IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 2, 2020, accepted December 15, 2020, date of publication December 17, 2020, date of current version January 4, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3045503

Beamforming Optimization and Power Allocation for User-Centric MIMO-NOMA IoT Networks

QIANG WANG AND ZHAO WU, (Member, IEEE)
Key Laboratory of Complex System Optimization and Big Data Processing, Yulin Normal University, Yulin 537000, China
School of Physics and Telecommunication Engineering, Yulin Normal University, Yulin 537000, China

Corresponding author: Zhao Wu (kianty@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61871466, in part by the Guangxi Natural Science Foundation Project under Grant 2019GXNSFBA245076 and Grant 2018GXNSFBA281124, in part by the Project of Education Department of Guangxi under Grant 2020KY14016, in part by the Opening Foundation of Yulin Research Institute of Big Data under Grant 2020YJKY02, in part by the Project of Yulin Normal University under Grant G2017017, and in part by the Guangxi University High Level Innovation Team and Distinguished Scholars Program of China under Document [2018] 35.

ABSTRACT Ultra high reliability and ultra low latency are the key objectives of the internet of things (IoT) with massive connectivity. To support these objectives, we investigate the resource allocation for the user-centric multi-cell multiple-input multiple-output non-orthogonal multiple access (MIMO-NOMA) based IoT networks. The macro base station (MBS) equipped with multiple antennas transmits signals to access points (APs) in the backhaul link, and each device can be served by multiple APs in the access link, and the APs serving the same device compose one AP group (APG). The NOMA is applied in each APG to reduce the intra-APG interference. In this paper, the resource allocation problem involving the beamforming optimization and power allocation is formulated as nonconvex optimization problem which is extremely difficult to tackle. In order to reduce the computational complexity, we decompose the resource allocation problem into two subproblems in terms of the beamforming optimization and power allocation. For the beamforming optimization subproblem, the zero-forcing beamforming (ZFBF) algorithm is applied to solve it. When the beamforming strategy is fixed, the power allocation subproblem is still a nonconvex optimization problem. We first transform it as a difference of two convex functions (DC) problem, and then the DC programming method is adopted to optimize it. We further prove that the solution obtained by the DC programming method is one of the local optimal solutions of the original optimization problem. Extensive simulation results are presented to demonstrate the effectiveness of the proposed resource allocation scheme for the user-centric MIMO-NOMA IoT networks.

INDEX TERMS Internet of Things (IoT), MIMO, NOMA, power allocation, beamforming.

I. INTRODUCTION

With the rapid development of the internet of things (IoT) networks (e.g. smart city, connected health, industrial internet, vehicle network), IoT is facing huge challenges in terms of the reliability and latency. Indeed, the number of the wireless enabled sensors increases with the explosive exponential growth. In according with the prediction of Cisco, the number of the terminations of IoT will reach 14.6 billion by 2022 [1]. In addition, the require of some particular scenarios such as self-driving and industrial internet for reliability and latency is tougher and tougher.

The user-centric wireless network has been proposed to improve the reliability and decrease the latency. In the user-centric wireless network, substantial low power and small coverage access points (APs) are deployed, and the number of the APs is much more than that of the terminals. Therefore, one terminal can be served by multiple APs simultaneously, and it is possible that each terminal can get sustainable high quality service through cooperation among APs. By shifting the network architecture from traditional cell-centric to user-centric, the user experience will be enhanced significantly. However, dense-deployed of the APs will also cause serious interference which may degrade the system performance.

Non-orthogonal multiple access (NOMA) is another promising technology for IoT to provide massive connectivity over limited radio resources. NOMA allows that multiple users can share the same spectrum resource by means of superposition coding and successive interference cancellation (SIC). Comparing with the orthogonal multiple access (OMA), NOMA can significantly improve the spectrum

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

efficiency by exploiting the power domain. Due to the spectrum multiplex, it is important to optimize the power allocation for NOMA system. In addition, multiple-input-multiple-output (MIMO) has been also widely investigated to boost the throughput and reliability of IoT networks. To make full use of these advantages, MIMO-NOMA has been proposed as a multiple access technique for the 5G mobile networks [2].

In the user-centric MIMO-NOMA IoT networks, ultra dense terminals and limited spectrum resource will cause serious interference, which will significantly degrade the system performance. Therefore, how to suppress the interference is crucial to improve the system performance. In this case, it is important for the user-centric MIMO-NOMA IoT to optimize the resource allocation to eliminate the interference. On the other hand, the resource allocation problem for the MIMO-NOMA IoT is excellently difficult, since both the beamforming strategy and power allocation are needed to be optimized. Furthermore, no matter the beamforming optimization problem or the power allocation optimization problem are nonconvex optimization problem due to the utilization of non-orthogonal spectrum resource.

In this paper, we investigate the joint resource allocation problem involving the beamforming optimization and power allocation for the user-centric MIMO-NOMA IoT networks in order to maximize the system throughput. We consider both the backhaul downlink (from the MBS to APs) and access downlink (from APs to devices) since the transmission rate of the access downlink is limited by the backhaul downlink. In the backhaul downlink, the MBS equipped with multiple antennas transmits signals to the single-antenna APs. The APs are grouped to serve devices, and the APs in the same AP group (APG) will share the same beamforming vector. That is, the MBS and each APG form a virtual MIMO system. In the access downlink, each terminal is served by multiple APs simultaneously to enhance the reliability and decrease the latency. In order to decrease the interference, the NOMA is applied both at the AP side and device side for the backhaul downlink and access downlink, respectively. In the backhaul downlink, the APs in the same APG will share the same beamforming vector, and the SIC is applied in each AP to mitigate the interference. In the access downlink, the APs in the same APG transmit information to the corresponding device using the same spectrum, and the co-channel interference will be reduced by the SIC. The joint resource allocation involving the beamforming optimization and power allocation is formulated as a nonconvex optimization problem which is extremely difficult to tackle. In order to reduce the computational complexity, we decompose the resource allocation problem into two subproblems in terms of the beamforming optimization and power allocation. For the beamforming optimization subproblem, a novel zero-forcing beamforming (ZFBF) algorithm is applied to solve it. However, the power allocation subproblem is still a nonconvex optimization problem due to the term of the interference. To solve it, we first reformulate the power allocation optimization subproblem as a DC programming problem, and then the DC programming method is adopted to optimize it. We further prove that the solution obtained by the DC programming method is one of the local optimal solutions of the original problem.

The main contributions of this paper are summarized as follows:

1) We investigate the joint resource allocation problem involving the beamforming optimization and power allocation for the user-centric MIMO-NOMA IoT networks to maximize the system throughput, and then the resource allocation is formulated as a nonconvex optimization problem.
2) The formulated resource allocation model is a large scale nonconvex optimization problem which is extremely difficult to solve. In order to reduce the computational complexity, we decompose the original optimization problem into two optimization subproblems in terms of the beamforming optimization and power allocation respectively. For the beamforming optimization subproblem, a novel zero-forcing beamforming ZFBF) algorithm is applied to solve it.
3) The power allocation is still a nonconvex optimization problem due to the intra-APG interference, and the problem can be transferred as a DC programming problem. The DC programming method is adopted to optimize the power allocation problem. We further prove that the solution obtained by the DC programming method is one of the local optimal solutions of the original problem.

