IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received March 31, 2019, accepted April 30, 2019, date of publication May 14, 2019, date of current version May 24, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2916921

Joint Power Allocation and User Scheduling for Device-to-Device-Enabled Heterogeneous Networks With Non-Orthogonal Multiple Access

JIAQI LIU[1], (Student Member, IEEE), GANG WU[1], (Member, IEEE), SA XIAO[1], (Member, IEEE), XIANGWEI ZHOU[2], (Senior Member, IEEE), GEOFFREY YE LI[3], (Fellow, IEEE), SHENGJIE GUO[2], AND SHAOQIAN LI[1], (Fellow, IEEE)

[1]National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China
[2]Division of Electrical and Computer Engineering, Louisiana State University, Baton Rouge, LA 70803, USA
[3]School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

Corresponding author: Gang Wu (wugang99@uestc.edu.cn)

This work was supported in part by the National Science Foundation of China under Grant 61771106, in part by the Research Funds for the International Cooperation Project of Sichuan Province under Grant 2018HH0138, in part by the National Science Foundation under Grant 1560437, and in part by the China Postdoctoral Science Foundation under Grant 2018M631075.

ABSTRACT *Device-to-device* (D2D) communications-enabled dense *heterogeneous networks* (HetNets) with *non-orthogonal multiple access* (NOMA) are promising solutions to meet the high-throughput requirement and support massive connectivity. In this paper, we propose a novel framework on the D2D-enabled HetNets with NOMA, where small cells underlay the uplink spectrum of macrocells to make full use of spectrum resources, NOMA technique is invoked to serve more downlink users simultaneously, and D2D-enabled multi-hop transmission is established to enhance signal reception of the *far users* (FUs) on cell edge. We investigate joint power allocation and user scheduling to maximize the ergodic sum rate of the *near users* (NUs) in the small cells while guaranteeing the quality-of-service requirements of the FUs and the macro-cell users. The optimal solution to this problem is complexity-prohibitive especially with large numbers of users and *small base stations* (SBSs) because it requires an exhaustive search over all possible combinations of SBSs, NUs, and FUs. To simplify the solution, we develop a two-step approach by decomposing the original problem into a power allocation problem and a user scheduling problem. We derive the closed-form solution of the power allocation problem via analyzing the objective function and constraints. The user scheduling problem is a joint user pairing and access point assignment problem. To solve it, we propose an SBS-NU-FU matching algorithm to obtain a near-optimal one-to-one three-sided matching of SBSs, NUs, and FUs. The simulation results show that the two-step method gets around 95% of the system throughput of the optimal one and can significantly improve the spectral efficiency of the D2D-enabled HetNets.

INDEX TERMS Device-to-device communications, heterogeneous networks, matching theory, non-orthogonal multiple access, resource allocation.

I. INTRODUCTION

Due to the exponential growth of wireless devices and data traffic in wireless networks, the forthcoming *fifth-generation* (5G) and future cellular networks will face daunting challenges of spectrum scarcity and massive connectivity [1]–[3]. To improve the spectrum utilization and provide more users with ubiquitous services, deploying ultra-dense *heterogeneous networks* (HetNets) has been included as a key feature in both existing and future cellular networks [4]–[6].

In HetNets, lots of low-cost and low-power *small-base stations* (SBSs) are densely deployed to provide high-data-rate services [7]. These SBSs can help offload the data traffic from *macro-base stations* (MBSs) and thus relieve traffic congestion problems and improve the *spectral efficiency* (SE) of cellular networks. Moreover, the SBSs can also act as relays to enhance the data transmission of the cell-edge users [8], [9].

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

However, deploying SBSs faces some challenges in practice. First, SBSs require additional infrastructure and high maintenance costs, especially with a large number of SBSs. Second, static SBSs may not be flexible enough for the dynamic environments [10]. To address these challenges, D2D communications has been proposed as an important complement of cellular communications in HetNets recently [11]–[13]. With D2D communications, data transmission can be directly established between user devices without traversing the *base stations* (BSs), which helps offload the data traffic from the BSs. Meanwhile, user devices can act as *D2D relay nodes* (DRNs) to enhance the transmission of cell-edge or deeply faded users. In comparison with SBSs, D2D communications are much easier to establish and require no additional infrastructure or maintenance cost [14], [15]. Therefore, D2D-enabled HetNets have attracted extensive research interests. Using D2D communications to offload data traffic from BSs is studied in [16], [17]. Popular contents are saved by cache-enabled users and shared among users via D2D communications in [16]. D2D communications are enabled to offload machine-type communications from cellular networks in [17]. Deploying DRNs to improve the connective experience of cell-edge or deeply faded users are studied in [18]–[20]. However, the aforementioned work only considers traditional *orthogonal multiple access* (OMA), without the use of *non-orthogonal multiple access* (NOMA).

NOMA allows the data transmission of multiple users to share the same time/frequency/code resource and thus has a great potential to support massive connectivity and provide higher SE [21]–[24]. However, resource allocation in NOMA systems becomes more challenging than that in OMA systems because power allocation among paired users needs to be carefully optimized to mitigate the co-channel interference among these users. It can be foreseen that introducing NOMA into D2D-enabled HetNets will greatly increase the complexity of resource allocation because different types of interference (i.e., co-channel interference of NOMA users, inter-tier interference of HetNets, and inter-link interference among D2D links) will be invoked. Therefore, the existing optimization methods for D2D-enabled HetNets with OMA are not applicable to the scenarios with NOMA and resource allocation becomes a crucial issue in D2D-enabled HetNets with NOMA.

Recently, there are some studies addressing the aforementioned challenges of applying NOMA in D2D-enabled HetNets. Spectrum allocation and power control for NOMA in HetNets is investigated in [25]. User scheduling for NOMA in HetNets is addressed in [26], [27]. Resource allocation in a two-tier D2D-enabled HetNet with NOMA is studied in [28], where each D2D transmitter serves multiple users with NOMA. However, the use of multi-hop D2D communications to achieve spatial diversity among NOMA users is not considered in the aforementioned work. In addition, efficient resource allocation algorithms have been proposed in [25] and [28] based on matching theory, but these algorithms are based on the assumption that the users have already been paired, without the consideration of user pairing.

There are also some studies on the usage of D2D communications in NOMA systems. For example, the spatial diversity gain achieved by D2D communications in NOMA networks is analyzed in [29], [30], *simultaneous wireless information and power transfer* (SWIPT) is applied at the DRNs to enhance the signal reception of cell-edge users with the energy harvested from BSs in [31], full-duplex technique is introduced at the DRNs to improve the system throughput in [32], [33], and mode switch in D2D-enabled NOMA networks is studied in [34]. However, the aforementioned work deploys only one user pair in single-cell networks and thus consider no user pairing. The trend of more and more intensive network deployment motivates us to deploy D2D-enabled NOMA in the scenarios with multiple user pairs and multi-tier HetNets and take user pairing into consideration.

In this paper, we propose a novel network framework of D2D-enabled HetNets with NOMA. We investigate joint power allocation and user scheduling to maximize the ergodic sum rate of small-cell *near users* (NUs) while satisfying the *Quality-of-Service* (QoS) requirements of small-cell *far users* (FUs) and *macro-cell users* (MUs). We develop a two-step approach by decomposing the original problem into a power allocation problem and a user scheduling problem. We derive the closed-form expressions of the optimal solution to the power allocation problem. To solve the user scheduling problem, we propose a three-sided matching algorithm [35], [36], instead of using the high-complexity exhaustive search algorithm, to deal with the matching of SBSs, FUs, and NUs. The three-sided matching algorithm has been exploited for resource allocation with the consideration of user pairing in wireless networks, such as subchannel allocation with uplink-downlink user pairing for full-duplex systems in [37] and subchannel allocation with vehicular-cellular user pairing for vehicular networks in [38].

The main contributions of this paper can be summarized as follows:

- We propose a new framework of D2D-enbled HetNets with NOMA, which improves the SE of D2D-enabled HetNets [20] with NOMA and extends the study on D2D-enabled NOMA [29], [39] to multi-user and multi-cell scenarios.
- We develop a three-sided matching algorithm to solve the resource allocation problem. Compared with the previous studies on two-sided matching between radio resources (e.g. access points and subchannels) and already existing user pairs in NOMA networks [25], [28], [40], [41], we further take user pairing into consideration.
- We consider a more practical scenario with partial *channel state information* (CSI) of interference channels and develop an algorithm robust to the uncertainty of interference channel.

TABLE 1. Notation.

Notation	Definition
$\mathcal{N}, \mathcal{F}, \mathcal{M}, \mathcal{B}$	Set of NUs, FUs, MUs, and SBSs
h	Channel coefficient of desired link
f, g	Channel coefficients of interference link
$\lambda_{m,j}, \nu_{m,k}$	Pathloss component of $g_{m,j}$ and $f_{m,k}$
$s_j^{\mathrm{N}}, s_k^{\mathrm{F}}, s_m^{\mathrm{M}}$	Desired signal for each NU, FU, and MU
$P_{m,j}^{\mathrm{N}}, P_{m,k}^{\mathrm{F}}$	SBS transmit power allocated for NU and FU
p_m	Uplink transmit power of MU
P_j	Relay transmit power of NU
w	Additive white Gaussian noise with variance σ^2
$\gamma_{m,j,k}^{\mathrm{N}}, \gamma_{m,j,k}^{\mathrm{N,F}}$	SINR for NU_j to decode s_j^{N} and s_k^{F}
$\gamma_{m,j,k}^{\mathrm{F}}$	SINR at NU on relay phase
$\gamma_{m,j,k}^{\mathrm{M,d}}, \gamma_{m,j,k}^{\mathrm{M,r}}$	SINR at MU on direct phase and relay phase
$\bar{\gamma}_k^{\mathrm{F}}, \bar{\gamma}_m^{\mathrm{M}}$	Minimum SINR threshold for FU and MU
p_0	Outage probability threshold for FUs
$x_{m,j,k}$	User scheduling indicator

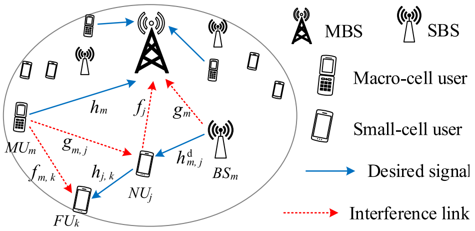


FIGURE 1. System model for a D2D-enabled cooperative NOMA HetNet.

The rest of this paper is organized as follows. In Section II, we present the system model and problem formulation. In Section III, we solve the power allocation problem for each SBS-NU-FU combination with closed-form expression. In Section IV, we formulate the user association problem as a one-to-one three-sided matching problem and propose a matching algorithm to solve the matching problem. In Section V, we present the simulation results. In Section VI, we conclude the paper. The notation of this paper is summarized in Table 1.

II. SYSTEM MODEL AND PROBLEM FORMULATION

We consider a two-tier HetNet shown as Figure 1, where M SBSs denoted as $\mathcal{B} = \{BS_1, \cdots, BS_M\}$ locate in a macro-cell to relieve the heavy traffic load of the MBS. The uplink spectrum of the MBS is divided into M orthogonal subchannels and each subchannel is designated for the uplink transmission of one MU. The M served MUs are denoted as $\mathcal{M} = \{MU_1, \cdots, MU_M\}$. J cell-center NUs, denoted as $\mathcal{N} = \{NU_1, \cdots, NU_J\}$, and K cell-edge FU,s denoted as $\mathcal{F} = \{FU_1, \cdots, FU_K\}$, are requiring downlink transmission from the SBSs. Due to physical obstacles or heavy shadowing, the FUs cannot establish good links to the BSs. Each SBS underlays the spectrum of one MU to serve an NU and an FU simultaneously by superimposing their signals in NOMA fashion. To help the signal reception of the FUs, we enable relay transmission of the NUs to forward the signals from the SBSs to the FUs. The whole downlink transmission of each SBS can be divided into two phases as shown in Figure 2. In the *direct phase*, the SBS transmits the superimposed signals to the NU. Then the NU decodes the FU's signal and its own signal sequentially with the *successive interference cancellation* (SIC) technique. In the *relay phase*, the NU forwards the FU's signal to the FU as a relay. We assume that NUs adopt half-duplex and decode-and-forward mode.

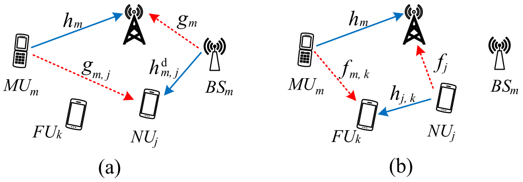


FIGURE 2. Signal model when BS_m serves NU_j and FU_k.

In the rest of this section, we present the channel model and the signal model and then formulate the joint power allocation and user scheduling problem.

