y}_k = h_k \mathbf{x} + z_k$, where z_k is the noise for user k . The channel between the user k and the BS is expressed as $h_k \in \mathbb{C}^{1 \times M}$ and $|h_i|^2 > |h_j|^2$, if $i > j$. In NOMA, the user can decode and subtract the signals to which the more power is allocated using SIC. We denote I_{nk} as the interference plus noise power for user k in cluster n , and it can be expressed as

$$I_{nk} = \sum_{i=1}^{k-1} |h_k w_n^T|^2 \sqrt{p_{ni}} + \sum_{i \in \mathcal{N} \setminus \{n\}} |h_k w_i^T|^2 \sqrt{p_{ik}} + z_k \quad (5)$$

where p_{nk} is the allocated power of user k in cluster n . The first term in Eq. (5) is sum of the intra-cell interference from users in the same cluster. Similarly, the second term in Eq. (5) is sum of the inter-cell interference from users that belong to

the neighboring clusters. The achievable data rate r_{nk} for user k in cluster n can be represented as below.

$$r_{nk} = t_{nk} \cdot B \log_2 \left(1 + \frac{|h_k w_n^T|^2 \sqrt{p_{nk}}}{I_{nk}} \right), \quad (6)$$

where t_{nk} is a binary variable that indicates the user clustering, $t_{nk} = 1$ if user k is associated with cluster n , otherwise 0. Data rate for user k is $r_k = \sum_{n \in \mathcal{N}} r_{nk}$. Based on the system model, utility U is defined as follows.

$$U = \sum_{k \in \mathcal{K}} v_k \log r_k \quad (7)$$

The utility is defined based on data rate and proportional fairness. To fairly clustering users, we set the users' weights and use the log function to prevent a small number of users with high data rates from having a large impact on the utility. We can formulate the problem as utility maximization while satisfying per user demand and power constraint.

$$\begin{aligned} & \max_{\mathbf{w}, \mathbf{p}, \mathbf{t}} U \\ & \text{s.t. } r_k \geq D_k, \forall k \in \mathcal{K}, \\ & \sum_{k=1}^K \|w_k\|_2^2 \leq P_t, \\ & \sum_{n=1}^N t_{nk} = 1, \quad \forall k \in \mathcal{K} \end{aligned} \quad (8)$$

where D_k and P_t are the demand (data rate) for user k and the total power budget, respectively. Constraint Eq.(8) indicates that client belongs to only one cluster. In addition, t_{nk} and p_{nk} satisfies $t_{nk} = 1 - \delta_{p_{nk}, 0}$, where δ is the Kronecker delta. We define the values p_n, \mathbf{p}, t_n , and \mathbf{t} as $p_n = [p_{n1} p_{n2} \dots p_{nK}]$, $\mathbf{p} = [p_1^T p_2^T \dots p_N^T]^T$, $t_n = [t_{n1} t_{n2} \dots t_{nK}]$ and $\mathbf{t} = [t_1^T t_2^T \dots t_N^T]^T$, respectively. We consider proportional fairness for the sum rate $v_k \log r_k$, where $v_k \propto \frac{1}{r_k}$ to strike a good balance between network throughput performance and fairness. In our problem, we assume system utility to be non-convex and non-decreasing function with respect to the beamforming. The goal is to jointly determine user clustering (\mathbf{t}) and find optimal beamforming (\mathbf{w} and \mathbf{p}).

$$(\mathbf{w}^*, \mathbf{p}^*, \mathbf{t}^*) = \arg \max_{\mathbf{w}, \mathbf{p}, \mathbf{t}} \sum_{k \in \mathcal{K}} v_k \log r_k \quad (9)$$

B. HARDNESS

Theorem 1: For a given beamforming, the JUCAB problem remains to be NP-hard.

Proof: We can find that our problem is the similar problem as joint multicell client association and beamforming in OFDMA [20] where multicell in OFDMA corresponds to the clusters in our problem. Suppose that the constraint of target rate is $r_k > 0, \forall k \in \mathcal{K}$, the beam interference of all users are same, there are two clusters A and B, and the total power constraint is close to infinite. We define \mathcal{K}_A and \mathcal{K}_B as the set of users in cluster A and B, respectively. The utility of cluster U_A and the user rate of cluster r_A can be defined as follows.

$$\begin{aligned} U_A &= \sum_{k \in \mathcal{K}_A} v_k \log r_k \\ r_A &= \frac{1}{\sum_{k \in \mathcal{K}_A} v_k} \end{aligned} \quad (10)$$

Then we get the utility of cluster A.

$$\sum_{k \in \mathcal{K}_A} v_k \log \frac{1}{\sum_{k \in \mathcal{K}_A} v_k} \quad (11)$$

Similarly, we have the utility of cluster B.

$$\sum_{k \in \mathcal{K}_B} v_k \log \frac{1}{\sum_{k \in \mathcal{K}_B} v_k} \quad (12)$$

After applying normalized weights, the sum of weights become 1, and we get the following equation.

$$\sum_{k \in \mathcal{K}_A} v_k + \sum_{k \in \mathcal{K}_B} v_k = 1 \quad (13)$$

By substituting the X for $\sum_{k \in \mathcal{K}_A} r_k$, the utility of cluster A and cluster B can be rewritten as $X \log(\frac{1}{X})$ and $(1 - X) \log(\frac{1}{1-X})$. The problem can be appeared as follows.

$$\max_X X \log \left(\frac{1}{X} \right) + (1 - X) \log \left(\frac{1}{1 - X} \right) \quad (14)$$

This problem is equal to the NP-hard problem in [20], which proves above equation as NP-complete using subset sum problem. ■

C. ALGORITHM

Several beamforming algorithms in NOMA determine user clustering based on the channel gain (channel difference) between users without considering beamforming effect. Although algorithms consider both the user clustering and beamforming, they are not jointly solved. As a result, the benefits of beamforming and NOMA have not been fully achieved. Our solution differs from the previous methods in the sense that our solution couples beamforming and user clustering in NOMA. Given that our problem is NP-hard, we propose the following heuristic algorithm.