The rest of the paper is organized as follows. In section II, the previous related works are reviewed. We describe the system model in section III. The beamforming optimization is introduced in section IV. In section V, we optimize the power allocation problem using DC programming method and analyze the performance the algorithm. In section VI, simulation results are provided to evaluate the performance of the proposed algorithm. Finally, we conclude this paper in section VII.

II. RELATED WORKS

As aforementioned, the user-centric wireless networks can effectively improve the reliability and decrease the latency since each terminal can be served by multiple APs simultaneously. However, the serious interference caused by the limited spectrum resource and massive nodes will degrade the system performance seriously. In addition, the complexity of the resource allocation problem of the user-centric wireless networks is extremely high since the resource allocation problem is usually formulated as a large scale nonconvex optimization problem. In order to decrease the co-channel interference, [3] proposes a fractional frequency reuse mechanism in which the bandwidth is partitioned into several sunbands, and each subband is independently allocated to APs. Reference [4] proposes a cluster-based spectrum resource allocation in which the massive APs are clustered as some

APGs, and each APG is regarded as the spectrum resource allocation unit. Reference [5] proposes a joint frequency bandwidth dynamic division, clustering and power control algorithm to control interference and improve spectrum efficiency. In [5], the authors divide the system bandwidth into three frequency bands (i.e. the reuse band, the femtocell dedicated band, and the macrocell dedicated band) dynamically according to the density and the location of femtocells. In [6], the authors present a novel clustering-based resource allocation framework for downlink transmission, and they jointly optimize the subchannel and power allocation and user traffic demand in terms of a large-scale network scenario, and formulate the resource allocation as a combinatorial optimization problem. To reduce the complexity, an interference-separation clustering-based scheme is proposed to divide the massive small cells into smaller groups with different priorities, which reduces the network scale. Reference [7] proposes a multi-cluster based dynamic channel assignment to improve system performance for the downlink. In [7], the graph coloring algorithm is adopted to group APs, and then a dynamic subchannel assignment scheme is proposed to allocate subchannels for maximizing the system throughput.

NOMA has been widely applied in IoT networks since it can significantly improve the spectrum efficiency by exploring the power domain multiplexing. Reference [8] studies the Narrowband-Internet of Things (NB-IoT) to support machine type communications (MTCs) in next generation mobile networks. The authors in [8] propose a power domain NOMA scheme with user clustering for an NB-IoT system. In particular, the MTC devices are assigned to different ranks within the NOMA clusters where they transmit over the same frequency resources. Then, an optimization problem is formulated to maximize the total throughput of the network by optimizing the resource allocation of MTC devices and NOMA clustering while satisfying the transmission power and quality of service requirements. Reference [9] proposes a power-domain NOMA scheme for the NB-IoT systems to enhance the connection density by allowing multiple IoT devices to simultaneously access one subcarrier. In [9], both single-tone and multi-tone transmission modes of the NB-IoT systems are considered, where each device can access a single subcarrier or a bond of contiguous subcarriers, respectively. The authors in [9] formulate joint subcarrier and power allocation problems for both transmission modes to maximize the connection density while taking the quality of service requirements and the transmit power constraints of IoT devices into account. Reference [10] investigates a NOMA-enhanced IoT network. In [10], a novel cluster strategy is proposed, where multiple devices can be served simultaneously. Then, the stochastic geometry is adopted to model the spatial randomness of both terrestrial and aerial devices. Reference [11] considers the radio frequency (RF) energy harvesting IoT relay system. In [11], the authors consider to transmit the data of IoT relay node along with source node data using non-orthogonal multiple access (NOMA) protocol in the presence of an interfering signal to their respective destinations.

MIMO is another critical technology for wireless networks since it can significantly improve the system performance. Reference [12] investigates the integration of massive MIMO enabled heterogeneous cellular networks and IoT networks. Considering that the human-to-human devices mainly concentrate on downlink throughput, but the IoT devices pay more attention to uplink power consumption, the authors in [12] design a device association mechanism to achieve a tradeoff between these two performance metrics under devices' association requirements. Reference [13] investigates the feasibility of improving the energy efficiency of massive MIMO orthogonal frequency division multiplexing (OFDM) systems applied to a battery-limited IoT networks. For the uplink aspect, the authors in [13] consider the uplink reference signal (RS) power control. Reducing uplink RS power could induce the battery saving of IoT devices but could cause an increase in channel estimation error. For the downlink aspect, they consider the peak-to-average power ratio reduction of the OFDM signal and downlink transmitter power control. In addition, [13] considers the utilization of radio frequency energy transfer using unmanned aerial vehicles to extend the operating time of battery-limited IoT devices. Reference [14] considers the access phase for IoT networks in a mixed-analog-to-digital converter distributed massive MIMO system, in which users are classified into light-load users and heavy-load users depending on their traffic load requirements. To meet the low-latency and low-cost demands in IoT, the access scheme for both types of users are designed in a grant-free fashion.

In order to make full user of the advantages of NOMA and MIMO, the integration of NOMA with MIMO has been drawn a lot of attentions. Reference [15] proposes a downlink scheme combining virtual MIMO and NOMA to support massive connectivity and boost spectral efficiency for IoT networks. In [15], the outage probability and goodput of the virtual MIMO-NOMA system are thoroughly investigated by considering the Kronecker model, which embraces both the transmit and receive correlations. Then, the goodput maximization problems is solved in closed form by the Karush-Kuhn-Tucker conditions, with which the joint power and rate selection is realized by using alternately iterating optimization. In [16], a new MIMO-NOMA scheme in IoT is designed to serve users quickly for small packet transmission, where one user is served with its quality of service requirement strictly met, and the other user is served opportunistically by using the NOMA concept. In [17], the authors propose a novel successive sub-array activation diversity scheme for a massive MIMO system in combination with NOMA, and a low-complexity two-stage beamformer which is constructed based only on the long-term channel statistical information is proposed. Then, The authors in [17] derive an exact closed-form expression for the outage probability through carrying out the in-depth analytical analysis. Reference [18] addresses multi-user multi-cluster massive MIMO systems with NOMA. In [18], the power optimization is simplified to a convex problem, and then an iterative algorithm

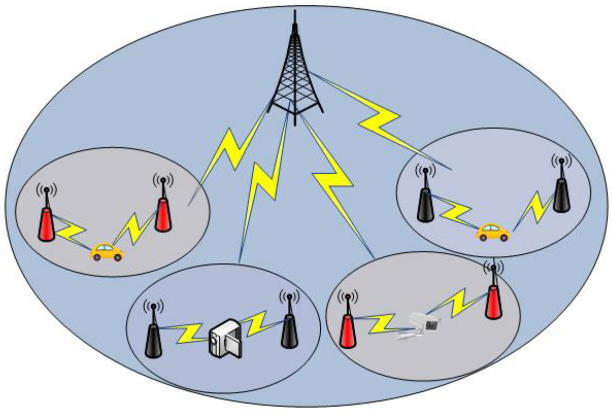

FIGURE 1. System model of the user-centric IoT network.

is proposed to provide fairness also among different sub-groups.