A. CHANNEL MODEL

In general, wireless channels have pathloss depending on the distance and fast fading due to multi-path propagation [20], [33]. The channel gain of the small-cell downlink $h^{\mathrm{d}}_{m,j}$ from BS_m to NU_j is

$$|h^{\mathrm{d}}_{m,j}|^2 = \xi_{m,j} A d_{m,j}^{-\alpha}, \tag{1}$$

where $\xi_{m,j}$ is the fast fading gain with exponential distribution, A is a pathloss constant determined by system parameters, d is the distance between BS_m and NU_j, and α is the decay exponent. Similarly, we can express the channel gains of the macro-cell uplink h_m from MU_m to the MBS, the relay link $h_{j,k}$ from NU_j to FU_k, the interference link g_m from BS_m to the MBS, the interference link f_j from NU_j to the MBS, the interference link $g_{m,j}$ from MU_m to NU_j, and the interference link $f_{m,k}$ from MU_m to FU_k.

Assume that the CSI of the above-mentioned links are shared among MBS and SBSs. Based on the availability of the CSI, we divide these links into two categories:

1) *Complete-CSI channels:* through the cooperation between the source and the destination of each link, the complete instantaneous CSI of these links, such as h_m, $h^{\mathrm{d}}_{m,j}$, $h_{j,k}$, g_m, and f_j, can be obtained with pilot-based channel estimation and CSI feedback;
2) *Partial-CSI channels:* due to the lack of cooperation, only the pathloss components of these links, such as $g_{m,j}$ and $f_{m,k}$,[1] can be computed according to the locations of users.

[1] Acquiring the complete CSI of $g_{m,j}$ and $f_{m,k}$ at the BSs consumes a large number of time/frequency resources. In particular, $g_{m,j}$ can be estimated by NU_j and then transmitted to the BSs through uplink; $f_{m,k}$ can be estimated by FU_k, transmitted to NU_j through D2D communications, and then forwarded to the BSs through uplink. Therefore, to make full use of the scarce radio resources, we reduce the feedback of $g_{m,j}$ and $f_{m,k}$ with the consumption of partial CSI on these links.

B. SIGNAL MODEL

Assuming that BS_m underlays the uplink spectrum of MU_m to serve NU_j and FU_k, we present the signal models of the direct and relay phases, which are shown in Figure 2 (a) and (b), respectively.

1) DIRECT PHASE

As Figure 2 (a) shows, BS_m transmits the superimposed signal of NU_j and FU_k to NU_j. Let s_j^{N} and s_k^{F} denote the signals for NU_j and FU_k, respectively. The transmitted signal of BS_m is

$$x_m = \sqrt{P_{m,j}^{\mathrm{N}}} s_j^{\mathrm{N}} + \sqrt{P_{m,k}^{\mathrm{F}}} s_k^{\mathrm{F}}, \tag{2}$$

where $P_{m,j}^{\mathrm{N}}$ and $P_{m,k}^{\mathrm{F}}$ denote the transmit powers for s_j^{N} and s_k^{F}, respectively. Meanwhile, MU_m transmits its signal s_m^{M} to the MBS over the same subchannel. Then the received signal of NU_j is

$$y_j^{\mathrm{N}} = h_{m,j}^{\mathrm{d}} x_m + g_{m,j}\sqrt{p_m} s_m^{\mathrm{M}} + w_j^{\mathrm{N}}, \tag{3}$$

where p_m is the transmit power of MU_m, and $w_j^{\mathrm{N}} \sim \mathcal{CN}(0, \sigma^2)$ is the *additive Gaussian white noise* (AWGN) at NU_j with variance σ^2.

With SIC, NU_j first decodes the signal s_k^{F} of FU_k and then decodes its own signal s_j^{N} after subtracting s_k^{F} from y_j^{N}. When decoding s_k^{F}, NU_j treats both s_j^{N} and s_m^{M} as interference. The signals are normalized such that $E[|s_j^{\mathrm{N}}|^2] = E[|s_k^{\mathrm{F}}|^2] = E[|s_m^{\mathrm{M}}|^2] = 1$. Then the *signal-to-interference-plus-noise ratio* (SINR) for NU_j to decode s_k^{F} is

$$\gamma_{m,j,k}^{\mathrm{N,F}} = \frac{P_{m,j}^{\mathrm{F}} |h_{m,j}^{\mathrm{d}}|^2}{P_{m,j}^{\mathrm{N}} |h_{m,j}^{\mathrm{d}}|^2 + p_m |g_{m,j}|^2 + \sigma^2}. \tag{4}$$

After NU_j successfully decodes s_k^{F}, it subtracts s_k^{F} from y_j^{N} and decodes s_j^{N}. Here s_m^{M} is still treated as interference. Then the SINR for NU_j to decode s_j^{N} is

$$\gamma_{m,j,k}^{\mathrm{N}} = \frac{P_{m,j}^{\mathrm{N}} \left|h_{m,j}^{\mathrm{d}}\right|^2}{p_m \left|g_{m,j}\right|^2 + \sigma^2}. \tag{5}$$

The MBS receives a superimposed signal of s_m^{M} and x_m as

$$y_m^{\mathrm{M,d}} = h_m \sqrt{p_m} s_m^{\mathrm{M}} + \sqrt{\kappa} g_m x_m + w_m^{\mathrm{d}}, \tag{6}$$

where $w_m^{\mathrm{d}} \sim \mathcal{CN}(0, \sigma^2)$ is the AWGN and κ characterizes the capability of the interference mitigation technique adopted by the MBS to mitigate the interference from the SBSs.[2] Then the SINR for the MBS to decode s_m^{M} is

$$\gamma_{m,j,k}^{\mathrm{M,d}} = \frac{p_m |h_m|^2}{\kappa(P_{m,j}^{\mathrm{N}} + P_{m,k}^{\mathrm{F}}) |g_m|^2 + \sigma^2}. \tag{7}$$

Since FU_k cannot establish a good link to BS_m, we ignore the received signal of FU_k in the direct phase.

2) RELAY PHASE

As Figure 2 (b) shows, after NU_j successfully decodes s_k^{F}, it transmits s_k^{F} to FU_k with power P_j over the same subchannel that BS_m uses in the direct phase. In consideration of the interference from MU_m, the received signal at FU_k is

$$y_k^{\mathrm{F}} = h_{j,k}\sqrt{P_j} s_k^{\mathrm{F}} + f_{m,k}\sqrt{p_m} s_m^{\mathrm{M}} + w_k^{\mathrm{F}}, \tag{8}$$

where $w_k^{\mathrm{F}} \sim \mathcal{CN}(0, \sigma^2)$ is the AWGN at FU_k. Then the SINR for FU_k to decode its signal s_k^{F} is

$$\gamma_{m,j,k}^{\mathrm{F}} = \frac{P_j \left|h_{j,k}\right|^2}{p_m |f_{m,k}|^2 + \sigma^2}. \tag{9}$$

The MBS receives a superimposed signal of s_m^{M} and s_k^{F} as

$$y_m^{\mathrm{M,r}} = h_m \sqrt{p_m} s_m^{\mathrm{M}} + f_j \sqrt{P_j} s_k^{\mathrm{F}} + w_m^{\mathrm{r}}, \tag{10}$$

where $w_m^{\mathrm{r}} \sim \mathcal{CN}(0, \sigma^2)$ is the AWGN. Then the SINR for the MBS to decode s_m^{M} is

$$\gamma_{m,j,k}^{\mathrm{M,r}} = \frac{p_m |h_m|^2}{P_j \left|f_j\right|^2 + \sigma^2}. \tag{11}$$

C. PROBLEM FORMULATION

In this subsection, we formulate the optimization problem based on the aforementioned system model.

Denote a user association indicator $x_{m,j,k}$, which is $x_{m,j,k} = 1$ when BS_m serves NU_j and FU_k and $x_{m,j,k} = 0$ otherwise. We aim to maximize the ergodic sum rate of the NUs while guaranteeing the QoS requirements of the FUs and the MUs. The ergodic rate of NU_j when it connects to BS_m and serves FU_j as a relay is

$$\mathbb{R}_{m,j,k} = \mathbb{E}\left[\frac{1}{2}\log_2\left(1 + \gamma_{m,j,k}^{\mathrm{N}}\right)\right], \tag{12}$$

where the expectation $\mathbb{E}[\cdot]$ is taken over the fast fading distribution and the factor $1/2$ accounts for the equal time partition between the direct and relay phases. Then the resource allocation problem can be formulated as

$$\max_{\substack{\{P_{m,j}^{\mathrm{N}}\},\{P_{m,k}^{\mathrm{F}}\},\\ \{P_j\},\{p_m\},\{x_{m,j,k}\}}} \sum_{m=1}^{M}\sum_{j=1}^{J}\sum_{k=1}^{K} x_{m,j,k}\,\mathbb{R}_{m,j,k} \tag{13}$$

$$\text{s.t. } \Pr\left\{\gamma_{m,j,k}^{\mathrm{N,F}} \le \bar{\gamma}_k^{\mathrm{F}}\right\} \le p_0, \text{ if } x_{m,j,k} = 1, \tag{13a}$$

$$\Pr\left\{\gamma_{m,j,k}^{\mathrm{F}} \le \bar{\gamma}_k^{\mathrm{F}}\right\} \le p_0, \text{ if } x_{m,j,k} = 1, \tag{13b}$$

$$\min\left\{\gamma_{m,j,k}^{\mathrm{M,d}}, \gamma_{m,j,k}^{\mathrm{M,r}}\right\} \ge \bar{\gamma}_m^{\mathrm{M}}, \ \forall m, \tag{13c}$$

$$\begin{aligned} &P_{m,j}^{\mathrm{N}} \ge 0, \ P_{m,k}^{\mathrm{F}} \ge 0, \\ &P_{m,j}^{\mathrm{N}} + P_{m,k}^{\mathrm{F}} \le P_{BS_m}^{\max}, \ \forall m\, \forall j\, \forall k, \end{aligned} \tag{13d}$$

$$0 \le p_m \le p_m^{\max}, \ \forall m, \tag{13e}$$

$$0 \le P_j \le P_{NU_j}^{\max}, \ \forall j, \tag{13f}$$

$$x_{m,j,k} \in \{0, 1\}, \ \forall m\, \forall j\, \forall k, \tag{13g}$$

$$\textstyle\sum_{m=1}^{M}\sum_{k=1}^{K} x_{m,j,k} = 1, \ \forall j, \tag{13h}$$

[2] It is practical to mitigate the inter-tier interference between the MBS and the SBSs with cooperative signal processing and beamforming [42].

$$\textstyle\sum_{m=1}^{M}\sum_{j=1}^{J} x_{m,j,k} = 1, \ \ \forall k, \tag{13i}$$

$$\textstyle\sum_{j=1}^{J}\sum_{k=1}^{K} x_{m,j,k} = 1, \ \ \forall m, \tag{13j}$$

where $\bar{\gamma}_k^{\mathrm{F}}$ and $\bar{\gamma}_m^{\mathrm{M}}$ are the minimum SINR thresholds for the FUs and the MUs, respectively, to establish reliable links, p_0 is the tolerable outage probability of the FUs, constraints (13b) and (13a) ensure the maximum outage probability bounds of the FUs on the direct and relay links, respectively, constraint (13c) ensures the minimum rate requirements of the MUs, constraint (13d) ensures the validity of the SBS power allocation for the NUs and the FUs, constraint (13e) is the peak uplink transmit power constraint for the MUs, constraint (13f) is the peak relay transmit power constraint for the NUs, constraint (13g) ensures that the user association indicators are binary, constraints (13h) and (13i) ensure that each FU can only be paired with at most one NU and vice versa, respectively, and constraint (13j) ensures that each SBS can serve at most one NU-FU user pair.

It can be easily seen that problem (13) is a mixed integer and nonlinear optimization problem, which cannot be directly solved. In the following sections, we solve problem (13) by decomposing it into two sub-problems, i.e., the power allocation problem and the user scheduling problem.

