



HAL
open science

Zero-forcing Oriented Power Minimization for Multi-cell MISO-NOMA Systems: A Joint User Grouping, Beamforming and Power Control Perspective

Yaru Fu, Mingshan Zhang, Lou Salaun, Chi Wan Sung, Chung Shue Chen

► To cite this version:

Yaru Fu, Mingshan Zhang, Lou Salaun, Chi Wan Sung, Chung Shue Chen. Zero-forcing Oriented Power Minimization for Multi-cell MISO-NOMA Systems: A Joint User Grouping, Beamforming and Power Control Perspective. IEEE Journal on Selected Areas in Communications, 2020. hal-02491316

HAL Id: hal-02491316

<https://inria.hal.science/hal-02491316v1>

Submitted on 26 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Zero-forcing Oriented Power Minimization for Multi-cell MISO-NOMA Systems: A Joint User Grouping, Beamforming and Power Control Perspective

Yaru Fu, Mingshan Zhang, Lou Salaün, Chi Wan Sung, and Chung Shue Chen

Abstract—Future wireless communication systems have been imposed high requirement on power efficiency for operator’s profitability as well as to alleviate information and communication technology (ICT) global carbon emission. To meet these challenges, the power consumption minimization problem for a generic multi-cell multiple input and single output non-orthogonal multiple access (MISO-NOMA) system is studied in this work. The associated joint user grouping, beamforming (BF) and power control problem is a mixed integer non-convex programming problem, which is tackled by an iterative distributed methodology. Towards this end, the near-optimal zero-forcing (ZF) BF is leveraged, wherein the semiorthogonal user selection (SUS) strategy is applied to select BF users. Based on these, the BF vectors and BF users are determined for each cell using only local information. Then, two distributed user grouping strategies are proposed. The first one, called channel condition based user clustering (CCUC), performs user grouping in each cell based on the channel conditions. This is conducted independently of the power control part and has low computational complexity. Another algorithm, called power consumption based user clustering (PCUC), uses both the channel conditions and inter-cell interference information to minimize each cell’s power consumption. In contrary to CCUC, PCUC is optimized jointly with the power control. Finally, with the obtained user grouping and BF vectors, the resultant power allocation problem is optimally solved via an iterative algorithm, whose convergence is mathematically proven given that the problem is feasible. We perform Monte-Carlo simulation and numerical results show that the proposed resource management methods outperform various conventional MISO schemes and the non-clustered MISO-NOMA strategy in several aspects, including power consumption, outage probability, energy efficiency, and connectivity efficiency.

Index Terms—Beamforming, connectivity efficiency, energy efficiency, inter-cell interference, multiple input single output (MISO), non-orthogonal multiple access (NOMA), optimal power control, user clustering.

This work was partially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU 11216416). A part of the work was carried out at LINC3 (www.lincs.fr). Yaru Fu and Mingshan Zhang contributed equally to this work. (Corresponding author: Chung Shue Chen)

C. S. Chen, L. Salaün, and M. Zhang are with Bell Labs, Nokia Paris-Saclay, 91620 Nozay, France (e-mail: chung_shue.chen@nokia-bell-labs.com, lou.salaun@nokia-bell-labs.com, mshzhang@bjtu.edu.cn). M. Zhang is now with Beijing Jiaotong University.

Y. Fu and C. W. Sung are with the Dept. of Electrical Engineering, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong SAR (e-mail: albert.sung@cityu.edu.hk, yaru_fu@sutd.edu.sg). Y. Fu is now with Singapore University of Technology and Design.

I. INTRODUCTION

The dramatically increased demand of data traffic as well as the explosive growth of innovative applications, e.g., intelligent transportation system, intelligent computing, and internet-of-things (IoT), have imposed great challenges for future wireless cellular networks to provide massive connectivity and satisfied user quality-of-service (QoS). Non-orthogonal multiple access (NOMA) has been included by the third generation partnership project long-term evolution advanced (3GPP-LTE-A) and been recognized as a promising radio access solution for addressing the massive access challenge [1] and for its superior spectrum efficiency compared to conventional orthogonal multiple access (OMA) strategies [2]. Specifically, in NOMA scheme, multiple users with diverse power levels can be accommodated within the same resource block simultaneously with the aid of advanced transmission scheme, i.e., superposition coding, and progressive reception technique, that is successive interference cancellation (SIC).

Therefore, transmit power of the multiplexed users should be assigned properly as it is crucial for successful interference elimination and signal decoding. The power allocation for uplink and downlink single-carrier NOMA systems are investigated in [3] and [4], respectively. The joint subcarrier and power allocation problem for multi-carrier NOMA is analyzed in [5]–[7], and its NP-hardness is shown in [8]. Additionally, there are some other works that study power allocation problems for NOMA enabled multi-access edge computing (MEC) networks [9], NOMA assisted unmanned aerial vehicle (UAV) systems [10], NOMA in IoT systems [11] and NOMA based wireless caching networks [12]–[14].

In the meantime, multiple-input and multiple-output (MIMO) is another promising technique to address 5G and beyond (B5G) challenges. In regard to the resource management for multi-antenna systems, we explore two works that are worth mentioning. In [15], the energy-efficient beamforming (BF) for a single-cell multiple-input and single-output (MISO) network is investigated. Thereof, the authors propose a branch-and-reduce-and-bound approach to attain the optimal solution. In order to decrease the computational complexity of the optimal algorithm, two time-efficient approximate methods are designed. One adopts zero-forcing BF (ZF-BF), with which the maximization problem recasts to be a concave-convex fractional programming. The other one is bases on

the sequential convex approximation algorithm to find a stationary point. In [16], the BF and power control problem is studied for MISO broadcast channel with the consideration of various linear constraints. Numerically efficient algorithms are provided to solve the general weighted sum rate problem under different BF policies, i.e., dirty paper coding (DPC) and ZF. Nevertheless, the designed algorithms cannot be applied to MIMO-NOMA networks directly due to the distinct principle of NOMA when compared to that of the conventional OMA based strategies. As far as MIMO-NOMA, it was shown by [17] that the application of MIMO in NOMA is able to achieve an enhanced system performance compared to pure NOMA and pure MIMO. The notion of MIMO-NOMA was first introduced in [18], which has attracted much attention [19]–[29] since then. Specifically, in [19], the ergodic capacity maximization problem for a single-cell MIMO-NOMA system is studied, wherein only statistical channel state information is available at the transmitter side. In [20], the authors investigate a NOMA beamforming system, in which each BF vector is capable of accommodating multiple users. The optimal precoding scheme for a two-user MISO-NOMA system is obtained in [21] via Newton’s iterative algorithm to minimize total power. A joint user scheduling and power allocation algorithm is designed to optimally solve the sum rate maximization problem for millimeter wave multiuser MISO-NOMA system in [22]. To reduce the computational complexity, a sub-optimal strategy is also developed. In addition, [23] investigates the user clustering, BF and power allocation problem for sum rate maximization of a single-cell MIMO-NOMA system, in which the number of antennas per user is more than that of the base station. The power minimization oriented precoding for a single-cell uplink MIMO-NOMA is studied in [24] with the consideration of jamming attacks, wherein the authors design the precoders at the transmitters and equalizers at the receiver jointly.

For multi-cell scenarios, [25] studies the sum throughput maximization problem for multi-cell MIMO-NOMA systems subject to the users’ QoS requirements, wherein only two users are superimposed using NOMA scheme. In [26], several interference alignment based coordinated BF strategies are developed for a two-cell MIMO-NOMA network to enhance the cell-edge users’ throughput and system fairness. Besides, [27] maximizes the weighted sum rate of the strongest users in a coordinated two-cell MIMO-NOMA system, where an efficient majorization-minimization oriented iterative algorithm is designed. A joint design of BF and power allocation for multi-cell MIMO-NOMA networks is proposed in [28], in which the base stations (BSs) adopt coordinated multipoint transmission in the downlink. A recent work [29] investigates the interference alignment based rate maximization problem for a three-cell MIMO-NOMA network, where each cell only has two subscribers. Therein, the non-convex optimization problem is solved by a successive convex approximation method. The above discussed multi-cell works mainly focus on rate maximization [25]–[29]. To the best of our knowledge, power minimization problem for generic multi-cell MISO-NOMA systems has not been addressed yet, which is also an important research direction for NOMA since future green

communication systems target at achieving a 10-fold enhancement in power efficiency.

The aforementioned discussion motivates us to investigate the power minimization problem for the downlink of a general multi-cell MISO-NOMA system with arbitrary number of cells and users, in which each cluster can multiplex arbitrary number of subscribers. We summarize the main contributions of our work as follows:

- 1) We first explain the impact of user grouping, BF and power control on the power cost of a generic multi-cell MISO-NOMA network. Then, the total power minimization problem is formulated taking into account each user’s data rate requirement and a maximum number of multiplexed users per cluster constraint. The formulated problem is a mixed integer non-convex programming problem, which is in general intractable to be solved directly.
- 2) To handle this challenging power minimization problem, an efficient distributed iterative resource allocation methodology is proposed. First, the BF vectors are computed by ZF-BF, and semiorthogonal user selection (SUS) algorithm is applied to select the BF users in each cell. At this point, each beam only serves one BF user. Each remaining user is then grouped with a BF user, such that the BS serves them on the same beam through signal superposition. Users in the same group (i.e., cluster) performs SIC to decode the superposed signal. We propose two user grouping strategies, namely CCUC and PCUC. Finally, with the obtained user grouping and BF vectors, the power control problem is solved iteratively. The proposed schemes (BF, user grouping and power control) are all distributed in the sense that they only require local information available at each cell, i.e., the link gains between each user and their serving BS, the inter-cell interference estimated at each user’s equipment.
- 3) The proposed user grouping strategies are metric-oriented, namely channel conditions based user clustering (CCUC) and power consumption based user clustering (PCUC). CCUC performs user grouping in each cell based on the channel conditions. This is conducted independently of the power control part and has low computational complexity. In contrast, PCUC uses both the channel conditions and inter-cell interference information to minimize each cell’s power consumption in a distributed manner. PCUC is based on the linear-sum assignment problem and is optimized jointly with the power control.
- 4) Given the user clustering and BF results, the original optimization problem can be transformed into a non-convex power control problem, denoted by \mathbb{P} . The feasibility of problem \mathbb{P} is characterized by an associated problem defined as \mathbb{Q} . Through rigorous analysis, we show that the optimal values of the two problems are equivalent and there exists a one-to-one mapping between their optimal solutions. Fortunately, problem \mathbb{Q} falls into the paradigm of classical power control theory and can be optimally solved by an iterative algorithm with provable

convergence property. Such an optimal iterative power control scheme as been proposed in our previous work [4] for single-input single output (SISO) NOMA systems. In this work, we will go further by generalizing it to MISO-NOMA systems, as well as developing user grouping and beamforming on top of the iterative power control.

- 5) Monte-Carlo simulation results show that our proposed joint user grouping, BF and power allocation algorithms outperform various conventional MISO schemes and the non-clustered MISO-NOMA strategy in several aspects, including outage rate, power consumption, energy efficiency and connectivity efficiency, suitable for future systems.

Note that a preliminary result of this work has been published in [30], in which power minimization problem for a specific two-cell MISO-NOMA system under distributed power control is studied. Here, we consider the more general problem with arbitrary number of cells and users, wherein each cluster can multiplex arbitrary number of subscribers. Note that cluster based multi-antenna NOMA is considered for supporting massive access in B5G systems [1], [17].

The rest of this paper is organized as follows. We describe the multi-cell MISO-NOMA system model and formulate the power minimization problem in Section II. Section III presents the proposed user grouping and BF methods. With the obtained user grouping and BF vectors, the original optimization problem is transformed into a power minimization problem, which is denoted as \mathbb{P} . In Section IV, we show the feasibility as well as the optimality of \mathbb{P} analytically. Based on which, the optimal power allocation algorithm is designed, whose convergence is proved as well. The performance of the proposed resource management schemes are evaluated by extensive computer simulations and are compared to various benchmarks in Section V. Finally, we conclude the paper in Section VI.

II. SYSTEM DESCRIPTION AND PROBLEM FORMULATION

In this section, we first describe the multi-cell MISO-NOMA system model and then the problem formulation.

A. System Description

The downlink of a MISO-NOMA system that consists of M cells is considered. Denote by $\mathcal{M} \triangleq \{1, 2, \dots, M\}$ the index set of all cells. In each cell $m \in \mathcal{M}$, there is one base station (BS) serving a number of K_m users. Each BS is equipped with an antenna array of N elements for the downlink transmission and each user is equipped with a single antenna. Define $\mathcal{K}_m \triangleq \{1, 2, \dots, K_m\}$ as the set of the indices of all subscribers who are associated with BS m . Without loss of generality, we assume that K_m is independent from (i.e., can be larger or smaller than) N for all $m \in \mathcal{M}$. For $m \in \mathcal{M}$, $m' \in \mathcal{M}$ and $i \in \mathcal{K}_m$, define $\mathbf{g}_{i,m,m'} \in \mathbb{C}^{1 \times N}$ as the channel gain vector between BS m' and user i who is attached to BS m . We focus on the scenario where each BS $m \in \mathcal{M}$ transmits $|\mathcal{J}_m|$ beams to provide multiuser downlink transmission. Therein, each beam $j \in \mathcal{J}_m$ serves one cluster with $L_{j,m}$ subscribers performing NOMA pattern, in which

\mathcal{J}_m indicates the index set of all the clusters in cell m . Let $\mathbf{w}_{j,m} \in \mathbb{C}^{N \times 1}$ be the corresponding beamforming (BF) vector of cluster j in cell m , where $m \in \mathcal{M}$ and $j \in \mathcal{J}_m$. Besides, denote by $\mathcal{U}_{j,m}$ the index set of users who are assigned to the j -th cluster by BS m . In addition, let $p_{i,m}$ be the allocated power to user i by BS m . Furthermore, we define $\mathbf{p}_{j,m} \triangleq (p_{i,m})_{i \in \mathcal{U}_{j,m}}$ as the power vector of users in the j -th cluster of cell m and let $\mathbf{p}_m \triangleq (\mathbf{p}_{1,m}, \mathbf{p}_{2,m}, \dots, \mathbf{p}_{|\mathcal{J}_m|,m})$ be the power allocation strategy of all the users in cell m .

