Fairness of User Clustering in MIMO Non-orthogonal Multiple Access Systems

Yuanwei Liu, *Student Member, IEEE,* Maged Elkashlan, *Member, IEEE,* Zhiguo Ding, *Senior Member, IEEE,* and George K. Karagiannidis, *Fellow, IEEE*

***Abstract*—In this paper, a downlink multiple-input-multiple-output (MIMO) non-orthogonal multiple access (NOMA) scenario is considered. We investigate a dynamic user clustering problem from a fairness perspective. In order to solve this optimization problem, three sub-optimal algorithms, namely top-down A, top-down B, and bottom up, are proposed to realize different tradeoffs of complexity and throughput of the worst user. In addition, for each given user clustering case, we optimize the power allocation coefficients for the users in each cluster by adopting a bisection search based algorithm. Numerical results show that the proposed algorithms can lower the complexity with an acceptable degradation on throughput compared with the exhaustive search method. It is worth noting that top-down B algorithm can achieve a good tradeoff between complexity and throughput among the three proposed algorithms.**

I. Introduction

Recently, non-orthogonal multiple access (NOMA) has attracted much research interest as a promising candidate for the fifth generation (5G) networks [1]. As an alternative from the existing multiple access (MA) techniques, NOMA exploits a new dimension—power domain to implement MA, which means that a base station (BS) can serve multiple users at the same time/frequency/code resource. In [2], the improvement of spectral efficiency of NOMA was demonstrated, from the perspective of system implementation, by using a two-user case. A more general scenario which considers M randomly deployed users was investigated in [3], by evaluating the ergodic sum rate and outage performance. Considering the fairness issue among single-antenna users, a power allocation optimization problem was investigated under perfect channel state information (CSI) and average CSI, in [4]. Furthermore, from the perspective of energy efficiency and spectrum efficiency, a cooperative simultaneously wireless information and power transfer NOMA protocol was proposed in [5].

All of the aforementioned NOMA works focus on single-antenna scenarios [2–5]. In order to further improve the performance of NOMA by using the spacial domain, several works assumed multiple-antenna techniques. Particularly, in [6], a multiple-input single-output (MISO) NOMA scenario was investigated, with a proposed two-stage beamforming approach to support multiple users. In [7], a multiple-input multiple-output (MIMO) NOMA scenario was considered, where the ergodic system capacity of the considered MIMO-NOMA systems was optimized using new power allocation schemes. In [8], the authors proposed a new design of precoding and detection matrices for a MIMO-NOMA system, and its performance was analyzed and demonstrated to outperform conventional MIMO orthogonal MA.

A key feature of NOMA is the balanced tradeoff between throughput and user fairness. Different from [4] which considered the fairness issue in single-antenna scenarios, a dynamic user allocation and power optimization problem is investigated in this paper, by considering the fairness issue in cluster-based MIMO-NOMA systems. Allocating users into different clusters is a non-deterministic polynomial-time (NP)-hard problem, where exhaustive search yields optimal performance but with prohibitive complexity. The main contributions of this paper are summarized in the following. From the standpoint of fairness, we propose a two-step sub-optimal method for solving the dynamic user allocation problem. In the first step, we optimize the power allocation coefficients by invoking a bisection search based algorithm in order to maximize the signal-to-interference-and-noise-ratio (SINR) of the worst user in each cluster. In the second step, we propose three efficient user allocation algorithms to seek a tradeoff between computational complexity and throughput of the worst user.

II. System Model

A downlink MIMO NOMA scenario is considered, where a BS is equipped with M antennas, while K users are equipped with N antennas each. In order to implement NOMA in the considered MIMO scenario, the K users are further grouped into M clusters, where each cluster includes at least two users. It is assumed that the number of users and clusters are fixed. This assumption is motivated by the fact that these number are predetermined by the load of the networks. The number of users in each cluster is denoted as $\mathbf{L} = \{L_m\}, m = 1, \ldots, M$ with $L_m \geq 2$, and $K = \sum_{m=1}^{M} L_m$. The signals transmitted by the BS are given by

$$\mathbf{x} = \mathbf{P}\tilde{\mathbf{s}}, \tag{1}$$

where $\mathbf{P}$ is a $M \times M$ identity precoding matrix and the $M \times 1$ vector $\tilde{\mathbf{s}}$ is given by