III. OPTIMAL POWER ALLOCATION FOR EACH SBS-NU-FU COMBINATION

In this section, we study the optimal power allocation for each SBS-NU-FU combination. Given an arbitrary user association pattern, where BS_m serves NU_j and FU_k, the power allocation problem in (13) is

$$\max_{P_{m,j}^{\mathrm{N}}, P_{m,j}^{\mathrm{F}}, P_j, p_m} \ \mathbb{R}_{m,j,k} \tag{14}$$

$$\text{s.t. } \Pr\left\{\gamma_{m,j,k}^{\mathrm{N,F}} \le \bar{\gamma}_k^{\mathrm{F}}\right\} \le p_0, \tag{14a}$$

$$\Pr\left\{\gamma_{m,j,k}^{\mathrm{F}} \le \bar{\gamma}_k^{\mathrm{F}}\right\} \le p_0, \tag{14b}$$

$$\min\left\{\gamma_{m,j,k}^{\mathrm{M,d}}, \gamma_{m,j,k}^{\mathrm{M,r}}\right\} \ge \bar{\gamma}_m^{\mathrm{M}}, \tag{14c}$$

$$\begin{aligned} &P_{m,j}^{\mathrm{N}} \ge 0, \ P_{m,k}^{\mathrm{F}} \ge 0, \\ &P_{m,j}^{\mathrm{N}} + P_{m,k}^{\mathrm{F}} \le P_{BS_m}^{\max}, \end{aligned} \tag{14d}$$

$$0 \le p_m \le p_m^{\max}, \tag{14e}$$

$$0 \le P_j \le P_{NU_j}^{\max}. \tag{14f}$$

Due to the recent advances on cellular communication technologies and the slowly varying positions of the SBSs, the MBS can suppress the interference from the SBSs effectively with zero-forcing receiver [43]. Therefore, we can have an approximation of $\kappa \approx 0$, which indicates $\gamma_{m,j,k}^{\mathrm{M,d}} \approx \frac{p_m|h_m|^2}{\sigma^2} < \gamma_{m,j,k}^{\mathrm{M,r}}$ and constraint (14c) can be simplified as

$$\gamma_{m,j,k}^{\mathrm{M,r}} = \frac{p_m|h_m|^2}{P_j\left|f_j\right|^2 + \sigma^2} \ge \bar{\gamma}_{m,j,k}^{\mathrm{M}}. \tag{15}$$

In the rest of this section, we will solve problem (14). To begin with, we introduce the following proposition to show our observations on the property of the ergodic rate $\mathbb{R}_{m,j,k}$ of NU_j.

Proposition 1: The ergodic rate of NU_j can be written as

$$\mathbb{R}_{m,j,k} = \frac{1}{2}\int_0^{+\infty} \log_2\left(1 + \frac{P_{m,j}^{\mathrm{N}}\left|h_{m,j}^{\mathrm{d}}\right|^2}{p_m|g_{m,j}|^2 + \sigma^2}\right) \times \frac{1}{\lambda_{m,j}} e^{-\frac{\left|h_{m,j}^{\mathrm{d}}\right|^2}{\lambda_{m,j}}} \mathrm{d}\left|h_{m,j}^{\mathrm{d}}\right|^2, \tag{16}$$

where $\lambda_{m,j}$ is the path loss on the interference link from MU_m to NU_j. From (16), we can easily obtain the following observations:

- *With fixed $P_{m,j}^{\mathrm{N}}$, $\mathbb{R}_{m,j,k}$ monotonically decreases with p_m;*
- *With fixed p_m, $\mathbb{R}_{m,j,k}$ monotonically increases with $P_{m,j}^{\mathrm{N}}$.*

Based on **Proposition 1**, we derive the optimal solution to problem (14) in **Theorem 1** using **Lemma 2** and **Lemma 1**, proved in Appendix A and B, respectively.

Lemma 1: According to constraint (14a), the optimal solution to problem (14) must satisfy $P_{m,k}^{\mathrm{F}} = f_1\left(P_{m,j}^{\mathrm{N}}, p_m\right)$, where

$$f_1\left(P_{m,j}^{\mathrm{N}}, p_m\right) \triangleq -p_m\frac{\bar{\gamma}_k^{\mathrm{F}}\lambda_{m,j}\ln p_0}{|h_{m,j}^{\mathrm{d}}|^2} + P_{m,j}^{\mathrm{N}}\bar{\gamma}_k^{\mathrm{F}} + \frac{\bar{\gamma}_k^{\mathrm{F}}\sigma^2}{|h_{m,j}^{\mathrm{d}}|^2}. \tag{17}$$

Lemma 2: According to constraints (14b) and (14c), problem (14) is feasible only if

$$|h_m|^2|h_{j,k}|^2 + \bar{\gamma}_m^{\mathrm{M}}\bar{\gamma}_k^{\mathrm{F}}\left|f_j\right|^2 \nu_{m,k}\ln p_0 > 0, \tag{18}$$

where $\nu_{m,k}$ is the path loss on the interference link from MU_m to FU_k, and the optimal solution to problem (14) must satisfy $p_m \ge p_m^{\mathrm{LB}}$ and $P_j = f_2(p_m)$, where

$$p_m^{\mathrm{LB}} \triangleq \frac{\bar{\gamma}_m^{\mathrm{M}}|h_{j,k}|^2\sigma^2 + \bar{\gamma}_k^{\mathrm{F}}\bar{\gamma}_m^{\mathrm{M}}\left|f_j\right|^2\sigma^2}{|h_m|^2|h_{j,k}|^2 + \bar{\gamma}_k^{\mathrm{F}}\bar{\gamma}_m^{\mathrm{M}}\left|f_j\right|^2\nu_{m,k}\ln p_0} \tag{19}$$

and

$$f_2(p_m) \triangleq p_m\frac{|h_m|^2}{\left|f_j\right|^2\bar{\gamma}_m^{\mathrm{M}}} - \frac{\sigma^2}{\left|f_j\right|^2}. \tag{20}$$

Ignoring the peak transmit power constraints, **Lemma 1** ensures that the QoS constraint of FU_k in the direct phase can be satisfied and **Lemma 2** ensures that the QoS constraints of MU_m and FU_k in the relay phase can be satisfied. Then we derive the optimal solution to problem (14) in **Theorem 1** in Appendix C, where we first obtain the feasible region of problem (14) by taking **Lemma 2**, **Lemma 1**, and the peak transmit power constraints into consideration and then derive the closed-form expression of the optimal solution.

Theorem 1: Problem (14) is feasible only if

$$P_{BS_m}^{\max} - \frac{\bar{\gamma}_k^{\mathrm{F}}\sigma^2}{|h_{m,j}^{\mathrm{d}}|^2} > 0, \tag{21}$$

$$p_m^{\mathrm{LB}} \le \min\left\{p_m^{\max}, A_1\right\}, \tag{22}$$

and (18) hold, where $A_1 = \frac{P_{BS_m}^{\max}|h_{m,j}^{\mathrm{d}}|^2 - \bar{\gamma}_k^{\mathrm{F}}\sigma^2}{\bar{\gamma}_k^{\mathrm{F}}\lambda_{m,j}\ln(1/p_0)}$*. The optimal solution* $\left\{P_{m,j}^{\mathrm{N}\,*}, P_{m,j}^{\mathrm{F}\,*}, P_j^*, p_m^*\right\}$ *to problem (14) is*

$$\left\{f_3(p_m^*), P_{BS_m}^{\max} - P_{m,j}^{\mathrm{N}\,*}, f_2(p_m^*), p_m^{\mathrm{LB}}\right\}, \tag{23}$$

where

$$f_3(p_m) = \frac{P_{BS_m}^{\max}|h_{m,j}^{\mathrm{d}}|^2 - p_m\bar{\gamma}_k^{\mathrm{F}}\lambda_{m,j}\ln(1/p_0) - \bar{\gamma}_k^{\mathrm{F}}\sigma^2}{(1+\bar{\gamma}_k^{\mathrm{F}})|h_{m,j}^{\mathrm{d}}|^2}. \tag{24}$$

Given **Theorem 1**, we set $\mathbb{R}_{m,j,k}^* = -\infty$ if problem (14) is not feasible; Otherwise, we first derive the optimal power allocation according to (23) and then compute the corresponding ergodic rate $\mathbb{R}_{m,j,k}^*$ of NU_j based on the following lemma, proved in Appendix D.

Lemma 3: When NU_j connects to BS_m and serves FU_j as a relay, the ergodic rate of NU_j is given by

$$\begin{aligned}\mathbb{R}_{m,j,k} =& \frac{1}{2}\log_2\left(1+\frac{P_{m,j}^{\mathrm{N}}\left|h_{m,j}^{\mathrm{d}}\right|^2}{\sigma^2}\right)\\ &-\frac{1}{2\ln 2}e^{\frac{P_{m,j}^{\mathrm{N}}\left|h_{m,j}^{\mathrm{d}}\right|^2+\sigma^2}{p_m\lambda_{m,j}}}\mathsf{Ei}\left(-\frac{P_{m,j}^{\mathrm{N}}\left|h_{m,j}^{\mathrm{d}}\right|^2+\sigma^2}{p_m\lambda_{m,j}}\right)\\ &+\frac{1}{2\ln 2}e^{\frac{\sigma^2}{p_m\lambda_{m,j}}}\mathsf{Ei}\left(-\frac{\sigma^2}{p_m\lambda_{m,j}}\right),\end{aligned} \tag{25}$$

where $\mathsf{Ei}(x) = \int_{-\infty}^{x}\frac{e^t}{t}\mathrm{d}t$ *is the exponential integral function.*

IV. EFFICIENT USER SCHEDULING WITH THE MATCHING ALGORITHM

In the previous section, we have obtained the optimal power allocation for each SBS-NU-FU combination and the corresponding NU ergodic rate $\mathbb{R}_{m,j,k}^*$. The resource allocation problem in (13) can be reduced to a user scheduling problem, which is a joint user pairing and access point assignment problem and can expressed as

$$\begin{aligned}&\max_{\{x_{m,j,k}\}}\sum_{m=1}^{M}\sum_{j=1}^{J}\sum_{k=1}^{K}x_{m,j,k}\mathbb{R}_{m,j,k}^*\\ &\text{s.t. (13g)} \;--\; \text{(13j)}.\end{aligned} \tag{26}$$

The optimal solution to problem (26) requires an exhaustive search, which is complexity-prohibitive. To efficiently solve problem (26), we formulate it as a one-to-one three-sided matching problem and propose a low-complexity *SBS-NU-FU matching algorithm* (SNFMA) to obtain a near-optimal solution to this matching problem. In the rest of this section, we will formulate the SBS-NU-FU matching problem, describe the details of the SNFMA, and analyze the stability, convergence, and complexity of the SNFMA.

A. ONE-TO-ONE THREE-SIDED SBS-NU-FU MATCHING PROBLEM

We formulate problem (26) as a one-to-one three-sided matching problem. To better describe the matching problem, we will first introduce some notations and definitions for the matching problem.

Definition 1: When NU_j connects to BS_m and serves FU_j as a relay, we say that BS_m, NU_j, and FU_k are matched with each other. They three together form a ***matching triple*** *denoted by $\boldsymbol{\Delta} = (BS_m, NU_j, FU_k)$. Any two of these three agents form an agent pair, i.e., (BS_m, NU_j), (BS_m, FU_k), and (NU_j, FU_k) are* ***agent pairs****.*

We denote the *triple utility* of any triple (BS_m, NU_j, FU_k) by the ergodic rate of the NU in this triple, i.e., $U_{m,j,k} = \mathbb{R}_{m,j,k}^*$. Apparently, each triple has unique channels among its agents and therefore yields a unique triple utility. When an agent is matched with different agent pairs and form different triples, these triples yield different utilities. Based on the differences among the triple utilities, we define the *preference* of each agent over the agent pairs. For example, we say that NU_j prefers agent pair (BS_m, FU_k) to agent pair (BS_p, FU_q) because triple (BS_m, NU_j, FU_k) yields a higher utility than triple (BS_p, NU_j, FU_q) does, which can be expressed as

$$(BS_m, FU_k) \succ_{NU_j} (BS_p, FU_q) \;\Leftrightarrow\; U_{m,j,k} > U_{p,j,q}. \tag{27}$$

By sorting the utilities of the triples, a strictly-ordered *preference list* [44] can be built up for each agent to record its preference over all agent pairs from the other two agent sets. The set of the preference lists of all agents is denoted as

$$\begin{aligned}\boldsymbol{P} = \{&\boldsymbol{P}(BS_1), \cdots, \boldsymbol{P}(BS_M), \boldsymbol{P}(NU_1), \cdots, \boldsymbol{P}(NU_J),\\ &\boldsymbol{P}(FU_1), \cdots, \boldsymbol{P}(FU_K)\},\end{aligned} \tag{28}$$

where $\boldsymbol{P}(BS_m)$, $\boldsymbol{P}(NU_j)$, and $\boldsymbol{P}(FU_k)$ are the preference lists of BS_m, NU_j, and FU_k, which contain at most JK NU-FU pairs, MK SBS-FU pairs, and MJ SBS-NU pairs, respectively. Note that the agent pairs that yield non-positive utilities are not listed in the preference lists. For example, the preference list $\boldsymbol{P}(NU_j)$ of NU_j contains all SBS-FU pairs that yield positive utilities with NU_j and these SBS-FU pairs are in descending order of NU_j's preference over them such that (BS_m, NU_j) is always in front of (BS_p, NU_q) if $(BS_m, FU_k) \succ_{NU_j} (BS_p, FU_q)$.

Next, we will formulate problem (26) as a matching problem with the notations and definitions of the matching triple and the preference list.

