Joint Optimization of BS Clustering and Power Control for NOMA-Enabled CoMP Transmission in Dense Cellular Networks

Yanpeng Dai, *Member, IEEE*, Junyu Liu, *Member, IEEE*, Min Sheng, *Senior Member, IEEE*, Nan Cheng, *Member, IEEE*, and Xuemin Shen, *Fellow, IEEE*

***Abstract*—Non-orthogonal multiple access (NOMA)-enabled coordinated multipoint (CoMP) transmission has great potential in balancing spectrum utilization and interference mitigation in dense cellular networks. However, NOMA reforms the spectrum sharing policy of CoMP transmission due to introducing additional interference among coordinated base stations (BSs), which deteriorates the CoMP transmission rate. In this paper, we investigate the joint optimization of BS clustering and power control for NOMA-enabled CoMP transmission in dense cellular networks to maximize system sum-rate. We first characterize the impact of interference among coordinated BSs on CoMP transmission rate and find that the BS clustering is dependent on NOMA user grouping and restricted by the NOMA decoding condition. We then derive a tight lower bound on the system sum-rate and exploit it to design a joint BS clustering and power control scheme. Specifically, a power control algorithm is designed by a penalty convex-concave procedure to satisfy the NOMA decoding condition and users' rate requirements. A BS clustering algorithm based on successive convex approximation is designed to iteratively update the BS clustering and NOMA user grouping to increase system sum-rate. Finally, two algorithms are alternately performed until all users successfully access and the system sum-rate converges. Simulation results show that the proposed scheme can efficiently alleviate interference among coordinated BSs to improve system sum-rate and spectrum efficiency of NOMA-enabled CoMP transmission even under high user density.**

***Index Terms*—Dense cellular networks, CoMP, NOMA, resource allocation.**

Manuscript received September 12, 2020; revised December 28, 2020; accepted January 25, 2021. Date of publication February 1, 2021; date of current version March 10, 2021. This work was supported in part by the Natural Science Foundation of China under Grants 62002042, 61725103, 61931005, and 62071356, in part by Young Elite Scientists Sponsorship Program by CAST, and in part by Cooperative Scientific Research Project of Chunhui Program by Ministry of Education of China. The review of this article was coordinated by Prof. K. Le. *(Corresponding author: Junyu Liu.)*

Yanpeng Dai is with the School of Information Science and Technology, Dalian Maritime University, Dalian 116086, China (e-mail: yanpdai@gmail.com).

Junyu Liu, Min Sheng, and Nan Cheng are with the State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China (e-mail: seeker.junyuliu@gmail.com; mshengxd@gmail.com; wmchengnan@gmail.com).

Xuemin Shen is with the Department of Electrical, and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: sshen@uwaterloo.ca).

Digital Object Identifier 10.1109/TVT.2021.3055769

I. Introduction

NETWORK densification is an irresistible trend for wireless networks in the era of fifth-generation (5G) and beyond to fulfill tremendous capacity requirement [1], [2]. In order to mitigate severe inter-cell interference, ultra-dense networks (UDN) employ the coordinated multipoint (CoMP) joint transmission which enables a cluster of BSs to simultaneously provide the signal transmission for a single user to enhance the desired signal power and eliminate interference power [3], [4]. However, since a single user occupies the spectrum of multiple adjacent cells, CoMP leads to low spectrum utilization and reduces the user-access capability. Therefore, how to satisfy the access requirements of massive users is a crucial problem for CoMP systems in UDN.

To enhance the user-access capability, non-orthogonal multiple access (NOMA) is recently introduced into CoMP systems [5], [6] to allow the CoMP user to form NOMA groups with non-CoMP users served by the single-point transmission. By utilizing NOMA, coordinated BSs can use identical spectrum resources to simultaneously serve the CoMP user and non-CoMP users, which thus efficiently improves spectrum utilization. Extensive research has been done on NOMA-enabled CoMP transmission [7]–[11]. Through performance analysis, J. Choi in [7] showed that NOMA could improve the transmission rate when two coordinated BSs serve one CoMP user and two non-CoMP users. Y. Tian *et al.* in [8] proposed an opportunistic NOMA scheme in CoMP systems where BSs are clustered in an opportunistic manner. M. S. Ali *et al.* in [9] and Y. Al-Eryani *et al.* in [10] designed the power control algorithms for NOMA-enabled CoMP transmission in distributed and centralized patterns, respectively.

The BS clustering for CoMP transmission is as important as power control of coordinated BSs, which however becomes more difficult in NOMA-UDN due to complicated interference. To be specific, the dense network deployment incurs the line-of-sight (LOS) path between BSs and users [3], [12]. Although the LOS path enhances the desired signal power, it increases the interference power as well. Thus, the association between BSs and CoMP users should be carefully designed to avoid severe interference over LOS paths. Furthermore, in NOMA-UDN, the BS clustering not only determines the BS selection of CoMP users but also decides the NOMA user grouping between CoMP

users and non-CoMP users. The interference in the NOMA group can spread to other users in the same BS cluster, which directly reduces their transmission rates. Thus, it is necessary to optimize the BS clustering for CoMP transmission in NOMA-UDN by considering the impact of NOMA user grouping.

The scheduling for CoMP systems in NOMA-UDN faces several challenges due to the coexistence of BS clusters and NOMA groups. First, the CoMP transmission suffers from both the interference in the BS cluster and interference from other clusters. This is fundamentally different with CoMP transmission in the orthogonal multiple access (OMA) networks where the interference is only from other clusters [13], [14]. Thus, the BS clustering scheme for OMA networks is intuitively not applicable to CoMP in NOMA networks. Second, severe interference via LOS paths in UDN makes the NOMA decoding condition unsatisfied to interrupt the signal transmission of CoMP users and non-CoMP users. Hence, the interference management for NOMA becomes more difficult compared to existing works [13], [15], [16]. Third, the power control should adapt to the optimization of BS clustering and be capable of tolerating severe interference in UDN to avoid the solving failure [2], [9], [17]. Therefore, it is essential to jointly design BS clustering and power control for releasing the potential of NOMA-enabled CoMP transmission in dense cellular networks.

In this paper, we investigate the BS clustering and power control problem for CoMP joint transmission in NOMA-UDN to increase system sum-rate while meeting the access requirements of all users. We depict the interference between the CoMP user and non-CoMP users in one NOMA group and characterize its impact on the CoMP transmission rate. Then, a system sum-rate maximization problem is formulated to jointly optimize the BS clustering and power control while maintaining the required transmission rates of all users. We analyze the structure of formulated system sum-rate maximization problem which can be decomposed into two subproblems, a power control subproblem and a BS clustering problem. We show that the power control subproblem with given BS clustering is a difference-of-convex (D.C.) programming problem and can be solved by the penalty convex-concave procedure (CCP). The BS clustering subproblem with the given transmit power is shown to be a quadratic binary problem that can be transformed into a concave problem via variable substitution and solved by successive convex approximation. Finally, a joint BS clustering and power control scheme is proposed based on the alternating direction method. The main contributions of our paper are summarized as follows.

- We analyze the power control algorithm to show that the penalty CCP method can effectively overcome the issue of the initial point infeasibility caused by severe interference in UDN. This issue results in the access failure of users especially when the user density is high.
- We derive a tight lower bound on the system sum-rate and exploit it to design a joint BS clustering and power control scheme. It is proved that this lower bound increases with the optimization of power control and BS clustering.
- We propose a joint BS clustering and power control scheme which is proved to obtain a series of non-decreasing values of system sum-rates to approach the optimal solution until convergence. The results show that when the user density is high, the proposed scheme can further increase the system sum-rate whereas the system sum-rates of traditional schemes decrease.

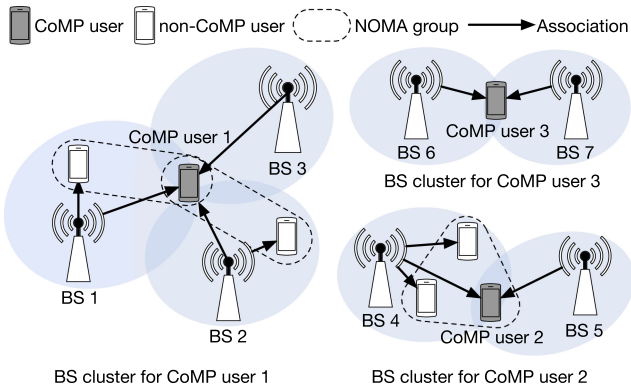

Fig. 1. The BS clusters to perform CoMP transmission in NOMA-UDN, where the CoMP user and non-CoMP user is allowed to coexist in one NOMA group.

Note that CoMP or CoMP transmission in the remainder of this paper refers to CoMP joint transmission for convenience. The remainder of the paper is organized as follows. The network scenario, NOMA-enabled CoMP transmission model, and problem formulation is given in Section II. Then, the problem decomposition and proposed algorithms are presented in Section III. Simulation results are given in Section IV. Finally, we conclude this paper in Section V.

II. System Model and Problem Formulation

In this section, we introduce the network model and the NOMA-enabled CoMP transmission model. Then, the problem formulation is given.

A. Network Model

We consider the CoMP transmission in downlink small-cell networks (see Fig. 1), where NOMA is adopted as the multiple access technique. Multiple BSs are densely deployed and share the same spectrum resource. Let $\mathcal{B} = \{1, 2, \ldots, K\}$ denote the set of all BSs. Users are classified into CoMP users and non-CoMP users according to the reference signal received power (RSRP) [9], [14], [18]. The user is labelled as a CoMP user if the differences of RSRPs between a user and near BSs are small enough, and otherwise the user is labeled as a non-CoMP user. Let $\mathcal{C} = \{1, 2, \ldots, N\}$ denote the set of all CoMP users and $\mathcal{M}_k = \{1, 2, \ldots, M_k\}$ denote the set of non-CoMP users associated with BS k. Each CoMP user is served by a cluster of BSs and can receive the same desired signal from multiple BSs in the cluster through joint transmission. Single-point transmission is performed between each non-CoMP user and its associated BS. Each BS can be associated with at most D non-CoMP users, i.e., $0 \leqslant M_k \leqslant D, \forall k \in \mathcal{B}$. The association between non-CoMP users and BSs is performed according to the nearest association principle [19], [20]. Note that each BS is typically allowed to serve at most one CoMP user in practical system. The clustering of BSs requires frequent signaling interactions between

coordinated BSs to finish the synchronization and information exchange [18]. Hence, we limit that the maximum size of a BS cluster is C to maintain a reasonable amount of signaling overheads [21], [22].

Let $h^{\mathrm{C}}_{n,k}$ denote the channel power gain between CoMP user n and BS k and $h^{\mathrm{N}}_{m,k}$ denote the channel power gain between a non-CoMP user m and BS k. The channel power gain includes the pathloss and channel fading. Especially, we adopt the multiple-slope pathloss model [3] to characterize the existence of LOS path and non-LOS (NLOS) path which is defined by $\zeta(\{\alpha_l\}_{l=0}^{L-1}; d) = A_l d^{-\alpha_n}$, $\bar{d}_l \leqslant d < \bar{d}_{l+1}$. d is the distance from the transmitter to the receiver and $\bar{d}_l$ denotes the corner distance. α_l denotes the pathloss exponent within $[\bar{d}_l, \bar{d}_{l+1}]$. A_l denotes pathlosses at a reference distance $d = 1$.

The receiver in downlink NOMA networks adopts the successive interference cancellation (SIC) decoding to detect the signals in the ascending order of channel quality coefficients from the BS to associated users [23]–[26]. Once the signal is successfully decoded, it would be subtracted from the superposed signal. Hence, the user with higher channel quality coefficient can cause interference for other users with lower channel quality coefficients in NOMA group. Different the single-cell NOMA networks, channel quality coefficients in multi-cell NOMA networks are influenced by channel power gain as well as inter-cell interference. The channel quality coefficient of non-CoMP user m associated with BS k is denoted as $g_{m,k} = \frac{h^{\mathrm{N}}_{m,k}}{I_{m,k}+\sigma^2}$, where $I_{m,k}$ denotes inter-cell interference received by non-CoMP user m, and σ^2 denotes the additive white Gaussian noise power. The detailed expression of $I_{m,k}$ is provided in latter. Let $\mathcal{G}_{m,k} = \{n \mid g_{m,k} < g_{n,k}, \forall n \in \mathcal{M}_k\}, \forall m \in \mathcal{M}_k, \forall k \in \mathcal{B}$ denote the set of non-CoMP users which are associated with BS k and cause interference to non-CoMP user m in the NOMA group.