We use $\|\mathbf{z}\|_1$ to represent the l_1 -norm of any given vector \mathbf{z} , which equals to the sum of all the components in \mathbf{z} . For $m \in \mathcal{M}$ and $j \in \mathcal{J}_m$, let $q_{j,m} \triangleq \|\mathbf{p}_{j,m}\|_1$ be the transmit power summation of all users in j -th cluster of cell m . In addition, denote by $\mathbf{q} \triangleq (q_{1,1}, q_{2,1}, \dots, q_{|\mathcal{J}_M|,M})$ the power vector of all the clusters. Furthermore, we define

$$\mathbf{q}_{-j,m} \triangleq (q_{1,1}, q_{2,1}, \dots, q_{j-1,m}, q_{j+1,m}, \dots, q_{|\mathcal{J}_M|,M}), \quad (1)$$

where $m \in \mathcal{M}$ and $j \in \mathcal{J}_m$.

In our MISO-NOMA system, the messages of users that belong to the same cluster are transmitted via superposition coding. For expression simplicity, we assume user i is assigned to the j -th cluster of cell m . In addition, we assume frequency-flat block fading channel. With aforementioned definitions, the received signal at user i in cell m , $y_{i,m}$, can be expressed as follows:

$$\begin{aligned} y_{i,m} &= \mathbf{g}_{i,m,m} \mathbf{w}_{j,m} x_{j,m} + \mathbf{g}_{i,m,m} \sum_{j' \in \mathcal{J}_m, j' \neq j} \mathbf{w}_{j',m} x_{j',m} \\ &+ \sum_{m'=1, m' \neq m}^M \sum_{l \in \mathcal{J}_{m'}} \mathbf{g}_{i,m,m'} \mathbf{w}_{l,m'} x_{l,m'} + n_{i,m}, \end{aligned} \quad (2)$$

where $n_{i,m} \sim \mathcal{CN}(0, \sigma_{i,m}^2)$ indicates the received noise signal at user i in cell m , $x_{j,m}$ represents the superposed signal of users in cluster j of cell m . Note that $x_{j,m} \triangleq \sum_{i \in \mathcal{U}_{j,m}} \sqrt{p_{i,m}} s_{i,m}$, where $s_{i,m}$ is the desired message of user i . Thereof, $E[|s_{i,m}|^2] = 1$ is satisfied for $i \in \mathcal{K}_m$ and $m \in \mathcal{M}$.

Denote by $\hat{I}_{i,m}$ the inter-cluster interference plus noise value of user i who is assigned to cluster j in cell m , and thus it can be obtained due to (2) as follows:

$$\begin{aligned} \hat{I}_{i,m} &\triangleq \sum_{j' \in \mathcal{J}_m, j' \neq j} q_{j',m} |\mathbf{g}_{i,m,m} \mathbf{w}_{j',m}|^2 \\ &+ \sum_{m'=1, m' \neq m}^M \sum_{l \in \mathcal{J}_{m'}} q_{l,m'} |\mathbf{g}_{i,m,m'} \mathbf{w}_{l,m'}|^2 + \sigma_{i,m}^2, \end{aligned} \quad (3)$$

where the first and second items on the right side of (3) refer to the intra-cell interference and inter-cell interference at user i in cell m , respectively. For notation simplicity, we use $I_{i,m}$ to represent the normalized inter-cluster interference plus noise value of user i who is attached to BS m , i.e.,

$$I_{i,m} \triangleq \hat{I}_{i,m} / |\mathbf{g}_{i,m,m} \mathbf{w}_{j,m}|^2, \quad (4)$$

where $\hat{I}_{i,m}$ is given by (3).

Since SIC is adopted among users within each cluster, the decoding order of the multiplexed users should be considered. For $m \in \mathcal{M}$ and $j \in \mathcal{J}_m$, we adopt $\Pi_{j,m}$ to indicate the

set of all possible permutations of the elements in $\mathcal{U}_{j,m}$. For instance, if $\mathcal{U}_{j,m} = \{1, 2, 3\}$, then we have $\Pi_{j,m} = \{(1, 2, 3), (2, 1, 3), (1, 3, 2), (3, 1, 2), (2, 3, 1), (3, 2, 1)\}$. Furthermore, we define an auxiliary vector $\hat{\pi}_{j,m} = (\hat{\pi}_{j,m}(1), \hat{\pi}_{j,m}(2), \dots, \hat{\pi}_{j,m}(L_{j,m})) \in \Pi_{j,m}$ as the decoding order of users in the j -th cluster of cell m . Let $\hat{\pi}_{j,m}(l)$, where $l \in \{1, 2, \dots, L_{j,m}\}$, be its l -th element, which indicates that subscriber $\hat{\pi}_{j,m}(l)$ first decodes the messages of $\hat{\pi}_{j,m}(1)$ to $\hat{\pi}_{j,m}(l-1)$, subtracting these signals and finally decodes its own message by treating the signals of the remaining users in this cluster as noise. For multi-cell NOMA scenario, the communication quality of each user depends not only on its link gain but also on its suffered interference. Therefore, the decoding order of subscribers in cluster j of cell m should be determined by their normalized interference plus noise values [4], i.e., $I_{i,m}$, $i \in \mathcal{U}_{j,m}$. Since for $i \in \mathcal{U}_{j,m}$ the value of $I_{i,m}$ is determined by $\mathbf{q}_{-j,m}$, $\pi_{j,m}$ is characterized as a function of $\mathbf{q}_{-j,m}$ and it is expressed as:

$$\pi_{j,m}(\mathbf{q}_{-j,m}) = (\pi_{j,m}(1), \pi_{j,m}(2), \dots, \pi_{j,m}(L_{j,m})) \in \Pi_{j,m}, \quad (5)$$

so that the following conditions are met:

- 1) The normalized interference plus noise values, i.e., $I_{\pi_{j,m}(l),m}$, $l \in \{1, 2, \dots, L_{j,m}\}$, are permuted in descending order as $I_{\pi_{j,m}(1),m} \geq I_{\pi_{j,m}(2),m} \geq \dots \geq I_{\pi_{j,m}(L_{j,m}),m}$.
- 2) When there is a tie, i.e., $I_{\pi_{j,m}(l),m} = I_{\pi_{j,m}(l'),m}$ for $l < l'$, then $\pi_{j,m}(l) < \pi_{j,m}(l')$.

For $m \in \mathcal{M}$, denote by $\mathbf{w}_m \triangleq (\mathbf{w}_{1,m}, \mathbf{w}_{2,m}, \dots, \mathbf{w}_{|\mathcal{J}_m|,m})$ and $\mathbf{w} \triangleq (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_M)$ the BF strategies of cell m as well as the system, respectively. In addition, let $\mathbf{p} \triangleq (\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_M)$. For $m \in \mathcal{M}$ and $i \in \mathcal{K}_m$, we define $R_{i,m}(\mathbf{p}, \mathbf{w})$ as the transmission data rate of subscriber i from cell m . In addition, let $R'_{i,m}(\hat{\pi}_{j,m}, \mathbf{p}, \mathbf{w})$ be an auxiliary function given below:

$$R'_{i,m}(\hat{\pi}_{j,m}, \mathbf{p}, \mathbf{w}) \triangleq W \log_2 \left(1 + \frac{p_{i,m}}{\sum_{l'=\hat{\pi}_{j,m}^{-1}(i)+1}^{L_{j,m}} p_{\hat{\pi}_{j,m}(l'),m} + I_{i,m}} \right), \quad (6)$$

in which W is the system bandwidth, $\hat{\pi}_{j,m} \in \Pi_{j,m}$, and $\hat{\pi}_{j,m}^{-1}(i)$ represents the decoding order of user i in $\hat{\pi}_{j,m}$. Specifically, $\hat{\pi}_{j,m}^{-1}(i) = l$ if $\hat{\pi}_{j,m}(l) = i$. With aforementioned definitions, we have

$$R_{i,m}(\mathbf{p}, \mathbf{w}) = R'_{i,m}(\pi_{j,m}(\mathbf{q}_{-j,m}), \mathbf{p}, \mathbf{w}), \quad (7)$$

where $m \in \mathcal{M}$, $i \in \mathcal{K}_m$. Note that $\mathbf{q}_{-j,m}$ is given by (1), which is a function of \mathbf{p} .

B. Problem Formulation

In this article, we aim to minimize the total transmit power of the system while taking into account each user's data rate requirement and a maximum number of multiplexed users per cluster. For $m \in \mathcal{M}$ and $i \in \mathcal{K}_m$, denote by $r_{i,m}$ the required minimum transmission rate of user i who is attached to BS

m . The power minimization problem is defined below:

$$\min \sum_{m \in \mathcal{M}} \sum_{i \in \mathcal{K}_m} p_{i,m} \quad (8)$$

subject to

$$\begin{aligned} C1: & R_{i,m}(\mathbf{p}, \mathbf{w}) \geq r_{i,m}, \quad m \in \mathcal{M}, i \in \mathcal{K}_m, \\ C2: & L_{j,m} \leq L, \quad m \in \mathcal{M}, j \in \mathcal{J}_m, \\ C3: & \mathcal{U}_{j,m} \cap \mathcal{U}_{j',m} = \emptyset, \quad m \in \mathcal{M}, j \in \mathcal{J}_m, j' \in \mathcal{J}_m, j \neq j', \\ C4: & p_{i,m} \geq 0, \quad m \in \mathcal{M}, i \in \mathcal{K}_m. \end{aligned} \quad (9)$$

Note that $C1$ indicates the minimum data rate requirement of each user, $C2$ restricts the number of users served in cluster j of cell m to be no greater than L , where L is a parameter which jointly considers the practical limitations of receiver design complexity and the error propagation due to SIC, $C3$ is used to make sure that each user is scheduled by a single beam at most, and $C4$ refers to the non-negativity of allocated power.

The minimization problem (8) is a mixed integer non-linear and non-convex programming problem, which is in general intractable to solve directly [22]. Therefore, in the following sections, a two-fold methodology is designated. Specifically, in step 1, the user clustering and BF are determined. Based on the obtained results, an optimal power allocation algorithm is developed in step 2 to solve the resultant power allocation problem.

III. BEAMFORMING AND USER CLUSTERING ALGORITHMS

In this section, we first introduce the proposed BF strategy. Based on which, two metric-oriented user clustering schemes are designed. One is based on the channel condition and the other rests with the metric of power cost.

A. Beamforming Design

It is known that dirty paper coding (DPC) is the optimal method to achieve the capacity region of multi-user MIMO broadcast channel systems. However, DPC is difficult to use in practical system as it employs brute-force searching [20]. Instead, zero-forcing BF (ZF-BF) is a near-optimal and easy-to-implement substitute [31]. Therefore, in this article, we apply ZF as the BF policy. In our NOMA system, the BF vectors of each cell are generated by some of its associated users' channel vectors. We call these users *BF users* and their multiplexed users in each beam are named as *matched users*. It is assumed that the channel state information of all users are perfectly known by the BSs.

For $m \in \mathcal{M}$, denote by $\mathcal{T}_{z,m}$ the index sets of the selected BF users in cell m . In addition, let $\mathbf{H}(\mathcal{T}_{z,m}) \triangleq \{(\mathbf{g}_{i,m,m}^T)_{i \in \mathcal{T}_{z,m}}\}^T$, where $(\cdot)^T$ corresponds to the transpose of a matrix. Since the beam vectors of the ZF-oriented BF strategy are determined by the channel information of BF users, the corresponding ZF-BF vectors that generated by the link gains of users in $\mathcal{T}_{z,m}$ can be expressed as follows:

$$\mathbf{w}_m = \mathbf{H}^\dagger(\mathcal{T}_{z,m}) = \mathbf{H}^*(\mathcal{T}_{z,m})(\mathbf{H}(\mathcal{T}_{z,m})\mathbf{H}^*(\mathcal{T}_{z,m}))^{-1}, \quad (10)$$

where $m \in \mathcal{M}$, the operators $(\cdot)^\dagger$ and $(\cdot)^*$ represent the Moore Penrose pseudoinverse and the conjugate transpose of a matrix, respectively.

In this work, semiorthogonal user selection (SUS) algorithm is adopted to select the BF users, which has been proven to achieve the asymptotic optimal performance of DPC [31]. SUS is conducted in an iterative manner. Its main idea is to select a BF user from a constructed user group denoted by \mathcal{T}_{SUS} during each iteration, and the set \mathcal{T}_{SUS} is renewed at the end of each iteration such that the users in this group are semiorthogonal to the BF user set. The procedure of BF user selection continues until N users are selected or the user group \mathcal{T}_{SUS} is empty. It is worth noting that the number of clusters is less than the number of antennas if the algorithm terminates according to the second condition. The orthogonality among BF users is controlled by the SUS factor α , which is a small positive constant between 0 and 1. In this work, the values of α for different system settings are chosen empirically by simulations, which will be discussed later in Section V-A. The implementation of the above SUS scheme is detailed in Algorithm 1. With the output of Algorithm 1, the BF vectors for each cell can be computed by (10).

Algorithm 1 SUS algorithm for BF users selection for cell m

Input: $\mathbf{g}_{i,m,m}$ for $i \in \mathcal{K}_m$.

Output: $\mathcal{T}_{z,m}$.