Definition 2: Given a set $\mathcal{B}$ of M SBSs, a set $\mathcal{N}$ of J NUs, and a set $\mathcal{F}$ of K FUs. A one-to-one three-sided matching Ψ is a function defined as $\Psi : \mathcal{B} \longmapsto \mathcal{N} \times \mathcal{F}, \mathcal{N} \longmapsto \mathcal{B} \times \mathcal{F}, \mathcal{F} \longmapsto \mathcal{B} \times \mathcal{N}$. Under Ψ, there are at most $L = \min\{M, J, K\}$ disjoint matching triples denoted by $\{\boldsymbol{\Delta}_1, \cdots, \boldsymbol{\Delta}_L\}$, which satisfy:

1) *$\Psi(BS_m) = (NU_j, FU_k)$, $\Psi(NU_j) = (BS_m, FU_k)$, and $\Psi(FU_k) = (BS_m, NU_j)$ if BS_m, NU_j, and FU_k are in matching triple $\boldsymbol{\Delta}_l$, $l \in \{1, \cdots, L\}$;*
2) *$\Psi(BS_m) = \varnothing$, if BS_m is not in any matching triple;*
3) *$\Psi(NU_j) = \varnothing$, if NU_j is not in any matching triple;*
4) *$\Psi(FU_k) = \varnothing$, if FU_k is not in any matching triple;*
5) *$\boldsymbol{\Delta}_p \cap \boldsymbol{\Delta}_q = \varnothing$, $\forall p, q \in \{1, \cdots, L\}andp \neq q$.*

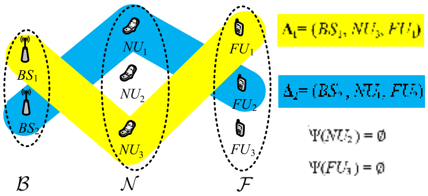


FIGURE 3. An example of a one-to-one three-sided SBS-NU-FU matching composed of two matching triples.

Figure 3 illustrates an example of a one-to-one three-sided SBS-NU-FU matching, where two SBSs, two NUs, and two FUs form two disjoint matching triples and NU_2 and FU_k are two unmatched agents.

The total utility of matching Ψ is defined as the sum of the triple utilities of the matching triples under Ψ, which can be expressed as

$$U_\Psi = \sum_{m=1}^{M} \sum_{j=1}^{J} \sum_{k=1}^{K} x_{m,j,k} \mathbb{R}^*_{m,j,k}. \tag{29}$$

We formulate problem (26) as a *one-to-one three-sided SBS-NU-FU matching problem*, which can be described as: we aim to find a one-to-one three-sided matching Ψ from a set of SBSs, a set of NUs, and a set of FUs that maximizes the total utility U_Ψ, subject to constraints (13g) – (13j). The optimization variables of this problem are $\{x_{m,j,k}\}$. Next, we will propose a matching algorithm to solve this problem in Section IV-B.

B. SBS-NU-FU MATCHING ALGORITHM

We solve the one-to-one three-sided SBS-NU-FU matching problem by utilizing matching theory, where each NU is an agent trying to be matched with its favorite SBS-FU pair and the SBSs and FUs tend to choose the NUs that maximizes the total utility. We give the NUs the priority to choose their favorite SBS-FU pairs, and thus the matching procedure begins with the NUs nominating themselves to their favorite SBS-FU pairs. After receiving the nominations from the NUs, the SBSs and the FUs will decide whether to accept or reject these nominations. It is likely that an SBS or an FU may receive multiple nominations from more NUs than one, which indicates the conflict of the nominations happens. To explain how the SBSs and FUs choose among the conflicting nominations from multiple NUs, we introduce the concept of *blocking triples*, which is defined as follows.

Definition 3: Given a matching Ψ and a triple $\mathbf{\Delta}_{\mathrm{B}}$ containing at least one matched agent and at least one unmatched agent under Ψ. We have:

1) *If only one agent of $\mathbf{\Delta}_{\mathrm{B}}$ is matched under Ψ, let a_1 denote this matched agent. $\mathbf{\Delta}_{\mathrm{B}}$ is a **blocking triple** under Ψ when $U_{\mathbf{\Delta}_{\mathrm{B}}} > U_{a_1,\Psi(a_1)}$;*
2) *If two agents of $\mathbf{\Delta}_{\mathrm{B}}$ are matched under Ψ, let a_1 and a_2 denote these matched agents. We have:*
   a) *If $a_1 \in \Psi(a_2)$, $\mathbf{\Delta}_{\mathrm{B}}$ is a **blocking triple** under Ψ when $U_{\mathbf{\Delta}_{\mathrm{B}}} > U_{a_1,\Psi(a_1)}$;*
   b) *If $a_1 \notin \Psi(a_2)$, $\mathbf{\Delta}_{\mathrm{B}}$ is a **blocking triple** under Ψ when $U_{\mathbf{\Delta}_{\mathrm{B}}} > U_{a_1,\Psi(a_1)} + U_{a_2,\Psi(a_2)}$.*

Denote a_1, a_2, and a_3 as the three agents in triple $\mathbf{\Delta}_{\mathrm{B}}$, where a_3 is unmatched and at least one of a_1 and a_2 is matched under Ψ. If a_3 nominates itself to (a_1, a_2) and (a_1, a_2) accepts a_3, at least one original triple under Ψ will be broken up because at least one agent of a_1 and a_2 must reject its currently matched agents as a result of being matched with a_3. Items 1) and 2a) imply the case that one original triple will be broken up, where $\mathbf{\Delta}_{\mathrm{B}}$ is a blocking triple if the utility of $\mathbf{\Delta}_{\mathrm{B}}$ is higher than that of the broken-up triple. Item 2b) implies the case that two original triples will be broken up, where $\mathbf{\Delta}_{\mathrm{B}}$ is a blocking triple if the utility of $\mathbf{\Delta}_{\mathrm{B}}$ is higher than the sum of the triple utilities of the broken-up triples.

The details of the SNFMA are described in **Algorithm 1**, which consists of a power allocation procedure in Stage I, an initialization procedure in Stage II, and a matching procedure in Stage III.

In Stage I, the power allocation for each SBS-NU-FU combination is determined, and the corresponding ergodic rate of each NU is computed.

In Stage II, the preference list of each NU is obtained by sorting the preference over the SBS-FU pairs of each NU, and three unmatched agent sets, $\mathcal{US}$, $\mathcal{UN}$, and $\mathcal{UF}$ are built to record whether or not each SBS, NU, and FU, respectively, are matched.

In Stage III, the matching procedure requires multiple iterations. In each iteration, each NU_j nominates itself to its favorite SBS-FU pair (BS_{m^*}, FU_{k^*}) that has never rejected NU_j in the previous iterations. If both BS_{m^*} and FU_{k^*} have never received any nomination in the previous iterations and receive no nomination from other NUs in this iteration, (BS_{m^*}, FU_{k^*}) will temporarily accept the nomination from NU_j and be match with NU_j. Otherwise, if BS_{m^*}, or FU_{k^*}, or both of them have already been matched before this iteration or have received nominations from other FUs in this iteration, we first check whether $(BS_{m^*}, NU_j, FU_{k^*})$ is a blocking triple under current matching: if $(BS_{m^*}, NU_j, FU_{k^*})$ is a blocking triple, BS_{m^*} and FU_{k^*} will accept NU_j and rejects their previously matched NUs; otherwise, they will reject NU_j. Note that this blocking triple only "blocks" the current matching, i.e., BS_{m^*} and FU_{k^*} may reject NU_j and be matched with another NU in a later iteration if the newly formed triple "blocks" $(BS_{m^*}, NU_j, FU_{k^*})$. The iterations stop when no NU nominate itself to the SBS-FU pairs, which means that each NU is either matched to an SBS-FU pair or rejected by all SBS-FU pairs. The matching pattern is recorded by setting $x_{m,j,k} = 1$ if BS_m, NU_j, and FU_k are matched and $x_{m,j,k} = 0$ otherwise.

C. ANALYSIS OF THE SBS-NU-FU MATCHING ALGORITHM

1) STABILITY AND CONVERGENCE

We illustrate that the SNFMA can finally converge to a stable matching. To begin with, we introduce the definition of a stable matching.

Algorithm 1 SBS-NU-FU Matching Algorithm

```
Input: Set of SBSs $\mathcal{B}$; Set of NUs $\mathcal{N}$; Set of FUs $\mathcal{F}$
Output: A stable one-to-one three-sided matching $\Psi$; Power
    allocation scheme.
1:  – Stage I: Power Allocation
2:  for $m = 1 : M$ do
3:      for $j = 1 : J$ do
4:          for $k = 1 : K$ do
5:              Compute the optimal power allocation and the
                corresponding NU ergodic rate according to Theorem 1 and Lemma 3, respectively;
6:          end for
7:      end for
8:  end for
9:  – Stage II: Initialization of Matching Algorithm
10: Form $M \times K$ SBS-FU pairs from $\mathcal{B} \times \mathcal{F}$;
11: Construct the preference lists of all NUs in $\mathcal{N}$ over all
    SBS-FU pairs: $\{\boldsymbol{P}(NU_1), \cdots, \boldsymbol{P}(NU_J)\}$;
12: Build up the sets of unmatched SBSs $\mathcal{US} = \mathcal{B}$,
    unmatched NUs $\mathcal{UN} = \mathcal{N}$, and unmatched UFs $\mathcal{UF} = \mathcal{F}$;
13: – Stage III: Matching
14: while $\mathcal{UN}$ is not empty and the preference list $\boldsymbol{P}(NU_j)$ of
    each $NU_j$ is not empty do
15:     $NU_j$ nominates itself to its currently favorite SBS-FU
        pair $(BS_{m^*}, FU_{k^*})$, namely the first one in $\boldsymbol{P}(NU_j)$;
16:     if $BS_{m^*} \in \mathcal{US}$ and $FU_{k^*} \in \mathcal{UF}$ then
17:         Match $NU_j$ with $(BS_{m^*}, FU_{k^*})$ and record this
            matching triple;
18:         Remove $NU_j$, $BS_{m^*}$ and $FU_{k^*}$ from $\mathcal{UN}$, $\mathcal{US}$, and
            $\mathcal{UF}$, respectively;
19:     else if $(BS_{m^*}, NU_j, FU_{k^*})$ is a blocking triple then
20:         Match $NU_j$ with $(BS_{m^*}, FU_{k^*})$ and record this
            matching triple;
21:         Remove $NU_j$ from $\mathcal{UN}$;
22:         $BS_{m^*}$ and $FU_{k^*}$ reject their currently matched
            agents;
23:         Add the rejected agents in their corresponding
            unmatched agent sets;
24:     end if
25:     Remove $(BS_{m^*}, FU_{k^*})$ from $\boldsymbol{P}(NU_j)$;
26: end while
```

*Definition 4: A matching Ψ is **stable** if there exists no blocking triple except the matching triples under Ψ.*

Lemma 4: If the SNFMA converges to a matching Ψ^*, Ψ^* is a stable matching.

Proof: Suppose that a matching Ψ^* is achieved by the end of Stage III of **Algorithm 1**. We prove that every triple except the matching triples under Ψ^* cannot be a blocking triple. As shown in Line 15 of **Algorithm 1**, every NU is willing to nominate itself to its favorite SBS-FU pair that has not previously rejected it, i.e., an NU will not nominate itself to an SBS-FU pair until it is rejected by all SBS-FU pairs that it prefers to this SBS-FU pair. Moreover, the condition of the while loop (Line 14 of the algorithm) indicates that each NU will continue the nomination until it is accepted by an SBS-FU pair or rejected by every SBS-FU pair. Suppose that there is a triple $\boldsymbol{\Delta}_t = (BS_m, NU_j, FU_k)$ except the matching triples under Ψ^*. Then there are two possible cases:

Case 1) NU_j has never nominated itself to (BS_m, FU_k). Based on the previous description, NU_j must prefer its currently matched SBS-FU pair $\Psi^*(NU_j)$ to (BS_m, FU_k), i.e., $\Psi^*(NU_j) \succ_{NU_j} (BS_m, FU_k)$ and $U_{NU_j,\Psi^*(NU_j)} > U_{\boldsymbol{\Delta}_t}$, which contradicts the definition of a blocking triple. Therefore, $\boldsymbol{\Delta}_t$ is not a blocking triple.

Case 2) NU_j has nominated itself to (BS_m, FU_k) at a certain iteration, but it is rejected by (BS_m, FU_k) later. This indicates that one agent in (BS_m, FU_k) or both of them can form another triple $\boldsymbol{\Delta}_s$ with an NU other than NU_j such that $U_{\boldsymbol{\Delta}_s} > U_{\boldsymbol{\Delta}_t}$, which also contradicts the definition of a blocking triple. Therefore, $\boldsymbol{\Delta}_t$ is not a blocking triple.

Since $\boldsymbol{\Delta}_t$ is arbitrarily chosen, the above conclusion is applicable to any triple except the matching triples under Ψ^*. Therefore, matching Ψ^* is stable. ■

Theorem 2: The SNFMA converges to a stable matching in a finite number of iterations.