III. SYSTEM MODEL

As shown in Fig.1, we consider a MIMO-NOMA system in an user-centric IoT that consists of one MBS equipped with N_t antennas overlapped a set of M single-antenna APs. There are N ($N<M$) IoT devices that need to access the network, and each IoT device is served by multiple APs simultaneously. It is assumed that M APs have been divided into N AP groups (APG), and the nth APG serves the nth device.

We assume that the transmission process includes the access links (from APs to IoT devices) and downlink backhaul links (from MBS to APs). For the backhaul downlinks, the MBS allocates the same beamformer to the APs belonging to one APG. Then, the SIC is applied in each AP to mitigate the interference. For the access downlink, the APs in the same APG transmit signals to the corresponding device using the same spectrum resource. In order to reduce the intra-APG interference, the NOMA is adopted in the access links, that is, multiple APs superpose multiple signals on the same frequency to the corresponding IoT device, and then the SIC is applied at the receiver to decode the superposed signals.

A. DOWNLINK SIGNALS MODEL

In this subsection, we provide the signal model of the backhaul downlink model and access downlink respectively for the user-centric MIMO-NOMA IoT.

1) BACKHAUL DOWNLINK SIGNAL MODEL

As aforementioned, the MBS equipped with N_t antennas, and there are M single-antenna APs which are divided into N APGs. It is assumed that the number of APs in the nth APG is M_n, $n = 1, \ldots, N$. Obviously, $\sum_{n=1}^{N} M_n = M$. Then, for the backhaul downlink, M_n APs in the nth APG shares the same beamformer, $n = 1, \ldots, N$, and N independent data streams can be given as:

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]^T \tag{1}$$

where $\mathbf{x}_n = \sqrt{p_1}s_1 + \ldots + \sqrt{p_{m_n}}s_{m_n} + \ldots + \sqrt{p_{M_n}}s_{M_n}$; p_{m_n} denotes the transmit power from the MBS to the m_nth AP in the nth APG, $m_n = 1, 2, \ldots, M_n$, $n = 1, 2, \ldots, N$.

The received signals of the APs can be written as:

$$\mathbf{y} = \mathbf{HWX} + z \tag{2}$$

where $\mathbf{H} = [H_1^T, \ldots, H_N^T]^T \in \mathbb{C}^{M\times N_t}$ is the channel matrix from the MBS to APs, and $H_n \in \mathbb{C}^{M_n\times N_t}$; $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_N) \in \mathbb{C}^{N_t\times N}$ denotes the beamforming matrix at the MBS, and $\|w_n\| = 1, n = 1, \ldots, N$; z is the additive Gaussian white noise (AWGN) with zero mean and variance δ^2.

The received signal of the mth AP in the nth APG can be written as:

$$\mathbf{y}_{m_n} = \mathbf{h}_{m_n}\mathbf{w}_n\sqrt{p_{m_n}}s_{m_n} + \sum_{i=1, i\neq m_n}^{N_r} \mathbf{h}_{m_n}\mathbf{w}_n\sqrt{p_{i_n}}s_{i_n} + \sum_{j=1, j\neq n} \mathbf{h}_{m_n}\mathbf{w}_j\mathbf{x}_j + z_{m_n} \tag{3}$$

where $\mathbf{h}_{m_n} \in \mathbb{C}^{1\times N_t}$ denotes the channel coefficient from the MBS to the m_nth AP in the nth APG; the first term is the desired signal, and the second term is the intra-APG interference from the other $M_n - 1$ APs of the nth APG, and the third term is the inter-APG interference from the other $N-1$ APGs; z_{m_n} is the AWGN with zero mean and variance δ^2 for the m_nth AP in the nth APG from the MBS.

In each APG, the NOMA is applied to reduce the interference caused by the beamformer sharing. Without loss of generality, we assume the channel coefficients from the MBS to APs can be ordered as follows: Without loss of generality, it is assume that the channel coefficients can be ordered as:

$$\|\mathbf{h}_{1_n}\| \geq \|\mathbf{h}_{2_n}\| \geq \ldots \geq \|\mathbf{h}_{M_n}\| \tag{4}$$

where $\|\bullet\|$ represents the 2-norm.

In accordance with the principle of NOMA, the AP with better channel condition can decode the signals of the APs with weaker channel condition and then proceeds to subtract it from the received signal and decode its own data. The received signal-to-interference-plus-noise ratio (SINR) of the m_nth AP in the nth APG is given by:

$$\mathbf{SINR}_{m_n} = \frac{|\mathbf{h}_{m_n}\mathbf{w}_n|^2 p_{m_n}}{\sum_{i=1}^{m_n-1} |\mathbf{h}_{m_n}\mathbf{w}_n|^2 p_{i_n} + \sum_{j=1, j\neq n} |\mathbf{h}_{m_n}\mathbf{w}_j|^2 p_j + \delta^2} \tag{5}$$

where $\mathbf{p}_j$ is the total transmit power of the jth APG.

Therefore, the achievable rate of the m_nth AP in APG n in the backhaul downlink can be written as:

$$R_{b,m_n} = \log(1 + \mathbf{SINR}_{m_n}) \tag{6}$$

The achievable backhaul rate of the nth user can be given as:

$$R_{b,n} = \sum_{m_n=1}^{M_n} R_{b,m_n} = \sum_{m_n=1}^{M_n} \log(1 + \mathbf{SINR}_{m_n}) \tag{7}$$

B. ACCESS DOWNLINK SIGNALS MODEL

In this subsection, we give the signal model of the access downlink model of the user-centric MIMO-NOMA IoT networks. For the access downlink, there exists intra-APG interference as a result that multiple APs in the same APG transmit signals to the corresponding device simultaneously. For simplification, it is assumed that the inter-APG interference can be avoided due to the proper APs grouping and resource allocation. The received signals of the nth device can be given as:

$$y_n = \sum_{m_n=1}^{N_r} h_{m_n n}\sqrt{p_{m_n n}} s_{m_n n} + z_{m_n n} \tag{8}$$

where $h_{m_n n}$ denotes the channel coefficient from the m_nth AP in the nth APG to the nth device; $p_{m_n n}$ represents the transmit power of the m_nth AP in the nth APG to the nth device; $z_{m_n n}$ means the AWGN with zero mean and variance δ^2 for the nth device from the m_nth AP in the nth APG.