- 1: Initialization: $\mathcal{T}_{z,m} = \emptyset$, $\mathcal{T}_{SUS} = \{1, \dots, K_m\}$ and $j = 1$.
 - 2: **while** $j \leq N$ and $\mathcal{T}_{SUS} \neq \emptyset$ **do**
 - 3: Search for a new BF user for cluster j of cell m .
 - 4: **for** $i \in \mathcal{T}_{SUS}$ **do**
 - 5: **if** $j == 1$ **then**
 - 6: $\mathbf{h}_i := \mathbf{g}_{i,m,m}$.
 - 7: **else**
 - 8: $\mathbf{h}_i := \mathbf{g}_{i,m,m} - \sum_{j'=1}^{j-1} \frac{\mathbf{g}_{i,m,m} \mathbf{h}_{(j')}^*}{\|\mathbf{h}_{(j')}\|^2} \mathbf{h}_{(j')}$, in which \mathbf{x}^* depicts the conjugate transpose of a vector \mathbf{x} .
 - 9: **end if**
 - 10: **end for**
 - 11: Denote $i^* := \arg \max_{i \in \mathcal{T}_{SUS}} \|\mathbf{h}_i\|$.
 - 12: Update $\mathbf{h}_{(j)} := \mathbf{g}_{i^*,m,m}$.
 - 13: Adjust $\mathcal{T}_{z,m} \leftarrow \mathcal{T}_{z,m} \cup \{i^*\}$.
 - 14: Update $\mathcal{T}_{SUS} \leftarrow \{i \in \mathcal{T}_{SUS}, i \neq i^*, \frac{|\mathbf{g}_{i,m,m} \mathbf{h}_{(j)}^*|}{\|\mathbf{g}_{i,m,m}\| \|\mathbf{h}_{(j)}\|} < \alpha\}$, where α is the SUS factor, which is a small positive constant.
 - 15: $j \leftarrow j + 1$.
 - 16: **end while**
-

B. User Clustering Methods

In this subsection, two user clustering approaches are designed with the prior information of BF vectors obtained in Section III-A.

1) *Channel Condition based User Clustering (CCUC)*: In this scheme, the clustering is determined by the link quality of each user. The interference information is not required, i.e., the user grouping decision as well as the power control optimization are mutually independent. CCUC algorithm is imple-

mented in an iterative way. At each iteration a matched user is assigned to a given BF user according to the channel quality oriented metric until all the users are clustered. Without loss of generality, we take cell m as an example. Before detailing the algorithm, we first give some definitions. Let $\mathcal{T}_{z,j,m}$ and $\mathcal{T}_{\bar{z},j,m}$ be the BF user and the matched users in the j -th cluster of cell m , respectively. Thereby, the index set of users in cluster j of cell m can be obtained as $\mathcal{U}_{j,m} = \mathcal{T}_{z,j,m} \cup \mathcal{T}_{\bar{z},j,m}$. Besides, we define $\mathcal{T}_{\bar{z},m} = \mathcal{T}_{\bar{z},1,m} \cup \dots \cup \mathcal{T}_{\bar{z},|\mathcal{J}_m|,m}$ as the index set of all the selected matched users. Moreover, let $\mathcal{T}_m = \mathcal{T}_{z,m} \cup \mathcal{T}_{\bar{z},m}$ be the index set of the already selected users. In the beginning of the algorithm, we assume the user index sets are empty, i.e., $\mathcal{T}_{z,m} = \emptyset$, $\mathcal{T}_{\bar{z},m} = \emptyset$. In each iteration, the cluster-matched user pair (j^*, i^*) with the highest correlation is chosen as stated, i.e., the matched user i^* is then grouped in cluster j^* . We repeat the aforementioned step until each of the subscribers is allocated into a cluster. The pseudo-code of the CCUC strategy in cell m is summarized in Algorithm 2.

Algorithm 2 The proposed CCUC algorithm for cell m

- 1: **Initialization:** $\mathcal{T}_{z,m} = \emptyset$, $\mathcal{T}_{\bar{z},m} = \emptyset$, $\mathcal{T}_m = \mathcal{T}_{z,m} \cup \mathcal{T}_{\bar{z},m}$, and $\mathcal{T}_{\bar{z},m} = \mathcal{T}_{\bar{z},1,m} \cup \dots \cup \mathcal{T}_{\bar{z},|\mathcal{J}_m|,m}$.
 - 2: Determine $\mathcal{T}_{z,m}$ via selecting BF users according to Algorithm 1.
 - 3: **while** $\mathcal{T}_m \neq \mathcal{K}_m$ **do**
 - 4: Match a new user to a cluster such that they have the highest correlation value, i.e.,
 - 5:
$$(j^*, i^*) \triangleq \underset{(j,i)_{j \in \mathcal{T}_{z,m} \setminus \{j' \mid |\mathcal{T}_{z,j',m}| = L-1, j' \in \mathcal{T}_{z,m}\}}, i \in \mathcal{K}_m \setminus \mathcal{T}_m}{\arg \max} \left(\frac{|\mathbf{g}_{i,m,m} \cdot \mathbf{g}_{\mathcal{T}_{\bar{z},j^*,m,m}}^*|}{\|\mathbf{g}_{i,m,m}\| \cdot \|\mathbf{g}_{\mathcal{T}_{\bar{z},j^*,m,m}}\|} \right)$$
 - 6: $\mathcal{T}_{\bar{z},j^*,m} \leftarrow \mathcal{T}_{\bar{z},j^*,m} \cup \{i^*\}$.
 - 7: **end while**
 - 8: Form the user indices set for each cluster in cell m , i.e.,
 - 9: **for** $j := 1, 2, \dots, |\mathcal{T}_{z,m}|$ **do**
 - 10: $\mathcal{U}_{j,m} \leftarrow \mathcal{T}_{z,j,m} \cup \mathcal{T}_{\bar{z},j,m}$.
 - 11: **end for**
-

2) *Power Consumption based User Clustering (PCUC)*: The main idea of PCUC algorithm is to assign users to different clusters such that the power consumption of the BS is minimized. Before expounding the details about PCUC, we first demonstrate the needed transmit power per cluster to meet the multiplexed users' data rate requirements under a predetermined decoding order. We take an arbitrary cluster j in cell m as an example. Suppose the decoding order $\hat{\pi}_{j,m}$ is given, and we aim to meet the minimum data rate constraint per subscriber in this cluster, which is expressed as follows:

$$R'_{i,m}(\hat{\pi}_{j,m}, \mathbf{p}) \geq r_{i,m}, i \in \mathcal{U}_{j,m}. \quad (11)$$

Let $\tilde{\mathbf{p}}_{j,m}(\hat{\pi}_{j,m}, \mathbf{q}_{-j,m})$ be the minimum power expenditure of cluster j in BS m , where the least data rate requirement per subscriber within this cluster is met. Therein, the prior information of the BF strategy \mathbf{w} , the decoding order as well as the transmit power of other clusters, $\mathbf{q}_{-j,m}$, are given. For

notation simplicity, we define

$$\gamma_{i,m} \triangleq 2^{\frac{r_{i,m}}{W}} - 1 \quad (12)$$

as the required SINR of subscriber i in cell m , where $i \in \mathcal{K}_m$ and $m \in \mathcal{M}$. Specifically, we have the following Lemma 1.

Lemma 1. For $m \in \mathcal{M}$, $j \in \mathcal{J}_m$, the minimum power consumption of cluster j in cell m under decoding order $\hat{\pi}_{j,m} \in \Pi_{j,m}$ is given as follows:

$$\tilde{p}_{j,m}(\hat{\pi}_{j,m}, \mathbf{q}_{-j,m}) = \sum_{l=1}^{L_{j,m}} \chi_{l,m}(\hat{\pi}_{j,m}) I_{\hat{\pi}_{j,m}(l),m}, \quad (13)$$

where

$$\chi_{l,m}(\hat{\pi}_{j,m}) \triangleq \left[\prod_{l' < l} (\gamma_{\hat{\pi}_{j,m}(l'),m} + 1) \right] \gamma_{\hat{\pi}_{j,m}(l),m}. \quad (14)$$

Proof. Obviously, the minimum sum power of all users in cluster j is obtained given that the data rate constraints in (11) achieves with equality. Firstly, we consider the power of users $p_{\hat{\pi}_{j,m}(L_{j,m}),m}$ and $p_{\hat{\pi}_{j,m}(L_{j,m}-1),m}$. According to (6), we have

$$p_{\hat{\pi}_{j,m}(L_{j,m}),m} = \gamma_{\hat{\pi}_{j,m}(L_{j,m}),m} I_{\hat{\pi}_{j,m}(L_{j,m}),m}, \quad (15)$$

$$p_{\hat{\pi}_{j,m}(L_{j,m}-1),m} = \gamma_{\hat{\pi}_{j,m}(L_{j,m}-1),m} \times (p_{\hat{\pi}_{j,m}(L_{j,m}),m} + I_{\hat{\pi}_{j,m}(L_{j,m}-1),m}). \quad (16)$$

Based on (15) and (16), we obtain that

$$\sum_{l=L_{j,m}-1}^{L_{j,m}} p_{\hat{\pi}_{j,m}(l),m} = \gamma_{\hat{\pi}_{j,m}(L_{j,m}-1),m} I_{\hat{\pi}_{j,m}(L_{j,m}-1),m} + (\gamma_{\hat{\pi}_{j,m}(L_{j,m}-1),m} + 1) \gamma_{\hat{\pi}_{j,m}(L_{j,m}),m} I_{\hat{\pi}_{j,m}(L_{j,m}),m}. \quad (17)$$

Then, we calculate the power of user $p_{\hat{\pi}_{j,m}(L_{j,m}-2),m}$, which is as follows:

$$p_{\hat{\pi}_{j,m}(L_{j,m}-2),m} = \gamma_{\hat{\pi}_{j,m}(L_{j,m}-2),m} \times (p_{\hat{\pi}_{j,m}(L_{j,m}-1),m} + p_{\hat{\pi}_{j,m}(L_{j,m}),m} + I_{\hat{\pi}_{j,m}(L_{j,m}-2),m}). \quad (18)$$

Based on (17) and (18), the required power of the last three users is obtained. We repeat the aforementioned steps, the summation of all the $L_{j,m}$ users' transmit power is achievable, i.e.,

$$\sum_{l=1}^{L_{j,m}} p_{\hat{\pi}_{j,m}(l),m} = \sum_{l=1}^{L_{j,m}} \left[\prod_{l' < l} (\gamma_{\hat{\pi}_{j,m}(l'),m} + 1) \right] \times \gamma_{\hat{\pi}_{j,m}(l),m} I_{\hat{\pi}_{j,m}(l),m}, \quad (19)$$

which completes the proof. \square

Now we go back to the PCUC algorithm, whose inputs are the link gains of each user in the cell, as well as the inter-cell interference estimated at each user equipment side. Since the inter-cell interference may vary in time due to the distributed iterative power control, PCUC is recomputed at the beginning of each power control iteration using the most up-to-date inter-cell interference information. To be more specific, at iteration t , the power consumptions of each user in cell

m under all the possible clustering policies are calculated. We represent the performance of each policy in a cost matrix $\mathbf{C}^m \in \mathbb{R}^{K_m \times L|\mathcal{J}_m|}$, whose $(i, L(j-1) + l)$ -th component, denoted by $\mathbf{C}_{i,L(j-1)+l}^m$, indicates the power cost of user i in terms of satisfying its data rate requirement, if it is assigned to the j -th cluster of cell m and is in the l -th decoding order. For simplicity, we assume that the data rate requirement of each user in the system is the same, and the corresponding SINR requirement is denoted by γ . In accordance with (19), the value of $\mathbf{C}_{i,L(j-1)+l}^m$ can be calculated as follows:

$$\mathbf{C}_{i,L(j-1)+l}^m = \gamma(\gamma + 1)^{(l-1)} I_{i,m}^{(t-1)}, \quad (20)$$

where $i \in \mathcal{K}_m$, $l = 1, 2, \dots, L$, and $j \in \mathcal{J}_m$. $I_{i,m}^{(t-1)}$ represents the normalized interference plus noise value of user i who is attached to BS m , and is assumed to be assigned to cluster j during the t -th iteration. It is worth noting that $I_{i,m}^{(t-1)}$ is computed based on the power control at the $(t-1)$ -th iteration.

Based on the aforementioned analysis, the user clustering problem can be transformed to the linear sum assignment (LSA) problem, wherein we aim to obtain a Boolean matrix \mathbf{X}^m to minimize the BS's power cost, i.e., $\sum_i \sum_{j,l} \mathbf{C}_{i,L(j-1)+l}^m \mathbf{X}_{i,L(j-1)+l}^m$, where \mathbf{X}^m indicates the clustering strategy and has the same dimensions as \mathbf{C}^m . In addition, $\mathbf{C}_{i,L(j-1)+l}^m$ represents the $(i, L(j-1) + l)$ -th entry of \mathbf{X}^m , whose value takes 1 if and only if user i is decoded in the l -th order among all the users in cluster j . \mathbf{X}^m is optimized such that the resultant clustering strategy induces the minimum power consumption. Hence, the clustering decision optimization can be formulated as the following LSA problem:

$$\min \sum_i \sum_{j,l} \mathbf{C}_{i,L(j-1)+l}^m \mathbf{X}_{i,L(j-1)+l}^m, \quad (21)$$

subjects to $\sum_{j,l} \mathbf{X}_{i,L(j-1)+l}^m = 1$ and $\sum_i \mathbf{X}_{i,L(j-1)+l}^m \leq 1$. These two constraints guarantee that the solution of Boolean matrix \mathbf{X}^m has the features that each row has exactly one '1', and each column has at most one '1'. In other words, each user is assigned to exactly one cluster, and each cluster contains at most L users, which correspond to the conditions C2 and C3 in (9). According to [32], problem (21) can be solved in polynomial time by Hungarian algorithm, which is also known as Kuhn-Munkres algorithm. For brevity, details of this algorithm are omitted here. The pseudo-code of the t -th iteration of our PCUC strategy in cell m is summarized in Algorithm 3.