B. NOMA-Enabled CoMP Transmission Model

In NOMA-UDN, a CoMP user is allowed to form NOMA groups with non-CoMP users which are associated with coordinated BSs. The BS cluster simultaneously provides the signal transmissions for the CoMP user and non-CoMP users on the same spectrum resource, which is shown in Fig. 1. Thus, a CoMP user could be included in multiple NOMA groups such as CoMP user 1 in Fig. 1. The decoding order of a CoMP user should be the same in all NOMA groups due to the simultaneous signal reception from multiple coordinated BSs [6], [10]. Furthermore, CoMP users are generally located at cell edge and is likely to have worse channel quality coefficients than non-CoMP users [27]. Therefore, we consider that the signal of the CoMP user is decoded prior to the signals of non-CoMP users in NOMA transmission, which has proven to be an effective and practical manner to conform the preference of NOMA to the distinct difference of the channel power gains between users [6], [9]. Fig. 2 shows an example of NOMA-enabled CoMP transmission. Specifically, the non-CoMP user first decodes and subtracts the signal of the CoMP user and then decodes its own signal without the interference from the CoMP user. The CoMP user directly decodes its signal with interference from the non-CoMP user.

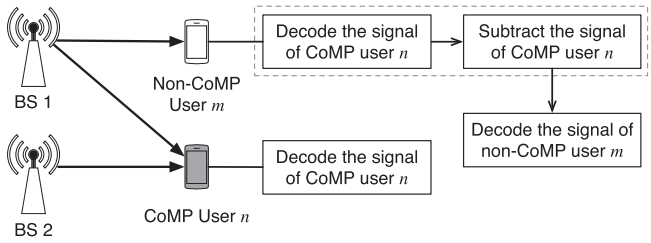

Fig. 2. An illustrative example of NOMA-enabled CoMP transmission.

The received signal at CoMP user n is expressed as

$$\begin{aligned} y^{\mathrm{C}}_n =& \sum_{m\in\mathcal{C}}\sum_{k\in\mathcal{B}} x_{m,k}\sqrt{h^{\mathrm{C}}_{n,k}p^{\mathrm{C}}_k}\alpha_m + \sum_{k\in\mathcal{B}}\sum_{m\in\mathcal{M}_k}\sqrt{h^{\mathrm{C}}_{n,k}p^{\mathrm{N}}_{m,k}}\beta_m + \sigma^2 \\ =& \sigma^2 + \underbrace{\sum_{k\in\mathcal{B}} x_{n,k}\sqrt{h^{\mathrm{C}}_{n,k}p^{\mathrm{C}}_k}\alpha_n}_{\text{desired signal}} + \underbrace{\sum_{k\in\mathcal{B}}\sum_{m\in\mathcal{M}_k}\sqrt{h^{\mathrm{C}}_{n,k}p^{\mathrm{N}}_{m,k}}\beta_m}_{\text{interference from non-CoMP users}} \\ &+ \underbrace{\sum_{m\in\mathcal{C}\setminus n}\sum_{k\in\mathcal{B}} x_{m,k}\sqrt{h^{\mathrm{C}}_{n,k}p^{\mathrm{C}}_k}\alpha_m}_{\text{interference from other CoMP users}} \end{aligned} \tag{1}$$

where p^{C}_k denotes the transmit power of BS k for CoMP transmission, and $p^{\mathrm{N}}_{m,k}$ denotes the transmit power of BS k for its non-CoMP m. α_n denotes the transmit signal of CoMP user n, and β_m denotes the transmit signal of non-CoMP user m. $x_{n,k}$ denotes the BS clustering indicator of CoMP transmission. $x_{n,k} = 1$ denotes that CoMP user n is associated with BS k, and $x_{n,k} = 0$ denotes the opposite. Each CoMP user receives the desired signal from multiple coordinated BSs and meanwhile suffers the interference from both non-CoMP users and other CoMP users. Hence, the signal-to-interference-and-noise-ratio (SINR) of CoMP user n is expressed as

$$\xi^{\mathrm{C}}_n = \frac{\sum_{k\in\mathcal{B}} x_{n,k}p^{\mathrm{C}}_k h^{\mathrm{C}}_{n,k}}{\sum_{k\in\mathcal{B}}\sum_{m\in\mathcal{M}_k} p^{\mathrm{N}}_{m,k}h^{\mathrm{C}}_{n,k} + \sum_{m\in\mathcal{C}\setminus n}\sum_{k\in\mathcal{B}} x_{m,k}p^{\mathrm{C}}_k h^{\mathrm{C}}_{n,k} + \sigma^2}, \tag{2}$$

The transmission rate of CoMP user n is expressed as $R^{\mathrm{C}}_n = \log_2(1 + \xi^{\mathrm{C}}_n)$.

The received signal at non-CoMP user m associated with BS k is expressed as

$$\begin{aligned} y^{\mathrm{N}}_{m,k} =& \sum_{j\in\mathcal{B}}\sum_{n\in\mathcal{M}_j}\sqrt{h^{\mathrm{N}}_{m,j}p^{\mathrm{N}}_{n,j}}\beta_n + \sum_{n\in\mathcal{C}}\sum_{j\in\mathcal{B}} x_{n,j}\sqrt{h^{\mathrm{N}}_{m,j}p^{\mathrm{C}}_j}\alpha_n + \sigma^2 \\ =& \sigma^2 + \underbrace{\sqrt{h^{\mathrm{N}}_{m,k}p^{\mathrm{N}}_{m,k}}\beta_m}_{\text{desired signal}} \\ &+ \sum_{n\in\mathcal{M}_k\setminus m}\sqrt{h^{\mathrm{N}}_{m,j}p^{\mathrm{N}}_{n,j}}\beta_n + \sum_{n\in\mathcal{C}} x_{n,k}\sum_{j\in\mathcal{B}} x_{n,j}\sqrt{h^{\mathrm{N}}_{m,j}p^{\mathrm{C}}_j}\alpha_n \\ &+ \underbrace{\sum_{j\in\mathcal{B}\setminus k}\sum_{n\in\mathcal{M}_j}\sqrt{h^{\mathrm{N}}_{m,j}p^{\mathrm{N}}_{n,j}}\beta_n + \sum_{n\in\mathcal{C}}(1-x_{n,k})\sum_{j\in\mathcal{B}} x_{n,j}\sqrt{h^{\mathrm{N}}_{m,j}p^{\mathrm{C}}_j}\alpha_n}_{\text{inter-cell interference}}. \end{aligned} \tag{3}$$

Thus, the inter-cell interference for non-CoMP user m is given by $I_{m,k} = \sum_{j\in\mathcal{B}\setminus k}\sum_{n\in\mathcal{M}_j} p^{\mathrm{N}}_{n,j}h^{\mathrm{N}}_{m,j} + \sum_{n\in\mathcal{C}}(1-x_{n,k})\sum_{j\in\mathcal{B}\setminus k} x_{n,j}p^{\mathrm{C}}_j h^{\mathrm{N}}_{m,j}$. Although non-CoMP user m receives the signals of all users in the NOMA group, it can adopt the SIC decoding to cancel a part of them and is only interfered with users in $\mathcal{G}_{m,k}$. Hence, the SINR of non-CoMP user m associated with BS k is expressed as

$$\xi^{\mathrm{N}}_{m,k} = \frac{p^{\mathrm{N}}_{m,k}h^{\mathrm{N}}_{m,k}}{\sum_{n\in\mathcal{G}_{m,k}} p^{\mathrm{N}}_{n,k}h^{\mathrm{N}}_{m,k} + I_{m,k} + \sigma^2}. \tag{4}$$

The transmission rate of non-CoMP user m is expressed as $R^{\mathrm{N}}_{m,k} = \log_2(1+\xi^{\mathrm{N}}_{m,k})$. However, the SIC decoding condition should be satisfied to realize the interference cancellation. If non-CoMP user m intends to decode and subtract the signal of CoMP user n supposing that both of them are associated with BS k, the following condition should be met [9], [10]:

$$\sum_{j\in\mathcal{B}} x_{n,j}h^{\mathrm{N}}_{m,j}p^{\mathrm{C}}_j \geqslant \theta\left(\sum_{j\in\mathcal{B}}\sum_{i\in\mathcal{M}_j} p^{\mathrm{N}}_{i,j}h^{\mathrm{N}}_{m,j} + \sum_{i\in\mathcal{C}\setminus n}\sum_{j\in\mathcal{B}} x_{i,j}p^{\mathrm{C}}_j h^{\mathrm{N}}_{m,j} + \sigma^2\right) \tag{5}$$

where θ denotes the SIC threshold. (5) means that the received signal strength of CoMP user n at non-CoMP user m must be at least θ times greater than the sum signal strength of all other users associated with BS k [6]. If non-CoMP user m intends to decode and subtract the signal of non-CoMP user m' supposing $m' \in \mathcal{M}_k \setminus \mathcal{G}_{m,k}$, the following condition should be satisfied:

$$p^{\mathrm{N}}_{m',k}h^{\mathrm{N}}_{m,k} \geqslant \theta\left(\sum_{n\in\mathcal{G}_{m',k}} p^{\mathrm{N}}_{n,k}h^{\mathrm{N}}_{m,k} + I_{m,k} + \sigma^2\right), \tag{6}$$

which shows the SIC decoding condition for non-CoMP user m to decode the signal of non-CoMP user m' [5], [17], [23], [24].

C. Problem Formulation

Let $\boldsymbol{X} = \{x_{n,k}, \forall n\in\mathcal{C}, \forall k\in\mathcal{B}\}$ denote the indicator variables for BS clustering. Let $\boldsymbol{P}^{\mathrm{C}} = \{p^{\mathrm{C}}_n, \forall n\in\mathcal{C}\}$ and $\boldsymbol{P}^{\mathbf{N}} = \{p^{\mathrm{N}}_{m,k}, \forall m\in\mathcal{M}_k, \forall k\in\mathcal{B}\}$ denote the transmit power variables for CoMP transmission and non-CoMP users, respectively. In order to design the BS clustering and power control scheme, we formulate a system sum-rate maximization (SRM) problem as follows.

$$(\mathbf{SRM}): \max_{\boldsymbol{X},\boldsymbol{P}^{\mathrm{C}},\boldsymbol{P}^{\mathbf{N}}} \sum_{k\in\mathcal{B}}\sum_{m\in\mathcal{M}_k} R^{\mathrm{N}}_{m,k} + \sum_{n\in\mathcal{C}} R^{\mathrm{C}}_n$$

$$\text{s.t. C1: } \sum_{n\in\mathcal{C}} x_{n,k} \leqslant 1, \forall k\in\mathcal{B}$$

$$\text{C2: } \sum_{k\in\mathcal{B}} x_{n,k} \leqslant C, \forall n\in\mathcal{C}$$

$$\text{C3: } x_{n,k}\in\{0,1\}, \forall n\in\mathcal{C}, \forall k\in\mathcal{B}$$

$$\text{C4: } \sum_{m\in\mathcal{M}_k} p^{\mathrm{N}}_{m,k} + p^{\mathrm{C}}_k \leqslant P_{\max}, \forall k\in\mathcal{B}$$

$$\text{C5: } \xi^{\mathrm{C}}_n \geqslant \xi_{\min}, \forall n\in\mathcal{C}$$

$$\text{C6: } \xi^{\mathrm{N}}_{m,k} \geqslant \xi_{\min}, \forall k\in\mathcal{B}, \forall m\in\mathcal{M}_k$$

$$\text{C7: } (5), \forall n\in\mathcal{C}, \forall m\in\mathcal{M}_k, \forall k\in\mathcal{B}$$

$$\text{C8: } (6). \forall m\in\mathcal{M}_k, \forall m'\in\mathcal{M}_k\setminus\mathcal{G}_{m,k}, \forall k\in\mathcal{B}$$

In Problem (SRM), C1 and C2 indicate that each BS serves at most one CoMP user, and each CoMP user is served by at most C BSs, respectively. C4 restricts the maximum transmit power $P_{\max}$ for each BS. C5 and C6 specify the SINR threshold of CoMP users and non-CoMP users, respectively. C7 and C8 specify the SIC threshold for CoMP users and non-CoMP users, respectively.

Problem (SRM) is a combinatorial and non-convex problem due to the binary variable $\boldsymbol{X}$ and non-convex objective function. In general, there is no computationally efficient and systematic method to optimally solve this kind of problems [28]. Thus, we focus on designing an efficient algorithm to solve it with moderate complexity.

III. Problem Solution and Algorithm Design

In this section, we propose a scheme to tackle Problem (SRM). It is observed that Problem (SRM) can be decomposed into two subproblems, namely, the BS clustering subproblem and the power control subproblem. We first solve the power control subproblem for any given BS clustering and then solve the BS clustering subproblem for any given transmit power. Finally, a joint optimization scheme of the BS clustering and power control is proposed to solve Problem (SRM) based on the alternating optimization method.