Without loss of generality, it is assumed that the channel coefficients can be ordered as:

$$h_{1n} \geq h_{2n} \geq \cdots \geq h_{M_n n} \tag{9}$$

According to the principle of NOMA, device n with SIC can successfully decode the signals of the APs with weaker channel condition. That is, the signal of the AP with best channel condition can be first decoded, however, it should experience interference from the other APs in the APG since the device cannot remove the signals from the other APs.

Thus, the received SINR of the nth device can be written as:

$$\mathbf{SINR}_{m_n n} = \frac{|h_{m_n n}|^2 p_{m_n n}}{\sum\limits_{d=m_n+1}^{M_n} |h_{dn}|^2 p_{dn} + \delta^2} \tag{10}$$

The achievable access rate of the nth device served by the nth APG can be given as:

$$R_{a,n} = \sum_{m_n=1}^{M_n} \log(1 + \mathbf{SINR}_{m_n n}) \tag{11}$$

Then, the total system throughput can be given as:

$$R = \sum_{n=1}^{N} R_{a,n} \tag{12}$$

$$R_{a,n} \leq R_{b,n} \tag{13}$$

C. PROBLEM FORMULATION

In this paper, we aim to maximize the system throughput of the user-centric MIMO-NOMA IoT networks through optimizing the beamforming strategy and power allocation. The optimization problem can be formulated as:

$$\mathbf{P1} \quad \max_{\mathbf{W},\mathbf{p}} \; R_{tot} = \sum_{n=1}^{N} R_{a,n} \tag{14}$$

$$s.t. \sum_{n=1}^{N} \sum_{m_n=1}^{M_n} p_{m_n} \leq P_B \tag{15}$$

$$p_{m_n n} \leq P_{ap}, \quad \forall m_n, \forall n \tag{16}$$

$$R_{a,n} \leq R_{b,n}, \quad \forall n \tag{17}$$

where (15) and (16) represent the maximum power constraint of the MBS and APs, respectively; (17) means that the achievable rate of each device should be less than the available data rate of backhaul links due to the limited backhual capacity.

It is obvious that the joint resource allocation problem **P1** is a nonconvex optimization problem with high computational complexity which is extremely difficult to solve. In order to reduce the computational complexity, we decompose problem **P1** into two optimization subproblems in terms of the beamforming optimization and power allocation respectively. For the beamforming optimization, we use a novel ZFBF algorithm in [19] to solve it. Then, the DC programming method is adopted to optimize the power allocation subproblem.

IV. BEAMFORMING OPTIMIZATION

In this section, we adopt a novel ZFBF algorithm in [19] to solve the beamforming optimization in the backhaul downlink. Next we briefly introduce the novel ZFBF algorithm, and the details of this algorithm can be referred to [19].

In the first, the channel coefficient matrix of the nth APG is obtained: $\mathbf{H}_n = [\mathbf{h}_{1_n}^T, \ldots, \mathbf{h}_{M_n}^T]^T \in \mathbb{C}^{M_n \times N_t}$, where $\mathbf{h}_{m_n} \in \mathbb{C}^{1\times N_t}$ is the channel coefficient from the MBS to the m_nth AP in the nth APG.

Then, the singular value decomposition is used to decompose the channel coefficient matrix $\mathbf{H}_n$ as:

$$H_n = U_n \Sigma_n V_n^T \tag{18}$$

where U_n and V_n are the left singular matrix and right singular matrix of $\mathbf{H}_n$, respectively. Σ_n is a diagonal matrix.

In the next, the equivalent matrix $\overline{\mathbf{H}}_n$ of $\mathbf{H}_n$ can be obtained:

$$\overline{\mathbf{H}}_n = U_n^{(1)T} \mathbf{H}_n \tag{19}$$

where $U_n^{(1)T} \in \mathbb{C}^{1\times M_n}$ denotes the transposition of the first column of matrix U_n.

The equivalent matrix $\overline{\mathbf{H}}$ of the channel coefficient matrix $\mathbf{H}$ of all APs is:

$$\overline{\mathbf{H}} = [\overline{\mathbf{H}}_1, \ldots, \overline{\mathbf{H}}_N] \tag{20}$$

Then, computing the pseudo inverse of $\overline{\mathbf{H}}$, and making $\overline{\mathbf{H}}$ unit, the beamforming matrix $\mathbf{W}$ can be obtained:

$$W = \frac{\overline{\mathbf{H}}^+}{\|\overline{\mathbf{H}}^+\|_F} = [\frac{\overline{\mathbf{w}}_1}{\|\overline{\mathbf{w}}_1\|_F}, \ldots, \frac{\overline{\mathbf{w}}_N}{\|\overline{\mathbf{w}}_N\|_F}] \tag{21}$$

where $\| \bullet \|_F$ represents the F-norm, and $\overline{\mathbf{w}}_n$ denotes the nth column of the pseudo inverse of $\overline{\mathbf{H}}$.

V. POWER ALLOCATION USING DC PROGRAMMING

In the above section, the beamforming matrix $\mathbf{W}$ has been optimized using the novel ZFBF algorithm. The original optimization problem can be reformulated as:

$$\mathbf{P2} \quad \max_{\mathbf{p}} \; R_{tot} = \sum_{n=1}^{N} R_{a,n} \tag{22}$$

$$s.t. \; (15), (16), (17) \tag{23}$$

It can be observed that problem **P2** is also difficult to solve for the power allocation subproblem since it is still a nonconvex optimization problem. In order to decrease the computational complexity, we first approximate the constraint (17) as a affine constraint. To achieve it, we introduce two additional constraints of problem **P2**:

$$\sum_{i=1}^{m_n-1} |\mathbf{h}_{m_n}\mathbf{w}_n^*|^2 p_{i_n} + \sum_{j=1, j\neq n} |\mathbf{h}_{m_n}\mathbf{w}_j^*|^2 p_j \leq \chi_b \tag{24}$$

$$\sum_{d=m_n+1}^{M_n} |\mathbf{h}_{dn}|^2 p_{dn} \leq \chi_a \tag{25}$$

where χ_b and χ_a are pre-set values, and by adjusting them, a good system performance can still be obtained.

Then, both the achievable backhaul downlink rate and the achievable access rate of user n are bounded by:

$$R_{b,n}^*(\mathbf{p}_1) = \sum_{m_n=1}^{M_n} \log(1 + \frac{|\mathbf{h}_{m_n}\mathbf{w}_n|^2 p_{m_n}}{\chi_b + \delta^2}) \tag{26}$$

$$R_{a,n}^*(\mathbf{p}_2) = \sum_{m_n=1}^{M_n} \log(1 + \frac{|h_{m_n n}|^2 p_{m_n n}}{\chi_a + \delta^2}) \tag{27}$$

where $\mathbf{p}_1 = \{p_{m_n}\}$, and $\mathbf{p}_2 = \{p_{m_n n}\}$.