$$\tilde{\mathbf{s}} = \begin{bmatrix} \sqrt{\alpha_{1,1}} s_{1,1} + \cdots + \sqrt{\alpha_{1,L_1}} s_{1,L_1} \\ \vdots \\ \sqrt{\alpha_{M,1}} s_{M,1} + \cdots + \sqrt{\alpha_{M,L_M}} s_{M,L_M} \end{bmatrix} \triangleq \begin{bmatrix} \tilde{\mathbf{s}}_1 \\ \vdots \\ \tilde{\mathbf{s}}_{\mathbf{M}} \end{bmatrix}, \tag{2}$$

where $s_{m,k}$ and $\alpha_{m,k}$ are defined as the transmitted information and the power allocation coefficient of the k-th user ($\{k = 1, \ldots, L_m\}$) in the m-th cluster, respectively.

The associate editor coordinating the review of this paper and approving it for publication was D. Wing Kwan Ng.

Y. Liu and M. Elkashlan are with the School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, UK. (email:{yuanwei.liu, maged.elkashlan}@qmul.ac.uk).

Z. Ding is with the School of Computing and Communications, Lancaster University, LA1 4WA, UK. (e-mail: z.ding@lancaster.ac.uk).

G. K. Karagiannidis is with the Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, 54 124 Thessaloniki, Greece (e-mail: geokarag@auth.gr).

The received signal at the k-th user in the m-th cluster is given by

$$\mathbf{y}_{m,k} = \mathbf{H}_{m,k}\mathbf{P}\tilde{\mathbf{s}} + \mathbf{n}_{m,k}, \tag{3}$$

where $\mathbf{H}_{m,k}$ is the $N \times M$ channel gain matrix from the BS to the k-th user in the m-th cluster, and $\mathbf{n}_{m,k}$ is an additive white Gaussian noise (AWGN) vector.

If we denote $\mathbf{w}_{m,k}$ as the detection vector at the receiver, then the signal model can be expressed as

$$\mathbf{w}_{m,k}^H\mathbf{y}_{m,k} = \mathbf{w}_{m,k}^H\mathbf{H}_{m,k}\mathbf{P}\tilde{\mathbf{s}} + \mathbf{w}_{m,k}^H\mathbf{n}_{m,k}. \tag{4}$$

Denote the i-th column of $\mathbf{P}$ by $\mathbf{p}_i$. The signal model in (4) can be written as

$$\begin{aligned}\mathbf{w}_{m,k}^H\mathbf{y}_{m,k} = &\sum_{i=1,i\neq m}^{M} \mathbf{w}_{m,k}^H\mathbf{H}_{m,k}\mathbf{p}_i\tilde{\mathbf{s}}_i + \mathbf{w}_{m,k}^H\mathbf{n}_{m,k} \\ &+ \mathbf{w}_{m,k}^H\mathbf{H}_{m,k}\mathbf{p}_m\left(\sqrt{\alpha_{m,1}}s_{m,1} + \cdots + \sqrt{\alpha_{m,L_m}}s_{m,L_m}\right).\end{aligned} \tag{5}$$

In NOMA systems, the ordering of channel conditions is important for canceling interference between users in the same cluster using successive interference cancellation (SIC). Without loss of generality, when implementing NOMA, channel gains and power allocation coefficients for the m-th cluster are assumed to be ordered as

$$\left|\mathbf{w}_{m,1}^H\mathbf{H}_{m,1}\right|^2 \geq \cdots \geq \left|\mathbf{w}_{m,L_m}^H\mathbf{H}_{m,L_m}\right|^2. \tag{6}$$

To completely remove the inter-cluster interference, the detection matrices need to satisfy $\mathbf{w}_{i,k}^H\mathbf{H}_{i,k}\mathbf{p}_m = 0$, where $i \in \{1,\ldots,M, i \neq m\}$, as described in [8]. As such, for the k-th user in the i-th cluster, we rewrite the constraints as follows:

$$\mathbf{w}_{\mathrm{i},k}^H\underbrace{\left[\begin{array}{ccccccc}\mathbf{h}_{1,ik} & \cdots & \mathbf{h}_{i-1,ik} & \mathbf{h}_{i+1,ik} & \cdots & \mathbf{h}_{M,ik}\end{array}\right]}_{\tilde{\mathbf{H}}_{i,k}} = \mathbf{0}, \tag{7}$$