A. Power Control for Multiple BS Clusters and Cells

In NOMA-UDN, the transmit power of each BS determines the desired signal strength as well as interference in NOMA groups and inter-cell interference. In this subsection, we investigate the power control subproblem with the given BS clustering. Let $\boldsymbol{P} = (\boldsymbol{P}^{\mathrm{C}}, \boldsymbol{P}^{\mathbf{N}})$ denote the vector of all transmit power variables. This subproblem is formulated as

$$\mathbf{P1}: \max_{\boldsymbol{P}} \sum_{k\in\mathcal{B}}\sum_{m\in\mathcal{M}_k} R^{\mathrm{N}}_{m,k} + \sum_{n\in\mathcal{C}} R^{\mathrm{C}}_n$$

$$\text{s.t. C4–C8.}$$

Although all the constraints in P1 are linear under given BS clustering, the interference among users makes the objective function non-convex.

Let $\mathcal{C}_k = \{j \mid x_{n,k}=1, x_{n,j}=1, \forall n\in\mathcal{C}, \forall j\in\mathcal{B}\setminus k\}$ denote the set of BSs which are in the same BS cluster with BS k. Let $\mathcal{B}_n = \{k \mid x_{n,k}=1, \forall k\in\mathcal{B}\}$ denote the BS cluster for CoMP user n. The objective function is rewritten by

$$Q\left(\boldsymbol{P}\right)=\sum_{k\in\mathcal{B}}\sum_{m\in\mathcal{M}_k}\left(g_m^{\mathrm{N}}\left(\boldsymbol{P}\right)-f_m^{\mathrm{N}}\left(\boldsymbol{P}\right)\right)+\sum_{n\in\mathcal{C}}\left(g_n^{\mathrm{C}}\left(\boldsymbol{P}\right)-f_n^{\mathrm{C}}\left(\boldsymbol{P}\right)\right), \tag{7}$$

$$g_m^{\mathrm{N}}\left(\boldsymbol{P}\right)=\log_2\left(p_{m,k}^{\mathrm{N}}h_{m,k}^{\mathrm{N}}+I_{m,k}^{\mathrm{N}}+\sum_{j\in\mathcal{B}\backslash\mathcal{C}_k}p_j^{\mathrm{C}}h_{m,j}^{\mathrm{N}}+\sigma^2\right), \tag{8}$$

$$f_m^{\mathrm{N}}\left(\boldsymbol{P}\right)=\log_2\left(I_{m,k}^{\mathrm{N}}+\sum_{j\in\mathcal{B}\backslash\mathcal{C}_k}p_j^{\mathrm{C}}h_{m,j}^{\mathrm{N}}+\sigma^2\right), \tag{9}$$

$$g_n^{\mathrm{C}}\left(\boldsymbol{P}\right)=\log_2\left(\sum_{k\in\mathcal{B}}\left(\sum_{m\in\mathcal{M}_k}p_{m,k}^{\mathrm{N}}+p_k^{\mathrm{C}}\right)h_{n,k}^{\mathrm{C}}+\sigma^2\right), \tag{10}$$

$$f_n^{\mathrm{C}}\left(\boldsymbol{P}\right)=\log_2\left(\sum_{k\in\mathcal{B}}\sum_{m\in\mathcal{M}_k}p_{m,k}^{\mathrm{N}}h_{n,k}^{\mathrm{C}}+\sum_{k\in\mathcal{B}\backslash\mathcal{B}_n}p_k^{\mathrm{C}}h_{n,k}^{\mathrm{C}}+\sigma^2\right), \tag{11}$$

where $g_m^{\mathrm{N}}(\boldsymbol{P})$, $f_m^{\mathrm{N}}(\boldsymbol{P})$, $g_n^{\mathrm{C}}(\boldsymbol{P})$, and $f_n^{\mathrm{C}}(\boldsymbol{P})$ are the concave functions. Hence, $Q(\boldsymbol{P})$ is a D.C. function, and P1 is a typical D.C. programming problem [29]–[31].

According to the first-order Taylor approximation, the following inequations are given

$$f_m^{\mathrm{N}}\left(\boldsymbol{P}\right)\leqslant f_m^{\mathrm{N}}\left(\boldsymbol{P}\left[t\right]\right)+\nabla f_m^{\mathrm{N}}\left(\boldsymbol{P}\left[t\right]\right)^{\mathrm{T}}\left(\boldsymbol{P}-\boldsymbol{P}\left[t\right]\right), \tag{12}$$

$$f_n^{\mathrm{C}}\left(\boldsymbol{P}\right)\leqslant f_n^{\mathrm{C}}\left(\boldsymbol{P}\left[t\right]\right)+\nabla f_n^{\mathrm{C}}\left(\boldsymbol{P}\left[t\right]\right)^{\mathrm{T}}\left(\boldsymbol{P}-\boldsymbol{P}\left[t\right]\right), \tag{13}$$

which respectively show the lower bounds of $f_m^{\mathrm{N}}(\boldsymbol{P})$ and $f_n^{\mathrm{C}}(\boldsymbol{P})$ at the point $\boldsymbol{P}[t]$. The equality holds when $\boldsymbol{P}=\boldsymbol{P}[t]$. Therefore, the optimal solution to P1 can be lower-bounded by the following problem:

$$\begin{aligned}\mathbf{P2}:\max_{\boldsymbol{P}}\ \ & Q\left(\boldsymbol{P};\boldsymbol{P}\left[t\right]\right)=\sum_{k\in\mathcal{B}}\sum_{m\in\mathcal{M}_k}g_m^{\mathrm{N}}\left(\boldsymbol{P}\right)+\sum_{n\in\mathcal{C}}g_n^{\mathrm{C}}\left(\boldsymbol{P}\right)\\&-\sum_{k\in\mathcal{B}}\sum_{m\in\mathcal{M}_k}\left(f_m^{\mathrm{N}}\left(\boldsymbol{P}\left[t\right]\right)+\nabla f_m^{\mathrm{N}}\left(\boldsymbol{P}\left[t\right]\right)^{\mathrm{T}}\left(\boldsymbol{P}-\boldsymbol{P}\left[t\right]\right)\right)\\&-\sum_{n\in\mathcal{C}}\left(f_n^{\mathrm{C}}\left(\boldsymbol{P}\left[t\right]\right)+\nabla f_n^{\mathrm{C}}\left(\boldsymbol{P}\left[t\right]\right)^{\mathrm{T}}\left(\boldsymbol{P}-\boldsymbol{P}\left[t\right]\right)\right)\\\text{s.t.}\ \ & \text{C4–C8},\end{aligned}$$

which is a typical convex programming problem and can be solved with the standard convex optimization methods, such as the interior point method [28]. A series of convex programming problems based on P2 can be constructed and solved in an iterative manner until the convergence is achieved. $t_{\max}$ denote the maximum number of iterations. The set $\mathcal{G}_{m,k},\forall m\in\mathcal{M}_k,\forall k\in\mathcal{B}$ is updated dynamically during the iteration procedure which can be determined at the beginning of each iteration according to the point $\boldsymbol{P}[t]$. Then, we give the following theorem.

Theorem 1: The sequence of optimal values $\{Q(\boldsymbol{P};\boldsymbol{P}[t])\}_{t=0}^{t_{\max}}$ for P2 is non-decreasing.

Proof: See Appendix A. ■

However, severe interference in UDN make it difficult for P2 to seek out an initial feasible point to start the iteration even though given BS clustering result is valid. The penalty CCP is an efficient technique to solve D.C. programming problem without the needs of the initial feasible point [32], [33]. The penalty CCP relaxes the problem by adding the slack variables and penalizing them in the objective function. For P2, we introduce the nonnegative variables $\{s_n^{\mathrm{c5}}\}$, $\{s_{m,k}^{\mathrm{c6}}\}$, $\{s_{m,n,k}^{\mathrm{c7}}\}$ and $\{s_{m,n,k}^{\mathrm{c8}}\}$ as slack variables to transform constraint C5, C6, C7, and C8, respectively. For example, after adding $\{s_n^{\mathrm{c5}}\}$, C5 transforms into C5a:$\xi_{\min}-\xi_n^{\mathrm{C}}\leqslant s_n^{\mathrm{c5}},\forall n\in\mathcal{C}$. Similarly, C6a, C7a, and C8a can be obtained from C6, C7, and C8, respectively. Furthermore, P2 is transformed into the following form:

$$\begin{aligned}\mathbf{P3}:\max_{\boldsymbol{P},\boldsymbol{S}}\ \ & Q\left(\boldsymbol{P};\boldsymbol{P}\left[t\right]\right)-\mu\sum_{i=1}^{4}S_i\\\text{s.t.}\ \ & \text{C4, C5a, C6a, C7a, C8a},\end{aligned}$$

where μ is a positive number as the penalty factor. In addition, the sums of slack variables are denoted by $S_1=\sum_{n\in\mathcal{C}}s_n^{\mathrm{c5}}$, $S_2=\sum_{k\in\mathcal{B}}\sum_{m\in\mathcal{M}_k}s_{m,k}^{\mathrm{c6}}$, $S_3=\sum_{k\in\mathcal{B}}\sum_{n<m,\{m,n\}\in\mathcal{M}_k}s_{m,n,k}^{\mathrm{c7}}$, and $S_4=\sum_{n\in\mathcal{C}}\sum_{k\in\mathcal{B}}\sum_{m\in\mathcal{M}_k}s_{m,n,k}^{\mathrm{c8}}$.

Algorithm 1: Power Control Algorithm.

1: **Initialization**
2: Given the BS clustering. Input μ and $\Delta>1$.
3: Set $t=0$, $\boldsymbol{P}[0]=0$ and $V[0]=+\infty$.
4: **repeat**
5: Update $\mathcal{G}_{m,k},\forall m\in\mathcal{M}_k,\forall k\in\mathcal{B}$ according to $\boldsymbol{P}[t]$.
6: Solve P3 based on $\boldsymbol{P}[t]$ to obtain the optimal solution $\boldsymbol{P}^*$ and the optimal value V^*.
7: Update $\boldsymbol{P}[t+1]=\boldsymbol{P}^*$ and $V[t+1]=V^*$.
8: $\varepsilon\leqslant|V[t+1]-V[t]|$.
9: Set $t=t+1$ and $\mu=\min\{\Delta\mu,\mu_{\max}\}$.
10: **until** $\varepsilon\leqslant\varepsilon_{\min}$ **or** $t\geqslant t_{\max}$
11: **if** $\mu\sum_{i=1}^{4}S_i>S_{\min}$ **then**
12: The problem is infeasible.
13: **end if**
14: **return** $\boldsymbol{P}^*$ if the problem is feasible.

According to P3, a power control algorithm is proposed which is shown in Algorithm 1. In Algorithm 1, steps 4–10 sequentially solve P3 to increase the system sum-rate until the convergence condition is met or $t_{\max}$ is achieved. $\mathcal{G}_{m,k},\forall m\in\mathcal{M}_k,\forall k\in\mathcal{B}$ is updated according to the point $\boldsymbol{P}[t]$ before solving P3. Since the penalty value in the objective function increases with the growth of μ, the maximization problem will make $\boldsymbol{S}$ gradually approach zero. If the penalty value is sufficiently small when the convergence is achieved, a suboptimal solution of power control subproblem P1 is found. Furthermore, we can know that the convergence of Algorithm 1 is guaranteed especially when $\mu=\mu_{\max}$ according to Theorem 1.

Remark 1: Algorithm 1 has the advantage of reducing solving failure because it can tolerate severe interference at the initial phase of iteration. Therefore, Algorithm 1 can be applicable in UDN with high densities of BSs and users.