1) CONSIDERED FACTORS FOR USER SWAP

Since both components of our problem (user clustering and beamforming vector) are hard, we must clarify the interaction between them to propose a solution. Unlike the beamforming problem, the user clustering problem, although hard, can be solved more efficiently. Therefore, JUCAB employs a simple but efficient user clustering algorithm as the core building block for solving the joint problem. At a high level, JUCAB solves the user clustering problem for a given beamforming configuration and uses the utility to manipulate the beamforming while reconfiguring clusters in an iterative manner until an efficient solution is attained. Specifically, JUCAB first initializes the user clustering, then swaps users in a way that utility increases.

To determine which cluster and which user need to be swapped, we observed various factors that significantly increased the network throughput. We found three factors for user swap; beamforming interference, signal power² and utility. To identify the impact of these factors, we first conducted experiments using each factor one by one, and quantified the effects. We denote that Case 1 is where beamforming interference is only considered for user swap, Case 2 is signal power and Case 3 is utility. In particular, given user topology and clusters, the user with the highest beamforming interference within the cluster is selected for user swap. The user with high beamforming interference means that user clustering is not correctly configured, therefore higher beamforming gain can be achieved if such high-interference user moves to another cluster. Similarly, users with low signal power and low utility are subject to user swap.

We conducted experiments consisting of 20 users with 5 clusters (4 users per each cluster). We repeated experiment 1000 times and varied the channel for each experiment in order to have diversity, however kept the other values remain same. We also varied the user's location. In each experiment, we measured the change (difference) in beamforming interference, signal power and utility of the swapped users. Figure 3a shows the CDF gain of beamforming interference when users with the highest beamforming interference are swapped. We can see that beamforming interference is decreased in 94% cases (out of 1000 experiments) when reconfiguring the clusters by swapping users with high beamforming interference. Similarly, we can confirm that gains increase in most case when switching users with low signal power as shown in Figure 3b (median gain is 21.5 dBm and the highest is 38.4 dBm). Figure 3c represents the CDF gain of network utility. We can see that there are utility benefits in 70% of cases when relocating low utility users. It is clear that swapping users based on each criterion improves its own factor significantly (e.g., beamforming interference, signal power and utility).

²The received signal power S_{nk} for user k in cluster n can be obtained by using the channel h_k and beamforming vector w_n for cluster n . It can be written as $S_{nk} = |h_k w_n|^2$.

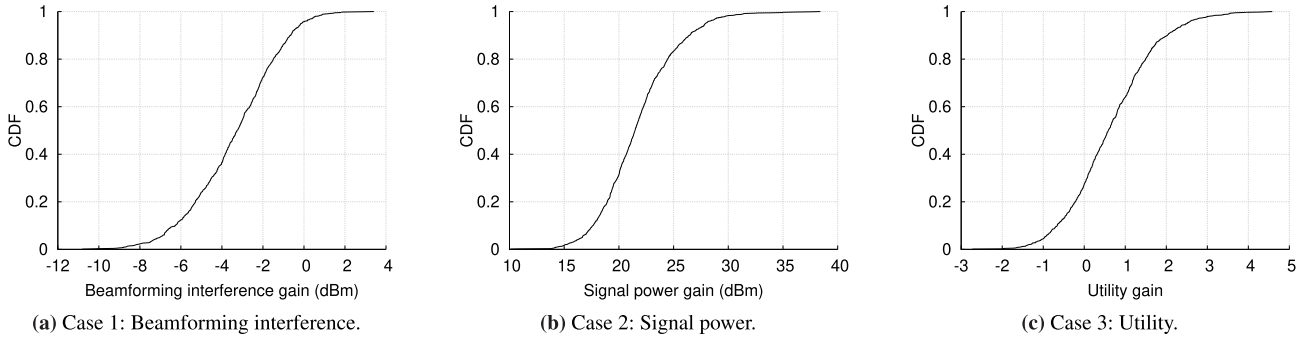


FIGURE 3. The CDF gain of each factor is presented. (a) shows beamforming interference is reduced in 94% of cases. (b) Signal power is improved in most cases. (c) There are utility benefits in 70% of cases when relocating low utility users.

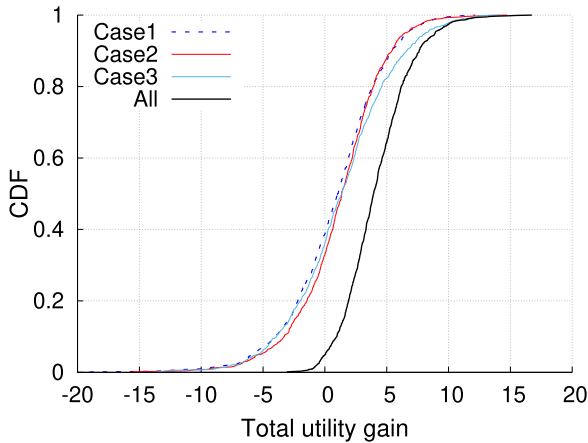


FIGURE 4. CDF of utility gain. The utility gains are significantly high when considering all three factors for user swap. For instance, we have observed total utility gain in 95% of cases.