where $\mathbf{h}_{i,ik}$ is the m-th column of $\mathbf{H}_{i,k}$, which has been removed. It is noted that $\tilde{\mathbf{H}}_{i,k}$ is a submatrix of $\mathbf{H}_{i,k}$ by removing column $\mathbf{h}_{i,ik}$. Consequently, $\mathbf{w}_{i,k}$ can be obtained from the null space of $\tilde{\mathbf{H}}_{i,k}$, i.e., $\mathbf{w}_{i,k} = \mathbf{U}_{i,k}\mathbf{z}_{i,k}$, where $\mathbf{U}_{i,k}$ contains all the left singular vectors of $\tilde{\mathbf{H}}_{i,k}$ corresponding to zero singular values, and $\mathbf{z}_{i,k}$ is a $(N-M+1)\times 1$ normalized vector[1]. The choice of $\mathbf{z}_{i,k}$, when the maximal radio combining (MRC) is used can be given by $\mathbf{z}_{i,k} = \frac{\mathbf{U}_{i,k}^H\mathbf{h}_{i,ik}}{\left|\mathbf{U}_{i,k}^H\mathbf{h}_{i,ik}\right|}$.

By adopting the detection vector $\mathbf{w}_{m,k}$ at the receiver, the inter-cluster interference can be removed. Note that the identity precoding scheme in this work does not require the users to feedback their channel matrices to the base station. Instead, each user only needs to feedback one effective channel gain which is a scalar value, and therefore the amount of the required CSI feedback can be significantly reduced. As such, in the m-th cluster, the SINR for the k-th user ($1 \leq k \leq L_m$) to detect the j-th user ($k \leq j \leq L_m$) is given by

$$SINR_{m,k}^j = \frac{\left|\mathbf{w}_{m,k}^H\mathbf{h}_{m,mk}\right|^2\alpha_{m,j}}{\sum_{l=1}^{j-1}\left|\mathbf{w}_{m,k}^H\mathbf{h}_{m,mk}\right|^2\alpha_{m,l} + \left|\mathbf{w}_{m,k}\right|^2\frac{1}{\rho}}, \tag{8}$$

[1]It is assumed that $N \geq M$ to ensure the existence of $\mathbf{w}_{i,k}$.

where ρ denotes the transmit signal-to-noise-ratio (SNR). For the special case $k = j = 1$, the SINR can be simplified as

$$SINR_{m,1}^1 = \rho\frac{\left|\mathbf{w}_{m,1}^H\mathbf{h}_{m,m1}\right|^2\alpha_{m,1}}{\left|\mathbf{w}_{m,1}\right|^2}. \tag{9}$$

III. Problem Formulation and Proposed Optimization Methods

A. Problem Formulation

The objective of this work is to maximize the throughput of the worst user among all K users, by dynamically allocating users into different clusters. For each given combination of user allocation, in order to further improve the performance of MIMO-NOMA within each cluster, the power allocation coefficients are optimized according to instantaneous channel conditions in each cluster. In addition, as aforementioned in Section II, each cluster accommodates at least two users. Taking into account above, the throughput of the system can be optimized by solving the following problem:

$$\max_{\Omega}\ \min_{\forall m}\left(\log_2\left(1 + \max_{\boldsymbol{\alpha_m}}\min_{\forall k,\forall j}\left(SINR_{m,k}^j\right)\right)\right), \tag{10a}$$

$$\begin{aligned} s.t.\quad &\sum_{j=1}^{L_m}\alpha_{m,j} \leq \frac{L_m}{K}, \\ &\alpha_{m,j} \geq 0, j \in L_m. \end{aligned} \tag{10b}$$

where Ω is defined as the set of all user allocation combinations, $\boldsymbol{\alpha_m} = \{\alpha_{m,1},\cdots,\alpha_{m,L_m}\}$, is the power allocation coefficient vector.

B. Proposed Optimization Methods

In order to solve the above non-convex optimization problem, we use the alternating optimization strategy, which splits the throughput over j, k, $\boldsymbol{\alpha_m}$, m, and Ω into two steps:

1) Step 1: Fixing one combination of user allocation in Ω, and updating j, k, and $\boldsymbol{\alpha_m}$, we can solve the following sub-optimal problem:

$$\max_{\boldsymbol{\alpha_m}}\min_{\forall k,\forall j}\left(\mathbf{SINR_m}\left(\boldsymbol{\alpha_m}\right)\right) \quad s.t.\ (10b), \tag{11}$$

where $\mathbf{SINR_m}$ denotes all possible values of $SINR_{m,k}^j$, $\forall k, \forall j$ in the m-th cluster. Note that the max-min problem in (11) is not convex, which motives us to seek good equivalent transformations to make it tractable. We define the set $\mathbb{Q}_\Delta = \{\mathbf{SINR_m}\left(\alpha_\mathbf{m}\right) \geq \Delta, \Delta \in \mathbb{R}\}$ as the set of $\alpha_\mathbf{m}$ when the objective function is not smaller than Δ. Using the similar approach in [4], we can prove problem (11) is quasi-concave. As such, we can transform (11) equivalently as

$$\begin{aligned} &\text{Find} \qquad \boldsymbol{\alpha_m} \\ &s.t.\ (10b) \quad \text{and} \quad \left|\mathbf{w}_{m,k}^H\mathbf{h}_{m,mk}\right|^2\alpha_{m,j} \geq \Delta J, \end{aligned} \tag{12}$$

where $J = \sum_{l=1}^{j-1}\left|\mathbf{w}_{m,k}^H\mathbf{h}_{m,mk}\right|^2\alpha_{m,l} + \left|\mathbf{w}_{m,k}\right|^2\frac{1}{\rho}$. It is implied that with the aid of appropriately bounding Δ, a bisection search based method can be effectively used to reduce the searching scope of possible SINRs for obtaining $\boldsymbol{\alpha_m}$. Note that (12) is linear programs can be solved with standard optimization solvers. In this work, the *cvx* tool is invoked to find the optimal $\boldsymbol{\alpha_m}$ by utilizing the *Mosek* solver. The details of solving (12) are illustrated in **Algorithm 1**. Here $SINR_{\min}$

is the minimum of $\mathbf{SINR_m}$, Δ_{UB} and Δ_{LB} are initialized as the upper bound and lower bound for $\mathbf{SINR_m}$, respectively, and ε is the tolerance. After obtaining the minimum value of $\mathbf{SINR_m}$ for the m-th cluster, we can calculate the throughput of the worst user $\forall m$ using (10a).

2) Step 2: The second step is to traverse all the combinations, which is a NP-hard problem. In order to lower the computational complexity, we propose three efficient algorithms, named **Top-down A**, **Top-down B**, and **Bottom-up**.

Top-down A: As shown in **Algorithm 2**, we begin from the cluster with largest number (denoted by $\bar{L}_1$) of users, where $\bar{L}_1$ users are randomly allocated to this cluster. Then we continue random user allocation for the cluster with the second maximal user number, until the calculated combinations is less than C_{th}. The advantage of this algorithm is that it can reduce the complexity with the least decision times. This algorithm is suitable for the case when all the clusters have similar number of users.

Top-down B: As shown in **Algorithm 3**, initially, n users are randomly allocated to the cluster with the maximum number of users. Consequently, the number of combinations, C_1, is calculated, and compared with C_{th}. If $C_1 < C_{th}$, exhaustive search is performed among the rest of the clusters and the maximal throughput is obtained as the return value. Otherwise, we allocate n random users to the second sorted cluster. If $C_2 < C_{th}$, the obtained maximal throughput is returned. Otherwise, we update $n = n + 1$ and perform the above user allocation scheme until the calculated combinations is less than C_{th}. Finally, the corresponding throughput and the related user allocation scheme are output as return values. The advantage of this algorithm is that its complexity is the most controllable among the three.

Bottom-up: As shown in **Algorithm 4**, each cluster is randomly allocated with two users, as the starting point. Then, the cluster with the minimum user number is filled up randomly. After that we calculate the possible combination C_1 and compare it with C_{th}. If $C_1 < C_{th}$, exhaustive search is used to allocate the remaining users to the remaining clusters in order to achieve maximal throughput for the worst users. Otherwise, we continue the user allocation for the following clusters until the calculated combinations is less than C_{th}. The advantage of this algorithm is to guarantee that all clusters have some random search, and this is suitable for the case where the size of each cluster is significantly different.

C. Complexity of the proposed algorithms

Note that the complexity of this formulated problem contains two parts: 1) It can be easily observed that the complexity of solving (12) in *step 1* is linear to the number of users [4]. As a consequence, the computational complexity is $O(L_m)$. 2) The complexity of solving (10a) in *step 2* is given by Table I at the top of the next page, where r is the number of cluster which includes random search.