Proof: We prove that the matching process of the SNFMA in **Algorithm 1** will end in a limited number of iterations. In each iteration, each unmatched NU nominates itself to its favorite SBS-FU pair (Line 15 of **Algorithm 1**) and remove this SBS-FU pair from its preference list (Line 25 of **Algorithm 1**). As shown in Line 14 of **Algorithm 1**, the iteration terminates once every NU is matched or the preference list of each NU is empty. Since the preference list of each NU containing no more than MK SBS-FU pairs is with finite length and gets shorter after each iteration, the number of nominations each NU can make is no more than MK and the total number of the while-loop iterations is no more than MK. Therefore, the SNFMA can converge to a final matching in a finite number of iterations and the final matching is stable according to **Lemma 4**. ■

2) COMPLEXITY

Here we analyze the computational complexity of the SNFMA and compare it with the *optimal exhaustive search algorithm* (OESA) and the *two-step two-sided matching algorithm* (TSTSMA). Through the OESA, all possible cases of user association are listed, and then the case that yields the highest sum ergodic rate of NUs is selected as the optimal user association. Through the TSTSMA, user pairing and access point association are decoupled as two separate two-sided matching problems solved with the Gale-Shapeley algorithm sequentially [37].

The SNFMA, the OESA, and the TSTSMA all need first to solve the power allocation problem for each SBS-NU-FU combination, shown as Stage I of **Algorithm 1**, which involves totally MJK times of calculating (23) and (25) to solve the power allocation problems and to derive the corresponding NU ergodic rates, respectively.

Then we analyze the computational complexity of solving the user scheduling problem with the OESA and the SNFMA, respectively. In the OESA, we first assume $\min\{M, J, K\} = M$, which indicates that each of the M SBSs is associated with one NU and one FU. By choosing M NUs from $\mathcal{N}$ and M FUs from $\mathcal{F}$, there are totally $\binom{J}{M} \times \binom{K}{M}$ agent selection patterns. For each agent selection pattern, we have $(M!)^2$ matching cases among the selected agents. The OESA needs to traverse every matching case for every agent selection pattern. Similarly, we can calculate the complexity for the cases of $\min\{M, J, K\} = J$ and $\min\{M, J, K\} = K$. Therefore, the computational complexity of solving the user scheduling problem with the OESA is given by

$$\begin{cases} O\left(\dfrac{J!K!(M!)^2}{(J-M)!(K-M)!}\right), & \text{if } \min\{M,J,K\} = M, \\ O\left(\dfrac{M!K!(J!)^2}{(M-J)!(K-J)!}\right), & \text{if } \min\{M,J,K\} = J, \\ O\left(\dfrac{M!J!(K!)^2}{(M-K)!(J-K)!}\right), & \text{if } \min\{M,J,K\} = K. \end{cases} \tag{30}$$

Next, we analyze the computational complexity of the SNFMA. In Stage II of **Algorithm 1**, each NU requires a sorting process to obtain its preference list, which has a computational complexity of $O(M^2K^2)$. Therefore, the total computational complexity of Stage II is $O(JM^2K^2)$. In Stage III of **Algorithm 1**, each NU nominates itself at most MK times and thus the computational complexity of matching procedure at Stage III is at most $O(JMK)$. Final, we analyze the computational complexity of the TSTSMA. To solve the user pairing problem, each NU obtains its preference list through a sorting process with a computational complexity of $O(K^2)$ and nominates itself at most K times. After user pairing, there exists at most $L = \min\{J, K\}$ user pairs. To solve the access point association problem, each user pair obtains its preference list through a sorting process with a computational complexity of $O(M^2)$ and nominates itself at most M times. Therefore, the total computational complexity of the TSTSMA is $O(JK^2 + LM^2)$ for the sorting procedure and $O(JK + LM)$ for the matching procedure.

As we have analyzed, the computational complexity of the OESA shows an exponential growth with the increasing numbers of SBSs, NUs, and FUs. In contrast, the computational complexity of the SNFMA shows a quadratic growth with the increasing numbers of SBSs and users. Therefore, the SNFMA has the advantage of a much lower complexity, especially when plenty of SBSs and users are deployed in the system. Our simulation results in Section V will demonstrate such an advantage by showing that the running time of the SNFMA is much shorter than that of the OESA.

V. SIMULATION RESULTS

In this section, we evaluate the performance of the SNFMA. We consider a three-sector cell with radius $r_0 = 250$ meters as shown in Figure 4, where the macro-BS is located in the center of the cell and three SBSs are deployed in the center of the three sectors [20]. The MUs are randomly located in the cell. The NUs are randomly located between $r_1 = 100$ m and r_0. The FUs are randomly allocated in the extended cell area between r_0 and $r_2 = 300$ m. The values of major simulation parameters are listed in Table 2.

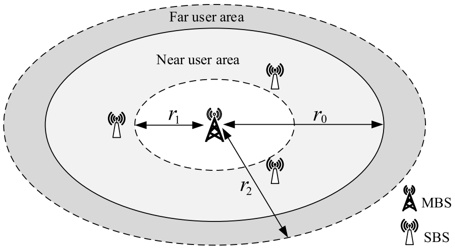


FIGURE 4. Illustration of system deployment.

TABLE 2. Simulation parameters.

Parameter	Value
Carrier center frequency	2 GHz
SBS bandwidth	5 MHz
Peak transmit power for each MU $p_m^{\max}$	23 dBm
Peak transmit power for each NU $P_{NU_j}^{\max}$	23 dBm
AWGN power spectral density	-174 dBm/Hz
Pathloss constant A	10^{-2}
Decay exponent α	3.76
SINR thresholds $\bar{\gamma}_k^{\mathrm{F}}$, $\bar{\gamma}_m^{\mathrm{M}}$	10 dB
Outage probability threshold p_0	0.01

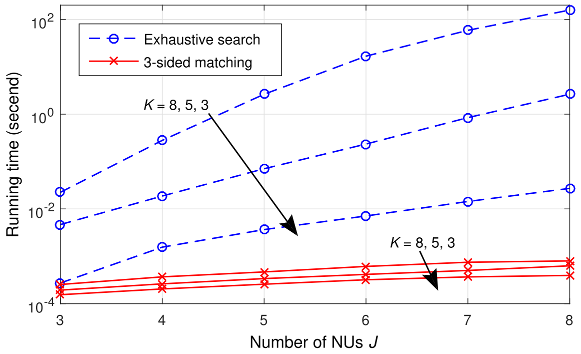


FIGURE 5. Comparison of running time between the OESA and the SNFMA.

Figure 5 shows the average running time of the SNFMA, in comparison with the OESA. The simulation is running with MATLAB 2015b on a personal computer with Intel Core i7-6770K CPU and 16 GB DDR3 RAM. From the figure, the average running time of the SNFMA is obviously much shorter than that of the OESA. The average running time of the OESA exhibits an exponential growth with the increasing number of users. In contrast, the running time of the SNFMA grows with the increasing number of users in a much lower speed.

The three subtables in Table 3 shows the average number of scheduled triples versus different $P_{BS_m}^{\max}$, J and K, and $\bar{\gamma}_k^{\mathrm{F}}$ and $\bar{\gamma}_m^{\mathrm{M}}$, respectively. In each subtable, the default parameters

TABLE 3. Average number of scheduled triples.

$P^{\max}_{BS_m}$ (dBm)	0	10	20	30	35	40
N_{triple}	1.09	1.61	1.80	2.12	2.13	2.14
$J = K$	3	4	5	6	7	8
N_{triple}	1.35	1.67	1.96	2.13	2.33	2.42
$\bar{\gamma}^{\text{F}}_k = \bar{\gamma}^{\text{M}}_m$ (dB)	0	4	8	12	16	20
N_{triple}	2.89	2.74	2.44	1.84	1.20	0.71

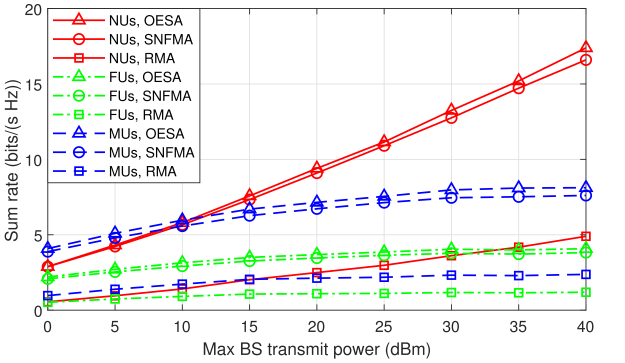


FIGURE 6. Average sum rate of users versus peak BS transmit power, with $J = K = 6$.

are $P^{\max}_{BS_m} = 35$ dBm, $J = K = 6$, and $\bar{\gamma}^{\text{F}}_k = \bar{\gamma}^{\text{M}}_m = 10$ dB. From the table, more triples are scheduled as $P^{\max}_{BS_m}$, J, and K increase, and fewer triples are scheduled as $\bar{\gamma}^{\text{F}}_k$ and $\bar{\gamma}^{\text{M}}_m$ increase. The reason is obvious: with higher $P^{\max}_{BS_m}$, the MUs' and FUs' QoS constraints are easier to satisfy; with larger J and K, more triples satisfying the MUs' and FUs' QoS constraints are available in the system; however, with higher $\bar{\gamma}^{\text{F}}_k$ and $\bar{\gamma}^{\text{M}}_m$, the MUs' and FUs' QoS constraints are more difficult to satisfy and therefore fewer triples can be scheduled.

Figure 6 shows the average sum rates of three sets of users (i.e., the NUs, the FUs, and the MUs) versus SBS peak transmit power $P^{\max}_{BS_m}$ in the case of $J = K = 6$. We compare the SNFMA with the OESA and the *random matching algorithm* (RMA). Through the RMA, the power allocation of each SBS-NU-FU combination is determined according to **Theorem 1** and the user scheduling is randomly determined. From the figure, the SNFMA performs very close to the OESA. In particular, when $P^{\max}_{BS_m} = 35$ dBm, the SNFMA gets around 96%, 94%, and 94% of the sum rates of the NUs, the FUs, and the MUs, respectively, by the OESA. In addition, the SNFMA significantly outperforms the RMA, which verifies the validity of the matching algorithm.

Figure 7 demonstrates the *cumulative distribution function* (CDF) of achievable individual user rate in the case of $J = K = 6$ and $P^{\max}_{BS_m} = 35$ dBm. We compare the NOMA-based network with an OMA-based network. For the OMA-based network, each time slot is equally separated into three phases, which are respectively used for the SBSs to transmit the signals of the NUs to the NUs, for the SBSs to transmit the signals of the FUs to the NUs, and for the NUs to forward the signals to the FUs. The optimization of the OMA-based network adopts the three-sided matching algorithm to maximize the NU's and the FUs' sum ergodic rates given the MUs' minimum SINR constraints, based on incomplete CSI of the interference links. From the figure, NOMA can efficiently improve the rates of the NUs due to the superimposed structure. The lower bound of the curves represents the probability of the idle state of each subchannel, where no user is scheduled on that subchannel, which is 30.6% for OMA and 20.6% for NOMA. This indicates that the performance gain of NOMA enables the network to schedule more users.

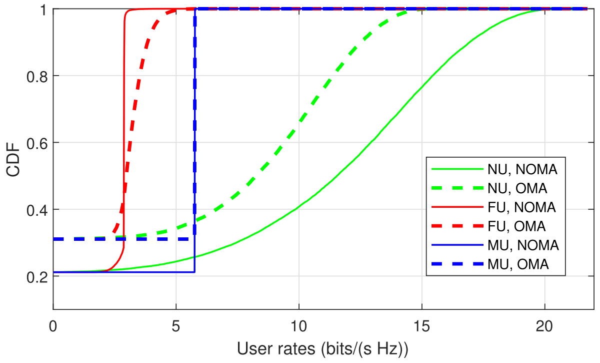


FIGURE 7. The CDF of achievable individual user rate, with $J = K = 6$ and $P^{\max}_{BS_m} = 35$ dBm.

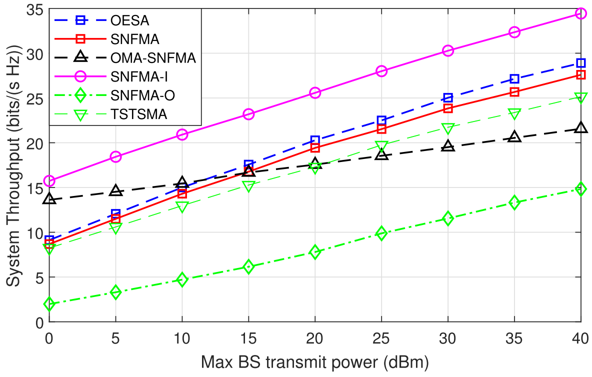
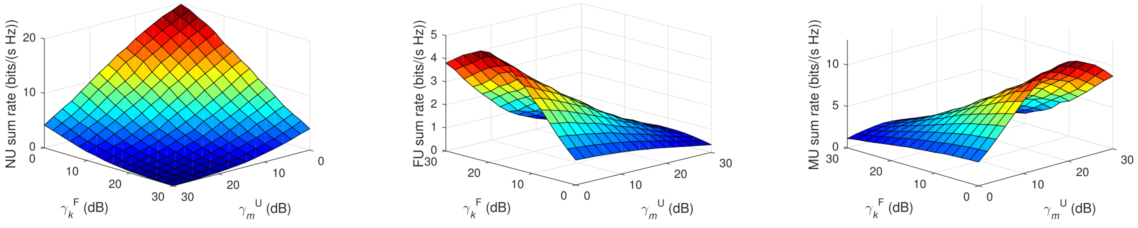


FIGURE 8. Average system throughput versus peak BS transmit power, with $J = K = 6$.