Next, to quantify the impact of three criteria on network utility, we present the CDF gain of network utility for Case 1, 2 and 3 in Figure 4. In all three cases, it can be seen an increase of 62% of the network utility (note that CDF of total utility gain in Case 1, 2 and 3 are nearly similar). Even if the utility of the swapped user increases, the network utility may decrease in some cases. This is because only a single factor that affects the network utility is considered. Thus, we consider all three factors and choose the users with high interference, low signal power and low utility in the cluster for swap. The network utility gain is presented in Figure 4 (labeled All) which clearly outperforms the result of Case 1, 2 and 3 where a single factor is considered. For example, we can see that network utility is improved in 95% of cases. Based on these results, it can be inferred that selecting two users with large gap will yield higher network utility gain.

It is important to identify which user’s swap leads to high utility gain in the network. As described earlier, we repeated 1000 experiments consisting of 5 clusters (4 users in each cluster). In the experiment, users in each cluster were ranked from 1 to 4 in the order of low beamforming interference, low signal power and low utility (e.g., rank 1 indicates the lowest beamforming interference and rank 4 indicates the highest beamforming interference). We counted the number

of users in each category (from 1 to 4) that were swapped in order to increase the network utility. The results are present in Table 4. For example, 70% users (in categories 3 and 4) were swapped due to high beamforming interference. Similarly, 68% users (in categories 1 and 2) and 71% users were relocated due to low signal power and low utility, respectively. In other words, we expect higher gain if the users experience low beamforming interference, high signal power and utility. Above-mentioned three criteria can be used to determine which user to relocate.

TABLE 4. To increase the network utility, we need to relocate the users with high beamforming interference, low signal power and low utility.

Factor	Order	1	2	3	4
	Beamforming interference		22	8	14
Signal power		23	45	31	1
Utility		35	36	24	5

In summary, we can infer that the following criteria can be used to determine which users and clusters to swap; 1) users with higher beamforming interference, 2) users with lower signal power and 3) users with lower utility than other users in the same cluster.

2) JUCAB ALGORITHM

Before designing a joint algorithm, we first define several functions used in the algorithm. The main goal of these functions is to find the best candidate for user swap. As mentioned, swapping the users with higher beamforming interference and lower signal power and utility in the cluster leads to increase of system utility. Based on these, we defined the following metrics and function.

- $f_{I_{nk}} = (I_{nk} - I_N)/|I_N|$ is a metric that calculates the rate of beamforming interference for user k in cluster n (assuming user k move into the cluster n), where $I_N = \sum_{k \in \mathcal{N}} I_k, k \in n, \forall n \in \mathcal{N}$.
- $f_{T_{nk}} = (T_{nk} - T_N)/|T_N|$ is a metric that calculates the rate of throughput for user k in cluster n , where $T_N = \sum_{k \in \mathcal{N}} T_k, k \in n, \forall n \in \mathcal{N}$.
- $f_{S_{nk}} = (S_{nk} - S_N)/|S_N|$ is a metric that calculates the rate of signal power for user k in cluster n , where $S_N = \sum_{k \in \mathcal{N}} S_k, k \in n, \forall n \in \mathcal{N}$.

- $F_{nk} = f_{I_{nk}} / (f_{T_{nk}} \cdot f_{S_{nk}})$ is the function that determines the best candidate user for swapping. The higher F_{nk} , the higher expected utility gain when swapping user k in cluster n .

Algorithm 1 JUCAB Algorithm

```

1: INPUT: channel matrix  $\mathbf{H} = \{h_1, h_2, \dots, h_k\}, \forall k \in \mathcal{K}$ 
2: OUTPUT: user clustering  $\mathbf{C} = \{C_1, C_2, \dots, C_n\}$ , beamforming vector  $\mathbf{w} = \{w_1, w_2, \dots, w_n\}$ , where  $\forall n \in \mathcal{N}$  and transmission power  $\mathbf{p}$ 
3: Initialize  $\mathbf{C}$  based on channel gain
4: Given  $\mathbf{C}$ , determine  $\mathbf{w}$  and  $\mathbf{p}$ 
5:  $U_c = \emptyset$ 
6: while  $i \leq M_I$  do
7:    $w_n \leftarrow$  beamforming vector,  $\forall n \in \mathcal{N}$ 
8:   Compute  $f_{I_{nk}}, f_{T_{nk}}, f_{S_{nk}}$  and  $F_{nk}$ , where  $\forall k \in \mathcal{K}, \forall n \in \mathcal{N}$ .
9:    $U_c \leftarrow U_c \cup \{k\}$  if user  $k$  has max.  $F_{nk}$  in cluster  $n$ 
10:   $(n_1, k_1, n_2, k_2) \leftarrow \max_{k_1 \in n_1, k_2 \in n_2, k_1, k_2 \in U_c} \exp(f_{S_{n_1 k_2}}) + \exp(f_{S_{n_2 k_1}})$ 
11:   $C_{swap} \leftarrow \text{swap}(k_1, k_2)$ 
12:   $u^* = \arg \max \{ \sum_{k \in C_{swap}} v_k \log r_k - \sum_{k \in C} v_k \log r_k \}^+$ 
13:  if  $u^* = \emptyset$  then
14:    Break
15:  end if
16:   $(\mathbf{w}^*, \mathbf{p}^*) = \arg \max_{\mathbf{w}, \mathbf{p}} U$ 
17:   $i \leftarrow i + 1$ 
18: end while

```

Algorithm 1 describes the clustering algorithm that jointly considers the multiple factors which effect system utility. At a high level, JUCAB solves the user clustering for a given beamforming configuration and uses system utility (i.e., network throughput) to adjust the beamforming until an efficient solution is obtained. The input is the channel matrix for all users in the network (line 1). JUCAB determines user clustering \mathbf{C} , beamforming vector \mathbf{w} for each cluster and the transmission power \mathbf{p} as output (line 2).