B. BS Clustering for CoMP Transmission

The association between CoMP users and BSs should be carefully designed under the practical limitation of the size of BS cluster. In this subsection, we investigate the BS clustering subproblem with the given transmit power which is formulated as

$$\mathbf{P4}: \max_{\boldsymbol{X}} \sum_{k\in\mathcal{B}}\sum_{m\in\mathcal{M}_k} R^{\mathrm{N}}_{m,k} + \sum_{n\in\mathcal{C}} R^{\mathrm{C}}_{n}$$
$$\text{s.t.} \quad \text{C1–C3 and C5–C8.}$$

Although the transmit power variables are fixed, P4 still cannot be solved directly due to the undetermined BS clustering and the interference among users. Specifically, the undetermined BS clustering results in the product of variables which makes P4 quadratic. To address this issue, we introduce the new binary variables $\bar{\boldsymbol{X}} = \{x_{n,k,j} \mid n \in \mathcal{C}, k \neq j, k \in \mathcal{B}, j \in \mathcal{B}\}$. $x_{n,k,j} = 1$ indicates that CoMP user n is simultaneously associated with BS k and BS j, and $x_{n,k,j} = 0$ indicates the opposite. Accordingly, $\xi^{\mathrm{N}}_{m,k}$ is equivalently rewritten as

$$\xi^{\mathrm{N}}_{m,k} = \frac{p^{\mathrm{N}}_{m,k} h^{\mathrm{N}}_{m,k}}{\bar{I}^{\mathrm{N}}_{m,k} - \sum_{n\in\mathcal{C}}\sum_{j\in\mathcal{B}\setminus k} x_{n,k,j} p^{\mathrm{C}}_{j} h^{\mathrm{N}}_{m,j} + \sigma^2}, \tag{14}$$

where $\bar{I}^{\mathrm{N}}_{m,k} = \sum_{n\in\mathcal{G}_{m,k}} p^{\mathrm{N}}_{n,k} h^{\mathrm{N}}_{m,k} + \sum_{j\in\mathcal{B}\setminus k} p^{\mathrm{T}}_{j} h^{\mathrm{N}}_{m,j}$. $p^{\mathrm{T}}_{j} = \sum_{m\in\mathcal{M}_j} p^{\mathrm{N}}_{m,j} + p^{\mathrm{C}}_{j}$ denotes the total transmit power consumed by BS j. Similarly, C7 and C8 can be rewritten as (15) and (16), respectively.

$$\mathrm{C9}: (1+\theta)\sum_{j\in\mathcal{B}} x_{n,j} p^{\mathrm{C}}_{j} h^{\mathrm{N}}_{m,j} \geqslant x_{m,k}\theta\left(\sum_{j\in\mathcal{B}} p^{\mathrm{T}}_{j} h^{\mathrm{N}}_{m,j} + \sigma^2\right), \quad \forall n \in \mathcal{C}, \forall m \in \mathcal{M}_k, \forall k \in \mathcal{B} \tag{15}$$

$$\mathrm{C10}: p^{\mathrm{N}}_{m,k} h^{\mathrm{N}}_{m,k} \geqslant \theta\left(\sum_{n\in\mathcal{G}_{m',k}} p_{n,k} h^{\mathrm{N}}_{m,k} + \sum_{j\in\mathcal{B}\setminus k} p^{\mathrm{T}}_{j} h^{\mathrm{N}}_{m,j} - \sum_{n\in\mathcal{C}}\sum_{j\in\mathcal{B}\setminus k} x_{n,kj} p^{\mathrm{C}}_{j} h^{\mathrm{N}}_{m,j} + \sigma^2\right). \quad \forall m \in \mathcal{M}_k, \forall m' \in \mathcal{M}_k \setminus \mathcal{G}_{m,k}, \forall k \in \mathcal{B} \tag{16}$$

It can be seen that the maximization problem can force all variables in $\bar{\boldsymbol{X}}$ to be 1 such that the SINR of CoMP users is increased. In order to be consistent with the BS clustering determined by $\boldsymbol{X}$, the following new constraints are added.

$$\mathrm{C11}: \sum_{j\in\mathcal{B}\setminus k} x_{n,k,j} = x_{n,k}, \forall n \in \mathcal{C}, \forall k \in \mathcal{B}, \tag{17}$$

$$\mathrm{C12}: \sum_{j\in\mathcal{B}\setminus k} x_{n,j,k} = x_{n,k}, \forall n \in \mathcal{C}, \forall k \in \mathcal{B}, \tag{18}$$

$$\mathrm{C13}: x_{n,k,j} \in \{0,1\}, \forall n \in \mathcal{C}, \forall \{k,j\} \in \mathcal{B}. \tag{19}$$

Accordingly, P4 is reformulated as

$$\mathbf{P5}: \max_{\boldsymbol{X},\bar{\boldsymbol{X}}} \sum_{k\in\mathcal{B}}\sum_{m\in\mathcal{M}_k} R^{\mathrm{N}}_{m,k} + \sum_{n\in\mathcal{C}} R^{\mathrm{C}}_{n}$$
$$\text{s.t.} \quad \text{C1–C3, C5, C6, C9–C13.}$$

Furthermore, the following proposition is given.

Proposition 1: With constraints C11, C12, and C13, the optimal solution of P5 suffices to be the optimal solution of P4.

Proof: Assume that there is a feasible solution for P5 where CoMP user n is associated with BS l and does not associate with BS k, i.e., $x_{n,j} = 1$ and $x_{n,k} = 0$. If $x_{n,l,k} = 1$ in this feasible solution, it leads to $\sum_{j\in\mathcal{B}\setminus k} x_{n,k,j} > x_{n,k}$ and $\sum_{j\in\mathcal{B}\setminus k} x_{n,j,k} > x_{n,k}$. This contradicts the initial assumption that the solution is feasible for P5. Hence, C11, C12 and C13 can ensure $\boldsymbol{X}$ and $\bar{\boldsymbol{X}}$ correspond to the same BS clustering result. In consequence, the optimal solution of P4 is equal to that of P5. ■

Besides the binary constraints C3 and C13, all the constraints of P5 are linear. However, the objective function is hard to be analyzed since it includes many variables. To tackle this problem, we introduce the slack variables $\boldsymbol{\Phi} = \{\Phi_{m,k} \mid \Phi_{m,k} > 0, k \in \mathcal{B}, m \in \mathcal{M}_k\}$ and reformulate P5 as the following problem:

$$\mathbf{P6}: \max_{\boldsymbol{X},\bar{\boldsymbol{X}},\boldsymbol{\Phi}} \sum_{k\in\mathcal{B}}\sum_{m\in\mathcal{M}_k} \log_2\left(1+\frac{1}{\Phi_{m,k}}\right) + \sum_{n\in\mathcal{C}} R^{\mathrm{C}}_{n}$$
$$\text{s.t.} \quad \text{C1–C3, C5, C9–C13,}$$
$$\mathrm{C14}: \frac{1}{\Phi_{m,k}} \geqslant \xi_{\min}, \forall k \in \mathcal{B}, \forall m \in \mathcal{M}_k$$
$$\mathrm{C15}: \xi^{\mathrm{N}}_{m,k} \geqslant \frac{1}{\Phi_{m,k}}, \forall k \in \mathcal{B}, \forall m \in \mathcal{M}_k$$

where C14 and C15 are the linear constraints. The slack variable $\frac{1}{\Phi_{m,k}}$ replaces $\xi^{\mathrm{N}}_{m,k}$ in the objective function such that each $R^{\mathrm{N}}_{m,k}$ includes only one variable. Furthermore, C14 replaces C6. After that, it is easy to see that the objective function of P6 is a concave function and can be further expressed as

$$\begin{aligned} O(\boldsymbol{X},\boldsymbol{\Phi}) &= \sum_{k\in\mathcal{B}}\sum_{m\in\mathcal{M}_k}\log_2\left(1+\frac{1}{\Phi_{m,k}}\right) + \sum_{n\in\mathcal{C}}\log_2\left(\sum_{k\in\mathcal{B}} p^{\mathrm{T}}_{k} h^{\mathrm{C}}_{n,k} + \sigma^2\right) \\ &\quad - \sum_{n\in\mathcal{C}}\log_2\left(\sum_{k\in\mathcal{B}}\sum_{m\in\mathcal{M}_k} p^{\mathrm{N}}_{m,k} h^{\mathrm{C}}_{n,k} + \sum_{m\in\mathcal{C}\setminus n}\sum_{k\in\mathcal{B}} x_{m,k} p^{\mathrm{C}}_{k} h^{\mathrm{C}}_{n,k} + \sigma^2\right). \end{aligned} \tag{20}$$

Then, we have the following proposition.

Proposition 2: By introducing the slack variables $\boldsymbol{\Phi}$, the optimal solution for P5 is always obtained by solving P6.

Proof: The slack variable $\boldsymbol{\Phi}$ actually expands the feasible region, because $\frac{1}{\Phi_{m,k}}$ is allowed to be smaller than or equal to $\xi^{\mathrm{N}}_{m,k}$. If $\xi^{\mathrm{N}}_{m,k} = \frac{1}{\Phi_{m,k}}, \forall k \in \mathcal{B}, \forall m \in \mathcal{M}_k$, the optimal value for P6 is equal to that of P5. It is easy to prove that the equality holds for all constraints in C15. Assume that the inequality holds for one constraint as $\xi^{\mathrm{N}}_{m,k} > \frac{1}{\Phi_{m,k}}, \exists k \in \mathcal{B}, \exists m \in \mathcal{M}_k$. We can

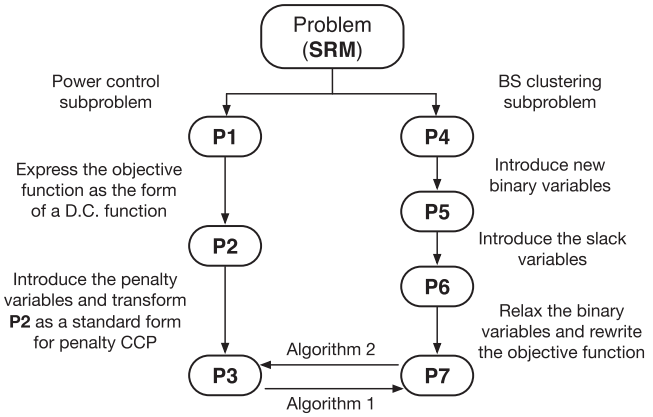

Fig. 3. Diagram to illustrate the problem decomposition and transformations.

reduce $\Phi_{m,k}$ to make the equality hold, which meanwhile increases the values of $\log_2(1+\frac{1}{\Phi_{m,k}})$ and the objective function. In other words, P6 always forces C15 to be strictly active so that the value of $\frac{1}{\Phi_{m,k}}$ and the objective function can be increased. Therefore, although the feasible region is expanded, the optimal solutions of P5 and P6 are identical. Now, we finish the proof of Proposition 2. ■

To relax binary constraints, we transform C3 and C13 into

$$x_{n,k} \in \{0,1\} \Leftrightarrow \begin{cases} 0 \leqslant x_{n,k} \leqslant 1 \\ x_{n,k} - x_{n,k}^2 \leqslant 0 \end{cases}, \tag{21}$$

$$x_{n,k,j} \in \{0,1\} \Leftrightarrow \begin{cases} 0 \leqslant x_{n,k,j} \leqslant 1 \\ x_{n,k,j} - x_{n,k,j}^2 \leqslant 0 \end{cases}. \tag{22}$$

According to all the above analysis, we use the successive convex approximation to solve P6. Let $(\boldsymbol{X}, \boldsymbol{\Phi})$ denote the vector of $\boldsymbol{X}$ and $\boldsymbol{\Phi}$. P7 is formulated to obtain a lower-bound solution of P6 at a given points, which is expressed as

$$\begin{aligned} \mathbf{P7}: \max_{\boldsymbol{X},\bar{\boldsymbol{X}},\boldsymbol{\Phi}} \quad & O\left(\boldsymbol{X}\left[t\right], \boldsymbol{\Phi}\left[t\right]\right) \\ & + \nabla O\left(\boldsymbol{X}\left[t\right], \boldsymbol{\Phi}\left[t\right]\right)^{\mathrm{T}} \left(\left(\boldsymbol{X}, \boldsymbol{\Phi}\right) - \left(\boldsymbol{X}\left[t\right], \boldsymbol{\Phi}\left[t\right]\right)\right) \\ \text{s.t.} \quad & \mathrm{C1}, \mathrm{C2}, \mathrm{C5}, \mathrm{C9}-\mathrm{C12}, \mathrm{C14}, \mathrm{C15}, \\ & \mathrm{C3a}: 0 \leqslant x_{n,k} \leqslant 1, \forall n \in \mathcal{C}, \forall k \in \mathcal{B} \\ & \mathrm{C3b}: -x_{n,k}^2\left[t\right] - 2x_{n,k}\left[t\right]\left(x_{n,k} - x_{n,k}\left[t\right]\right) \\ & \qquad + x_{n,k} \leqslant 0, \forall n \in \mathcal{C}, \forall k \in \mathcal{B} \\ & \mathrm{C13a}: 0 \leqslant x_{n,k,j} \leqslant 1, \forall n \in \mathcal{C}, \forall \{k,j\} \in \mathcal{B} \\ & \mathrm{C13b}: -x_{n,k,j}^2\left[t\right] - 2x_{n,k,j}\left[t\right]\left(x_{n,k,j} - x_{n,k,j}\left[t\right]\right) \\ & \qquad + x_{n,k,j} \leqslant 0, \forall n \in \mathcal{C}, \forall \{k,j\} \in \mathcal{B} \end{aligned}$$

where $\boldsymbol{X}[t]$, $\bar{\boldsymbol{X}}[t]$, and $\boldsymbol{\Phi}[t]$ are the given points. P7 is a convex problem and can be solved by the standard convex programming method [28]. Then, we propose the BS clustering algorithm and show the detailed procedure in Algorithm 2. Furthermore, we give the following theorem.

Theorem 2: Algorithm 2 can obtain a series of non-decreasing optimal values for P7 before achieving the convergence.