Figure 8 depicts the average system throughput versus $P^{\max}_{BS_m}$ in the case of $J = K = 6$. We compare the SNFMA with three benchmarks: 1) the OMA-based network (OMA-SNFMA), 2) an NOMA-based D2D-enabled HetNet with complete CSI for interference links (SNFMA-CCSI), and 3) an NOMA-based D2D-enabled HetNet with complete but outdated CSI for interference links (SNFMA-OCSI). SNFMA-CCSI and SNFMA-OCSI adopt the three-sided matching algorithm to maximize the NUs' sum instant rate instead of the NUs' sum ergordic rate, given the FUs' minimum SINR constraints instead of the FUs' maximum outage probability constraints. Compared to the OMA network, the NOMA network can achieve higher system throughput when $P^{\max}_{BS_m}$ is higher than around 12 dBm in our scenario. For example, when $P^{\max}_{BS_j} = 35$ dBm, the SNFMA achieves



(a) (b) (c)

FIGURE 9. Average achievable sum rate of users versus SINR thresholds, with $J = K = 6$ and $P^{\max}_{BS_m} = 35$ dBm. (a) Average achievable sum rate of NUs. (b) Average achievable sum rate of FUs. (c) Average achievable sum rate of MUs.

roughly a 42% higher average system throughput than the OMA-SNFMA. Compared with the instant-rate optimization based on complete CSI, our proposed method performs closely with the instant-rate optimization with complete CSI and outperforms the instant-rate optimization with outdated CSI significantly. In particular, when $P^{\max}_{BS_j} = 35$ dBm, the SNFMA gets around 80% of the average system throughput achieved by the SNFMA-CCSI and around 193% of the average system throughput achieved by the SNFMA-OCSI. In additions, the SNFMA achieves roughly a 10% higher average system throughput than the TSTSMA described in Section IV-C.2 when $P^{\max}_{BS_j} = 35$ dBm.

Figure 9 shows of the average achievable sum rates of the three sets of users versus the minimum SINR thresholds for MUs and FUs in the case of $J = K = 6$ and $P^{\max}_{BS_m} = 35$ dBm. In Figure 9(a), the NUs' average achievable sum rate decreases as either $\bar{\gamma}^{\mathrm{F}}_k$ or $\bar{\gamma}^{\mathrm{M}}_m$ increases. This is because increasing $\bar{\gamma}^{\mathrm{F}}_k$ encourages the SBSs to allocate more transmit power to the FUs and thus to allocate less transmit power to the NUs. And increasing $\bar{\gamma}^{\mathrm{M}}_m$ encourages the MUs to increase their transmit powers, which introduces more interference to the NUs. In Figure 9(b), the FUs' average achievable sum rate decreases as $\bar{\gamma}^{\mathrm{M}}_m$ increases because increasing $\bar{\gamma}^{\mathrm{M}}_m$ encourages the MUs to increase their transmit powers, which introduces more interference to the FUs. On the other hand, the FUs' average achievable sum rate first increases and then decreases as $\bar{\gamma}^{\mathrm{F}}_k$ increases. This is because increasing $\bar{\gamma}^{\mathrm{F}}_k$ encourages the SBSs to allocate more transmit power to the FUs. However, the outage probability of the FUs also increases with increasing $\bar{\gamma}^{\mathrm{F}}_k$. In Figure 9(c), the MUs' average achievable sum rate decreases as $\bar{\gamma}^{\mathrm{F}}_k$ increases because increasing $\bar{\gamma}^{\mathrm{F}}_k$ encourages the NUs to increase the relay transmit powers, which introduces more interference to the MBS. On the other hand, the MUs' average achievable sum rate first increases and then decreases as $\bar{\gamma}^{\mathrm{M}}_m$ increases. This is because increasing $\bar{\gamma}^{\mathrm{M}}_m$ encourage the MUs to increase their transmit powers. However, the outage probability of the MUs also increases with increasing $\bar{\gamma}^{\mathrm{M}}_m$.

VI. CONCLUSIONS

In this paper, we propose a novel framework on D2D-enabled HetNets with NOMA, where the SBSs can serve multiple users simultaneously with NOMA and D2D-enabled multi-hop transmission is established to enhance the signal reception of the FUs. We study joint power allocation and user scheduling to maximize the ergodic rate of the NUs while guaranteeing the QoS requirements of the FUs and the MUs. To solve the joint optimization problem, we develop a two-step approach by decomposing the original problem into two sub-problems, which are the power allocation problem and the user scheduling problem. We obtain the closed-form expression of the optimal solution to the power allocation problem. We formulate the user scheduling problem as a one-to-one three-sided matching problem and propose a low-complexity matching algorithm to obtain a near-optimal solution to the matching problem. Simulation results show that the two-step method performs very closely with the optimal one and can significantly improve the SE of D2D-enabled HetNets.

APPENDIX

A. PROOF OF LEMMA 1

According to the channel model, the *probability density function* (PDF) of $|g_{m,j}|^2$ is $f_{|g_{m,j}|^2}(x) = \frac{1}{\lambda_{m,j}} e^{-\frac{x}{\lambda_{m,j}}}$. Based on (4), we have

$$\begin{aligned}
&\Pr\left\{\gamma^{\mathrm{N,F}}_{m,j,k} \le \bar{\gamma}^{\mathrm{F}}_k\right\}\\
&\quad= \Pr\left\{\frac{P^{\mathrm{F}}_{m,j}|h^{\mathrm{d}}_{m,j}|^2}{P^{\mathrm{N}}_{m,j}|h^{\mathrm{d}}_{m,j}|^2 + p_m|g_{m,j}|^2 + \sigma^2} \le \bar{\gamma}^{\mathrm{F}}_k\right\}\\
&\quad= \Pr\left\{|g_{m,j}|^2 \ge \frac{P^{\mathrm{F}}_{m,j}|h^{\mathrm{d}}_{m,j}|^2 - \bar{\gamma}^{\mathrm{F}}_k P^{\mathrm{N}}_{m,j}|h^{\mathrm{d}}_{m,j}|^2 - \bar{\gamma}^{\mathrm{F}}_k\sigma^2}{\bar{\gamma}^{\mathrm{F}}_k p_m}\right\}\\
&\quad= \int_{\frac{P^{\mathrm{F}}_{m,j}|h^{\mathrm{d}}_{m,j}|^2 - \bar{\gamma}^{\mathrm{F}}_k P^{\mathrm{N}}_{m,j}|h^{\mathrm{d}}_{m,j}|^2 - \bar{\gamma}^{\mathrm{F}}_k\sigma^2}{\bar{\gamma}^{\mathrm{F}}_k p_m}}^{\infty} \frac{1}{\lambda_{m,j}} e^{-\frac{x}{\lambda_{m,j}}}\,\mathrm{d}x\\
&\quad= \begin{cases} e^{-\frac{P^{\mathrm{F}}_{m,j}|h^{\mathrm{d}}_{m,j}|^2 - \bar{\gamma}^{\mathrm{F}}_k P^{\mathrm{N}}_{m,j}|h^{\mathrm{d}}_{m,j}|^2 - \bar{\gamma}^{\mathrm{F}}_k\sigma^2}{\bar{\gamma}^{\mathrm{F}}_k p_m \lambda_{m,j}}}, \\ \qquad P^{\mathrm{F}}_{m,j} \ge P^{\mathrm{N}}_{m,j}\bar{\gamma}^{\mathrm{F}}_k + \frac{\bar{\gamma}^{\mathrm{F}}_k\sigma^2}{|h^{\mathrm{d}}_{m,j}|^2}, \\ 1, \quad P^{\mathrm{F}}_{m,j} < P^{\mathrm{N}}_{m,j}\bar{\gamma}^{\mathrm{F}}_k + \frac{\bar{\gamma}^{\mathrm{F}}_k\sigma^2}{|h^{\mathrm{d}}_{m,j}|^2}. \end{cases} \qquad (31)
\end{aligned}$$

Substituting (31) in constraint (14a), we have

$$\begin{aligned}
P^{\mathrm{F}}_{m,k} &\ge -p_m\frac{\bar{\gamma}^{\mathrm{F}}_k\lambda_{m,j}\ln p_0}{|h^{\mathrm{d}}_{m,j}|^2} + P^{\mathrm{N}}_{m,j}\bar{\gamma}^{\mathrm{F}}_k + \frac{\bar{\gamma}^{\mathrm{F}}_k\sigma^2}{|h^{\mathrm{d}}_{m,j}|^2}\\
&= f_1\left(P^{\mathrm{N}}_{m,j}, p_m\right). \qquad (32)
\end{aligned}$$

Let $(P^{\mathrm{N}*}_{m,j}, P^{\mathrm{F}*}_{m,k}, P^*_j, p^*_m)$ denote the optimal solution to problem (14) and $\mathbb{R}^*_{m,j,k}$ denote the corresponding optimal ergodic rate of NU_j. We assume

$$P^{\mathrm{F}*}_{m,k} > f_1\left(P^{\mathrm{N}*}_{m,j}, p^*_m\right). \tag{33}$$

Since $f_1\left(P^{\mathrm{N}}_{m,j}, p_m\right)$ is a monotonically increasing and continuous function of $P^{\mathrm{N}}_{m,j}$ with fixed p_m, there must exist another power allocation

$$(P^{\mathrm{N}*}_{m,j}+\epsilon, P^{\mathrm{F}*}_{m,k}-\epsilon, P^*_j, p^*_m), \tag{34}$$

which also satisfies (33), i.e., $P^{\mathrm{F}*}_{m,k}-\epsilon > f_1\left(P^{\mathrm{N}*}_{m,j}+\epsilon, p^*_m\right)$, where ϵ is a positive value small enough. Since $\left(P^{\mathrm{N}*}_{m,j}+\epsilon\right)+\left(P^{\mathrm{F}*}_{m,k}-\epsilon\right) = P^{\mathrm{N}*}_{m,j}+P^{\mathrm{F}*}_{m,k} \le P^{\max}_{BS_m}$, power allocation (34) satisfies constraint (14d). It is easy to prove that the power allocation (34) satisfies the other constraints of problem (14) because these constraints are only w.r.t. P_j and p_m. Let $\mathbb{R}^{\dagger}_{m,j,k}$ denote the ergodic rate of NU_j with power allocation (34). Since $\mathbb{R}_{m,j,k}$ is a monotonically increasing function of $P^{\mathrm{N}}_{m,j}$ with fixed p_m according to on **Proposition 1**, it is easy to prove that $\mathbb{R}^{\dagger}_{m,j,k} > \mathbb{R}^*_{m,j,k}$, which contradicts the optimality of $\mathbb{R}^*_{m,j,k}$. Therefore, the assumption (33) is invalid, which indicates $P^{\mathrm{F}}_{m,j} = f_1\left(P^{\mathrm{N}}_{m,j}, p_m\right)$ according to (32).