First, JUCAB initializes \mathbf{C} , set of cluster, based only on channel gains (line 3). The initial set of cluster is not critical because JUCAB algorithm constantly changes the clustering toward maximizing the system utility. Once user clustering is determined, the algorithm finds the best beamforming and its transmission power (line 4). The value i is iteration index, and M_I is the maximum number of iterations (line 6).

User clustering \mathbf{C} is reconfigured iteratively by swapping two users in different clusters in the direction of increasing utility (line 6 ~ 17). The beamforming vector for each cluster w_n is determined at the beginning of the iteration (line 7). The value I_{nk} , T_{nk} , and S_{nk} represent the beamforming interference, throughput and the signal power of user k in cluster n , respectively. Once these three values are obtained, metrics $f_{I_{nk}}, f_{T_{nk}}$ and $f_{S_{nk}}$ and function F_{nk} are calculated (line 8). These values are the rate for the users in the cluster to which the swapped users belong in order to measure the impact of

cluster reconfiguration. The algorithm finds the set of users U_c that user k has the maximum F_{nk} in each cluster n (line 9). Users in U_c experience the highest beamforming interference in each cluster, therefore they have the lowest power and utility. Higher utility gains are expected when swapping users in U_c . Specifically, we could mitigate beamforming interference and increase signal power by replacing users and adjusting beamforming accordingly. We find a pair of users (e.g., k_1 and k_2) in U_c that maximize the addition of the exponential of $f_{S_{n_1 k_2}}$ and $f_{S_{n_2 k_1}}$ (line 10). C_{swap} is the reconfigured cluster sets after swapping the users k_1 and k_2 from the clusters n_1 and n_2 (line 11). u^* is arg max of positive incremental utility, it returns only if the incremental utility is greater than 0, otherwise it terminates the iteration because there is no more utility gain obtained from cluster reconfiguration (line 12 ~ 14). Optimal beamforming vector and transmission power can be obtained in a way that maximizes the system utility U (line 15).

Specifically, once user clustering is configured, beamforming vector for cluster n can be determined by taking SVD of the channel matrix channel $\mathbf{H}_n = \{h_{n,1}, h_{n,2}, \dots, h_{n,k}\}$, $\forall k \in \mathcal{K}$, $\mathbf{H}_n = U_n \Sigma_n V_n^H$ [1]. Then the equivalent channel matrix $\tilde{\mathbf{H}}_n = U_n^{(1)H} \mathbf{H}_n$, where $U_n^{(1)H}$ is the Hermitian transpose of first column of U_n . The, beamforming vector w_n can be determined as follows;

$$w = \frac{\tilde{\mathbf{H}}^\dagger}{\|\tilde{\mathbf{H}}^\dagger\|_F} \quad (15)$$

where \dagger indicates right pseudo-inverse, and $\|\cdot\|_F$ is Frobenius norm. After determining the clustering and beamforming vectors, power allocation must be calculated in order to maximize the NOMA gain while mitigating the NOMA interference. The problem of maximizing sum-rate is proposed below.

$$\max_P \sum_{k \in K_n} r_{nk}$$

$$\text{s.t. } \sum_{k \in K_n} \sqrt{P_{nk}} \leq P_{nt}, \quad \forall k \in K_n, \quad (16)$$

$$r_{nk} \geq D_k, \quad \forall k \in K_n, \quad (17)$$

$$|h_{k-1} w_n^T|^2 \sqrt{P_{nk}} - \sum_{j=1}^{k-1} |h_{k-1} w_n^T|^2 \sqrt{P_{nj}} \geq P_{tol},$$

$$\forall k \in K_n \setminus \{1\}, \quad (18)$$

where K_n is the set of users in cluster n . P_{nt} is the maximum power that cluster n has and P_{tol} is the minimum power difference for NOMA. Eq. (16) represents the total power constraint, Eq. (17) is the minimum data rate constraint and Eq. (18) is the minimum power difference constraint, respectively. This problem is similar to the downlink optimal power allocation problem of [4]. However, our problem differs from that of [4] in the sense that we have to consider beamforming interference. The authors in [4] assume that the beamforming interference does not exist in their problem because it has already been determined. In our problem,