Before ending this section, we claim that the PCUC user clustering method needs to be executed jointly with the power control algorithm during each iteration, which will be presented in the following section. Indeed, inter-cell interference may change from one power control iteration to the other, therefore PCUC is recomputed at the beginning of each power control iteration using the most up-to-date inter-cell interference information. These information can be measured or estimated locally and be conducted through some periodical or broadcast messages similar to that suggested in [33]. The signaling transmission associated energy consumption is neglected compared with that of data packet transmission [34] and thus is not considered in this paper. The same assumption

Algorithm 3 The t -th iteration of PCUC algorithm for cell m

Input: $I_{i,m}^{(t-1)}$ for $i \in \mathcal{K}_m$, and $\mathcal{T}_{z,m}$.

Output: $\mathcal{U}_{j,m}$ for $j \in |\mathcal{J}_m|$.

- 1: **Initialization:** $\mathcal{U}_{1,m}, \dots, \mathcal{U}_{|\mathcal{J}_m|,m} = \emptyset$.
- 2: Formulate the LSA cost matrix \mathbf{C} , i.e.,
- 3: **for** $i := 1, 2, \dots, K_m$ **do**
- 4: **for** $j := 1, 2, \dots, |\mathcal{J}_m|$ **do**
- 5: **for** $l := 1, 2, \dots, L$ **do**
- 6: Calculate the required power of user i assuming that user i is assigned to a cluster j and is the l -th decoded user, i.e.,

$$\mathbf{C}_{i,L(j-1)+l}^m = \gamma(\gamma + 1)^{(l-1)} I_{i,m}^{(t-1)},$$

- 7: where $I_{i,m}^{(t-1)}$ is calculated based on (4).

8: **end for**

9: **end for**

10: **end for**

- 11: Decide the boolean matrix \mathbf{X}^m of user clustering by Hungarian algorithm, such that the BS has the minimum transmit power, i.e.,

$$\min \sum_i \sum_{j,l} \mathbf{C}_{i,L(j-1)+l}^m \mathbf{X}_{i,L(j-1)+l}^m.$$

- 12: **for** $j := 1, 2, \dots, |\mathcal{J}_m|$ **do**

- 13: Determine the user set of each cluster j based on \mathbf{X}^m .

$$\mathcal{U}_{j,m} \leftarrow \mathcal{U}_{j,m} \cup \{i \mid \sum_l \mathbf{X}_{i,L(j-1)+l}^m = 1\}.$$

14: **end for**

- 15: Iterative power control of t -th iteration according to Algorithm 4.

was made in existing work [25]–[29] as well.

IV. POWER ALLOCATION ALGORITHM WITH FIXED USER GROUPING AND BF

In this section, we investigate the optimal power allocation algorithm for the transmit power minimization problem with given user grouping and the BF strategy \mathbf{w} . This power control problem is then expressed as follows:

$$\min \sum_{m \in \mathcal{M}} \sum_{i \in \mathcal{K}_m} p_{i,m} \quad (22)$$

subject to

$$C1': R_{i,m}(\mathbf{p}) \geq r_{i,m}, \quad m \in \mathcal{M}, i \in \mathcal{K}_m, \quad (23)$$

$$C4: p_{i,m} \geq 0, \quad m \in \mathcal{M}, i \in \mathcal{K}_m. \quad (24)$$

Note that \mathbf{w} is ignored in the expression of $R_{i,m}$ since it is given. We denote the above problem by \mathbb{P} , which is still non-convex due to the non-convexity of $R_{i,m}(\mathbf{p})$. In addition, define $\mathcal{P} \subseteq \mathbb{R}^{n^\dagger}$ as the feasible power region of \mathbb{P} , wherein $n^\dagger \triangleq \sum_{m=1}^M \sum_{j=1}^{|\mathcal{J}_m|} L_{j,m}$.

A. The Optimal Users Decoding

In this subsection, we prove that the decoding order defined in (5) is optimal, which requires the least power cost when

compared to that of the other decoding orders. Denote by $\mathfrak{p}_{j,m}(\mathbf{q}_{-j,m})$ the minimum power expenditure of cluster j to meet the data rate requirements of its managed subscribers. According to the foregoing analysis, we have

$$\mathfrak{p}_{j,m}(\mathbf{q}_{-j,m}) = \tilde{\mathfrak{p}}_{j,m}(\boldsymbol{\pi}_{j,m}(\mathbf{q}_{-j,m}), \mathbf{q}_{-j,m}), \quad (25)$$

where $\tilde{\mathfrak{p}}_{j,m}$ is well defined in Lemma 1.

Lemma 2. $\boldsymbol{\pi}_{j,m}(\mathbf{q}_{-j,m})$ is the optimal decoding order for the multiplexed users in cluster j of cell m since it has the minimum power cost among all the possible decoding orders, i.e.,

$$\mathfrak{p}_{j,m}(\mathbf{q}_{-j,m}) \triangleq \min_{\hat{\boldsymbol{\pi}}_{j,m}^\dagger \in \Pi_{j,m}} \tilde{\mathfrak{p}}_{j,m}(\hat{\boldsymbol{\pi}}_{j,m}^\dagger, \mathbf{q}_{-j,m}), \quad m \in \mathcal{M}, j \in \mathcal{J}_m. \quad (26)$$

Proof. Let $\hat{\boldsymbol{\pi}}_{j,m}^* \triangleq \boldsymbol{\pi}_{j,m}(\mathbf{q}_{-j,m})$. We need to prove that $\tilde{\mathfrak{p}}_{j,m}(\hat{\boldsymbol{\pi}}_{j,m}^*, \mathbf{q}_{-j,m}) \leq \tilde{\mathfrak{p}}_{j,m}(\hat{\boldsymbol{\pi}}_{j,m}^\dagger, \mathbf{q}_{-j,m})$ is satisfied for $\hat{\boldsymbol{\pi}}_{j,m}^\dagger \in \Pi_{j,m}$. Without loss of generality, it is assumed that

$$\hat{\boldsymbol{\pi}}_{j,m}^\dagger = (\hat{\pi}_{j,m}^\dagger(1), \dots, \hat{\pi}_{j,m}^\dagger(L_{j,m})) \in \Pi_{j,m} \setminus \{\boldsymbol{\pi}_{j,m}^*\}$$

achieves the minimum power cost in (26), i.e.,

$$\tilde{\mathfrak{p}}_{j,m}(\hat{\boldsymbol{\pi}}_{j,m}^\dagger, \mathbf{q}_{-j,m}) < \tilde{\mathfrak{p}}_{j,m}(\hat{\boldsymbol{\pi}}_{j,m}^*, \mathbf{q}_{-j,m}). \quad (27)$$

Based on the definition of decoding order in Section II-A, there must exist a $l < L_{j,m}$ such that

$$I_{\hat{\pi}_{j,m}^\dagger(l),m} < I_{\hat{\pi}_{j,m}^\dagger(l+1),m}, \quad (28)$$

since $\hat{\boldsymbol{\pi}}_{j,m}^\dagger$ and $\hat{\boldsymbol{\pi}}_{j,m}^*$ are two distinct decoding orders.

By exchanging the orders of users $\hat{\pi}_{j,m}^\dagger(l)$ and $\hat{\pi}_{j,m}^\dagger(l+1)$, a new decoding order is obtained, which is denoted as $\hat{\boldsymbol{\pi}}_{j,m}^\circ = (\hat{\pi}_{j,m}^\circ(1), \dots, \hat{\pi}_{j,m}^\circ(l-1), \hat{\pi}_{j,m}^\circ(l+1), \hat{\pi}_{j,m}^\circ(l), \hat{\pi}_{j,m}^\circ(l+2), \dots, \hat{\pi}_{j,m}^\circ(L_{j,m}))$. In the following, we reveal that

$$\tilde{\mathfrak{p}}_{j,m}(\hat{\boldsymbol{\pi}}_{j,m}^\circ, \mathbf{q}_{-j,m}) < \tilde{\mathfrak{p}}_{j,m}(\hat{\boldsymbol{\pi}}_{j,m}^\dagger, \mathbf{q}_{-j,m}), \quad (29)$$

which arises a contradiction to that $\hat{\boldsymbol{\pi}}_{j,m}^\dagger$ achieves the minimum power in (26).

It is easy to observe that, the needed transmit power of users $\hat{\pi}_{j,m}^\dagger(\tilde{l})$ for $\tilde{l} > l+1$ where $\hat{\pi}_{j,m}^\dagger(\tilde{l}) \in \mathcal{U}_{j,m}$, will not be affected after changing the orders of these two adjacent users, i.e., users $\hat{\pi}_{j,m}^\dagger(l)$ and $\hat{\pi}_{j,m}^\dagger(l+1)$. Define p^\dagger and p° as the total required power of users $\hat{\pi}_{j,m}^\dagger(l)$ and $\hat{\pi}_{j,m}^\dagger(l+1)$ in $\boldsymbol{\pi}_{j,m}^\dagger$ and $\boldsymbol{\pi}_{j,m}^\circ$, respectively. Therefore, we only need to prove that $p^\circ < p^\dagger$, since the power cost of user $\hat{\pi}_{j,m}^\dagger(\tilde{l})$ under the order $\boldsymbol{\pi}_{j,m}^\circ$ is strictly less than that under the decoding order $\boldsymbol{\pi}_{j,m}^\dagger$ for $\tilde{l} < l$ with the assumption that $p^\circ < p^\dagger$. By some manipulations, it can be obtained from (6) that

$$\begin{aligned} p^\dagger &= \gamma_{\hat{\pi}_{j,m}^\dagger(l),m} \left(\sum_{l'=l+2}^{L_{j,m}} p_{\hat{\pi}_{j,m}^\dagger(l'),m} + I_{\hat{\pi}_{j,m}^\dagger(l),m} \right) \\ &+ (\gamma_{\hat{\pi}_{j,m}^\dagger(l),m} + 1) \left(\sum_{l'=l+2}^{L_{j,m}} p_{\hat{\pi}_{j,m}^\dagger(l'),m} + I_{\hat{\pi}_{j,m}^\dagger(l+1),m} \right) \gamma_{\hat{\pi}_{j,m}^\dagger(l+1),m}, \end{aligned}$$

whereas

$$p^\circ = \gamma_{\hat{\pi}_{j,m}^\dagger(l+1),m} \left(\sum_{l'=l+2}^{L_{j,m}} p_{\hat{\pi}_{j,m}^\dagger(l'),m} + I_{\hat{\pi}_{j,m}^\dagger(l+1),m} \right) \\ + (\gamma_{\hat{\pi}_{j,m}^\dagger(l+1),m} + 1) \left(\sum_{l'=l+2}^{L_{j,m}} p_{\hat{\pi}_{j,m}^\dagger(l'),m} + I_{\hat{\pi}_{j,m}^\dagger(l),m} \right) \gamma_{\hat{\pi}_{j,m}^\dagger(l),m}.$$

Consequently, we have

$$p^\circ - p^\dagger = \gamma_{\hat{\pi}_{j,m}^\dagger(l+1),m} \gamma_{\hat{\pi}_{j,m}^\dagger(l),m} (I_{\hat{\pi}_{j,m}^\dagger(l),m} - I_{\hat{\pi}_{j,m}^\dagger(l+1),m}),$$

which is negative according to (28). This completes the proof. \square

B. Cluster Power Control Problem, \mathbb{Q}

In this subsection, we first give the formulation of \mathbb{Q} and show its property of standard. Some technical details are required. Then, we analyze the feasibility of \mathbb{Q} . Finally, we prove that the optimal solution to \mathbb{Q} is unique.

1) *The Formulation of \mathbb{Q}* : The related cluster power control problem \mathbb{Q} has the objective function $\min \sum_{m \in \mathcal{M}} \sum_{j \in \mathcal{J}_m} q_{j,m}$ and subjects to the following constraints:

$$q_{j,m} \geq \mathfrak{p}_{j,m}(\mathbf{q}_{-j,m}), \quad m \in \mathcal{M}, j \in \mathcal{J}_m, \quad (30)$$

$$q_{j,m} \geq 0, \quad m \in \mathcal{M}, j \in \mathcal{J}_m, \quad (31)$$

whose feasible power region is defined as $\mathcal{Q} \subseteq \mathbb{R}^{n^\ddagger}$, wherein $n^\ddagger \triangleq \sum_{m \in \mathcal{M}} |\mathcal{J}_m|$.