IV. Numerical Results

In the simulations, it is assumed that the total number of users is $K = 9$, the number of clusters is $M = 3$, the number of antennas at the BS and the receiver are assumed to be the same as $M = N = 3$, and each cluster contains $\mathbf{L} = \{4, 3, 2\}$ users. For simplicity, large scale path loss is not considered in this work and all the channel gains between the BS and users are assumed to be independent and identically complex Gaussian distributed, which is valid for many indoor scenarios. The initial values in Algorithm 1 are, $\Delta_{LB} = 0$, $\Delta_{UB} = 1000$, $\varepsilon = 10^{-4}$, and C_{th}= 200. The initial value of n is set as, $n = 1$, in Algorithm 3.

Fig. 1 shows the comparison of throughput for the worst case user among the three proposed algorithms and the exhaustive search with different transmit SNR. The solid curves rep-

Algorithm 1 Optimization Algorithm for Solving (12)

Input:
Δ_{LB}, Δ_{UB}, ε.
while $\Delta_{UB} - \Delta_{LB} \geq \varepsilon$ **do**
  Update $\Delta = (\Delta_{UB} + \Delta_{LB})/2$;
  Calculate $\boldsymbol{\alpha_m}$ with the constraints in (12), by solving the convex problem;
  if feasible **then**
    $\boldsymbol{\alpha}^*_{\boldsymbol{m}} = \boldsymbol{\alpha_m}$; $SINR_{\min} = \Delta$;
    Update $\Delta_{LB} = \Delta$;
  else
    Update $\Delta_{UB} = \Delta$;
  end if
end while
Return: Output $SINR_{\min}$ and $\boldsymbol{\alpha}^*_{\boldsymbol{m}}$.

Algorithm 2 Top-down A Algorithm

Input:
K, $\mathbf{L}$, $\bar{\boldsymbol{L}} = sort(\mathbf{L}, descend)$, C_{th}, C_0, $m = 1$.
while $C_{m-1} > C_{th}$ **do**
  Allocate $\bar{L}_m$ users randomly from $K - \sum_{p=1}^{m-1} L_p$ users to the cluster with m-th maximal number of users;
  Update $m = m + 1$;
  if $C_m < C_{th}$ **then**
    Exhaustive search among the rest clusters;
    Record the user allocation scheme, break;
  end if
end while
Return: The corresponding user allocation scheme.

Algorithm 3 Top-down B Algorithm

Input:
K, $\mathbf{L}$, $\bar{\boldsymbol{L}} = sort(\mathbf{L}, descend)$, C_{th}, C_0, $m = 1$, $0 < n < min(\mathbf{L})$.
while $C_{m-1} > C_{th}$ **do**
  for $m = 1$ *to* M **do**
    Allocate n users randomly to the cluster with m-th maximal users number from $K - (m-1) \times n$ users;
    Update $m = m + 1$;
    if $C_m < C_{th}$ **then**
      Exhaustive search among the rest clusters;
      Record the user allocation scheme, break;
    end if
  end for
  Update $n = n + 1$;
end while
Return: The corresponding user allocation scheme.

TABLE I
COMPLEXITY OF THE PROPOSED ALGORITHMS

Algorithm	Exhaustive Search	Top-down A	Top-down B	Bottom-up
Complexity	$O\left(\frac{K!}{\prod_{m=1}^{M} L_m!}\right)$	$O\left(\frac{\left(K-\sum_{m=1}^{r}\tilde{L}_m\right)!}{\prod_{m=r+1}^{M}\tilde{L}_m!}\right)$	$O\left(\frac{(K-r\times n)!}{\prod_{m=1}^{r}(\tilde{L}_m-n)!\prod_{m=r+1}^{M}\tilde{L}_m!}\right)$	$O\left(\frac{\left(K-\sum_{m=1}^{r}\tilde{L}_m-2(M-r)\right)!}{\prod_{m=r+1}^{M}\tilde{L}_m!}\right)$

Algorithm 4 Bottom-up Algorithm

Input:

K, $\mathbf{L}$, $\tilde{\boldsymbol{L}} = sort\,(\mathbf{L}, ascend)$, C_{th}, C_0, $m = 1$.

1: **while** $C_{m-1} > C_{th}$ **do**
2: Allocate every two users into each cluster randomly.
3: Allocate $\tilde{L}_m - 2$ users randomly from $K - \sum_{p=1}^{m-1} L_p$ users to the cluster with m-th minimal number of users;
4: Update $m = m + 1$;
5: **if** $C_m < C_{th}$ **then**
6: Exhaustive search among the rest clusters;
7: Record the user allocation scheme, break;
8: **end if**
9: **end while**

Return: The corresponding user allocation scheme.