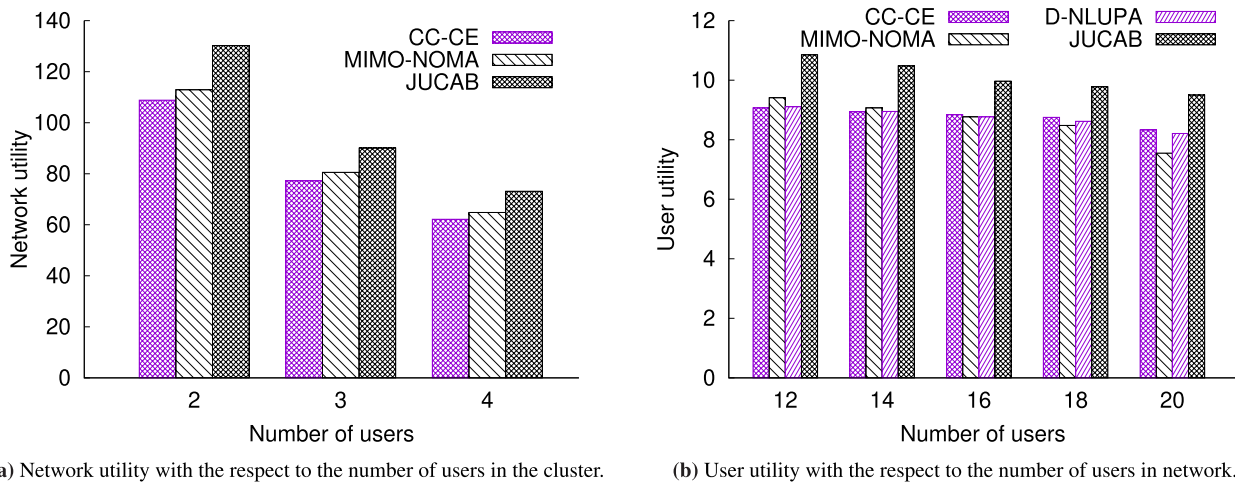


FIGURE 5. The evaluation results show the utility of CC-CE, MIMO-NOMA, D-NLUPA and JUCAB. We can clearly see that JUCAB outperforms other schemes.

beamforming interference in each cluster is considered noise, because beamforming is already fixed. To determine power allocation, a closed form solution is applied.

D. COMPLEXITY ANALYSIS

The most complicated parts of the JUCAB algorithm are maximum function (line 10) and beamforming vector function (line 15). In maximum function, the number of possible cases for selecting clusters is $\frac{n \cdot (n-1)}{2}$ and the number of possible cases for selecting two users in the clusters K^2 . Thus, the computational complexity becomes $O(N^2 K^2)$. For beamforming vector, the complexity of obtaining $\bar{\mathbf{H}}_n$ is $O(MK)$, where M is the number of antenna arrays. In addition, the complexity of the right pseudo-inverse and norm are $O(N^3)$ and $O(K^{\frac{3}{2}})$, respectively. Therefore, finding beamforming vector requires $O(MN^3 K^{\frac{5}{2}})$. In the case of power allocation, the complexity of assigning power for user k in cluster n , P_{nk} , is $O(K)$. Similarly, P_n and P require $O(K)$ and $O(N)$, respectively. The complexity of allocating transmit power is $O(NK^2)$.

In theoretical analysis, the convergence is not guaranteed, however the number of iterations in the while loop stops the algorithm and the termination condition based on the positive incremental utility breaks out from the iteration. For instance, the JUCAB algorithm converges within less than 20 iterations in 1000 experiments.

V. EVALUATION

A. SETUP

We evaluate the performance of JUCAB in large-scale simulations. The simulations were conducted on a single cell with a cell radius of 300m, and multiple users were randomly deployed in the cell with Rayleigh fading. Table 5 summarizes the parameters used in the simulation. The maximum transmission power at the BS is 41 dBm and the noise density at the receiver is -169 dBm/Hz. Power budget is

TABLE 5. Parameter settings.

Parameter	Value
Cell radius	300m
Bandwidth	6.8 MHz
Power budget	41 dBm
Path-loss exponent	3.8
Noise density	-169 dBm/Hz
Number of antennas	10
Fading	Rayleigh fading

used to measure JUCAB’s performance and compare it with other schemes. For performance comparison, CC-CE and MIMO-NOMA were used for the baseline strategies. We also consider the D-NLUPA [13] scheme for comparison. Note that, D-NLUPA configures only 2-user cluster, therefore the experiment results of D-NLUPA are omitted as the number of users in the cluster increases.

B. SIMULATION RESULTS

1) UTILITY

First, we measure the aggregated network utility as defined in Eq. (7). Figure 5a shows the network utility of three schemes with respect to the number of users in the cluster. In this experiment, we set the number of users in the network to 12, and then increase the number of users in each cluster from 2 to 4.³ It is natural that increasing number of users leads to higher NOMA interference in the cluster, therefore the network utility decreases. It can be seen that JUCAB outperforms others (CC-CE and MIMO-NOMA) regardless the number of users. Moreover, JUCAB provides higher utility when fewer users are in the cluster. This is because the beamforming gain

³In the experiment, we manually set the number of users in each cluster. Having the optimal number of users in a cluster is beyond our scope. Indeed, it is an interesting topic for our future work.