Define $\mathbf{I}(\mathbf{q}) \triangleq (\mathcal{I}_{j,m}(\mathbf{q}))_{m \in \mathcal{M}, j \in \mathcal{J}_m}$ as the interference function of \mathbb{Q} , each item of which can be taken as the total interference that the corresponding cluster needs to overcome. According to (30) and Lemma 2, we have $\mathcal{I}_{j,m}(\mathbf{q}) \triangleq \mathfrak{p}_{j,m}(\mathbf{q}_{-j,m}) = \min_{\hat{\pi}_{j,m} \in \Pi_{j,m}} \tilde{\mathfrak{p}}_{j,m}(\hat{\pi}_{j,m}, \mathbf{q}_{-j,m})$, where $m \in \mathcal{M}$ and $j \in \mathcal{J}_m$. The standard property of $\mathbf{I}(\mathbf{q})$ is discussed below:

Lemma 3. *The interference function $\mathbf{I}(\mathbf{q})$ is standard.*

Proof. In accordance with Lemma 1, $\tilde{\mathfrak{p}}_{j,m}(\hat{\pi}_{j,m}, \mathbf{q}_{-j,m})$ is expressed below via substituting (4) into (13), i.e.,

$$\tilde{\mathfrak{p}}_{j,m}(\hat{\pi}_{j,m}, \mathbf{q}_{-j,m}) = \sum_{l=1}^{L_{j,m}} \beta_{l,m}(\hat{\pi}_{j,m}) \left(\sum_{j' \in \mathcal{J}_m, j' \neq j} q_{j',m} |\mathbf{g}_{\hat{\pi}_{j,m}(l),m,m} \mathbf{w}_{j',m}|^2 + \sum_{m=1, m' \neq m}^M \sum_{l' \in \mathcal{J}_{m'}} q_{l',m'} |\mathbf{g}_{\hat{\pi}_{j,m}(l),m,m'} \mathbf{w}_{l',m'}|^2 + \sigma_{\hat{\pi}_{j,m}(l),m}^2 \right), \quad (32)$$

where $\beta_{l,m}(\hat{\pi}_{j,m}) \triangleq \frac{\chi_{l,m}(\hat{\pi}_{j,m})}{|\mathbf{g}_{\hat{\pi}_{j,m}(l),m,m} \mathbf{w}_{j,m}|^2}$ and $\chi_{l,m}(\hat{\pi}_{j,m})$ is given by (14). Intuitively, $\tilde{\mathfrak{p}}_{j,m}(\hat{\pi}_{j,m}, \mathbf{q}_{-j,m})$ is an affine function to $\mathbf{q}_{-j,m}$, satisfying the three properties of standard [35]. Since the minimization operation preserves the property of standard [35], $\mathcal{I}_{j,m}(\mathbf{q})$ is standard for $m \in \mathcal{M}$ and $j \in \mathcal{J}_m$. This completes the proof. \square

2) *The Feasibility of \mathbb{Q}* : In this paragraph, we show the feasibility of \mathbb{Q} . Define $\hat{\pi} \triangleq (\hat{\pi}_{1,1}, \hat{\pi}_{2,1}, \dots, \hat{\pi}_{1,M}, \dots, \hat{\pi}_{|\mathcal{J}_M|,M})$ and $\Pi \triangleq \Pi_{1,1} \times \Pi_{2,1} \times \dots \times \Pi_{1,M} \times \dots \times \Pi_{|\mathcal{J}_M|,M}$. The power optimization subproblem with $\hat{\pi} \in \Pi$ is considered:

$$\min \sum_{m \in \mathcal{M}} \sum_{j \in \mathcal{J}_m} q_{j,m} \quad (33)$$

subject to

$$q_{j,m} \geq \tilde{\mathfrak{p}}_{j,m}(\hat{\pi}_{j,m}, \mathbf{q}_{-j,m}), \quad m \in \mathcal{M}, j \in \mathcal{J}_m, \quad (34)$$

$$q_{j,m} \geq 0, \quad \forall m \in \mathcal{M}, j \in \mathcal{J}_m.$$

Since $\hat{\pi}_{j,m} \in \Pi_{j,m}$ and $|\Pi_{j,m}| = L_{j,m}!$, we have $\prod_{m=1}^M \prod_{j=1}^{|\mathcal{J}_m|} (L_{j,m}!)$ such power control subproblems in total. Based on (32), we see that

$$q_{j,m} \geq \sum_{l=1}^{L_{j,m}} \beta_{l,m}(\hat{\pi}_{j,m}) \times \left(\sum_{j' \in \mathcal{J}_m, j' \neq j} q_{j',m} |\mathbf{g}_{\hat{\pi}_{j,m}(l),m,m} \mathbf{w}_{j',m}|^2 + \sum_{m' \in \mathcal{M}, m' \neq m} \sum_{l' \in \mathcal{J}_{m'}} q_{l',m'} |\mathbf{g}_{\hat{\pi}_{j,m}(l),m,m'} \mathbf{w}_{l',m'}|^2 + \sigma_{\hat{\pi}_{j,m}(l),m}^2 \right). \quad (35)$$

The optimization problem (33) can be rewritten into the following problem with matrix form:

$$\min \sum_{m \in \mathcal{M}} \sum_{j \in \mathcal{J}_m} q_{j,m} \quad (36)$$

subject to

$$(\mathbb{I} - \mathbf{D}(\hat{\pi})) \mathbf{q} \geq \mathbf{d}(\hat{\pi}), \quad (37)$$

$$\mathbf{q} \geq \mathbf{0}, \quad (38)$$

where \mathbb{I} represents the identity matrix, $\mathbf{d}(\hat{\pi}) \triangleq \left(\sum_{l=1}^{L_{j,m}} \beta_{l,m}(\hat{\pi}_{j,m}) \sigma_{\hat{\pi}_{j,m}(l),m}^2 \right)_{m \in \mathcal{M}, j \in \mathcal{J}_m}^T$, in which $(\mathbf{x})^T$ indicates the transpose of vector \mathbf{x} , and $\mathbf{D}(\hat{\pi})$ is an off-diagonal matrix whose (u, v) -th element, where $u \neq v$, is expressed as follows:

$$\sum_{l=1}^{L_{a,m}} \beta_{l,m}(\hat{\pi}_{a,m}) |\mathbf{g}_{\hat{\pi}_{a,m}(l),m,m'} \mathbf{w}_{b,m'}|^2,$$

so that the listed two criteria are met:

- 1) $\sum_{l=1}^{m-1} |\mathcal{J}_l| < u \leq \sum_{l=1}^m |\mathcal{J}_l|$,
- 2) $\sum_{l=1}^{m'-1} |\mathcal{J}_l| < v \leq \sum_{l=1}^{m'} |\mathcal{J}_l|$,

where $a \triangleq u - \sum_{l=1}^{m-1} |\mathcal{J}_l|$ and $b \triangleq v - \sum_{l=1}^{m'-1} |\mathcal{J}_l|$.

The above problem (36) is characterized by the decoding orders of all clusters, i.e., $\hat{\pi}$. For $\hat{\pi} \in \Pi$, define $\mathcal{F}(\hat{\pi})$ as the feasible power region of the corresponding power control subproblem. It is clear that $\mathcal{F}(\hat{\pi}) = \bigcap_{m \in \mathcal{M}, j \in \mathcal{J}_m} \mathcal{F}_{j,m}(\hat{\pi}_{j,m})$, where $\mathcal{F}_{j,m}(\hat{\pi}_{j,m}) \triangleq \{\mathbf{q} \geq \mathbf{0} : q_{j,m} \geq \tilde{\mathfrak{p}}_{j,m}(\hat{\pi}_{j,m}, \mathbf{q}_{-j,m})\}$. Therefore, we have the following Lemma 4.

Lemma 4. *The feasible region of problem \mathbb{Q} is given as follows:*

$$\mathcal{Q} = \bigcup_{\hat{\pi} \in \Pi} \mathcal{F}(\hat{\pi}). \quad (39)$$

Proof. See Appendix B. \square

3) *The Optimality of \mathbb{Q} :* In this part, we depict that the optimal solution to \mathbb{Q} is unique. We observe that the optimal power strategy to \mathbb{Q} achieves when the inequalities in (30) satisfy with equalities. Thereby, the following result is obtained.

Theorem 5. *Given that \mathcal{Q} is non-empty, \mathbb{Q} has a unique optimal solution.*

Proof. Since \mathbb{Q} is feasible, by Lemma 4, there must exist some of the power control subproblems which are feasible. In addition, the optimal solution to each subproblem is unique [36]. Therein, the one who needs the least power consumption is the optimal solution to \mathbb{Q} .

In the following, we show the uniqueness. Denote by $\mathbf{q}^* \triangleq (q_{1,1}^*, q_{2,1}^*, \dots, q_{|\mathcal{J}_M|,M}^*)$ and $\mathbf{q}^\dagger \triangleq (q_{1,1}^\dagger, q_{2,1}^\dagger, \dots, q_{|\mathcal{J}_M|,M}^\dagger)$ two diverse optimal strategies. Intuitively, $\|\mathbf{q}^*\|_1 = \|\mathbf{q}^\dagger\|_1$. Due to the fact that \mathbf{q}^\dagger and \mathbf{q}^* are different, an $\alpha > 1$ must exist so that $\alpha\mathbf{q}^* \geq \mathbf{q}^\dagger$ as well as $\alpha q_{j,m}^* = q_{j,m}^\dagger$ are satisfied for some $m \in \mathcal{M}$ and $j \in \mathcal{J}_m$. As a consequence, we have $\alpha q_{j,m}^* = \alpha \mathfrak{p}_{j,m}(\mathbf{q}_{-j,m}^*) > \mathfrak{p}_{j,m}(\alpha \mathbf{q}_{-j,m}^*) \geq \mathfrak{p}_{j,m}(\mathbf{q}_{-j,m}^\dagger) = q_{j,m}^\dagger$, which arises a contradiction. \square

C. Characterization of \mathbb{P}

In this subsection, we first analyze the feasibility of \mathbb{P} . Based on which, we show that problems \mathbb{P} and \mathbb{Q} have the same optimal values and one-to-one mapping relationship.

1) *The Feasibility of \mathbb{P} :* A mapping Φ from domain \mathcal{Q} to domain \mathcal{P} is defined. For any $\mathbf{q} \in \mathcal{Q}$, we must have $q_{j,m} \geq \mathfrak{p}_{j,m}(\mathbf{q}_{-j,m})$ for $m \in \mathcal{M}$, $j \in \mathcal{J}_m$. By letting the data rate requirements of users in $\mathcal{U}_{j,m}$ hold with strict equalities, we can find $\mathbf{p}_{j,m}^\dagger$ such that $\|\mathbf{p}_{j,m}^\dagger\|_1 = \mathfrak{p}_{j,m}(\mathbf{q}_{-j,m}) \leq q_{j,m}$. Denote by $\varepsilon \triangleq q_{j,m}/\|\mathbf{p}_{j,m}^\dagger\|_1 \geq 1$ and $\mathbf{p}_{j,m} = \varepsilon \mathbf{p}_{j,m}^\dagger$. It is easy to observe that $\mathbf{p}_{j,m} \in \mathcal{P}$, and then we have $\|\mathbf{p}_{j,m}\|_1 = q_{j,m}$. The mapping $\Phi : \mathcal{Q} \rightarrow \mathcal{P}$ is then defined as $\Phi(\mathbf{q}) \triangleq (\mathbf{p}_{1,1}, \mathbf{p}_{2,1}, \dots, \mathbf{p}_{|\mathcal{J}_M|,M})$, where $\mathbf{p}_{j,m}$ for $m \in \mathcal{M}$ and $j \in \mathcal{J}_m$ is characterized by the aforementioned procedure. We note that $\Phi(\mathbf{q}) \in \mathcal{P}$ for $\mathbf{q} \in \mathcal{Q}$. Similarly, another mapping Θ from domain \mathcal{P} to domain \mathcal{Q} is defined as $\Theta(\mathbf{p}_{1,1}, \mathbf{p}_{2,1}, \dots, \mathbf{p}_{|\mathcal{J}_M|,M}) \triangleq (q_{1,1}, q_{2,1}, \dots, q_{|\mathcal{J}_M|,M})$, where $q_{j,m} = \|\mathbf{p}_{j,m}\|_1$ for $m \in \mathcal{M}$ and $j \in \mathcal{J}_m$. It is easy to see that $\Theta(\mathbf{p}_{1,1}, \mathbf{p}_{2,1}, \dots, \mathbf{p}_{|\mathcal{J}_M|,M}) \in \mathcal{Q}$ if $(\mathbf{p}_{1,1}, \mathbf{p}_{2,1}, \dots, \mathbf{p}_{|\mathcal{J}_M|,M}) \in \mathcal{P}$. With the above definitions, we have the following result.

Theorem 6. *\mathcal{P} is non-empty if and only if there exists some $\hat{\pi} \in \Pi$ such that the Perron-Frobenius eigenvalue of $\mathbf{D}(\hat{\pi})$ is less than 1.*

Proof. Based on the definitions of Φ and Θ , it is easy to check that \mathbb{P} is feasible when and only when \mathbb{Q} is feasible. According to [36], the sufficient and necessary condition for a power control problem having feasible solutions is that the Perron-Frobenius eigenvalue of $\mathbf{D}(\hat{\pi})$ is less than 1. Given the above two facts as well as Lemma 4, we have Theorem 6. \square

2) *The Optimality of \mathbb{P} :* The optimal power allocation to problem \mathbb{P} obtains when the inequalities in (23) hold with equalities. We will show that the optimal solution to \mathbb{P} is

Algorithm 4 The t -th power iteration of cluster j in cell m

Input: $I_{i,m}^{(t-1)}$ for $i \in \mathcal{U}_{j,m}$.

Output: $q_{j,m}^{(t)}$ and $\mathbf{p}_{j,m}^{(t)}$.

- 1: Calculate the optimal decoding order, $\pi_{j,m}$, based on (5), with inputs $I_{i,m}^{(t-1)}$ for $i \in \mathcal{U}_{j,m}$.
- 2: **for** $l := 1, 2, \dots, L_{j,m}$, **do**
- 3: Calculate $\chi_{l,m}$ by (14), i.e., $\chi_{l,m} := \left[\prod_{l' < l} (\gamma_{\pi_{j,m}(l'),m} + 1) \right] \gamma_{\pi_{j,m}(l),m}$.
- 4: **end for**
- 5: Determine the least required transmit power of cluster j at BS m as follows:

$$q_{j,m}^{(t)} := \sum_{l=1}^{L_{j,m}} \chi_{l,m} I_{\pi_{j,m}(l),m}^{(t-1)}.$$

- 6: Determine the power allocation for user $\pi_{j,m}(L_{j,m})$, which is quoted below:

$$p_{\pi_{j,m}(L_{j,m}),m}^{(t)} := \gamma_{\pi_{j,m}(L_{j,m}),m} I_{\pi_{j,m}(L_{j,m}),m}^{(t)}.$$

- 7: **for** $l := L_{j,m} - 1, L_{j,m} - 2, \dots, 1$, **do**
- 8: Calculate the allocated power to user $\pi_{j,m}(l)$ in cell m as follows:

$$p_{\pi_{j,m}(l),m}^{(t)} := \gamma_{\pi_{j,m}(l),m} \left(\sum_{l' > l} p_{\pi_{j,m}(l'),m}^{(t)} + I_{\pi_{j,m}(l),m}^{(t-1)} \right).$$

- 9: **end for**

- 10: **return** $q_{j,m}^{(t)}$ and $(p_{i,m}^{(t)})_{i \in \mathcal{U}_{j,m}}$.
-

unique and there is a one-to-one mapping property between the optimal solutions to \mathbb{P} and \mathbb{Q} .