(26) and (27) can be approximated by the first Taylor expansion:

$$R_{b,n}^*(\mathbf{p}_1) \approx \overline{R}_{b,n}^*(\mathbf{p}_1) = R_{b,n}^*(\mathbf{p}_1^v) + \nabla R_{b,n}^*(\mathbf{p}_1^v)(\mathbf{p}_1 - \mathbf{p}_1^v) \tag{28}$$

$$R_{a,n}^*(\mathbf{p}_2) \approx \overline{R}_{a,n}^*(\mathbf{p}_2) = R_{a,n}^*(\mathbf{p}_2^v) + \nabla R_{a,n}^*(\mathbf{p}_2^v)(\mathbf{p}_2 - \mathbf{p}_2^v) \tag{29}$$

where $\mathbf{p}_1^v$ and $\mathbf{p}_2^v$ denote the solutions of $\mathbf{p}_1$ and $\mathbf{p}_2$ from the vth iteration.

Therefore, constraint (17) can be rewritten as a affine form:

$$\overline{R}_{b,n}^*(\mathbf{p}_1) - \overline{R}_{a,n}^*(\mathbf{p}_2) \geq 0 \tag{30}$$

In accordance the property of logarithmic function, the objective function of problem **P2** can be rewritten as:

$$\begin{aligned}\sum_{n=1}^{N} R_{a,n} &= \sum_{n=1}^{N}\sum_{m_n=1}^{M_n} \log(1+\frac{|h_{m_n n}|^2 p_{m_n n}}{\sum_{d=m_n+1}^{M_n} |h_{dn}|^2 p_{dn}+\delta^2}) \\ &= \sum_{n=1}^{N}\sum_{m_n=1}^{M_n} \log(\sum_{d=m_n+1}^{M_n} |h_{dn}|^2 p_{dn}+|h_{m_n n}|^2 p_{m_n n}+\delta^2) \\ &\quad -\sum_{n=1}^{N}\sum_{m_n=1}^{M_n} \log(\sum_{d=m_n+1}^{M_n} |h_{dn}|^2 p_{dn}+\delta^2) \end{aligned} \tag{31}$$

Let $H(\mathbf{p}_2^1) = \sum_{n=1}^{N}\sum_{m_n=1}^{M_n} \log(\sum_{d=m_n+1}^{M_n} |\mathbf{h}_{dn}|^2 p_{dn} + \delta^2)$, where $\mathbf{p}_2^1 = \{p_{(m_n+1)n}, \ldots, p_{M_n n}\}$. $H(\mathbf{p}_2^1)$ can be approximated by the first Taylor expansion as:

$$H(\mathbf{p}_2^1) = H(\mathbf{p}_2^{1v}) + \nabla H(\mathbf{p}_2^{1v})(\mathbf{p}_2^1 - \mathbf{p}_2^{1v}) \tag{32}$$

Thus, problem **P2** can be rewritten as:

$$\begin{aligned}\textbf{P3} \quad \max \sum_{n=1}^{N}\sum_{m_n=1}^{M_n} &\log(\sum_{d=m_n+1}^{M_n} |\mathbf{h}_{dn}|^2 p_{dn} + \delta^2 + |\mathbf{h}_{m_n n}|^2 p_{m_n n}) \\ &-H(\mathbf{p}_2^{1v}) - \nabla H(\mathbf{p}_2^{1v})(\mathbf{p}_2^1 - \mathbf{p}_2^{1v}) \end{aligned} \tag{33}$$

$$s.t. \ (15), (16), (30) \tag{34}$$

It can be observed that problem **P3** is a convex optimization problem, since the objective function is concave, and all the constraints are affine. In the following, we use the Lagrangian dual method to optimize it.

The Lagrangian function of problem **P3** can be given as:

$$\begin{aligned} L = &\sum_{n=1}^{N}\sum_{m_n=1}^{M_n} \log(\sum_{d=m_n+1}^{M_n} |\mathbf{h}_{dn}|^2 p_{dn} + \delta^2 + |\mathbf{h}_{m_n n}|^2 p_{m_n n}) \\ &-H(\mathbf{p}_2^{1v}) - \nabla H(\mathbf{p}_2^{1v})(\mathbf{p}_2^1 - \mathbf{p}_2^{1v}) \\ &-\lambda(\sum_{n=1}^{N}\sum_{m_n=1}^{M_n} p_{m_n} - P_B) - \sum_{n=1}^{N}\sum_{m_n=1}^{M_n} \mu_{m_n n}(p_{m_n n} - P_{ap}) \\ &-\eta_n(\overline{R}_{b,n}^*(\mathbf{p}_1) - \overline{R}_{a,n}^*(\mathbf{p}_2)) \end{aligned} \tag{35}$$

where λ, $\mu_{m_n n}(m_n = 1, 2, \ldots, M_n, n = 1, 2, \ldots, N)$ and $\eta_n(n = 1, 2, \ldots, N)$ are Lagrangian multipliers with respect to constraints (15), (16) and (30), respectively.

Then, we can obtain the dual function of problem **P3**:

$$g(\lambda, \overrightarrow{\mu}, \overrightarrow{\eta}) = \max_{\mathbf{p}} L(\mathbf{p}, \lambda, \overrightarrow{\mu}, \overrightarrow{\eta}) \tag{36}$$

Setting the the partial derivation of (35) with respect to $p_{m_n n}, \forall n \ \forall m_n$:

$$\frac{\partial L}{\partial p_{m_n} n} = \frac{|\mathbf{h}_{m_n n}|^2}{\sum_{d=m_n+1}^{M_n} |\mathbf{h}_{dn}|^2 p_{dn} + |\mathbf{h}_{m_n n}|^2 p_{m_n n} + \delta^2} - \mu_{m_n n} - \eta_n = 0 \tag{37}$$

Then, we can obtain M equations which are not linear. To address it, we transfer these equations into linear. (37) can be rewritten as:

$$\sum_{d=m_n+1}^{M_n} |\mathbf{h}_{dn}|^2 p_{dn} + |\mathbf{h}_{m_n n}|^2 p_{m_n n} + \delta^2 = \frac{|\mathbf{h}_{m_n n}|^2}{\mu_{m_n n} + \eta_n} \quad \forall n, \ \forall m_n \tag{38}$$

Let A denote the coefficient matrix of the above equations, and $Y = (p_1 n, \ldots, p_{M_N} n)$. Let $b = (\frac{|\mathbf{h}_{1n}|^2}{\mu_{1n}+\eta_n}, \ldots, \frac{|\mathbf{h}_{M_N n}|^2}{\mu_{M_N n}+\eta_n})^T$. The above equations can be rewritten as:

$$AY = b \tag{39}$$

Thus, we can obtain the solution of (39):

$$\widehat{\mathbf{p}}_2 = A^{-1} b \tag{40}$$

The iterative formula of $\mathbf{p}_2^v$ can be given as:

$$\mathbf{p}_2^{v+1} = \mathbf{p}_2^v + \theta(\widehat{\mathbf{p}}_2 - \mathbf{p}_2^v) \tag{41}$$

where θ is the stepsize.