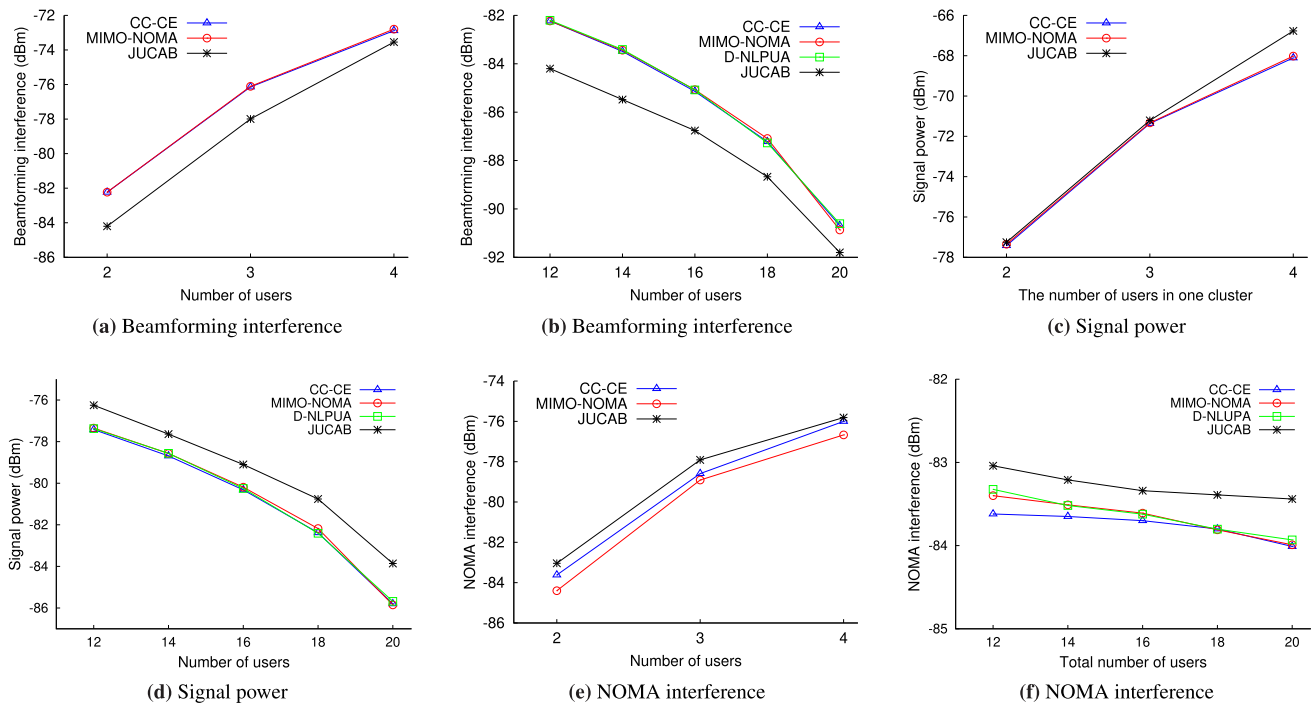


FIGURE 6. Simulation results show the average beamforming interference, signal power and NOMA interference of four schemes. JUCAB's beamforming interference is lower than other three schemes, while it has the highest signal power.

is lower when a large number of users are applied to the same beamforming vector. Specifically, JUCAB's utility is 20% (at top) higher than that of CC-CE when two users are assigned to the cluster. This indicates that JUCAB takes full advantages of beamforming. If many users are assigned to a cluster, clustering requires sophisticated algorithm that handles interference well. Otherwise, beamforming and NOMA benefits will be reduced.

Increasing the number of users in the network causes high NOMA interference, thus reducing the utility for each user. Such high interference hinders beamforming gain, and hence it is required to jointly use beamforming and clustering in order to maximize beamforming gain while minimizing NOMA interference. To quantify the impact of number of clusters in the network, we set the cluster size to 2 (i.e., each cluster has 2 users) then increase the number of users (from 12 to 20 in step of 2). in Figure 5b, we present per-user utility⁴ of four schemes as the number of users in the network increases. We can see that per-user utility of three schemes decreases as the number of users increases, however JUCAB still maintains higher utility than others. The performance of D-NLUPA is similar to that of CC-CE scheme. MIMO-NOMA scheme also takes into account beamforming along with NOMA, however it does not provide a higher utility compared to JUCAB. Even worse, in some cases utility of MIMO-NOMA is lower than CC-CE which does not use beamforming. This shows that MIMO-NOMA

⁴The user utility is the average utility for user k , $U_k = v_k \log r_k$, whereas the network utility is an aggregated utility for all users in the network as defined in Eq. (7).

uses beamforming and NOMA separately, therefore it is not able to achieve the maximum gains of both. NOMA-based method pairs users only considering NOMA gain therefore diverse clustering gain is ignore. This illustrates that our joint approach can handle complex problem well.

In this work, we have a fixed number of users per cluster. Having a variable number of users per cluster is a harder problem in optimizing network utility. We plan to investigate it in future work.

2) INTERFERENCE AND SIGNAL POWER

We have confirmed the utility gains of JUCAB in the previous subsection. Here, we take a microscopic view of the results to identify the root cause of JUCAB's performance improvements. In designing JUCAB algorithm in Section IV-C2, three factors were taken into account; beamforming interference, NOMA interference and signal power. We have measured three factors and present them in Figure 6 to quantify the impact of them.

As the number of users in the cluster increases, beamforming interference increases as shown in Figure 6a. This is because the more users in a cluster, the higher the NOMA interference. Moreover, covering more users with the same beamforming vector reduces the beamforming gains. In Figures 6a and 6b, beamforming interference of JUCAB is the lowest compared to others. This demonstrates that JUCAB handles beamforming interference well even in crowded environments, and hence it leads to utility gains.

In addition, signal power of each scheme is presented in Figures 6c and 6d. Signal power does not differ

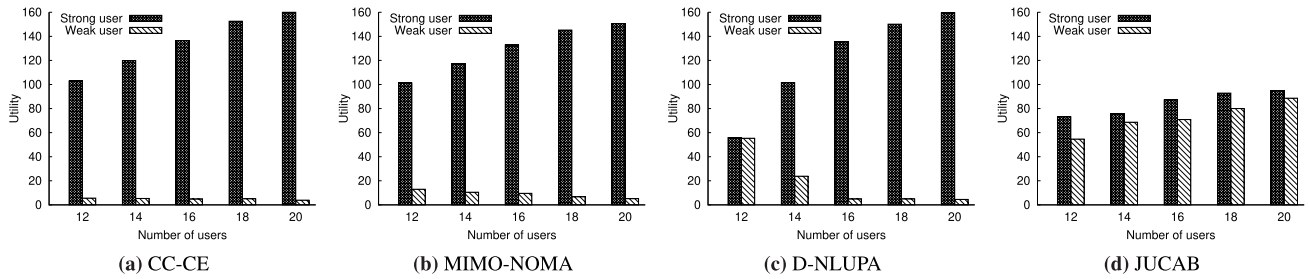


FIGURE 7. Results present the network utility of strong and weak users. We can see that both CC-CE and MIMO-NOMA have significant unfairness among users, whereas JUCAB provides higher fairness. D-NLUPA maintains good fairness up to 12 users in the network.

significantly between the four methods as presented in Figure 6c, unless there are 4 users in the cluster. The main reason for this is that total power budget of each scheme is identical and the clustering diversity does not exist due to the small number of users. Even if the impact of signal power is small, Figure 6d shows that when the network is crowded, JUCAB can handle signal power more effectively than other methods (by 1.5 dBm). Higher signal power is one of the contributing factors to JUCAB’s utility gain.