B. PROOF OF LEMMA 2

According to the channel model, the PDF of $|f_{m,k}|^2$ is $f_{|f_{m,k}|^2}(x) = \frac{1}{\nu_{m,k}} e^{-\frac{x}{\nu_{m,k}}}$. Based on (9), we have

$$\begin{aligned}
\Pr\left\{\gamma^{\mathrm{F}}_{m,j,k} \le \bar{\gamma}^{\mathrm{F}}_k\right\} &= \Pr\left\{\frac{P_j\left|h_{j,k}\right|^2}{p_m|f_{m,k}|^2+\sigma^2} \le \bar{\gamma}^{\mathrm{F}}_k\right\}\\
&= \Pr\left\{|f_{m,k}|^2 \ge \frac{P_j\left|h_{j,k}\right|^2-\bar{\gamma}^{\mathrm{F}}_k\sigma^2}{p_m\bar{\gamma}^{\mathrm{F}}_k}\right\}\\
&= \int_{\frac{P_j|h_{j,k}|^2-\bar{\gamma}^{\mathrm{F}}_k\sigma^2}{p_m\bar{\gamma}^{\mathrm{F}}_k}}^{\infty} \frac{1}{\nu_{m,k}} e^{-\frac{x}{\nu_{m,k}}}\,\mathrm{d}x\\
&= \begin{cases} e^{-\frac{P_j|h_{j,k}|^2-\bar{\gamma}^{\mathrm{F}}_k\sigma^2}{p_m\bar{\gamma}^{\mathrm{F}}_k\nu_{m,k}}}, & P_j \ge \frac{\bar{\gamma}^{\mathrm{F}}_k\sigma^2}{|h_{j,k}|^2},\\ 1, & P_j < \frac{\bar{\gamma}^{\mathrm{F}}_k\sigma^2}{|h_{j,k}|^2}. \end{cases}
\end{aligned} \tag{35}$$

Substituting (35) in constraint (14b), we have

$$P_j \ge -p_m\frac{\bar{\gamma}^{\mathrm{F}}_k\nu_{m,k}\ln p_0}{|h_{j,k}|^2} + \frac{\bar{\gamma}^{\mathrm{F}}_k\sigma^2}{|h_{j,k}|^2}. \tag{36}$$

Constraint (14c) can be transformed into

$$P_j \le p_m\frac{|h_m|^2}{|f_j|^2\,\bar{\gamma}^{\mathrm{M}}_m} - \frac{\sigma^2}{|f_j|^2} = f_2(p_m). \tag{37}$$

Combining (36) and (37), we have

$$-p_m\frac{\bar{\gamma}^{\mathrm{F}}_k\nu_{m,k}\ln p_0}{|h_{j,k}|^2}+\frac{\bar{\gamma}^{\mathrm{F}}_k\sigma^2}{|h_{j,k}|^2} \le P_j \le p_m\frac{|h_m|^2}{|f_j|^2\,\bar{\gamma}^{\mathrm{M}}_m}-\frac{\sigma^2}{|f_j|^2}, \tag{38}$$

which is valid only if

$$-p_m\frac{\bar{\gamma}^{\mathrm{F}}_k\nu_{m,k}\ln p_0}{|h_{j,k}|^2} + \frac{\bar{\gamma}^{\mathrm{F}}_k\sigma^2}{|h_{j,k}|^2} \le p_m\frac{|h_m|^2}{|f_j|^2\,\bar{\gamma}^{\mathrm{M}}_m} - \frac{\sigma^2}{|f_j|^2}. \tag{39}$$

Then we transform (39) into

$$p_m\left(\frac{|h_m|^2}{|f_j|^2\,\bar{\gamma}^{\mathrm{M}}_m} + \frac{\bar{\gamma}^{\mathrm{F}}_k\nu_{m,k}\ln p_0}{|h_{j,k}|^2}\right) \ge \frac{\sigma^2}{|f_j|^2} + \frac{\bar{\gamma}^{\mathrm{F}}_k\sigma^2}{|h_{j,k}|^2}. \tag{40}$$

Since p_m is non-negative, the inequity in (40) can hold only if $\frac{|h_m|^2}{|f_j|^2\bar{\gamma}^{\mathrm{M}}_m} + \frac{\bar{\gamma}^{\mathrm{F}}_k\nu_{m,k}\ln p_0}{|h_{j,k}|^2} > 0$, which is equivalent to (18).

When (18) holds, (40) can be transformed into

$$p_m \ge \frac{\bar{\gamma}^{\mathrm{M}}_m|h_{j,k}|^2\sigma^2 + \bar{\gamma}^{\mathrm{F}}_k\bar{\gamma}^{\mathrm{M}}_m\left|f_j\right|^2\sigma^2}{|h_m|^2|h_{j,k}|^2 + \bar{\gamma}^{\mathrm{F}}_k\bar{\gamma}^{\mathrm{M}}_m\left|f_j\right|^2\nu_{m,k}\ln p_0} = p^{\mathrm{LB}}_m. \tag{41}$$

Let $(P^{\mathrm{N}*}_{m,j}, P^{\mathrm{F}*}_{m,k}, P^*_j, p^*_m)$ denote the optimal solution to problem (14) and $\mathbb{R}^*_{m,j,k}$ denote the corresponding optimal ergodic rate of NU_j. We assume

$$P^*_j < f_2(p^*_m). \tag{42}$$

Since $f_2\left(p_m\right)$ is a monotonically increasing and continuous function of p_m, there must exist another power allocation

$$(P^{\mathrm{N}*}_{m,j}, P^{\mathrm{F}*}_{m,k}, P^*_j+\epsilon, p^*_m-\epsilon), \tag{43}$$

which satisfies (42), i.e., $P^*_j+\epsilon < f_2\left(p^*_m-\epsilon\right)$, where ϵ is a positive value small enough. It is easy to prove that the power allocation (43) satisfies the other constraints of problem (14) since these constraints are easier to be satisfied with a higher P_j and a lower p_m. Let $\mathbb{R}^{\mathrm{N}\dagger}_{m,j,k}$ denote the ergodic rate of NU_j with power allocation (43). Since $\mathbb{R}_{m,j,k}$ is a monotonically decreasing function of p_m with fixed $P^{\mathrm{N}}_{m,j}$ according to on **Proposition 1**, we have $\mathbb{R}^{\mathrm{N}\dagger}_{m,j,k} > \mathbb{R}^*_{m,j,k}$, which contradicts the optimality of $\mathbb{R}^*_{m,j,k}$. Therefore, assumption (42) is invalid, which indicates $P^*_j = f_2(p^*_m)$ according to (37).

C. PROOF OF THEOREM 1

Substituting (17) into (14d), we have

$$P^{\mathrm{N}}_{m,j}\left(1+\bar{\gamma}^{\mathrm{F}}_k\right) + p_m\frac{\bar{\gamma}^{\mathrm{F}}_k\lambda_{m,j}\ln\frac{1}{p_0}}{|h^{\mathrm{d}}_{m,j}|^2} \le P^{\max}_{BS_m} - \frac{\bar{\gamma}^{\mathrm{F}}_k\sigma^2}{|h^{\mathrm{d}}_{m,j}|^2}. \tag{44}$$

Due to the non-negative values of $P^{\mathrm{N}}_{m,j}$ and p_m, the right-hand side of inequity (44) must be non-negative, i.e., (21) must hold.

According to (44), **Lemma 2**, and (14e), we can plot the feasible region for problem (14) depending on $P^{\mathrm{N}}_{m,j}$ and p_m in two cases, based on whether $p^{\max}_m < A_1$ or not as shown in Figure 10 (a) and (b), where $A_1 = \frac{P^{\max}_{BS_m}|h^{\mathrm{d}}_{m,j}|^2-\bar{\gamma}^{\mathrm{F}}_k\sigma^2}{\bar{\gamma}^{\mathrm{F}}_k\nu_{m,j}\ln(1/p_0)}$ and $A_2 = \frac{P^{\max}_{BS_m}|h^{\mathrm{d}}_{m,j}|^2-\bar{\gamma}^{\mathrm{F}}_k\sigma^2}{(1+\bar{\gamma}^{\mathrm{F}}_k)|h^{\mathrm{d}}_{m,j}|^2}$. According to **Proposition 1**, the optimal solution to problem (14) can only reside at point $P = \left(p^{\mathrm{LB}}_m, f_3\left(p^{\mathrm{LB}}_m\right)\right)$.

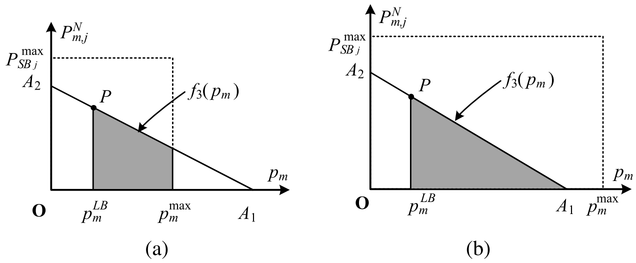


FIGURE 10. Two cases of feasible regions for problem (14) depending on $P_{m,j}^{\mathrm{N}}$ and p_m. (a) Case I: $p_m^{\max} < A_1$. (b) Case II: $p_m^{\max} \geq A_1$.

D. PROOF OF LEMMA 3

Let $z = \gamma_{m,j,k}^{\mathrm{F}}$. Then the ergodic rate of NU_j is

$$\begin{aligned}\mathbb{R}_{m,j,k} &= \mathbb{E}_Z\left[\frac{1}{2}\log_2\left(1+z\right)\right]\\ &= \frac{1}{2}\int_0^{\infty}\log_2\left(1+z\right)f_Z(z)\mathrm{d}z\\ &= \frac{1}{2\ln 2}\int_0^{\infty}\frac{1-F_Z(z)}{1+z}\mathrm{d}z, \end{aligned} \tag{45}$$

where $f_Z(z)$ and $F_Z(z)$ are the PDF and the *cumulative distribution function* (CDF) of the random variable Z, respectively.

Then $F_Z(z)$ in (45) is given by

$$\begin{aligned}F_Z(z) &= \Pr\left\{\frac{P_{m,j}^{\mathrm{N}}\left|h_{m,j}^{\mathrm{d}}\right|^2}{p_m|g_{m,j}|^2+\sigma^2} < z\right\}\\ &= \Pr\left\{|g_{m,j}|^2 > \frac{P_{m,j}^{\mathrm{N}}\left|h_{m,j}^{\mathrm{d}}\right|^2}{zp_m} - \frac{\sigma^2}{p_m}\right\}\\ &= \int_{\frac{P_{m,j}^{\mathrm{N}}\left|h_{m,j}^{\mathrm{d}}\right|^2}{zp_m}-\frac{\sigma^2}{p_m}}^{\infty}\frac{1}{\lambda_{m,j}}e^{-\frac{x}{\lambda_{m,j}}}\mathrm{d}x\\ &= \begin{cases} e^{-\frac{P_{m,j}^{\mathrm{N}}\left|h_{m,j}^{\mathrm{d}}\right|^2}{zp_m\lambda_{m,j}}+\frac{\sigma^2}{p_m\lambda_{m,j}}}, & 0 < z < \dfrac{P_{m,j}^{\mathrm{N}}\left|h_{m,j}^{\mathrm{d}}\right|^2}{\sigma^2},\\ 1, & z \geq \dfrac{P_{m,j}^{\mathrm{N}}\left|h_{m,j}^{\mathrm{d}}\right|^2}{\sigma^2}. \end{cases}\end{aligned} \tag{46}$$

Substituting (46) in (45), we have

$$\begin{aligned}\mathbb{R}_{m,j,k} =& \frac{1}{2}\int_0^{\frac{P_{m,j}^{\mathrm{N}}\left|h_{m,j}^{\mathrm{d}}\right|^2}{\sigma^2}}\frac{1-e^{-\frac{P_{m,j}^{\mathrm{N}}\left|h_{m,j}^{\mathrm{d}}\right|^2}{zp_m\lambda_{m,j}}+\frac{\sigma^2}{p_m\lambda_{m,j}}}}{1+z}\mathrm{d}z\\ =& \frac{1}{2}\log_2\left(1+\frac{P_{m,j}^{\mathrm{N}}\left|h_{m,j}^{\mathrm{d}}\right|^2}{\sigma^2}\right)\\ &-\frac{1}{2\ln 2}e^{\frac{\sigma^2}{p_m\lambda_{m,j}}}\underbrace{\int_0^{\frac{P_{m,j}^{\mathrm{N}}\left|h_{m,j}^{\mathrm{d}}\right|^2}{\sigma^2}}\frac{e^{-\frac{P_{m,j}^{\mathrm{N}}\left|h_{m,j}^{\mathrm{d}}\right|^2}{zp_m\lambda_{m,j}}}}{1+z}\mathrm{d}z}_{Q_1}, \end{aligned} \tag{47}$$

where

$$\begin{aligned}Q_1 &\overset{\theta=\frac{1}{z}}{=\!=} \int_{\frac{\sigma^2}{P_{m,j}^{\mathrm{N}}\left|h_{m,j}^{\mathrm{d}}\right|^2}}^{\infty}\frac{e^{-\frac{P_{m,j}^{\mathrm{N}}\left|h_{m,j}^{\mathrm{d}}\right|^2}{p_m\lambda_{m,j}}\theta}}{1+\frac{1}{\theta}}\left(-\frac{1}{\theta^2}\right)\mathrm{d}\theta\\ &= \int_{\frac{\sigma^2}{P_{m,j}^{\mathrm{N}}\left|h_{m,j}^{\mathrm{d}}\right|^2}}^{\infty}\frac{e^{-\frac{P_{m,j}^{\mathrm{N}}\left|h_{m,j}^{\mathrm{d}}\right|^2}{p_m\lambda_{m,j}}\theta}}{\theta+1}-\frac{e^{-\frac{P_{m,j}^{\mathrm{N}}\left|h_{m,j}^{\mathrm{d}}\right|^2}{p_m\lambda_{m,j}}\theta}}{\theta}\mathrm{d}\theta. \end{aligned} \tag{48}$$

According to equations (3.352.4) and (3.351.6) in [45], we can obtain (25) by substituting (48) in (47).

REFERENCES

[1] X. You, C. Zhang, X. Tan, S. Jin, and H. Wu, "AI for 5G: Research directions and paradigms," *Sci. China Inf. Sci.*, vol. 62, no. 2, p. 21301, 2019.

[2] C.-X. Wang *et al.*, "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 122–130, Feb. 2014.

[3] L. Li *et al.*, "mmWave communications for 5G: Implementation challenges and advances," *Sci. China Inf. Sci.*, vol. 61, no. 2, 2018, Art. no. 021301.