Lemma 7. *Problems \mathbb{Q} and \mathbb{P} have the same optimal values.*

Proof. Let \mathbf{p}^* and \mathbf{q}^* be the optimal power strategies to \mathbb{P} and \mathbb{Q} , respectively. Thereby, we have $\|\mathbf{q}^*\|_1 = \|\Phi(\mathbf{q}^*)\|_1 \geq \|\mathbf{p}^*\|_1$ and $\|\mathbf{q}^*\|_1 \leq \|\Theta(\mathbf{p}^*)\|_1 = \|\mathbf{p}^*\|_1$. This completes the proof. \square

Theorem 8. *Given that \mathbb{P} is feasible, then its optimal solution, denoted by \mathbf{p}^* , is unique, which satisfies $\mathbf{p}^* = \Phi(\mathbf{q}^*)$, wherein \mathbf{q}^* is the optimal power allocation policy of \mathbb{Q} .*

Proof. Since \mathbb{P} is feasible, \mathbb{Q} is also feasible. By Theorem 5, the optimal strategy of \mathbb{Q} , \mathbf{q}^* , is unique. Based on Lemma 7, $\mathbf{p}^* = \Phi(\mathbf{q}^*)$ is an optimal solution. Let \mathbf{p}^\ddagger be another optimal strategy to \mathbb{P} . According to Lemma 7, $\Theta(\mathbf{p}^\ddagger)$ is also optimal to \mathbb{Q} and equals to \mathbf{q}^* , which means $\|\mathbf{p}_{j,m}^\ddagger\|_1 = \|\mathbf{p}_{j,m}^*\|_1 = q_{j,m}^*$ is satisfied for $m \in \mathcal{M}$ and $j \in \mathcal{J}_m$. Since the optimal power allocation for \mathbb{P} obtains when each of the users achieves exactly its minimum data rate requirement, we have $p_{i,m}^\ddagger = p_{i,m}^*$ under any fixed $\mathbf{q}_{-j,m}^*$, which represents $\mathbf{p}^\ddagger = \mathbf{p}^*$. \square

The above obtained results uncover that we can get the optimal allocation policy to \mathbb{P} via solving \mathbb{Q} . In next subsection, the optimal power control algorithm for \mathbb{Q} is designed.

D. Optimal Power Allocation Algorithm

In this subsection, we describe the proposed optimal power allocation to solve \mathbb{P} , which has guaranteed convergence performance and optimality given that \mathbb{P} is feasible, i.e., \mathcal{P} is non-empty.

Let $q_{j,m}^{(t)}$ be the total transmit power of all users in the j -th cluster of cell m at iteration t , where $m \in \mathcal{M}$ and $j \in \mathcal{J}_m$. In addition, define $\mathbf{q}^{(t)} = (q_{1,1}^{(t)}, q_{2,1}^{(t)}, \dots, q_{|\mathcal{J}_M|,M}^{(t)})$ and $\mathbf{q}_{-j,m}^{(t)} = (q_{1,1}^{(t)}, q_{2,1}^{(t)}, \dots, q_{j-1,m}^{(t)}, q_{j+1,m}^{(t)}, \dots, q_{|\mathcal{J}_M|,M}^{(t)})$. In the t -th iteration, the transmit power allocation for subscribers within each cluster is conducted by their associated BS. With aforementioned definitions, the iterative algorithm is given below:

$$q_{j,m}^{(t)} = p_{j,m}(\mathbf{q}_{-j,m}^{(t-1)}) \text{ for } m \in \mathcal{M}, j \in \mathcal{J}_m. \quad (40)$$

We re-write (40) in vector form, i.e.,

$$\mathbf{q}^{(t)} = \mathbf{I}(\mathbf{q}^{(t-1)}), \quad (41)$$

whose convergence is presented in the following theorem.

Theorem 9. *Given that \mathbb{P} is feasible, i.e., \mathcal{P} is non-empty, the iterative power control algorithm stated in (40) converges to the unique optimal power strategy \mathbf{q}^* with any fixed initial point.*

Proof. Since \mathcal{P} is non-empty, \mathbb{P} is feasible and thus \mathbb{Q} is also feasible according to the definitions of previously defined mappings. By Theorem 5, the optimal strategy of \mathbb{Q} , \mathbf{q}^* , is unique. In accordance with Lemma 3, (41) is standard. Therefore, the iterative power allocation converges to a unique fixed point with any predetermined initial power values [35]. The proof is completed since the optimal solution to \mathbb{Q} is a fixed point. \square

With the optimal strategy \mathbf{q}^* , the optimal solution of the original minimization problem can be obtained via the mapping Φ . For expression simplicity, we summarize the pseudo code of the t -th iteration of our designed optimal power allocation in Algorithm 4.

So far, the design for the joint user clustering, BF and power control algorithm is completed. Before ending this section, we declare that our proposed resource management algorithms are applicable to the usage scenarios with or without BS's transmit power budget taken into consideration. To be more specific, for each BS, if the total required power in terms of satisfying all its associated users' minimum data rate requirements is larger than the budget, outage occurs, as demonstrated in Section V, wherein a maximum value of total transmit power per BS is set to analyze the outage performance of our proposed versatile methodologies.

V. SIMULATION RESULTS

In this section, Monte-Carlo simulation is used to evaluate the performance of our proposed joint user grouping, BF and power allocation strategies for multi-cell MISO-NOMA system. In the simulation, we consider a 7-cell ($M = 7$) network adopting wrap-around technique with 12 transmit antennas ($N = 12$) for each BS. The inter-site distance is set to 500 meters, and K_m users are uniformly distributed within

TABLE I
SIMULATION PARAMETERS

Parameters	Values
Inter-site distance	500 m
Minimum distance from user to BS	35 m
Distance-dependent path loss	$128.1 + 37.6 \log_{10}(d)$ dB, d in km
Shadowing	Log-normal, standard deviation 8 dB
Small scale fading	Rayleigh fading with variance 1
Noise power spectral density	-174 dBm/Hz
Throughput calculation	Shannon's capacity formula
System bandwidth, W	10 MHz
Carrier frequency	2 GHz
Maximum number of iterations	100
Transmit power budget of a BS	0.1 W
Number of simulated BSs, M	7
Number of antennas for each BS, N	12
Maximum number of users per cluster, L	K
Number of simulation instances for each case	10000
Number of users per cell, K	8, 9, ..., 13
User data rate requirement, R	(0.2, 0.4, ..., 1.4) Mbits/s

the hexagonal cell m , where $m \in \mathcal{M}$. The carrier frequency is set to be 2 GHz, and the system bandwidth W is set to be 10 MHz. The noise power spectral density at each receiver is assumed equal to -174 dBm/Hz. The radio propagation model follows [37], which includes three components, i.e., the distance-dependent path loss, shadow fading as well as the small-scale fading. In detail, the distance-dependent path loss is given by $128.1 + 37.6 \log_{10}(d)$, where d is the Euclidean distance between the BS (transmitter) and the user (receiver) in km. The shadow fading follows a log-normal shadowing model with a standard deviation of 8 dB. Each user also experiences independent Rayleigh fading with variance 1 as small scale fading. In addition, we assume all the subscribers have the same data rate requirement, i.e., $r_{i,m} = R$ for $m \in \mathcal{M}$ and $i \in \mathcal{K}_m$. Besides, each cell is assumed to have the same number of subscribers, i.e., $K_m = K$ for $m \in \mathcal{M}$. Moreover, the maximum number of users per cluster, i.e., L , is assumed to be equal to K . Note that our proposed schemes are also applicable to the scenarios with different values of K_m and $r_{i,m}$ for $m \in \mathcal{M}$ and $i \in \mathcal{K}_m$. It is worth mentioning that the plots are obtained by simulating 10000 independent instances of randomly generated network realization. We summarize the simulation parameters in Table I.

We use conventional ZF assisted OMA based MISO scheme as a baseline, named as MISO-OMA-ZF, to compare its performance with the proposed clustering enabled MISO-NOMA schemes, which are denoted by MISO-NOMA-CCUC and MISO-NOMA-PCUC, respectively. Specifically, in MISO-OMA-ZF scheme, users in the same cell are grouped to different clusters exclusively and ZF-BF is applied. In MISO-NOMA-CCUC, the CCUC approach is adopted to perform user clustering, which is optimized independently of the power control. In contrast, MISO-NOMA-PCUC scheme requires the power control information to perform user clustering at each iteration, and thus is a joint user clustering and power allocation paradigm. Note that in the proposed two strategies, ZF-BF and the optimal power control algorithm (i.e., Algorithm 4) are used for the beamforming and power allocation, respectively. Therein, SUS algorithm is applied to do the ZF user selection.

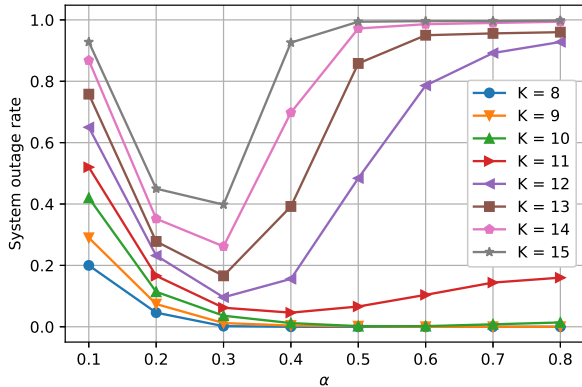
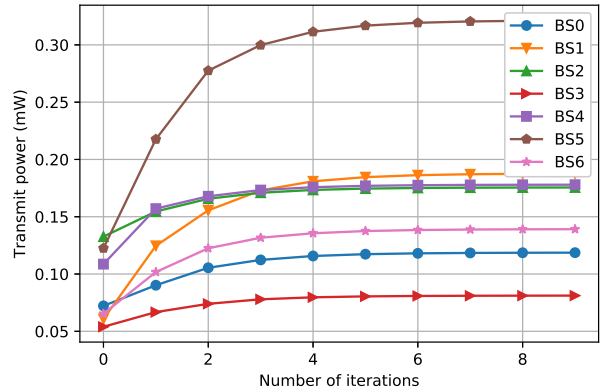


Fig. 1. System outage rate of MISO-NOMA-CCUC scheme under different α and K with $N = 12$ and $R = 1.0$ Mbits/s.

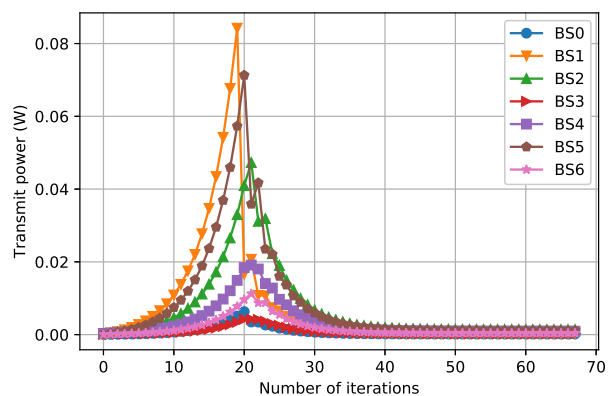
To show the effectiveness of our clustering based MISO-NOMA strategies, the non-clustered MISO-NOMA approach (i.e., $|\mathcal{J}_M| = 1$) [4] is also taken into consideration. We declare that the proposed ZF-BF as well as the power control algorithms are also applicable for the non-clustered scheme, which is denoted by MISO-NOMA. To be more specific, in MISO-NOMA strategy, the BS transmits the superposed signals of all users within each cell. And therein the BF vector is calculated based on the user with the highest channel gain. Nevertheless, it should be noted that, this approach is less practical for implementation due to the limitations of receiver design complexity and the error propagation issue induced by SIC, especially when the user number is large. Moreover, for comprehensive comparison, the performance of other BF techniques for MISO-NOMA system such as matched filter (MF) [30] and minimum mean square error (MMSE) [38] are also evaluated. They are named respectively as MISO-NOMA-MF and MISO-NOMA-MMSE.

A. Optimal Value of α in SUS

In this subsection, we explain how the SUS factor α is chosen, as it significantly influences the performance of our proposed MISO-NOMA strategies. Indeed, if α is too large, the inter-cluster interference will be large due to the high correlation between clusters, while if α is too small, more users will be grouped in a cluster. In this work, the near-optimal values of the SUS factor α for different scenarios are searched through empirical simulations for both MISO-NOMA-CCUC and MISO-NOMA-PCUC schemes, wherein α varies from 0.1 to 0.8 with step size equal to 0.1. The best value of α is chosen such that the minimum system outage rate can be achieved, which is usually in our first concern. Note that the system outage rate is defined by the proportion of the number of instances where the total power consumption of users in each cell exceeds the power budget of their associated BS. An example of the system outage rate under various α and K when $R = 1.0$ Mbits/s is shown in Fig. 1. The optimal choice of α for different K can be taken at the minimum point of each curve. It can be seen that the optimal α decreases



(a) Without user outage.



(b) With at least one user in outage.

Fig. 2. Convergence behavior of MISO-NOMA-PCUC with or without user outage.

with K , from 0.6 to 0.3 for K increasing from 8 to 15. In the following simulations, the found best empirical value of α is adopted. It should be noted that the optimality of α can be further improved by more exhaustive search.