For the transmit power $\mathbf{p}_1$, we adopt the maximum transmit power P_B, and all APs are averagely allocated the power. That is,

$$p_{m_n}^* = \frac{P_B}{M}, \quad \forall n, \ \forall m_n \tag{42}$$

Then, the Lagrangian multiplier λ can be removed.

Algorithm 1 The Power Allocation Optimization Using DC Programming

Initialization
set $v = 0$.
Initialize the power allocation variables $\mathbf{p}^0$, the Lagrangian multipliers $\overrightarrow{\mu}(0)$, and $\overrightarrow{\eta}(0)$.
Set the convergence threshold τ_{thr}.
Updating
while $\mathbf{p}^v - \mathbf{p}^{v-1} \succeq \tau_{thr}$ **do**
  $v = v + 1$;
  Compute the optimal power allocation of **P3** $\hat{\mathbf{p}}_2$ according to (40).
  Updating power allocation $\mathbf{p}_2^v$ according to (41);
  Solve the dual problem (43) and updating Lagrangian multiplier $\mu_{m_n n}, \forall n, \forall m_n$ and $\eta_n, \forall n$ according to (46) and (47) respectively.
end while

A. OPTIMIZING THE DUAL PROBLEM

The dual optimization problem of problem **P3** is given as:

$$\begin{aligned} &\min g(\overrightarrow{\mu}, \overrightarrow{\eta}) \\ &s.t. \ \overrightarrow{\mu} \succeq 0, \quad \overrightarrow{\eta} \succeq 0 \end{aligned} \tag{43}$$

The dual problem is a convex optimization problem, and the sub-gradient method can be used to optimize it.

The sub-gradient at the point $\overrightarrow{\mu}$, $\overrightarrow{\eta}$ can be given:

$$\Delta\mu_{m_n n} = P_{ap} - \sum_{n=1}^{N}\sum_{m_n=1}^{M_n} p_{m_n n}, \quad \forall n, \ \forall m_n \tag{44}$$

$$\Delta\eta_n = \overline{R}_{b,n}^*(\mathbf{p}_1^*) - \overline{R}_{a,n}^*(\mathbf{p}_2^*), \quad \forall n \tag{45}$$

The Lagrangian multipliers can be updated by the following formulas:

$$\mu_{m_n n}(i+1) = \mu_{m_n n}(i) - \upsilon_{m_n n}\Delta\mu_{m_n n} \tag{46}$$

$$\eta_n(i+1) = \eta_n(i) - \iota_n \Delta\eta_n \tag{47}$$

where $\upsilon_{m_n n}$ $(\forall n, \forall m_n)$ and ι_n are the iteration stepsizes corresponding to $\mu_{m_n n}$, $\forall n, \forall m_n$ and η_n, $\forall n$, respectively. The details of the proposed power allocation scheme for the user-centric MIMO-NOMA IoT networks is shown in Algorithm 1.

Theorem 1: The solution obtained by the DC programming algorithm is one of the local optimal solutions of the original problem.

Proof: Suppose $\widehat{\mathbf{p}}_2(\mathbf{p}_2^v)$ is the optimal solution of problem **P**3. There is, $\nabla L(\widehat{\mathbf{p}}_2(\mathbf{p}_2^v)) = \mathbf{0}$. As proved in [20]

$$\lim_{v\to\infty} \widehat{\mathbf{p}}_2(\mathbf{p}_2^v) = \lim_{v\to\infty} \mathbf{p}_2^v \tag{48}$$

In accordance with the property of the Taylor expansion, the gradient of the Taylor expansion at the expansion point is equal to that of the original problem at the expansion point. Therefore, while $v \to \infty$, the gradient of the original problem at the point $\widehat{\mathbf{p}}_2(\mathbf{p}_2^v)$ is equal to $\nabla L(\widehat{\mathbf{p}}_2(\mathbf{p}_2^v)) = \mathbf{0}$. As a result, $\lim_{v\to\infty} \widehat{\mathbf{p}}_2(\mathbf{p}_2^v)$ is a local optimal solution of the original problem.

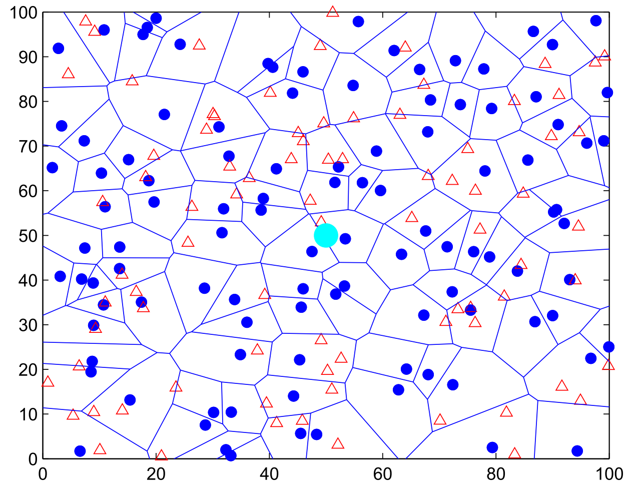

FIGURE 2. The APs and devices are modeled as independent PPP. The green solid point, blue solid point and the red triangle represent the MBS, APs and devices, respectively.

VI. SIMULATED RESULTS

In this section, we intend to evaluate the performance of the proposed resource allocation scheme for the user-centric MIMO-NOMA IoT. The parameters of the user-centric MIMO-NOMA IoT are set as follows unless otherwise specified. It is assumed that the coverage region of the BS is a square with area of 0.1 Km × 0.1 Km, and the MBS with 3 antennas is deployed in the center of the cell. The APs and users are modeled as independent Poisson points process (PPP) such as Fig.2, respectively. The large scale path loss is modeled according to the WINNER model [21], and the small scale fading is modeled as Rayleigh fading. The transmit power of the MBS and APs are set as 46 dBm and 30 dBm respectively. The system bandwidth is set to 20 MHz, and the spectrum density of the noise is set to −174 dBm/Hz.

In order to illustrate the effectiveness of the proposed resource allocation scheme, we compare it with a few of benchmarks as follows: 1) the traditional cell-centric access scheme with NOMA, in which each device can be served by at most one AP in the access link, and the NOMA is applied in each cell; 2) the user-centric access scheme with OMA, in which the NOMA will not be applied in each APG both for the backhaul downlink and access downlink; 3) the traditional cell-centric access scheme with OMA, in which each device can be served by at most one AP in the access link, and the NOMA is not applied in each cell.