Algorithm 2: BS Clustering Algorithm.

1: **Initialization**
2: Given the transmit powers of all BSs. Set $t = 0$.
3: Input the current BS clustering as initial point $\boldsymbol{X}[0]$. Calculate $\bar{\boldsymbol{X}}[0]$ and $\boldsymbol{\Phi}[0]$.
4: **repeat**
5: Update $\mathcal{G}_{m,k}, \forall m \in \mathcal{M}_k, \forall k \in \mathcal{B}$ according to $\boldsymbol{X}[t]$.
6: Solve P7 based on $\boldsymbol{X}[t]$, $\bar{\boldsymbol{X}}[t]$, and $\boldsymbol{\Phi}[t]$ and get the optimal solution $\boldsymbol{X}^*$, $\bar{\boldsymbol{X}}^*$, and $\boldsymbol{\Phi}^*$.
7: Update $\boldsymbol{X}[t+1] = \boldsymbol{X}^*$, $\bar{\boldsymbol{X}}[t+1] = \bar{\boldsymbol{X}}^*$, and $\boldsymbol{\Phi}[t+1] = \boldsymbol{\Phi}^*$.
8: $\varepsilon = |O(\boldsymbol{X}[t+1], \boldsymbol{\Phi}[t+1]) - O(\boldsymbol{X}[t], \boldsymbol{\Phi}[t])|$.
9: Set $t = t + 1$.
10: **until** $\varepsilon \leqslant \varepsilon_{\min}$ **or** $t \geqslant t_{\max}$
11: **return** $\boldsymbol{X}^*$.

Proof: See Appendix B. ■

In Algorithm 2, steps 4–10 iteratively solve P7 until the convergence condition is met or $t_{\max}$ is achieved. $\mathcal{G}_{m,k}, \forall m \in \mathcal{M}_k, \forall k \in \mathcal{B}$ of each iteration is obtained according to the point $\boldsymbol{X}[t]$ before solving P7. Afterwards, a suboptimal solution for BS clustering is obtained. Because the system sum-rates obtained by Algorithm 2 are non-decreasing and must be bounded by the optimal solution of P4, the convergence is guaranteed.

C. Joint BS Clustering and Power Control Scheme

Based on results of Sections III-A and III-B, we design a joint BS clustering and power control (JOBCPC) scheme to solve Problem (SRM). The detailed procedure is presented in Algorithm 3. Fig. 3 shows the decomposition of Problem (SRM) and the relationship from P1 to P7.

Specifically, Algorithm 3 initializes the BS clustering by performing Algorithm 4, which is shown in steps 1–3. In Algorithm 4, the CoMP user n^* with maximum $\xi_{\min} - \xi_n^{\mathrm{C}}$ is prior selected and allocated with BS k^* which can maximize the SINR of CoMP user n. The initial BS clustering is obtained until all the BSs or CoMP users are allocated with a CoMP user or BSs. After the initialization, steps 4–11 in Algorithm 3 alternately performs the power control algorithm and BS clustering algorithm until the convergence condition is met or the maximum number of iteration $i_{\max}$ is achieved. The convergence condition is that the gap between the system sum-rates of two adjacent iterations is sufficiently small or the BS clustering keeps unchanged between two adjacent iterations. Finally, we can get the solution of Problem (SRM) if P3 with initial BS clustering is feasible. Based on Theorems 1 and 2, we give the following corollary.

Corollary 1: The convergence of JOBCPC scheme is guaranteed and its system sum-rates are non-decreasing at each iteration.

Proof: See Appendix C. ■

To analyze the complexity of JOBCPC algorithm, we first analyze the complexity of Algorithm 1 and Algorithm 2. The complexity of Algorithm 1 is $O(N^{3.5}D^{3.5}K^{3.5} \times t_{\max})$, if the interior point method is applied for convex programming [28].

Algorithm 3: Joint BS Clustering and Power Control (JOBCPC) Scheme.

Initialization
Initialize the BS clustering by performing Algorithm 4
Set iteration $i = 0$ and system sum-rate $R[0] = 0$.
repeat
  Update the tranmist power with given BS clustering by performing Algorithm 1 to obtain $R[i+1]$.
  $e = |R[i+1] - R[i]|$.
  if $e \leqslant e_{\min}$ **or** power control subproblem is infeasible.
  then
    break
  end if
  Update BS clustering with given tranmist power by performing Algorithm 2.
  $i = i + 1$.
until BS clustering keeps unchanged **or** $i \geqslant i_{\max}$
return $\boldsymbol{X}^*$ and $\boldsymbol{P}^*$ if feasible.

Algorithm 4: Initialize BS Clustering.

Set the BS cluster $\mathcal{B}_n = \emptyset$ for any CoMP user $n \in \mathcal{C}$.
Set $p_k^{\mathrm{C}} = p_{m,k}^{\mathrm{N}} = \frac{P_{\max}}{(M_k+1)}, \forall k \in \mathcal{B}, \forall m \in \mathcal{M}_k$.
repeat
  Calculate $\xi_n^{\mathrm{C}}, \forall n \in \mathcal{C}$ and $\xi_{m,k}^{\mathrm{N}}, \forall k \in \mathcal{B}, \forall m \in \mathcal{M}_k$ under the current BS clustering.
  Select CoMP user $n^* = \arg\max_{n \in \mathcal{C}} \xi_{\min} - \xi_n^{\mathrm{C}}$
  for all $k \in \mathcal{B}$ **do**
    Calculate $\zeta_{n,k}^{\mathrm{C}} = \xi_n^{\mathrm{C}}$ if CoMP user n^* is associated with BS k (i.e., BS k is in $\mathcal{B}_n$).
  end for
  Select BS $k^* = \arg\max_{k \in \mathcal{B}} \zeta_{n,k}^{\mathrm{C}}$.
  Set $x_{n^*,k^*} = 1$ and $\mathcal{B}_n = \mathcal{B} \cup k^*$.
  Remove n^* from $\mathcal{C}$ and k^* from $\mathcal{B}$.
until $\mathcal{B} = \emptyset$ **or** $\mathcal{C} = \emptyset$
returninitial BS clustering $\boldsymbol{X}$.

The complexity of Algorithm 2 is $O(N^7 K^{3.5} \times t_{\max})$. Therefore, the complexity of JOBCPC is $O(N^7 K^{3.5} \times t_{\max} \times i_{\max})$ considering $D \ll N$ in the practical systems.

IV. Simulation Results

In this section, we present the simulation results to evaluate the performance of our proposed BS clustering and power control scheme for CoMP transmission in NOMA-UDN. The BSs and users are uniformly distributed in the simulation scenario. The pathloss parameter is characterized by a dual-slope pathloss model [3], [34] to acquire the impact of the coexistence of LOS and NLOS paths. The RSRP is used to classify the users into CoMP users and non-CoMP users, and the association between non-CoMP users and BSs is performed based on the nearest association principle [9], [14], [18], [19]. The default simulation parameters are summarized in Table I. Furthermore, we compare the performance of proposed JOBCPC scheme with those of OMA-based CoMP (OMA) scheme and other four benchmark schemes. The OMA-based CoMP scheme does not allow the BSs to simultaneously transmit signals for CoMP users and non-CoMP users, and other procedures are same with JOBCPC scheme. The four benchmark schemes are listed as follows.

TABLE I
Default Simulation Parameters

Parameters	Values
Network area	100 m × 100 m
Size of CoMP cluster, C	3
Max number of non-CoMP users in a cell, D	1
Spectrum bandwidth	180 kHz
Max transmit power, $P_{\max}$	27 dBm
SINR threshold, $\xi_{\min}$	5 dB
SIC threshold, θ	0 dB [9], [17]
Noise power, σ^2	-174 dBm/Hz × 180 kHz
Path loss model	$10^{-10.38} d[\mathrm{km}]^{-2.09}$, $d < 0.01$ km $10^{-14.54} d[\mathrm{km}]^{-3.75}$, $d \geqslant 0.01$ km
Multiple-path fading	Exponential distribution with unit mean
Shadowing	Log-normal distribution with standard deviation of 8 dB

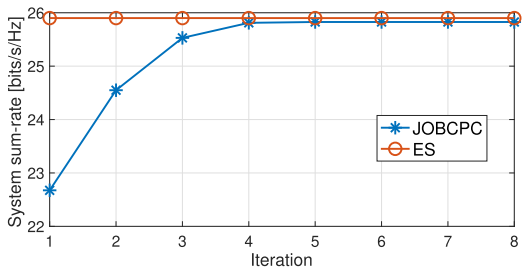

Fig. 4. Convergence of JOBCPC scheme.

- Exhaustive search (ES) scheme: The scheme searches all possible results of BS clustering and performs Algorithm 1 to calculate the system sum-rate of each result. Then, the scheme selects the result with maximum system sum-rate as the output.
- Swap matching-based (Matching) scheme: The scheme optimizes the BS clustering through the swap matching method [35], and other procedures are same with JOBCPC scheme. Its complexity is $O(N^{3.5} D^{3.5} K^{3.5} \times t_{\max} \times i_{\max})$.
- Greedy clustering (GC) scheme: The scheme uses the best channel gain criterion proposed in [13] to obtain the BS clustering in a greedy way and adopt Algorithm 1 for power control.
- Non-CoMP scheme: The scheme does not implement CoMP in NOMA-UDN and only uses the signal-point transmission. The other procedures are same with JOBCPC scheme.

A. Convergence of Proposed Algorithms

To evaluate the convergence of proposed JOBCPC scheme, we compare its performance with ES scheme, where the BS density is 6×10^2/km^2, the user density is 4×10^2/km^2 including CoMP users and non-CoMP users, and $D = 2$. Fig. 4 shows the

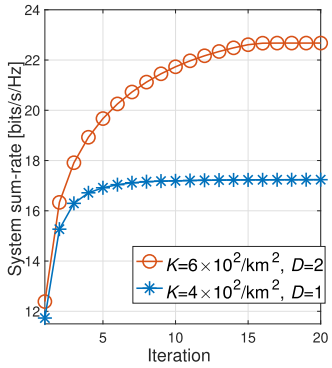

Fig. 5. Convergence of power control algorithm.

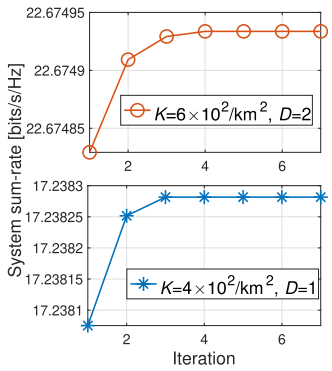

Fig. 6. Convergence of BS clustering algorithm.

performance comparison of JOBCPC scheme and ES scheme. It can be seen that the JOBCPC scheme converges within the limited number of iterations, which validates Corollary 1. Furthermore, the performance of JOBCPC scheme is gradually close to that of ES scheme as the number of iterations increases. However, there still exists a gap between JOBCPC scheme and ES scheme. The reasons is explained as follows. On one hand, Problem (SRM) is a non-convex and combinatorial optimization problem. On the other hand, although the problem decomposition makes Problem (SRM) solvable, it makes the feasible region of original problem unable to be globally searched. Furthermore, we evaluate the convergence of our proposed power control algorithm (Algorithm 1) and our proposed BS clustering algorithm (Algorithm 2). Figs. 5 and 6 show the convergence of Algorithms 1 and 2, respectively. It can be seen that both of them converge within a limited number of iterations where the average error in system sum-rate after convergence is in $10^{-5} \sim 5 \times 10^{-5}$ bits/s/Hz. More iterations are required when the BS density increases. The results validate Theorems 1 and 2.

B. Impact of BS Density and User Density on Network Performance

We evaluate the impact of BS density and user density on the network performance as follows. First, the impact of user density on the network performance is evaluated, where the BS density is set as $1.8 \times 10^3/\text{km}^2$, and the user density includes the densities of CoMP users and non-CoMP users. The density of non-CoMP users is set as $3 \times 10^2/\text{km}^2$, and the density of CoMP users increases from $1 \times 10^2/\text{km}^2$ to $9 \times 10^2/\text{km}^2$. Then, the impact of BS density on the network performance is evaluated, where the densities of CoMP users and non-CoMP users are set as $6 \times 10^2/\text{km}^2$ and $3 \times 10^2/\text{km}^2$, respectively.

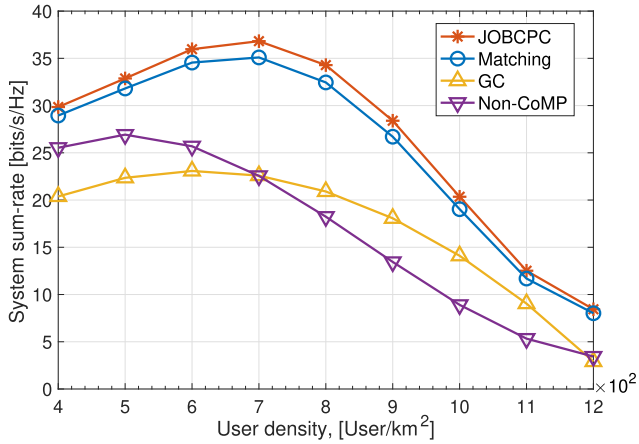

Fig. 7. System sum-rate versus user density with different schemes.