B. Convergence Performance

In most communication systems, the transmit power of each BS is usually limited. Given that the power consumption of BS m exceeds its power limit, we say that an outage occurs. Generally, the outage is due to the fact that the power control problem is infeasible, i.e., some of the users require very high transmit power¹ for meeting their data rate requirements. In this subsection, we investigate the convergence behavior of our designed joint user clustering and power control algorithm, as illustrated in Fig. 2(a) and Fig. 2(b). We plot the transmit power of each BS during the iterations. These results are taken from two typical instances in the simulation of MISO-NOMA-PCUC scheme. Specifically, Fig. 2(a) depicts an instance where there is no outage in the system. In this case, the power

¹The required transmit power for some users may be larger than the power budget of their BS. In such case, we say that these users are in outage and cannot be served. In practice, their allocated power will be then simply set to zero during the power control iterations.

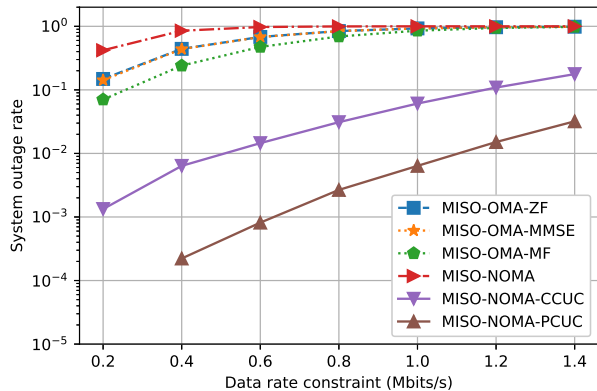
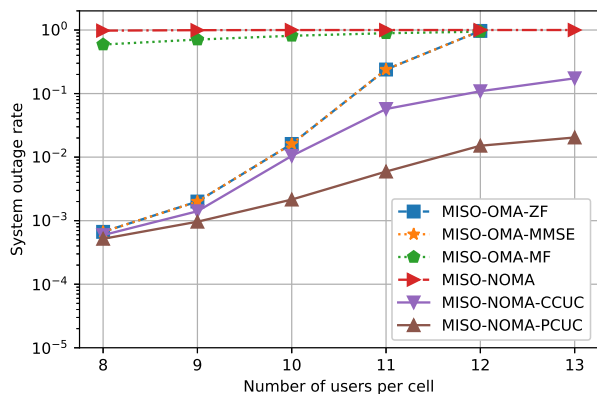
(a) System outage rate vs. data rate requirement, $K = 12$.(b) System outage rate vs. number of users per cell, $R = 1.4$ Mbits/s.

Fig. 3. System outage rate performance.

of all the 7 BSs can converge to their fine values rapidly. The power allocation is feasible, i.e., the data rate requirement of each user in the system can be satisfied. On the other hand, Fig. 2(b) illustrates an instance where there is at least one user in outage. It can be seen that the transmit power of the 7 BSs tends to diverge in the first 19 iterations. At the 19-th iteration, the power of BS 1 exceeds the set maximum power limit 0.1 Watt. We abandon the outage users (i.e., set the transmit power for them to zero) according to common practice since their data rate requirements cannot be satisfied given the power budget of the base stations. As a result, in the example, after the 19-th iteration, the transmit power of each BS in the system decreases naturally and finally converges to a feasible solution which satisfies the data rate requirements of the rest of the users.

C. Outage Performance

In this subsection, we evaluate the system outage rate and the user outage rate performance of all the schemes. The user outage rate is defined by the number of users in outage divided by the total number of users in the system. On the other hand, as aforementioned, the system outage rate is defined by the percentage of the number of instances where there exists at least one user in outage.

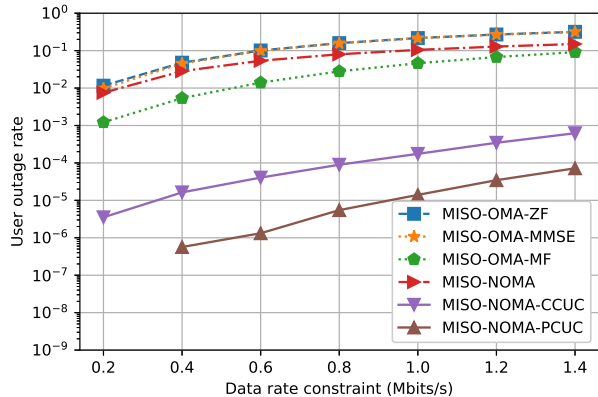
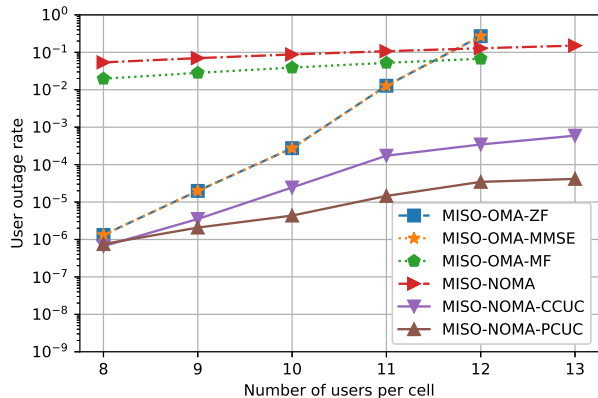
(a) User outage rate vs. data rate requirement, $K = 12$.(b) User outage rate vs. number of users per cell, $R = 1.4$ Mbits/s.

Fig. 4. User outage rate performance.

The system outage rate and user outage rate under various user data rate requirements are shown in Fig. 3(a) and Fig. 4(a), respectively. Thereof, it is worth mentioning that, the outage probabilities of MISO-NOMA-PCUC under $R = 0.2$ Mbits/s are zero in both figures. Since the logarithmic y -axis is adopted, the values of MISO-NOMA-PCUC at $R = 0.2$ Mbits/s in both Fig. 3(a) and Fig. 4(a) are neglected. From both figures, we can see that the three conventional OMA based MISO schemes have higher system outage rate and user outage rate when compared to that of our designed MISO-NOMA strategies. Since in the three MISO-OMA baselines, the chosen BF vectors could lead to great intra-cell and also inter-cell interference given that users have highly correlated channels, resulting in high outage probability. Besides, the non-clustered MISO-NOMA method has the highest system outage rate than all the other schemes. This is due to the higher power requirement with respect to decode the superposed signals by SIC. In contrast, the clustered MISO-NOMA enabled schemes would group the highly correlated users in a same cluster and use SIC at the receivers to separate their signals. This would reduce the interference and achieve lower outage rate. From Fig. 3(a) and Fig. 4(a), we can also see that MISO-NOMA-PCUC outperforms MISO-NOMA-CCUC in the outage rate. This is because MISO-NOMA-PCUC considers the interfer-

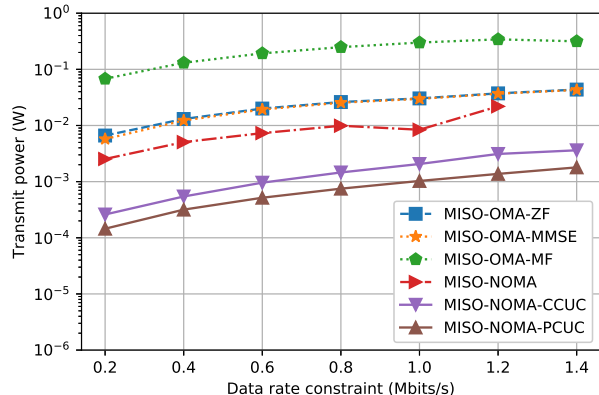
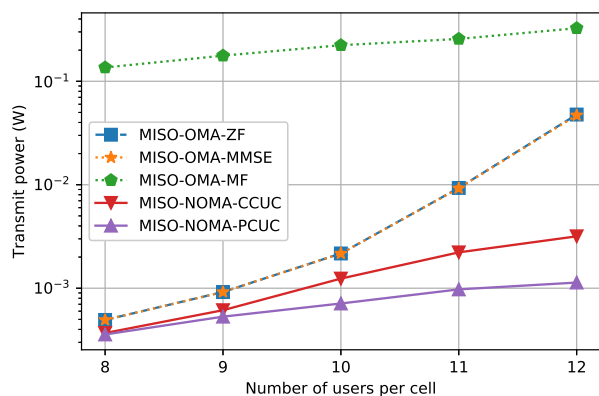
(a) Power consumption vs. data rate requirement, $K = 12$.(b) Power consumption vs. number of users per cell, $R = 1.4$ Mbits/s.

Fig. 5. Power consumption.

ence of each user and also their corresponding transmit power in order to decide the user clustering, while MISO-NOMA-CCUC is only based on the channel link gains.

The system outage rate and user outage rate under different number of users are plotted in Fig. 3(b) and Fig. 4(b), respectively. In the conventional OMA based MISO schemes, at most N users can be supported at the same time, where N is the number of antennas. Note that when more users are active in a cell, time scheduling would be required. Nevertheless, MISO-NOMA schemes can support more users per cell, i.e., $K > N$ is manageable under MISO-NOMA. In addition, we can see that MISO-NOMA-PCUC outperforms all other schemes in the outage rate performance, while the user outage rate of MISO-NOMA and MISO-OMA-MF tends to one when R is large. Besides, the performance gain of MISO-NOMA-PCUC compared to the other schemes is small when there are fewer users per cell. This indicates that when K is small, the effect of user grouping would be less important.

D. Power Consumption

In this subsection, the power consumption of the system is calculated for different schemes. It is worth noting that, to be fair in comparison, only the instances in the simulation which have feasible solution in all the schemes are considered in the

calculation. Each point in Fig. 5(a) and Fig. 5(b) is obtained by averaging over all the feasible instances.

The power consumption under different user data rate requirement R is shown in Fig. 5(a). It can be seen that the power cost of all schemes increase with the increasing R . In addition, for any given R , the conventional MISO-OMA-MF scheme requires the highest power consumption, while MISO-NOMA-PCUC has the lowest power consumption. MISO-OMA-ZF consumes slightly higher power than MISO-OMA-MMSE, because MMSE takes into account the effect of noise [30], [38]. From Fig. 5(a), on average MISO-NOMA-PCUC saves about 47% and 94% of the power when compared to MISO-NOMA-CCUC and MISO-OMA-ZF, respectively.

The power consumption under different number of users is shown in Fig. 5(b).² Therein, the data rate requirement of each user is set to be 1.4 Mbits/s. It is shown that MISO-NOMA-PCUC outperforms the other schemes in the power consumption, especially when the number of users increases. For example, when $K = 12$, MISO-NOMA-PCUC helps to save power by 51% and 91% when compared to MISO-NOMA-CCUC and MISO-OMA-ZF, respectively. It should be noted that the performance gain is more remarkable for larger K , because the performance of OMA based MISO strategies are degraded dramatically due to the inter-cluster interference when K is large.

E. Energy Efficiency and Connectivity Efficiency

Energy efficiency (EE) is an important metric for future wireless communication systems [4]. Here, we follow the definition in [39] and consider the following notion:

$$EE = \frac{\text{Total amount of transmitted data}}{\text{Total radiation energy consumption}} \text{ (bits/joule)}. \quad (42)$$

The energy efficiency of the schemes under different data rate requirement is shown in Fig. 6(a). We can see that the energy efficiency of all the schemes decreases with the increasing R . In addition, all the NOMA based schemes outperform the OMA aided strategies. Moreover, MISO-NOMA-PCUC still has the highest energy efficiency among all the schemes.

Besides, we define the following metric, namely connectivity efficiency (CE), to indicate the number of users that can be supported per unit power consumption. The number of supported users in the system is normalized by the total transmit power consumption as follows:

$$CE = \frac{\text{Total number of supported users}}{\text{Total transmit power consumption}} \text{ (W}^{-1}\text{)}. \quad (43)$$

Fig. 6(b) shows the connectivity efficiency of the different schemes under different data rate requirement. From Fig. 6(b), we can see that the connectivity efficiency for each scheme decreases with the increasing R . Among which, MISO-NOMA-PCUC outperforms the other five schemes as expected. This is also due to the fact that MISO-NOMA-PCUC has both lower user outage rate and lower total power consumption. Furthermore, NOMA enabled strategies achieve higher connectivity

²The power consumption of MISO-NOMA is not evaluated because its system outage rate tends to one when $R = 1.4$ Mbits/s.

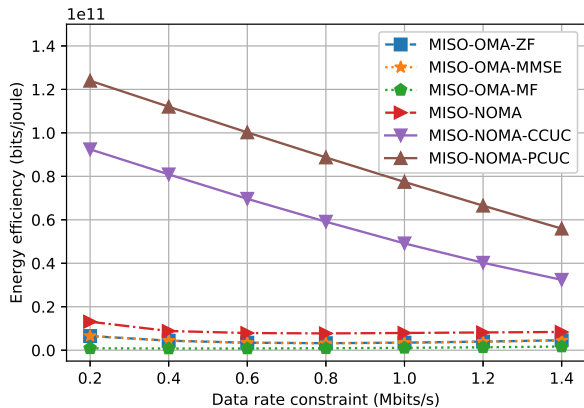
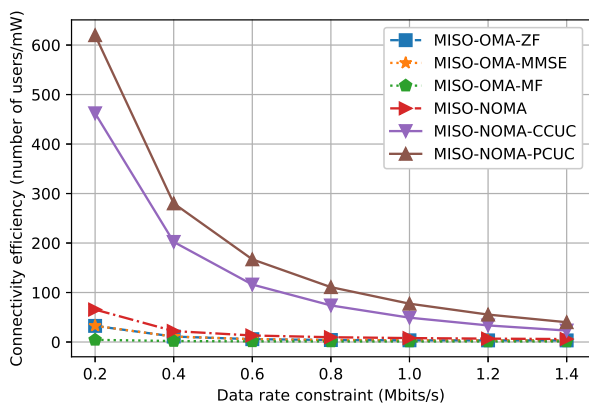
(a) Energy efficiency vs. data rate requirement, $K = 12$.(b) Connectivity efficiency vs. data rate requirement, $K = 12$.

Fig. 6. Energy efficiency and connectivity efficiency.

efficiency compared to all the OMA oriented schemes, showing the effectiveness of NOMA.