Fig.3 compares the system throughput of the MIMO system with that of the single-antenna system. From Fig.3, the system throughput increases with the increase of the number of APs for both the MIMO system and single-antenna system. Furthermore, the MIMO system achieves higher system throughput than the single-antenna system for the same number of the APs, and this result shows that the MIMO

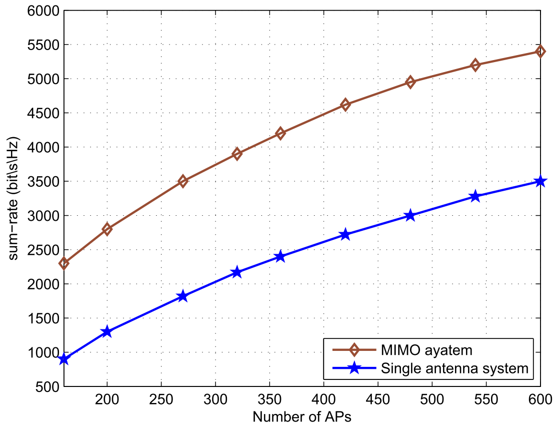

FIGURE 3. **Comparison of the system throughput using MIMO system with that of the single-antenna system.**

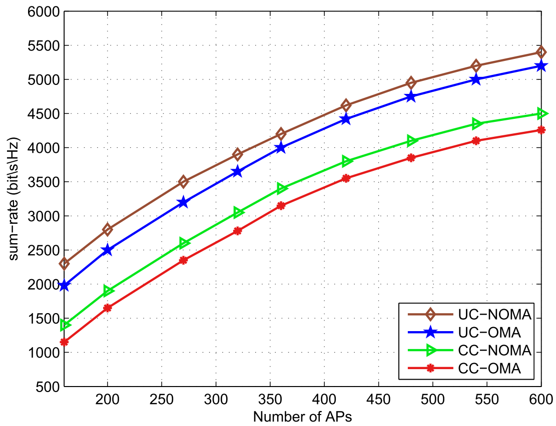

FIGURE 4. **System throughput vs the number of the APs.**

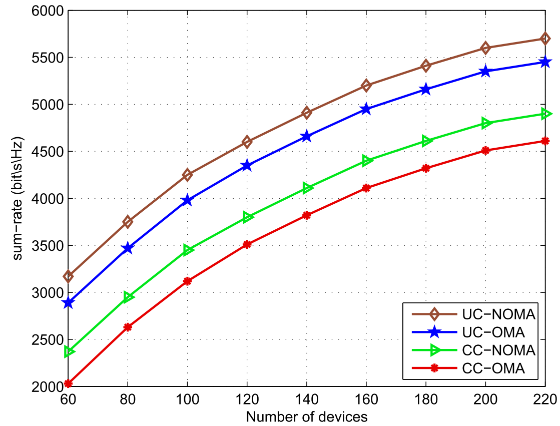

FIGURE 5. **System throughput vs the number of the devices.**

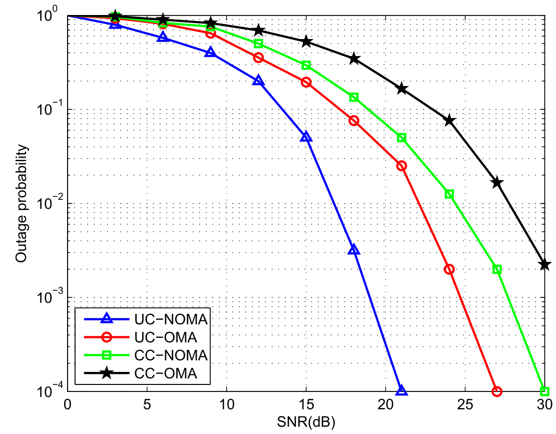

FIGURE 6. **Outage of the probability vs the number of the SNR.**

system can significantly improve the spectrum efficiency. On the other hand, Fig.Fig.3 illustrates that the combination of the MIMO and NOMA has great advantages. Indeed, grouping the APs can improve the performance of the beamforming since the number of the beamforming vectors is reduced, and then the NOMA is applied in each APG in order to decrease the intra-APG interference caused by the beamforming vector sharing.

Fig.4 shows the system throughput with the increase of the number of the APs. It can be observed from Fig.4 that the system throughput increases with the increase of the number of the APs. Furthermore, the proposed resource allocation scheme outperforms the other benchmarks in terms of the system throughput. In addition, it can also be seen that the user-centric access scheme with OMA has higher throughput than the other two benchmarks, and this result depicts the effectiveness of the user-centric access scheme. Fig.5 presents the system throughput with the increase of the number of devices, and it shows that the system throughput increases with the increase of the number of the devises due to the multi-user diversity. Fig.5 also shows similar curve trends with Fig.4 which illustrates the effectiveness of the proposed resource allocation scheme for the user-centric MIMO-NOMA IoT.

Fig.6 and Fig.7 depict the outage probability with different number of the APs and devices. The results of Fig.6 and Fig.7 show that the proposed resource allocation scheme for the user-centric MIMO-NOMA IoT has lower outage probability than the other benchmarks. The fact can be explained that the diversity gain can be achieved since multiple APs can serve the same devise simultaneously. On the other hand, it can also be seen from Fig.6 and Fig.7 that NOMA system can significantly decrease the outage probability comparing with the traditional OMA access scheme. Indeed, each beamforming vector is allocated to multiple APs simultaneously in order to improve the efficient. However, it will cause serious intra-APG co-channel interference. NOMA can reduce the intra-APG interference by removing the interference from the APs with better channel condition through utilizing the SIC. Furthermore, it can be observed that the curves decrease slow with the increase of the number of the APs. This result can be explained as that multiple APs serving the same device can improve the SNR at the side of the device. However, when the number of the APs is large enough, the performance of the NOMA will be reduced.

In the following, we intend to evaluate the convergence performance of the proposed resource allocation scheme for the user-centric MIMO-NOMA IoT, and the results is shown

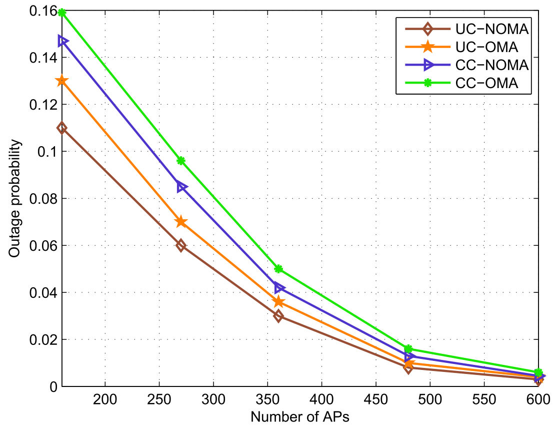

FIGURE 7. Outage of the probability vs the number of the APs.

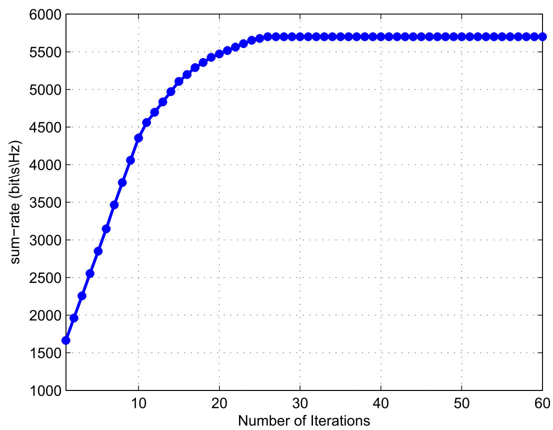

FIGURE 8. System throughput vs the number of iterative index.

in Fig.8. From Fig.8, it can be observed that the proposed resource allocation scheme has good convergence, and the algorithm converges with 25 iterations.