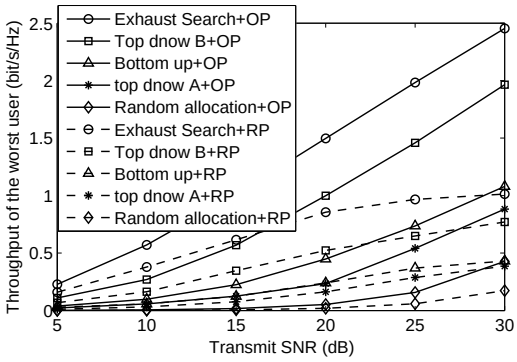

Fig. 1. Optimized throughput of the worst user with different algorithms for user allocations under different transmit SNR.

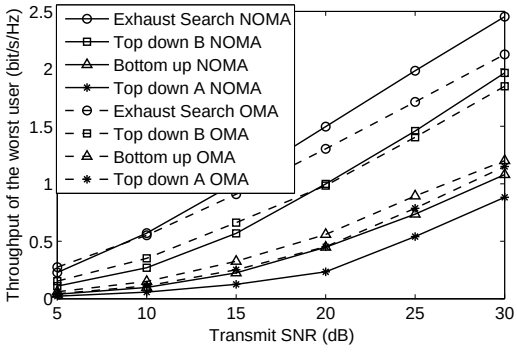

Fig. 2. Comparison between NOMA and OMA of the proposed algorithms.

resent the case with the optimal power allocation for each user within the cluster, while the dashed curves represent random power allocation which follows increasing order for each user. One can observe that the optimal power allocation achieves much better performance than random power allocation, which demonstrates the effectiveness of the proposed **Algorithm 1**.

Fig. 2 shows the fairness comparison between NOMA and orthogonal multiple access (OMA) for the exhaustive search and the three proposed algorithms. The solid curves represent NOMA, while the dashed curves represent OMA. It is observed that the exhaustive search and Top down B algorithms with NOMA outperform the corresponding ones with OMA, while the Bottom up and Top down B algorithms with OMA outperforms the corresponding ones with NOMA. This behavior can be explained as follows: the fairness in NOMA is more sensitive to the times of searching and system parameters. However, for the OMA scheme, since all the users are with the same power, reducing the time of searching will not affect much of the fairness performance. Therefore, we can conclude that by carefully designing the system parameters and choosing appropriate algorithm, NOMA can outperform OMA in term of fairness.

V. CONCLUSION

In this paper, a dynamic clustering optimization problem considering the fairness in MIMO-NOMA systems, was investigated. In order to solve this non-convex problem, a two-step optimization method was proposed. Three efficient suboptimal algorithms were proposed to reduce the computational complexity. To further improve the performance of the worst user in each cluster, power allocation coefficients were optimized by using bi-section search. Numerical results demonstrated that the proposed algorithms can achieve a good tradeoff between throughput and system complexity.

ACKNOWLEDGMENT

The authors would like to thank Dr. Muhammad Fainan Hanif for his suggestions.

REFERENCES

[1] Z. Ding and *et al.*, "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Commun. Mag.*, submitted. [Online]. Available: http://arxiv.org/abs/1511.08610
[2] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, "System-level performance evaluation of downlink non-orthogonal multiple access (NOMA)," in *Proc. IEEE Annual Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, London, UK, Sep. 2013.
[3] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1501–1505, Dec. 2014.
[4] S. Timotheou and I. Krikidis, "Fairness for non-orthogonal multiple access in 5G systems," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1647–1651, Oct. 2015.
[5] Y. Liu, Z. Ding, M. Elkashlan, and H. V. Poor, "Cooperative non-orthogonal multiple access with simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, Apr. 2016.
[6] J. Choi, "Minimum power multicast beamforming with superposition coding for multiresolution broadcast and application to NOMA systems," *IEEE Trans. Commun.*, vol. 63, no. 3, pp. 791–800, Mar. 2015.
[7] Q. Sun, S. Han, C.-L. I, and Z. Pan, "On the ergodic capacity of MIMO NOMA systems," *IEEE Wireless Commun. Lett.*, vol. 4, no. 4, pp. 405–408, Aug. 2015.
[8] Z. Ding, F. Adachi, and H. V. Poor, "The application of MIMO to non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 537–552, Jan. 2016.