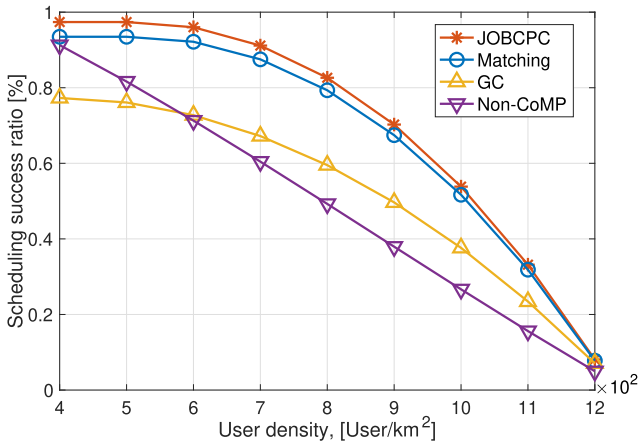

Fig. 8. Scheduling success ratio versus user density with different schemes.

Fig. 7 shows the system sum-rate versus the user density with different schemes. It can be seen that the system sum-rate is first increasing and then decreasing with the growth of user density. This is because the high user density aggravates the interference, which becomes more severe in UDN due to LoS paths between interfering BSs and users. Although the high user density degrades the system sum-rate, the proposed JOBCPC scheme outperforms all other schemes. Compared with the Matching scheme and GC scheme, JOBCPC scheme can find better BS clustering such that each CoMP user is served by the suitable BS cluster and harvests higher transmission rate. Compared with Non-CoMP scheme, JOBCPC scheme effectively restrains inter-cell interference and enhances the SINRs of CoMP users due to the exploitation of CoMP transmission.

Fig. 8 shows the scheduling success ratio versus the user density with different schemes. The scheduling success ratio is the probability that the SINR thresholds of all users are met. It can be seen that the scheduling success ratio gradually decreases with the growth of user density. This is because the severe interference makes it difficult to meet the SINR threshold for

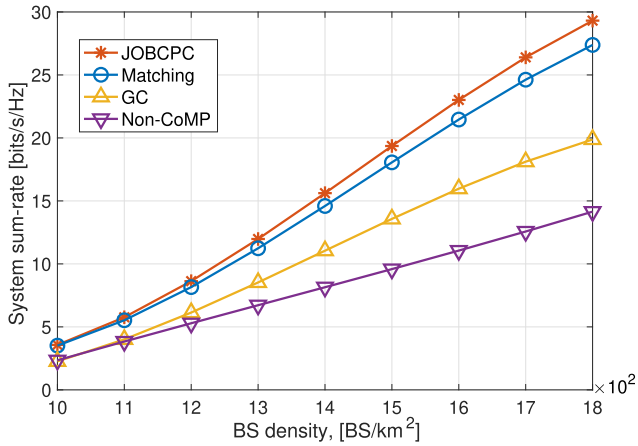

Fig. 9. System sum-rate versus BS density different schemes.

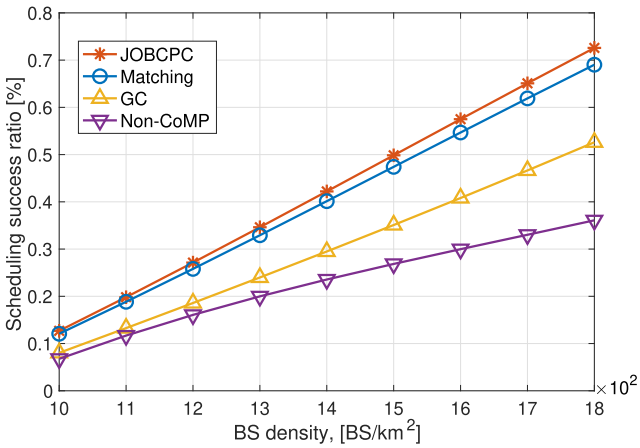

Fig. 10. Scheduling success ratio versus BS density with different schemes.

all users as the user density increases. It can be also seen that the proposed scheme achieves higher scheduling success ratio than other schemes. This is because the proposed scheme can effectively mitigate the interference through better BS clustering and power control method such that more users can achieve the SINR threshold.

Fig. 9 shows the system sum-rate versus the BS density with different schemes. It can be seen that the system sum-rate gradually increases with the increase of BS density. This is because the increase of BS density can provide more available BSs for the CoMP user to form the BS clusters with higher transmission rate. It is worth noting that the gaps between JOBCPC scheme and other schemes are enlarged as the BS density increases. This demonstrates that the proposed JOBCPC scheme can exploit the benefit of BS diversity to significantly improve the system sum-rate. Fig. 10 shows the scheduling success ratio versus the BS density with different schemes. It can be seen that the scheduling success ratio increases with the growth of BS density. This is because larger BS density makes CoMP users acquire more suitable coordinated BSs to enhance the CoMP transmission rate. Furthermore, JOBCPC scheme achieves higher scheduling success ratio than other schemes, since it more effectively optimizes the BS clustering.

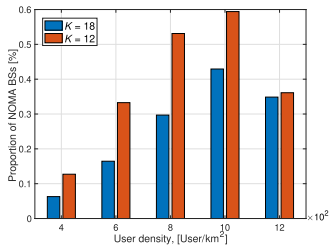

Fig. 11. Proportion of NOMA BSs versus user density.

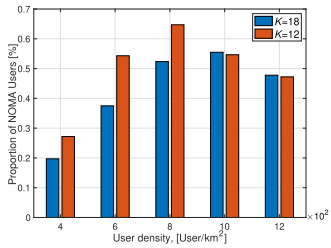

Fig. 12. Proportion of NOMA users versus user density.

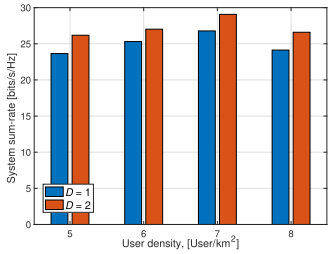

Fig. 13. System sum-rate versus user density with different D.

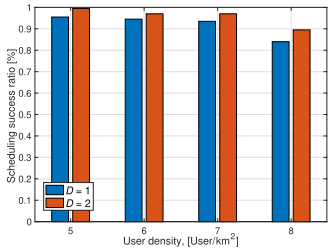

Fig. 14. Scheduling success ratio versus user density with different D.

To evaluate the effect of proposed JOBCPC scheme on NOMA transmission, Figs. 11 and 12 plot the proportion of BSs operating NOMA (i.e., NOMA BSs) and that of NOMA users under different K versus the user density. We can see from Fig. 11 that when the user density is $10 \times 10^2/\text{km}^2$, the proportions of NOMA BSs with $K = 12$ and $K = 18$ can achieve over 50% and over 40%, respectively. It demonstrates that our proposed scheme can efficiently employ NOMA to improve the performance of CoMP transmission. However, the proportion of NOMA BSs decreases when the user density is high. The reason is that the number of CoMP users is generally greater than that of non-CoMP users, since the growth of user density decreases the difference of RSRPs between users and near BSs [13], [18]. Partial CoMP users cannot find a non-CoMP user to form a NOMA group such that their BS clusters cannot operate NOMA. There are already related works to investigate that multiple CoMP users coexist in one NOMA group to enhance NOMA

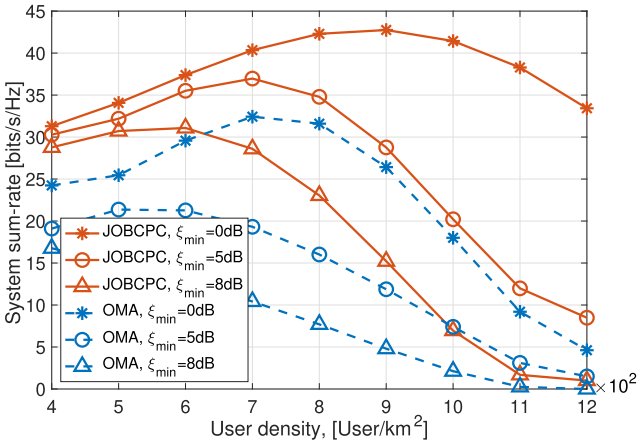

Fig. 15. System sum-rate under different SINR thresholds $\xi_{\min}$ versus user density.

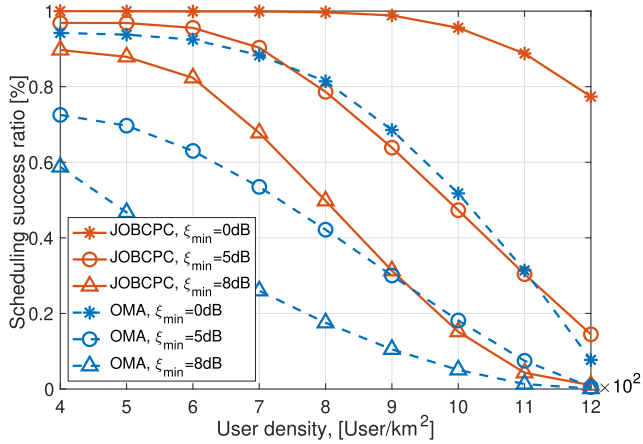

Fig. 16. Scheduling success ratio under different SINR thresholds $\xi_{\min}$ versus user density.

operation in CoMP systems [10]. Furthermore, we can see from Fig. 12 that the proportions of NOMA users with $K = 12$ and $K = 18$ achieve over 60% and 50%, respectively, when the user density is $8 \times 10^2/\text{km}^2$. High user density can decrease the proportion of NOMA users due to the limited number of BSs and non-CoMP user.

The impact of the number of non-CoMP users on network performance is also evaluated. Figs. 13 and 14 show the system sum-rate and the scheduling success ratio with different D versus the user density, where the BS density is set as $1 \times 10^3/\text{km}^2$, the density of non-CoMP users is set as $4 \times 10^2/\text{km}^2$, and the density of CoMP users increases from $1 \times 10^2/\text{km}^2$ to $4 \times 10^2/\text{km}^2$. It can be seen from Figs. 13 and 14 that the growth of D can increase the system sum-rate and scheduling success ratio. This is because the growth of D can improve the multiplexing gain of NOMA to efficiently alleviate interference in UDN and increase the spectrum efficiency of NOMA-enabled CoMP transmission.

C. Impact of SINR Threshold on Network Performance

The impact of SINR threshold $\xi_{\min}$ on the network performance is evaluated in the subsection. The BS density is set as $1.8 \times 10^3/\text{km}^2$. The user density is varying, where the density of non-CoMP users is set as $3 \times 10^2/\text{km}^2$, and the density of CoMP users increases from $1 \times 10^2/\text{km}^2$ to $9 \times 10^2/\text{km}^2$.

Fig. 15 shows the system sum-rate under different $\xi_{\min}$ versus the user density. It can be seen the increase of $\xi_{\min}$ leads to the decline of system sum-rate. When $\xi_{\min} = 8$ dB, the system sum-rate under high user density approaches zero. This is because the higher $\xi_{\min}$ makes it more difficult for multiple access, which results in the scheduling failure. More importantly, the proposed JOBCPC scheme significantly improves the system sum-rate compared with OMA-based CoMP scheme. This is because NOMA can utilize the power-domain multiplexing to support the spectrum reuse between CoMP users and non-CoMP users, which improves the spectrum utilization and enhances the user-access capability.

Fig. 16 shows the scheduling success ratio under different $\xi_{\min}$ versus the user density. It can be seen that the scheduling success ratio gradually decreases with the growth of $\xi_{\min}$. Furthermore, JOBCPC scheme always achieves higher scheduling success ratio than the OMA-based CoMP scheme. This is because NOMA can improve the spectrum utilization to make all SINRs of CoMP users and non-CoMP users achieve $\xi_{\min}$. In this regard, even though NOMA fails to avoid performance degradation caused by high user density, the proposed JOBCPC scheme significantly improves the network capacity and user-access capability in UDN through efficient interference management.