VII. CONCLUSION

In this paper, we have investigated the user-centric access framework for the IoT to enhance the system performance, in which each device are served by multiple APs simultaneously. In order to further improve the spectrum efficiency, the MIMO and NOMA are integrated due to the complementarity of these two issues. Then, the resource allocation involving the beamforming strategy optimization and the power allocation for the user-centric MIMO-NOMA IoT has been investigated, and we have formulated the resource allocation problem as a nonconvex optimization problem which is extremely difficult to tackle. To reduce the computational complexity, we decompose the original optimization problem into two optimization subproblems in terms of the beamforming stragegy optimization and power allocation. For the beamforming strategy optimization subproblem, a novel ZFBF algorithm is applied to solve it. The power allocation is still a nonconvex optimization problem due to the intra-APG interference, and the problem has been transferred as a DC programming problem. The DC programming approach has been adopted to optimize the power allocation problem. We have further proved that the solution obtained by the DC programming method is one of the local optimal solutions of the original optimization problem.

REFERENCES

[1] *Cisco Visual Networking Index: Forecast and Trends, 2017-2022*, Cisco, San Jose, CA, USA, 2019.

[2] J. Ding and Jun Cai, "Two-side coalitional matching approach for joint MIMO-NOMA clustering and bs selection in multi-cell MIMO-NOMA systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 2006–2021, Dec. 2019.

[3] P. Huang, H. Kao, and W. Liao, "Cross-tier cooperation for optimal resource utilization in ultra-dense heterogeneous networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 12, pp. 11193–11207, Dec. 2017.

[4] Y. Liu, X. Li, H. Ji, and H. Zhang, "A multi-user access scheme for throughput enhancement in UDN with NOMA," in *Proc. IEEE Int. Conf. Commun. Workshops*, May 2017, pp. 1364–1369.

[5] H. Li, X. Xu, D. Hu, X. Tao, P. Zhang, S. Ci, and H. Tang, "Clustering strategy based on graph method and power control for frequency resource management in femtocell and macrocell overlaid system," *J. Commun. Netw.*, vol. 13, no. 6, pp. 664–677, 2011.

[6] Q. Junfei, S. Youming, W. Ducheng, W. Qihui, and X. Yuhua, "Demand-aware resource allocation for ultra-dense small cell networks: An interference-separation clustering-based solution," *Eur. Trans. Telecommun.*, vol. 27, no. 8, pp. 1071–1086, 2016.

[7] S. J. Kim, I. Cho, B. Lee, S. H. Bae, and C. H. Cho, "Multi-cluster based dynamic channel assignment for dense femtocell networks," *Ksii Trans. Internet Inf. Syst.*, vol. 10, no. 4, pp. 1535–1554, 2016.

[8] A. Shahini, "AnsariNOMA aided narrowband IoT for machine type communications with user clustering," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 7183–7191, May 2019.

[9] A. E. Mostafa, Y. Zhou, and V. W. S. Wong, "Connection density maximization of narrowband IoT systems with NOMA," *IEEE Trans. Wireless Commun.*, vol. 18, no. 10, pp. 4708–4722, Jul. 2019.

[10] T. Hou, Y. Liu, Z. Song, X. Sun, and Y. Chen, "NOMA-enhanced terrestrial and aerial IoT networks with partial CSI," *IEEE Internet Things J.*, vol. 7, no. 4, pp. 3254–3266, Jan. 2020.

[11] A. Rauniyar, P. E. Engelstad, and O. N. ØsterbØ, "Performance analysis of RF energy harvesting and information transmission based on NOMA with interfering signal for IoT relay systems," *IEEE Sensors J.*, vol. 19, no. 17, pp. 7668–7682, May 2019.

[12] T. Zhou, "Joint device association and power coordination for H2H and IoT communications in massive MIMO enabled HCNs," *IEEE Access*, vol. 8, pp. 72971–72984, 2020.

[13] B. M. Lee, "Improved energy efficiency of massive MIMO-OFDM in battery-limited IoT networks," *IEEE Access*, vol. 6, pp. 38147–38160, 2018.

[14] J. Yuan, Q. He, M. Matthaiou, T. Q. S. Quek, and S. Jin, "'Toward massive connectivity for IoT in mixed-ADC distributed massive MIMO," *IEEE Internet Things J.*, vol. 7, no. 3, pp. 1841–1856, Dec. 2020.

[15] Z. Shi, "Zero-forcing-based downlink virtual MIMO–NOMA communications in IoT networks," *IEEE Internet Things J.*, vol. 7, no. 4, pp. 2716–2737, Dec. 2020.

[16] Z. Ding, L. Dai, and H. V. Poor, "MIMO-NOMA design for small packet transmission in the Internet of Things," *IEEE Access*, vol. 4, pp. 1393–1405, 2016.

[17] A. Sousa de Sena, D. B. da Costa, Z. Ding, P. H. J. Nardelli, U. S. Dias, and C. B. Papadias, "Massive MIMO-NOMA networks with successive sub-array activation," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1622–1635, Dec. 2020.

[18] A. S. de Sena, "Massive MIMO-NOMA networks with imperfect SIC: Design and fairness enhancement," *IEEE Trans. Wireless Commun.*, vol. 19, no. 9, pp. 6100–6115, Jun. 2020.

[19] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Jan. 2014.

[20] H. A. L. Thi, V. N. Huynh, and P. D. Tao, "DC programming and DCA for general DC programs," in *Proc. Adv. Intell. Syst. Comput.*, 2014, vol. 282, pp. 15–35.

[21] W. I. D1.1.2. (Sep. 2007). *Winner II Channel Models*. [Online]. Available: https://www.istwinner.org/deliverables.html

QIANG WANG received the B.S. degree from Qufu Normal University, China, in 2010, and the M.Sc. and Ph.D. degrees from the Guangdong University of Technology, China. He is currently working with the Key Laboratory of Complex System Optimization and Big Data Processing, Yulin Normal University, Yulin, China. His research interests include wireless communications, signal processing, and evolutionary algorithm.

ZHAO WU (Member, IEEE) was born in Guangxi, China, in 1987. He received the B.E. degree in electronic and information engineering and the Ph.D. degree in electromagnetic fields and microwave technology from Xidian University, Xi'an, China, in 2011 and 2016, respectively. From October 2016 to March 2017, he was with Huawei Technologies Company Ltd. Since April 2017, he has been working with the School of Physics and Telecommunication Engineering, Yulin Normal University, as a Lecturer. His research interests include metamaterials, novel antennas, and reconfigurable antennas design and applications.