It is natural that NOMA interference increases as the number of users in the cluster increases. JUCAB’s NOMA interference is slightly higher than other schemes as shown in Figures 6e and 6f. However, the difference in NOMA interference between JUCAB and others is smaller than the difference in beamforming interference. The gains of high signal power and low beamforming interference in JUCAB outweigh the disadvantages of high NOMA interference. Therefore, as the number of users increases, the effect of beamforming on higher signal power and less beamforming interference in the cluster must be taken into account.

Since CC-CE only considers channel difference, it has the lowest NOMA interference as seen in Figure 6f. However, the beamforming effect is not considered, resulting in higher beamforming interference and low utility. MIMO-NOMA takes into account channel coefficient and channel difference simultaneously (but not jointly), therefore it has lower beamforming interference than CC-CE. Unlike others, JUCAB balances beamforming interference and NOMA interference while maintaining higher signal power.

By analyzing the results, we can see that JUCAB considers three factors for both clustering and beamforming, thus it can maximize the utility.

3) FAIRNESS

We have seen the network utility (i.e., throughput) gains of JUCAB, however such gain should not be obtained at the cost of fairness among the users in the network. Fairness between users is important to meet the demands of user, thus algorithm must ensure fairness. To quantify fairness, the utility of strong and weak users is shown in Figure 7, respectively (strong users are those who receive relatively high signal power, usually cell-center user). Strong (i.e., CC) user maintains high utility at all times in CC-CE, MIMO-NOMA and D-NLUPA

schemes. The utility difference between strong and weak user is significantly high in above-mentioned schemes. Specifically, in CC-CE scheme, strong user receives 48.1 times higher utility (throughput) than weak user (when 20 users are in the network as shown in Figure 7a). As the number of users in the network increases, the utility gap (i.e., unfairness) grows. This indicates that there is high level of unfairness among users in both schemes because its clustering algorithms intend to increase the NOMA gain but the fairness. To strike a good balance between network performance (i.e., utility) and fairness, we design the utility function to be non-negative in addition to concave and non-decreasing. In particular, we adopt proportional fairness where the utility function is $U = v_k \log r_k$, where $v_k \propto \frac{1}{r_k}$. This accounts for fairness as can be seen in Figure 7d, where utility gap between strong and weak users is very small.

Both CC-CE and MIMO-NOMA algorithms initially divide into CC and CE users, therefore there is inherent unfairness. In other words, lack of flexibility in configuring cluster leads to unbalanced fairness among users. Although D-NLUPA adopts CC-CE clustering, it focuses on clustering fairness. This characteristic maintains high fairness when the number of users is less than 12, however it fails to provide good fairness as the number of users increases. Since JUCAB jointly considers both features of beamforming and NOMA regardless of user’s location (either CC or CE), it can provide a certain level of fairness among users (e.g., less than 1.2:1 ratio for strong and weak users, respectively).

4) CLUSTERING

For both CC-CE and MIMO-NOMA schemes divides users into CC and CE, then construct the cluster by selecting users from each set to increase the channel gain difference for NOMA. However, as shown in previous Section, JUCAB employ flexible clustering algorithm that aims to maximize channel gain difference while minimizing NOMA interference. Table 6 shows the percentage of clusters without CC and

TABLE 6. JUCAB provides high utility even though its clustering consists of the same types of users (e.g., CC or CE).

Number of users	12	14	16	18	20
% of clusters	79.2	77.7	75.6	73.7	70.4

CE user pairs in JUCAB. Unlike other schemes, more than 70% of clusters in JUCAB have users in the same category (either CC or CE). This result demonstrates that clustering users in vicinity is not harmful, and can enhance performance when jointly using beamforming. In other words, clustering based on CC and CE sets is not an efficient method in beamforming NOMA. Configuring clusters of CC and CE sets to obtain high channel difference and low NOMA interference does not always guarantee high utility. Joint beamforming and flexible clustering are critical to achieve high utility (i.e., throughput).

VI. CONCLUSIONS

NOMA is a promising technology for increasing spectral efficiency, and carefully designed beamforming in NOMA can further improve performance. However using both NOMA and beamforming is a challenging and complex problem. We formulate the utility maximization problem for both NOMA and beamforming that leverages the benefits of them. We propose the joint algorithm taking into account multiple factors that affect system utility. Experiment results show 20% increase in utility compared to other conventional schemes, maintaining fairness among users regardless of the channel conditions and deployment in the system. Our solution can be extended to multicell environments and serve as the basis for addressing the problem of adaptive beamforming.