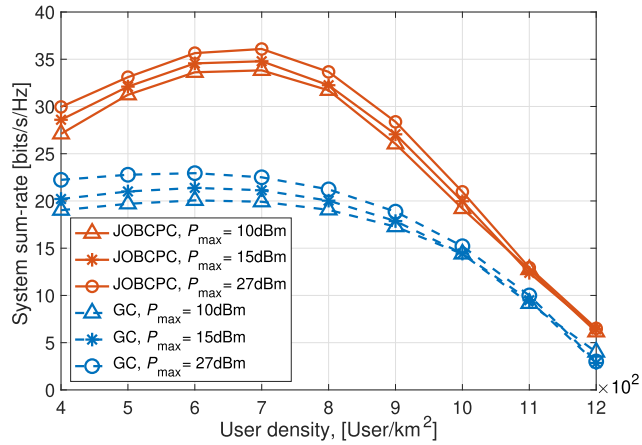

Fig. 17. System sum-rate under different maximum transmit powers $P_{\max}$ versus user density.

D. Impact of Maximum Transmit Power on Network Performance

We evaluate the network performance under different maximum transmit power $P_{\max}$. The settings of the densities of BSs and users are similar with Section IV-C.

Fig. 17 shows the system sum-rate under different $P_{\max}$. It can be seen that higher maximum transmit power makes the NOMA-UDN achieve higher system sum-rate. However, the gain from increasing $P_{\max}$ is gradually diminished with the growth of user density. This is because LOS paths between BSs and users make the inter-cell interference extremely severe in UDN, and hence BSs cannot use high transmit power to

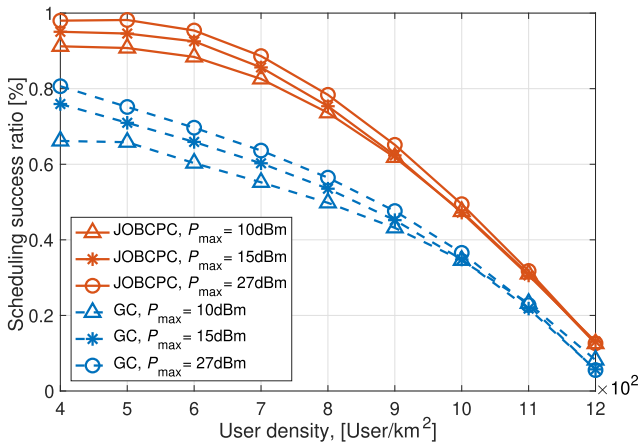

Fig. 18. Scheduling success ratio under different maximum transmit powers $P_{\max}$ versus user density.

serve the associated users. Furthermore, since the LOS path brings high channel power gains for transmission links, the SINR threshold can be met even by low transmit power. According to the above results, JOBCPC scheme effectively alleviates severe interference in UDN to achieve higher system sum-rate than GC scheme.

Fig. 18 shows the scheduling success ratio under different $P_{\max}$. It can be seen that the increase of $P_{\max}$ improves the scheduling success ratio for NOMA-UDN. Similarly, the performance gaps caused by different $P_{\max}$ between the JOBCPC scheme and the GC scheme diminish as the user density increases. Furthermore, the proposed scheme achieves higher scheduling success ratio than the GC scheme through using more efficient BS clustering algorithm.

V. Conclusion

This paper has investigated joint optimization problem of BS clustering and power control problem for NOMA-enabled CoMP transmission in UDN. First, the power control algorithm has been designed based on penalty CCP method to coordinate interference among coordinated BSs and meet NOMA decoding condition and users' rate requirements. Second, the BS clustering algorithm has been designed by successive convex approximation to increase the system sum-rate. Finally, a joint BS clustering and power control scheme has been proposed by alternately addressing the power control and BS clustering. It has been shown through the analysis that the proposed scheme can efficiently overcome the issue of initial point infeasibility caused by severe interference in UDN. Simulation results show that even in the case of high user density, the proposed scheme can further increase the system sum-rate while the system sum-rates of traditional schemes decrease. In future work, we investigate the application of machine learning to the CoMP transmission in UDN.

Appendix

A. Proof of Theorem 1

Assume that $\boldsymbol{P}[t-1]$ is the given point of t-th iteration which is also the optimal solution of $t-1$-th iteration. At the t-th iteration, Problem P2 with $\boldsymbol{P}[t-1]$ is formulated

$$\begin{aligned}\max_{\boldsymbol{P}} \ & \sum_{k\in\mathcal{B}}\sum_{m\in\mathcal{M}_k} g_m^{\mathrm{N}}(\boldsymbol{P}) + \sum_{n\in\mathcal{C}} g_n^{\mathrm{C}}(\boldsymbol{P}) \\ & - \sum_{k\in\mathcal{B}}\sum_{m\in\mathcal{M}_k} f_m^{\mathrm{N}}(\boldsymbol{P}[t-1]) - \sum_{n\in\mathcal{C}} f_n^{\mathrm{C}}(\boldsymbol{P}[t-1]) \\ & - \sum_{k\in\mathcal{B}}\sum_{m\in\mathcal{M}_k} \nabla f_m^{\mathrm{N}}(\boldsymbol{P}[t-1])^{\mathrm{T}}(\boldsymbol{P}-\boldsymbol{P}[t-1]) \\ & - \sum_{n\in\mathcal{C}} \nabla f_n^{\mathrm{C}}(\boldsymbol{P}[t-1])^{\mathrm{T}}(\boldsymbol{P}-\boldsymbol{P}[t-1]) \qquad (23)\end{aligned}$$

$$\text{s.t.} \quad \text{C4-C8}, \qquad (24)$$

which is a typical convex problem. Through solving the problem in (24), the optimal value $Q(\boldsymbol{P}[t])$ and the optimal solution $\boldsymbol{P}[t]$ are obtained. Because $\boldsymbol{P}[t-1]$ is the optimal solution of $t-1$-th iteration, it always meets C4–C8. $Q(\boldsymbol{P}[t-1])$ is the the optimal value of $t-1$-th iteration. Note that if $\boldsymbol{P}[t-1]$ is an arbitrary point and does not meet C4–C8, Problem (24) cannot be solved so that we cannot continue the proof.

For the convenience, we define the functions $G(\boldsymbol{P})$, $H(\boldsymbol{P})$, and $\hat{H}(\boldsymbol{P};\boldsymbol{P}[t])$ as follows.

$$G(\boldsymbol{P}) = \sum_{k\in\mathcal{B}}\sum_{m\in\mathcal{M}_k} g_m^{\mathrm{N}}(\boldsymbol{P}) + \sum_{n\in\mathcal{C}} g_n^{\mathrm{C}}(\boldsymbol{P}), \qquad (25)$$

$$H(\boldsymbol{P}) = \sum_{k\in\mathcal{B}}\sum_{m\in\mathcal{M}_k} f_m^{\mathrm{N}}(\boldsymbol{P}) + \sum_{n\in\mathcal{C}} f_n^{\mathrm{C}}(\boldsymbol{P}), \qquad (26)$$

$$\begin{aligned}\hat{H}(\boldsymbol{P};\boldsymbol{P}[t]) = & \sum_{k\in\mathcal{B}}\sum_{m\in\mathcal{M}_k} f_m^{\mathrm{N}}(\boldsymbol{P}[t]) \\ & + \sum_{k\in\mathcal{B}}\sum_{m\in\mathcal{M}_k} \nabla f_m^{\mathrm{N}}(\boldsymbol{P}[t])^{\mathrm{T}}(\boldsymbol{P}-\boldsymbol{P}[t]) \\ & + \sum_{n\in\mathcal{C}} \left(f_n^{\mathrm{C}}(\boldsymbol{P}[t]) + \nabla f_n^{\mathrm{C}}(\boldsymbol{P}[t])^{\mathrm{T}}(\boldsymbol{P}-\boldsymbol{P}[t])\right), \qquad (27)\end{aligned}$$

where $G(\boldsymbol{P})$ and $H(\boldsymbol{P})$ are the concave functions. $\hat{H}(\boldsymbol{P};\boldsymbol{P}[t])$ are the first-order Taylor approximation of $H(\boldsymbol{P})$ at the point $\boldsymbol{P}[t]$.

We have the following formula

$$\begin{aligned}& G(\boldsymbol{P}[t-1]) - H(\boldsymbol{P}[t-1]) \\ & \qquad = G(\boldsymbol{P}[t-1]) - \hat{H}(\boldsymbol{P}[t-1];\boldsymbol{P}[t-1]) \qquad (28) \\ & \qquad \overset{A}{\leqslant} G(\boldsymbol{P}[t]) - \hat{H}(\boldsymbol{P}[t];\boldsymbol{P}[t-1]). \qquad (29)\end{aligned}$$

The reason why A holds is that P2 as a maximization problem must find an optimal solution to maximize $G(\boldsymbol{P}) - \hat{H}(\boldsymbol{P};\boldsymbol{P}[t-1])$ at each iteration. At least, P2 can choose $\boldsymbol{P}[t] = \boldsymbol{P}[t-1]$ to ensure that the optimal value of t-th iteration is equal to that of $t-1$-th iteration. Therefore, we have that

$$\begin{aligned}& Q(\boldsymbol{P}[t-1]) \\ & \quad \leqslant G(\boldsymbol{P}[t]) - \hat{H}(\boldsymbol{P}[t];\boldsymbol{P}[t-1]) \leqslant Q(\boldsymbol{P}[t]). \qquad (30)\end{aligned}$$

Furthermore, according to the Taylor's formula, we know that

$$\begin{aligned}\hat{H}\left(\boldsymbol{P}\left[t-1\right];\boldsymbol{P}\left[t-1\right]\right) &= H\left(\boldsymbol{P}\left[t-1\right]\right)\\ &\overset{B}{\geqslant} \hat{H}\left(\boldsymbol{P}\left[t\right];\boldsymbol{P}\left[t-1\right]\right),\end{aligned} \tag{31}$$

where the equality at B holds when $\boldsymbol{P}[t]=\boldsymbol{P}[t-1]$. Hence, although $G(\boldsymbol{P}[t])=G(\boldsymbol{P}[t-1])$, we can still obtain the relationship between (28) and (29).

Therefore, the sequence of optimal values for P2 is non-decreasing. The convergence is achieved until the iterative algorithm cannot find a better solution than the solution of last iteration. Now, we finish the proof of Theorem 1.

B. Proof of Theorem 2

To prove that Algorithm 2 can obtain a series of non-decreasing optimal values for P7, it should be first proved that the t-th optimal solution is feasible for the $t+1$ iteration. Because $\boldsymbol{\Phi}$ in any constraint is not affected by the iterative point, we only consider $\boldsymbol{X}$ and $\bar{\boldsymbol{X}}$. Specifically, we should prove whether $\boldsymbol{X}[t+1]$ and $\bar{\boldsymbol{X}}[t+1]$ can be a feasible solution to P6 and P7 in $t+2$-th iteration, if $\boldsymbol{X}[t]$ and $\bar{\boldsymbol{X}}[t]$ are the optimal solution of t-th iteration.

Assume $\boldsymbol{X}[t]$ and $\bar{\boldsymbol{X}}[t]$ are the optimal solution of the t-th iteration. We have that

$$\begin{aligned}&x_{n,k}\left[t\right]-x_{n,k}^{2}\left[t\right]-2x_{n,k}\qquad\left[t\right]\left(x_{n,k}\left[t\right]-x_{n,k}\left[t\right]\right)\\ &=x_{n,k}\left[t\right]-x_{n,k}^{2}\left[t\right]\leqslant 0.\end{aligned} \tag{32}$$

Due to the convexity of $x_{n,k}^2$, i.e., $x_{n,k}^2 \geqslant x_{n,k}^2[t]+2x_{n,k}[t](x_{n,k}-x_{n,k}[t])$, we have that

$$\begin{aligned}&x_{n,k}-x_{n,k}^{2}\\ &\quad\leqslant x_{n,k}-x_{n,k}^{2}\left[t\right]-2x_{n,k}\left[t\right]\left(x_{n,k}-x_{n,k}\left[t\right]\right),\end{aligned} \tag{33}$$

where the equality holds if and only if $x_{n,k}=x_{n,k}[t]$. Based on $\boldsymbol{X}[t]$, we can obtain $\boldsymbol{X}[t+1]$ by solving P7. Hence, the following formula must hold

$$\begin{aligned}&x_{n,k}\left[t+1\right]-x_{n,k}^{2}\left[t+1\right]\\ &\qquad-2x_{n,k}\left[t+1\right]\left(x_{n,k}\left[t+1\right]-x_{n,k}\left[t+1\right]\right)\\ &\qquad\leqslant x_{n,k}\left[t+1\right]-x_{n,k}^{2}\left[t\right]\\ &\qquad\qquad-2x_{n,k}\left[t\right]\left(x_{n,k}\left[t+1\right]-x_{n,k}\left[t\right]\right)\leqslant 0.\end{aligned} \tag{34}$$

Hence, $\boldsymbol{X}[t+1]$ is a feasible solution to P6 and P7 in $t+2$ iteration. Similarly, we can also prove that $\bar{\boldsymbol{X}}[t+1]$ is feasible to P6 and P7 in $t+2$ iteration.