VI. CONCLUSION

In this paper, we investigate the power minimization problem for a generic multi-cell MISO-NOMA system through a joint user grouping, BF and power control perspective. The mixed integer and non-convex programming problem is solved by a two-fold methodology. Firstly, the BF and user clustering are determined by the SUS algorithm and the proposed user clustering strategies (PCUC and CCUC), respectively. We can see that PCUC outperforms CCUC, since it uses the information of the transmit power and interference level during iterations when conducting user grouping while CCUC only considers user's channel link gain. Subsequently, with the obtained user clustering and BF, we solve the power allocation problem optimally via an iterative algorithm with provable convergence guarantee. Numerical results show that our proposed schemes outperform conventional MISO-OMA strategies and non-clustered MISO-NOMA policy in several aspects, including power consumption, outage rate, energy efficiency, and connectivity efficiency. One may also consider in the future to use NOMA and multi-antenna techniques for

mobile edge computing, fog radio access, mission-critical IoT, and massive machine type communications in B5G systems.

APPENDIX A PROOF OF LEMMA 4

Based on (30) and (31), \mathcal{Q} is as follows:

$$\mathcal{Q} = \bigcap_{m \in \mathcal{M}, j \in \mathcal{J}_m} \{ \mathbf{q} \geq \mathbf{0} : q_{j,m} \geq \mathfrak{p}_{j,m}(\mathbf{q}_{-j,m}) \}. \quad (44)$$

According to Lemma 2, the above equation can be transformed into

$$\begin{aligned} \mathcal{Q} &= \bigcap_{m \in \mathcal{M}, j \in \mathcal{J}_m} \bigcup_{\hat{\boldsymbol{\pi}}_{j,m} \in \Pi_{j,m}} \{ \mathbf{q} \geq \mathbf{0} : q_{j,m} \geq \tilde{\mathfrak{p}}_{j,m}(\hat{\boldsymbol{\pi}}_{j,m}, \mathbf{q}_{-j,m}) \} \\ &= \bigcap_{m \in \mathcal{M}, j \in \mathcal{J}_m} \bigcup_{\hat{\boldsymbol{\pi}}_{j,m} \in \Pi_{j,m}} \mathcal{F}_{j,m}(\hat{\boldsymbol{\pi}}_{j,m}). \end{aligned} \quad (45)$$

Since $\mathcal{F}_{j,m}(\hat{\boldsymbol{\pi}}_{j,m})$ is purely determined by the decoding order of users in cluster j of cell m , (45) can be further expressed as follows:

$$\mathcal{Q} = \bigcap_{m \in \mathcal{M}, j \in \mathcal{J}_m} \bigcup_{\hat{\boldsymbol{\pi}} \in \Pi} \mathcal{F}_{j,m}(\hat{\boldsymbol{\pi}}_{j,m}). \quad (46)$$

Besides, intersection of unions is in general a superset of union of intersections, i.e.,

$$\begin{aligned} \bigcup_{\hat{\boldsymbol{\pi}} \in \Pi} \mathcal{F}(\hat{\boldsymbol{\pi}}) &= \bigcup_{\hat{\boldsymbol{\pi}} \in \Pi} \bigcap_{m \in \mathcal{M}, j \in \mathcal{J}_m} \mathcal{F}_{j,m}(\hat{\boldsymbol{\pi}}_{j,m}) \\ &\subseteq \bigcap_{m \in \mathcal{M}, j \in \mathcal{J}_m} \bigcup_{\hat{\boldsymbol{\pi}} \in \Pi} \mathcal{F}_{j,m}(\hat{\boldsymbol{\pi}}_{j,m}) = \mathcal{Q}. \end{aligned} \quad (47)$$

Given any feasible power vector of problem \mathcal{Q} , $\mathbf{q} \in \mathcal{Q}$, in accordance with (44) and Lemma 2, there must exist $\hat{\boldsymbol{\pi}}^* \triangleq (\hat{\boldsymbol{\pi}}_{1,1}^*, \hat{\boldsymbol{\pi}}_{2,1}^*, \dots, \hat{\boldsymbol{\pi}}_{|\mathcal{J}_M|,M}^*) \in \Pi$ such that

$$q_{j,m} \geq \mathfrak{p}_{j,m}(\mathbf{q}_{-j,m}) = \tilde{\mathfrak{p}}_{j,m}(\hat{\boldsymbol{\pi}}_{j,m}^*, \mathbf{q}_{-j,m}), \quad m \in \mathcal{M}, j \in \mathcal{J}_m. \quad (48)$$

In accordance with (44), we have

$$\mathbf{q} \in \bigcap_{m \in \mathcal{M}, j \in \mathcal{J}_m} \mathcal{F}_{j,m}(\hat{\boldsymbol{\pi}}_{j,m}^*) \subseteq \bigcup_{\hat{\boldsymbol{\pi}} \in \Pi} \bigcap_{m \in \mathcal{M}, j \in \mathcal{J}_m} \mathcal{F}_{j,m}(\hat{\boldsymbol{\pi}}_{j,m}),$$

which uncovers that

$$\mathcal{Q} \subseteq \bigcup_{\hat{\boldsymbol{\pi}} \in \Pi} \mathcal{F}(\hat{\boldsymbol{\pi}}). \quad (49)$$

Follows from (47) and (49), the proof is completed.

REFERENCES

- [1] Y. Liu, Z. Qin, M. ElKashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Non-orthogonal multiple access for 5G and beyond," *Proc. IEEE*, vol. 105, no. 12, pp. 2347–2381, Dec. 2017.
- [2] 3rd Generation Partnership Project (3GPP), "Study on downlink multi-user superposition transmission for LTE," Mar. 2015.
- [3] E. Balevi and R. D. Gitlin, "Pareto optimization for uplink NOMA power control," in *Proc. IEEE Wireless Telecommunications Symposium (WTS)*, Apr. 2018.
- [4] Y. Fu, Y. Chen, and C. W. Sung, "Distributed power control for the downlink of multi-cell NOMA systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 9, pp. 6207–6220, Sep. 2017.
- [5] L. Lei, D. Yuan, C. K. Ho, and S. Sun, "Power and channel allocation for non-orthogonal multiple access in 5G systems: tractability and computation," *IEEE Transactions on Wireless Communications*, vol. 15, no. 12, pp. 8580–8594, Dec. 2016.

- [6] Y. Fu, L. Salaun, C. W. Sung, C. S. Chen, and M. Coupechoux, "Double iterative waterfilling for sum rate maximization in multicarrier NOMA systems," *IEEE International Conference on Communications (ICC)*, May. 2017.
- [7] Y. Fu, L. Salaun, C. W. Sung, and C. S. Chen, "Subcarrier and power allocation for the downlink of multicarrier NOMA systems," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 12, pp. 11 833–11 847, Dec. 2018.
- [8] L. Salaun, C. S. Chen, and M. Coupechoux, "Optimal joint subcarrier and power allocation in NOMA is strongly NP-hard," in *Proc. IEEE International Conference on Communications (ICC)*, May. 2018.
- [9] Z. Yang, Y. Liu, Y. Chen, and N. Al-Dhahir, "Cache-aided NOMA mobile edge computing: A reinforcement learning approach," Aug. 2019. [Online]. Available: arXiv preprint arXiv:1906.08812.
- [10] Y. Liu, Z. Qin, Y. Cai, Y. Gao, G. Y. Li, and A. Nallanathan, "UAV communications based on non-orthogonal multiple access," *IEEE Wireless Communications*, vol. 26, no. 1, pp. 52–57, Feb. 2019.
- [11] Z. Shi, H. Wang, Y. Fu, G. Yang, S. Ma, F. Hou, and T. A. Tsiftsis, "Zero-forcing based downlink virtual MIMO-NOMA communications in IoT networks," *IEEE Internet of Things Journal*, 2019.
- [12] Y. Fu, Y. Liu, H. Wang, Z. Shi, and Y. Liu, "Mode selection between index coding and superposition coding in cache-based NOMA networks," *IEEE Communications Letters*, vol. 23, no. 3, pp. 478–481, Mar. 2019.
- [13] Y. Fu, W. Wen, Z. Zhao, T. Q. S. Quek, S. Jin, and F.-C. Zheng, "Dynamic power control for NOMA transmissions in wireless caching networks," *IEEE Wireless Communications Letters*, vol. 8, no. 5, pp. 1485–1488, Oct. 2019.
- [14] Y. Fu, K. Shum, C. W. Sung, and Y. Liu, "Optimal user pairing in cache-based NOMA systems with index coding," in *Proc. IEEE International Conference on Communications (ICC)*, May. 2019.
- [15] O. Tervo, L.-N. Tran, and M. Juntti, "Optimal energy-efficient transmit beamforming for multi-user MISO downlink," *IEEE Transactions on Signal Processing*, vol. 63, no. 20, pp. 5574–5588, 2015.
- [16] H. Huh, H. C. Papadopoulos, and G. Caire, "Multiuser MISO transmitter optimization for intercell interference mitigation," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4272–4285, 2010.
- [17] Z. Wei, J. Yuan, D. W. K. Ng, E. Maged, and Z. Ding, "A survey of downlink non-orthogonal multiple access for 5G wireless communication networks," *ZTE Communications*, vol. 14, no. 4, pp. 17–25, Oct. 2016.
- [18] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access NOMA for future radio access," *IEEE Vehicular Technology Conference (VTC)*, Jun. 2013.
- [19] Q. Sun, S. Han, Z. Xu, S. Wang, I. Chih-Lin, and Z. Pan, "Sum rate optimization for MIMO non-orthogonal multiple access systems," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 747–752, Mar. 2015.
- [20] B. Kim, S. Lim, H. Kim, S. Suh, and J. Kwun, "Non-orthogonal multiple access in a downlink multiuser beamforming system," *IEEE Military Communications Conference*, pp. 1278–1283, Nov. 2013.
- [21] Z. Chen, Z. Ding, P. Xu, and X. Dai, "Optimal precoding for a QoS optimization problem in two-user MISO-NOMA downlink," *IEEE Communications Letters*, vol. 20, no. 6, pp. 1263–1266, Jun. 2016.
- [22] J. Cui, Y. Liu, Z. Ding, P. Fan, and A. Nallanathan, "Optimal user scheduling and power allocation for millimeter wave NOMA systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 3, pp. 1502–1517, Mar. 2018.
- [23] M. S. Ali, E. Hossian, 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, Mar. 2017.
- [24] H. Wang, Y. Fu, R. Song, Z. Shi, and X. Sun, "Power minimization precoding in uplink multi-antenna NOMA systems with jamming," *IEEE Transactions on Green Communications and Networking*, vol. 3, no. 3, pp. 591–602, Sep. 2019.
- [25] V. D. Nguyen, H. D. Tuan, T. Q. Duong, H. V. Poor, and O. S. Shin, "Precoder design for signal superposition in MIMO-NOMA multicell networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 12, pp. 2681–2695, Dec. 2017.
- [26] W. Shin, M. Vaezi, B. Lee, D. J. Love, J. Lee, and H. V. Poor, "Coordinated beamforming for multi-cell MIMO-NOMA," *IEEE Communications Letters*, vol. 21, no. 2, pp. 84–87, Jan. 2018.
- [27] M. Jiang, Y. Li, Q. Zhang, Q. Li, and J. Qin, "MIMO beamforming design in nonorthogonal multiple access downlink interference channels," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 8, pp. 6951–6959, Aug. 2018.
- [28] X. Sun, N. Yang, S. Yan, Z. Ding, D. W. K. Ng, C. Shen, and Z. Zhong, "Joint beamforming and power allocation in downlink NOMA multiuser MIMO networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 8, pp. 5367–5381, Aug. 2018.
- [29] W. Shao, S. Zhang, X. Zhang, J. Ma, and N. Zhao, "Suppressing interference and power allocation over the multi-cell MIMO-NOMA networks," *IEEE Communications Letters*, vol. 23, no. 8, pp. 1397–1400, Aug. 2019.
- [30] Y. Fu, L. Salaun, C. W. Sung, and C. S. Chen, "Distributed power allocation for the downlink of a two-cell MISO-NOMA system," in *Proc. IEEE Vehicular Technology Conference (VTC Spring)*, May. 2018.
- [31] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 528–541, Mar. 2006.
- [32] H. W. Kuhn, "Variants of the hungarian method for assignment problems," *Naval Research Logistics Quarterly*, vol. 3, no. 4, pp. 253–258, Jan. 1956.
- [33] J. Huang, R. A. Berry, and M. L. Honig, "Distributed interference compensation for wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 5, pp. 1074–1084, May 2006.
- [34] L. Tian, C. Yan, W. Li, Z. Yuan, W. Cao, and Y. Yuan, "On uplink non-orthogonal multiple access for 5G: Opportunities and challenges," *China Communications*, vol. 14, no. 12, pp. 142–152, Dec. 2017.
- [35] R. D. Yates, "A framework for uplink power control in cellular radio systems," *IEEE Transactions on Vehicular Technology*, vol. 13, no. 7, pp. 1341–1347, Sep. 1995.
- [36] C. W. Sung and W. S. Wong, "Mathematical aspects of the power control problem in mobile communication systems," *Lectures on Systems, Control and Information: Lectures at the Morningside Center of Mathematics*, vol. 17, pp. 127–172, 2000.
- [37] GreenTouch, *Mobile communications WG architecture doc2: Reference scenarios*, May. 2013.
- [38] A. I. Perez-Neira and M. R. Campalans, *Cross-Layer Resource Allocation in Wireless Communications: Techniques and Models from PHY and MAC Layer Interaction*. Academic Press, 2008.
- [39] Y. Li, Z. Xu, X. Cong, C. Yang, S. Zhang, C. Yan, and S. Xu, "Energy-efficient wireless communications: Tutorial, survey, and open issues," *IEEE Wireless Communications*, vol. 18, no. 6, pp. 28–35, Dec. 2011.