REFERENCES

- [1] M. 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. 2017.
- [2] B. Kim, S. Lim, H. Kim, S. Suh, J. Kwun, S. Choi, C. Lee, S. Lee, and D. Hong, "Non-orthogonal multiple access in a downlink multiuser beamforming system," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, Nov. 2013, pp. 1278–1283.
- [3] Z. Chen, Z. Ding, and X. Dai, "Beamforming for combating inter-cluster and intra-cluster interference in hybrid NOMA systems," *IEEE Access*, vol. 4, pp. 4452–4463, Aug. 2016.
- [4] M. S. Ali, H. Tabassum, and E. Hossain, "Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (NOMA) systems," *IEEE Access*, vol. 4, pp. 6325–6343, Aug. 2016.
- [5] X. Diao, J. Zheng, Y. Cai, X. Dong, and X. Zhang, "Joint user clustering, resource allocation and power control for NOMA-based mobile edge computing," in *Proc. 10th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2018, pp. 1–6.
- [6] Y. Ma, H. Gao, T. Lv, and Y. Lu, "Novel user clustering schemes for downlink NOMA system," in *Proc. MILCOM-IEEE Mil. Commun. Conf. (MILCOM)*, Oct. 2017, pp. 191–196.
- [7] Z. Liu, L. Lei, N. Zhang, G. Kang, and S. Chatzinotas, "Joint beamforming and power optimization with iterative user clustering for MISO-NOMA systems," *IEEE Access*, vol. 5, pp. 6872–6884, May 2017.
- [8] J. Kim, J. Koh, J. Kang, K. Lee, and J. Kang, "Design of user clustering and precoding for downlink non-orthogonal multiple access (NOMA)," in *Proc. MILCOM-IEEE Mil. Commun. Conf.*, Oct. 2015, pp. 1170–1175.
- [9] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6010–6023, Aug. 2016.
- [10] A. Benjebbour, Y. Saito, Y. Kishiyama, A. Li, A. Harada, and T. Nakamura, "Concept and practical considerations of non-orthogonal multiple access (NOMA) for future radio access," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst.*, Naha, Japan, Nov. 2013, pp. 770–774.
- [11] T. Ohgane, Y. Ogawa, and K. Itoh, "A study on a channel allocation scheme with an adaptive array in SDMA," in *Proc. IEEE 47th Veh. Technol. Conference. Technol. Motion*, vol. 2, May 1997, pp. 725–729.
- [12] J. Choi, "A suboptimal approach for minimum transmit power NOMA beamforming," in *Proc. IEEE 86th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2017, pp. 1–5.
- [13] S. M. R. Islam, M. Zeng, O. A. Dobre, and K.-S. Kwak, "Resource allocation for downlink NOMA systems: Key techniques and open issues," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 40–47, Apr. 2018.
- [14] M. Zeng, A. Yadav, O. A. Dobre, G. I. Tsiropoulos, and H. V. Poor, "Capacity comparison between MIMO-NOMA and MIMO-OMA with multiple users in a cluster," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2413–2424, Oct. 2017.
- [15] M. Zeng, A. Yadav, O. A. Dobre, G. I. Tsiropoulos, and H. V. Poor, "On the sum rate of MIMO-NOMA and MIMO-OMA systems," *IEEE Wireless Commun. Lett.*, vol. 6, no. 4, pp. 534–537, Aug. 2017.
- [16] J. Ding, J. Cai, and C. Yi, "An improved coalition game approach for MIMO-NOMA clustering integrating beamforming and power allocation," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1672–1687, Feb. 2019.
- [17] J. Ding and J. Cai, "Efficient MIMO-NOMA clustering integrating joint beamforming and power allocation," in *Proc. GLOBECOM-IEEE Global Commun. Conf.*, Dec. 2017, pp. 1–6.
- [18] A. Sayed-Ahmed and M. Elsabrouty, "User selection and power allocation for guaranteed SIC detection in downlink beamforming non-orthogonal multiple access," in *Proc. IEEE Wireless Days*, Mar. 2017, pp. 188–193.
- [19] S. Dhakal, P. A. Martin, and P. J. Smith, "NOMA with guaranteed weak user QoS: Design and analysis," *IEEE Access*, vol. 7, pp. 32884–32896, 2019.
- [20] J. Yoon, K. Sundaresan, M. A. A. Khojastepour, S. Rangarajan, and S. Banerjee, "Joint multicell beamforming and client association in OFDMA small-cell networks," *IEEE Trans. Mobile Comput.*, vol. 15, no. 9, pp. 2260–2274, Sep. 2016.



HA-RYUNG KIM received the B.S. degree in computer science and engineering from Hanyang University, South Korea, in 2016, where she is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering. Her research interests include cellular networks, mobile networking, and embedded systems.



JIASHI CHEN (Member, IEEE) received the B.S. degree from Columbia University, New York, NY, USA, with internships at AT&T Labs Research, Florham Park, NJ, USA, and NEC Labs America, Princeton, NJ, and the Ph.D. degree from Princeton University, Princeton. She is currently an Assistant Professor with the Department of Computer Science and Engineering, University of California at Riverside, Riverside, CA, USA. Her current research interests include edge computing,

wireless and mobile systems, and multimedia networking, with a recent focus on machine learning at the network edge to aid augmented reality (AR)/virtual reality (VR) applications. She was a recipient of the Hellman Fellowship and the UCR Regents Faculty Fellowship.



JONGWON YOON (Member, IEEE) received the B.S. degree in computer science from Korea University, in 2007, and the M.S. and Ph.D. degrees from the University of Wisconsin–Madison, in 2012 and 2014, respectively. He is currently an Assistant Professor with the Department of Computer Science and Engineering, Hanyang University, South Korea. He is the Founding Director of the Mobile Systems Laboratory which broadly focuses on research in

wireless communication, mobile systems, and HCI. He was a recipient of the Lawrence Landweber Fellowship during his Ph.D. study.

• • •