The similar approach in the proof of Theorem 1 can be used to prove that the system sum-rates obtained by Algorithm 2 in the a series of iteration are non-decreasing and will converge.

C. Proof of Corollary 1

Let $Q[i]$ and $O[i]$ denote the system sum-rates of power control algorithm and BS clustering algorithm at the i-th iteration, respectively. Let $\boldsymbol{P}^*[i]$ and $\boldsymbol{X}^*[i]$ denote the solution of power control and BS clustering at the i-th iteration, respectively.

We first prove $O[i] \geqslant Q[i]$. We based on Theorem 1 know that with the given $\boldsymbol{X}^*[i-1]$, $\boldsymbol{P}^*[i]$ is obtained by power control algorithm at the i-th iteration and achieves a non-decreasing value $Q[i]$. Furthermore, since $\boldsymbol{X}^*[i-1]$ is as the initial point in Algorithm 2 to obtain $\boldsymbol{X}^*[i]$, $\boldsymbol{P}^*[i]$ is feasible for the BS clustering subproblem in the i-th iteration. Thus, we based on Theorem 2 know that $\boldsymbol{X}^*[i]$ achieves a non-decreasing value $O[i]$ compared with $Q[i]$, i.e., $O[i] \geqslant Q[i]$.

We next prove $Q[i+1] \geqslant O[i]$. Because $\boldsymbol{X}^*[i]$ is obtained by the BS clustering algorithm with given $\boldsymbol{P}^*[i]$, $\boldsymbol{X}^*[i]$ is feasible for the power control subproblem in the $i+1$-th iteration. We can based on Theorem 1 know that $Q[i+1] \geqslant O[i]$. Therefore, the system sum-rates at each iteration are non-decreasing. Since the system sum-rate is upper-bounded by the optimal value of Problem (SRM), JOBCPC scheme can converge within a limited number of iterations.

REFERENCES

[1] S. Andreev, V. Petrov, M. Dohler, and H. Yanikomeroglu, "Future of ultra-dense networks beyond 5G: Harnessing heterogeneous moving cells," *IEEE Commun. Mag.*, vol. 57, no. 6, pp. 86–92, Jun. 2019.

[2] J. Liu, M. Sheng, L. Liu, and J. Li, "Interference management in ultra-dense networks: Challenges and approaches," *IEEE Netw.*, vol. 31, no. 6, pp. 70–77, Nov. 2017.

[3] M. Ding, P. Wang, D. López-Pérez, G. Mao, and Z. Lin, "Performance impact of LoS and NLoS transmissions in dense cellular networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 2365–2380, Mar. 2016.

[4] S. Chen, T. Zhao, H. Chen, Z. Lu, and W. Meng, "Performance analysis of downlink coordinated multipoint joint transmission in ultra-dense networks," *IEEE Netw.*, vol. 31, no. 5, pp. 106–114, Aug. 2017.

[5] W. Shin, M. Vaezi, B. Lee, D. J. Love, J. Lee, and H. V. Poor, "Non-orthogonal multiple access in multi-cell networks: Theory, performance, and practical challenges," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 176–183, Oct. 2017.

[6] M. S. Ali, E. Hossain, and D. I. Kim, "Coordinated multipoint transmission in downlink multi-cell NOMA systems: Models and spectral efficiency performance," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 24–31, Apr. 2018.

[7] J. Choi, "Non-orthogonal multiple access in downlink coordinated two-point systems," *IEEE Commun. Lett.*, vol. 18, no. 2, pp. 313–316, Feb. 2014.

[8] Y. Tian, A. R. Nix, and M. Beach, "On the performance of opportunistic NOMA in downlink CoMP networks," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 998–1001, May 2016.

[9] M. S. Ali, E. Hossain, A. Al-Dweik, and D. I. Kim, "Downlink power allocation for CoMP-NOMA in multi-cell networks," *IEEE Trans. Commun.*, vol. 66, no. 9, pp. 3982–3998, Sep. 2018.

[10] Y. Al-Eryani, E. Hossain, and D. I. Kim, "Generalized coordinated multi-point (GCoMP)-enabled NOMA: Outage, capacity, and power allocation," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 7923–7936, Nov. 2019.

[11] Y. Sun, Z. Ding, X. Dai, and O. A. Dobre, "On the performance of network NOMA in uplink CoMP systems: A. stochastic geometry approach," *IEEE Trans. Commun.*, vol. 67, no. 7, pp. 5084–5098, Jul. 2019.

[12] F. Lyu *et al.*, "Characterizing urban vehicle-to-vehicle communications for reliable safety applications," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 6, pp. 2586–2602, Jun. 2020.

[13] W. Liao, M. G. Kibria, G. P. Villardi, O. Zhao, K. Ishizu, and F. Kojima, "Coordinated multi-point downlink transmission for dense small cell networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 431–441, Jan. 2019.

[14] G. Zhao, S. Chen, L. Zhao, and L. Hanzo, "Joint energy-spectral-efficiency optimization of CoMP and BS deployment in dense large-scale cellular networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4832–4847, Jul. 2017.

[15] L. P. Qian, Y. Wu, H. Zhou, and X. Shen, "Joint uplink base station association and power control for small-cell networks with non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5567–5582, Sep. 2017.

[16] W. Wu, N. Zhang, N. Cheng, Y. Tang, K. Aldubaikhy, and X. Shen, "Beef up mmwave dense cellular networks with D2D-assisted cooperative edge caching," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 3890–3904, Apr. 2019.

[17] Z. Yang, Z. Ding, P. Fan, and N. Al-Dhahir, "A general power allocation scheme to guarantee quality of service in downlink and uplink NOMA systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7244–7257, Nov. 2016.

[18] S. Bassoy, H. Farooq, M. A. Imran, and A. Imran, "Coordinated multi-point clustering schemes: A survey," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 743–764, Feb. 2017.
[19] J. Liu, M. Sheng, and J. Li, "Improving network capacity scaling law in ultra-dense small cell networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 9, pp. 6218–6230, Sep. 2018.
[20] K. Feng and M. Haenggi, "A location-dependent base station cooperation scheme for cellular networks," *IEEE Trans. Commun.*, vol. 67, no. 9, pp. 6415–6426, Sep. 2019.
[21] W. Sun and J. Liu, "2-to-M coordinated multipoint-based uplink transmission in ultra-dense cellular networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8342–8356, Dec. 2018.
[22] C. Pan, M. Elkashlan, J. Wang, J. Yuan, and L. Hanzo, "User-centric C-RAN architecture for ultra-dense 5G networks: Challenges and methodologies," *IEEE Commun. Mag.*, vol. 56, no. 6, pp. 14–20, Jun. 2018.
[23] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1501–1505, Dec. 2014.
[24] Z. Zhang, H. Sun, and R. Q. Hu, "Downlink and uplink non-orthogonal multiple access in a dense wireless network," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2771–2784, Dec. 2017.
[25] Y. Liu, X. Li, F. R. Yu, H. Ji, H. Zhang, and V. C. M. Leung, "Grouping and cooperating among access points in user-centric ultra-dense networks with non-orthogonal multiple access," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2295–2311, Oct. 2017.
[26] X. Chen, R. Jia, and D. W. K. Ng, "On the design of massive non-orthogonal multiple access with imperfect successive interference cancellation," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 2539–2551, Mar. 2019.
[27] M. Moltafet, R. Joda, N. Mokari, M. R. Sabagh, and M. Zorzi, "Joint access and fronthaul radio resource allocation in PD-NOMA-based 5G networks enabling dual connectivity and CoMP," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6463–6477, Dec. 2018.
[28] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[29] L. T. H. An and P. D. Tao, "The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems," *Ann. Operations Res.*, vol. 133, no. 1, pp. 23–46, Jan. 2005.
[30] Y. Dai, M. Sheng, J. Liu, N. Cheng, X. Shen, and Q. Yang, "Joint mode selection and resource allocation for D2D-enabled NOMA cellular networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 6721–6733, Jul. 2019.
[31] N. Cheng *et al.*, "Space/aerial-assisted computing offloading for IoT applications: A learning-based approach," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 5, pp. 1117–1129, May 2019.
[32] T. Lipp and S. Boyd, "Variations and extension of the convex-concave procedure," *Optim. Eng.*, vol. 17, no. 2, pp. 263–287, Jun. 2016.
[33] M. Sheng, Y. Dai, J. Liu, N. Cheng, X. Shen, and Q. Yang, "Delay-aware computation offloading in noma mec under differentiated uploading delay," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2813–2826, Apr. 2020.
[34] J. Liu, M. Sheng, L. Liu, and J. Li, "Performance of small cell networks under multislope bounded pathloss model: From sparse to ultradense deployment," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 11022–11034, Nov. 2018.
[35] J. Zhao, Y. Liu, K. K. Chai, A. Nallanathan, Y. Chen, and Z. Han, "Spectrum allocation and power control for non-orthogonal multiple access in HetNets," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5825–5837, Sep. 2017.

Yanpeng Dai (Member, IEEE) received the B.Eng. degree from Shandong Normal University, Jinan, China, in 2014, and the Ph.D. degree from Xidian University, Xi'an, China, in 2020. He is currently an Assistant Professor and a Postdoctoral Researcher with the School of Information Science and Technology, Dalian Maritime University, Dalian, China. From September 2017 to September 2018, he was a Visiting Student with the University of Waterloo, Waterloo, ON, Canada. His research interests include resource management and optimization for heterogeneous wireless networks and maritime wireless communications.

Junyu Liu (Member, IEEE) received the B.Eng. and Ph.D. degrees in communication and information systems from Xidian University, Xi'an, China, in 2007 and 2016, respectively. He is currently an Associate Professor with the State Key Laboratory of Integrated Service Networks, Institute of Information and Science, Xidian University. His research interests include interference management and performance evaluation of wireless heterogeneous networks and ultra-dense wireless networks.

Min Sheng (Senior Member, IEEE) received the M.S. and Ph.D. degrees in communication and information systems from Xidian University, Xi'an, China, in 2000 and 2004, respectively. She is currently a Full Professor and the Director with the State Key Laboratory of Integrated Service Networks, Xidian University. Her general research interests include mobile ad hoc networks, 5G mobile communication systems, and satellite communications networks. She was the recipient of the Distinguished Young Researcher Award from NSFC and the Changjiang Scholar from the Ministry of Education, China.

Nan Cheng (Member, IEEE) received the B.E. and M.S. degrees from the Department of Electronics and Information Engineering, Tongji University, Shanghai, China, in 2009 and 2012, respectively, and the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, in 2016. From 2017 to 2019, he was a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada. He is currently a Professor with the State Key Lab of ISN and the School of Telecommunications Engineering, Xidian University, Xi'an, China. His current research focuses on B5G or 6G, space-air-ground integrated network, big data in vehicular networks, self-driving system, performance analysis, MAC, opportunistic communication, and application of AI for vehicular networks.

Xuemin Shen (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 1990. He is currently a University Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. His research interests include network resource management, wireless network security, Internet of Things, 5G and beyond, and vehicular ad hoc and sensor networks.

He was the recipient of the R.A. Fessenden Award from the IEEE, Canada, in 2019, the Award of Merit from the Federation of Chinese Canadian Professionals, Ontario, in 2019, the James Evans Avant Garde Award in 2018 from the IEEE Vehicular Technology Society, the Joseph LoCicero Award in 2015, the Education Award in 2017 from the IEEE Communications Society, the Technical Recognition Award from Wireless Communications Technical Committee, in 2019, and the AHSN Technical Committee, in 2013. He was also the recipient of the Excellent Graduate Supervision Award in 2006 from the University of Waterloo and the Premiers Research Excellence Award in 2003 from the Province of Ontario, Canada. He was the Technical Program Committee Chair or Co-Chair for the IEEE Globecom' 16, the IEEE Infocom' 14, the IEEE VTC' 10 Fall, the IEEE Globecom' 07, and the Chair for the IEEE Communications Society Technical Committee on Wireless Communications. He is currently the President Elect of the IEEE Communication Society, the elected IEEE Communications Society Vice President for Technical & Educational Activities, the Vice President for Publications, a Member-at-Large on the Board of Governors, the Chair of the Distinguished Lecturer Selection Committee, and a Member of IEEE ComSoc Fellow Selection Committee. He was the Editor-in-Chief of the IEEE INTERNET OF THINGS JOURNAL, the IEEE NETWORK, and the *IET Communications*. He is a registered Professional Engineer of Ontario, Canada, a Fellow of the Engineering Institute of Canada, a Fellow of the Canadian Academy of Engineering, a Fellow of the Royal Society of Canada, a Foreign Member of the Chinese Academy of Engineering, and a Distinguished Lecturer of the IEEE Vehicular Technology Society and Communications Society.