[4] Z. Ma, Z. Zhang, Z. Ding, P. Fan, and H. Li, "Key techniques for 5G wireless communications: Network architecture, physical layer, and MAC layer perspectives," *Sci. China Inf. Sci.*, vol. 58, no. 4, pp. 1–20, Apr. 2015.

[5] Z. Wang, H. Li, and Z. Xu, "Real-world traffic analysis and joint caching and scheduling for in-RAN caching networks," *Sci. China Inf. Sci.*, vol. 60, no. 6, 2017, Art. no. 062302.

[6] F. Zhou, Y. Wu, R. Q. Hu, Y. Wang, and K. K. Wong, "Energy-efficient NOMA enabled heterogeneous cloud radio access networks," *IEEE Netw.*, vol. 32, no. 2, pp. 152–160, Mar. 2018.

[7] Q. Cui, Z. Cui, W. Zheng, R. Jäntti, and W. Xie, "Energy-aware deployment of dense heterogeneous cellular networks with QoS constraints," *Sci. China Inf. Sci.*, vol. 60, no. 4, 2017, Art. no. 042303.

[8] Q. Li, R. Q. Hu, G. Wu, and Y. Qian, "On the optimal mobile association in heterogeneous wireless relay networks," in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 1359–1367.

[9] H. Zhang, S. Huang, C. Jiang, K. Long, V. C. M. Leung, and H. V. Poor, "Energy efficient user association and power allocation in millimeter-wave-based ultra dense networks with energy harvesting base stations," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1936–1947, Sep. 2017.

[10] D. Liu *et al.*, "User association in 5G networks: A survey and an outlook," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1018–1044, 2nd Quart. 2016.

[11] B. Ma, H. Zhang, and Z. Zhang, "Joint power allocation and mode selection for D2D communications with imperfect CSI," *China Commun.*, vol. 12, no. 7, pp. 73–81, Jul. 2015.

[12] R. Zhang, Y. Li, C.-X. Wang, Y. Ruan, and H. Zhang, "Energy efficient power allocation for underlaying mobile D2D communications with peak/average interference constraints," *Sci. China Inf. Sci.*, vol. 61, no. 8, 2018, Art. no. 089301.

[13] Y. Yang, G. Song, W. Zhang, X. Ge, and C. Wang, "Neighbor-aware multiple access protocol for 5G mMTC applications," *China Commun.*, vol. 13, no. 2, pp. 80–88, 2016.

[14] D. Feng *et al.*, "Mode switching for energy-efficient device-to-device communications in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6993–7003, Dec. 2015.

[15] X. Lin, J. G. Andrews, and A. Ghosh, "Spectrum sharing for device-to-device communication in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 12, pp. 6727–6740, Dec. 2014.

[16] L. Zhang, M. Xiao, G. Wu, and S. Li, "Efficient scheduling and power allocation for D2D-assisted wireless caching networks," *IEEE J. Sel. Areas Commun.*, vol. 64, no. 6, pp. 2438–2452, Jun. 2016.

[17] S. Xu, H. Zhang, J. Tian, and P. T. Mathiopoulos, "Pilot reuse and power control of D2D underlaying massive MIMO systems for energy efficiency optimization," *Sci. China Inf. Sci.*, vol. 60, no. 10, 2017, Art. no. 100303.

[18] H. Nishiyama, M. Ito, and N. Kato, "Relay-by-smartphone: Realizing multihop device-to-device communications," *IEEE Commun. Mag.*, vol. 52, no. 4, pp. 56–65, Apr. 2014.

[19] A. A. Abdellatif, A. Mohamed, and C.-F. Chiasserini, "Concurrent association in heterogeneous networks with underlay D2D communication," in *Proc. 13th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jun. 2017, pp. 56–61.

[20] S. Xiao, X. Zhou, D. Feng, Y. Yuan-Wu, G. Y. Li, and W. Guo, "Energy-efficient mobile association in heterogeneous networks with device-to-device communications," *IEEE Trans. Wireless Commun.*, vol. 15, no. 8, pp. 5260–5271, Aug. 2016.

[21] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *Proc. IEEE 77th Veh. Technol. Conf. (VTC)*, Jun. 2013, pp. 1–5.

[22] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181–2195, Oct. 2017.

[23] Y. Cai, Z. Qin, F. Cui, G. Y. Li, and J. A. McCann, "Modulation and multiple access for 5G networks," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 629–646, 1st Quart., 2018.

[24] X. Liu, X. Wang, and Y. Liu, "Power allocation and performance analysis of the collaborative NOMA assisted relaying systems in 5G," *China Commun.*, vol. 14, no. 1, pp. 50–60, Jan. 2017.

[25] 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.

[26] Z. Qin, X. Yue, Y. Liu, Z. Ding, and A. Nallanathan, "User association and resource allocation in unified NOMA enabled heterogeneous ultra dense networks," *IEEE Commun. Mag.*, vol. 56, no. 6, pp. 86–92, Jun. 2018.

[27] Y. Liu, Z. Qin, M. Elkashlan, A. Nallanathan, and J. A. McCann, "Non-orthogonal multiple access in large-scale heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2667–2680, Dec. 2017.

[28] J. Zhao, Y. Liu, K. K. Chai, Y. Chen, and M. Elkashlan, "Joint subchannel and power allocation for NOMA enhanced D2D communications," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 5081–5094, Nov. 2017.

[29] Z. Ding, M. Peng, and H. V. Poor, "Cooperative non-orthogonal multiple access in 5G systems," *IEEE Commun. Lett.*, vol. 19, no. 8, pp. 1462–1465, Aug. 2015.

[30] M. Xu, F. Ji, M. Wen, and W. Duan, "Novel receiver design for the cooperative relaying system with non-orthogonal multiple access," *IEEE Commun. Lett.*, vol. 20, no. 8, pp. 1679–1682, Aug. 2016.

[31] Y. Liu, Z. Ding, M. Elkashlan, and H. V. Poor, "Cooperative non-orthogonal multiple access with simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 938–953, Apr. 2016.

[32] Z. Zhang, Z. Ma, M. Xiao, Z. Ding, and P. Fan, "Full-duplex device-to-device-aided cooperative nonorthogonal multiple access," *IEEE Trans. Veh. Technol.*, vol. 66, no. 5, pp. 4467–4471, May 2017.

[33] L. Zhang, J. Liu, M. Xiao, G. Wu, Y.-C. Liang, and S. Li, "Performance analysis and optimization in downlink NOMA systems with cooperative full-duplex relaying," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2398–2412, Oct. 2017.

[34] T. Do *et al.*, "Improving the performance of cell-edge users in NOMA systems using cooperative relaying," *IEEE Trans. Commun.*, vol. 66, no. 5, pp. 1883–1901, May 2018.

[35] V. Danilov, "Existence of stable matchings in some three-sided systems," *Math. Social Sci.*, vol. 46, pp. 145–148, Oct. 2003.

[36] E. Boros, V. Gurvich, S. Jaslar, and D. Krasner, "Stable matchings in three-sided systems with cyclic preferences," *Discrete Math.*, vol. 289, nos. 1–3, pp. 1–10, 2004.

[37] B. Di *et al.*, "Joint user pairing, subchannel, and power allocation in full-duplex multi-user OFDMA networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8260–8272, Dec. 2016.

[38] L. Liang, S. Xie, G. Y. Li, Z. Ding, and X. Yu, "Graph-based resource sharing in vehicular communication," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4579–4592, Jul. 2018.

[39] J. Liu, S. Xiao, X. Zhou, G. Y. Li, G. Wu, and S. Li, "Optimal mobile association and power allocation in device-to-device-enable heterogeneous networks with non-orthogonal multiple access protocol," in *Proc. IEEE Int. Conf. Commun.*, May 2018, pp. 1–6.

[40] B. Di, L. Song, and Y. Li, "Sub-channel assignment, power allocation, and user scheduling for non-orthogonal multiple access networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7686–7698, Nov. 2016.

[41] S. Zhang, B. Di, L. Song, and Y. Li, "Sub-channel and power allocation for non-orthogonal multiple access relay networks with amplify-and-forward protocol," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2249–2261, Apr. 2017.

[42] M. Peng, Y. Li, Z. Zhao, and C. Wang, "System architecture and key technologies for 5G heterogeneous cloud radio access networks," *IEEE Netw.*, vol. 29, no. 2, pp. 6–14, Mar. 2015.

[43] C. Wang, E. K. S. Au, R. D. Murch, W. H. Mow, R. S. Cheng, and V. Lau, "On the performance of the MIMO zero-forcing receiver in the presence of channel estimation error," *IEEE Trans. Wireless Commun.*, vol. 6, no. 3, pp. 805–810, Mar. 2007.

[44] C.-C. Huang, "Two's company, three's a crowd: Stable family and threesome roommates problems," in *Algorithms—ESA*. Berlin, Germany: Springer, 2007, pp. 558–569.

[45] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*. 7th ed. New York, NY, USA: Academic, 2007.

JIAQI LIU received the B.Eng. degree in electronic and information engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2013, where he is currently pursuing the Ph.D. degree with the National Key Laboratory of Science and Technology on Communications. He was a Visiting Student with the Division of Electrical and Computer Engineering, Louisiana State University, Baton Rouge, LA, USA, from 2017 to 2018. His research interests include multiple-access techniques and multi-carrier waveforms toward 5G, heterogeneous networks, and D2D communications.

GANG WU received the B.Eng. and M.Eng. degrees from the Chongqing University of Post and Telecommunications, Chongqing, China, in 1996 and 1999, respectively, and the Ph.D. degree from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2004, where he is currently a Professor with the National Key Laboratory of Science and Technology on Communications. In 2004, he joined UESTC. He was a Research Fellow of the Positioning and Wireless Technology Centre, Nanyang Technological University, Singapore, from 2005 to 2007. He was a Visiting Professor with the Georgia Institute of Technology, Atlanta, GA, USA, from 2009 to 2010. His research interest includes PHY/MAC techniques for 5G. He was a co-recipient of the IEEE GLOBECOM 2012 Best Paper Award. He is currently an Associate Editor for *Science China Information Sciences*.

SA XIAO (M'18) received the B.S.E., M.S.E., and Ph.D. degrees from the University of Electronic Science and Technology of China, Chengdu, China, in 2009, 2012, and 2017, respectively, where he is currently a Postdoctoral Researcher with the Center for Intelligent Communications and Networking (CINC). In 2015, he was a Visiting Student with the Department of Electrical and Computer Engineering, Southern Illinois University, Carbondale, IL, USA. He was a Visiting Student with the Division of Electrical and Computer Engineering, Louisiana State University, Baton Rouge, LA, USA, from 2015 to 2017. His research interests include device-to-device communications, full-duplex communications, heterogeneous networks, and the Internet of Things communications.

XIANGWEI ZHOU received the B.S. degree in communication engineering from the Nanjing University of Science and Technology, Nanjing, China, in 2005, the M.S. degree in information and communication engineering from Zhejiang University, Hangzhou, China, in 2007, and the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2011. Since 2015, he has been an Assistant Professor with the Division of Electrical and Computer Engineering, Louisiana State University, Baton Rouge, LA, USA. Prior to that, he was an Assistant Professor with the Department of Electrical and Computer Engineering, Southern Illinois University, Carbondale, IL, USA, from 2013 to 2015, and a Senior Systems Engineer with Marvell Semiconductor, Santa Clara, CA, USA, from 2011 to 2013. His research interests include wireless communications, statistical signal processing, and cross-layer optimization, with current emphasis on spectrum-efficient, energy-efficient and secure communications, coexistence of wireless systems, and machine learning for intelligent communications. He was a recipient of the Best Paper Award from the 2014 International Conference on Wireless Communications and Signal Processing. He has served as an Editor for the IEEE Transactions on Wireless Communications, from 2013 to 2018.

GEOFFREY YE LI (S'93–M'95–SM'97–F'06) received the B.S.E. and M.S.E. degrees from the Department of Wireless Engineering, Nanjing Institute of Technology, Nanjing, China, in 1983 and 1986, respectively, and the Ph.D. degree from the Department of Electrical Engineering, Auburn University, Auburn, AL, in 1994.

He was a Teaching Assistant and then a Lecturer with Southeast University, Nanjing, China, from 1986 to 1991, a Research and Teaching Assistant with Auburn University, from 1991 to 1994, and a Postdoctoral Research Associate with the University of Maryland at College Park, MD, USA, from 1994 to 1996. He was with AT&T Labs-Research, Red Bank, NJ, as a Senior and then a Principal Technical Staff Member, from 1996 to 2000. Since 2000, he has been with the School of Electrical and Computer Engineering, Georgia Institute of Technology, as an Associate Professor and then a Full Professor. In his research areas, he has published over 500 journal and conference papers. In addition, he held 40 granted patents. His publications have been cited over 35 000 times, and he has been recognized as the World's Most Influential Scientific Mind, also known as a Highly-Cited Researcher, by Thomson Reuters almost every year. His general research interests include statistical signal processing and machine learning for wireless communications.

Dr. Li was an IEEE Fellow for his contributions to signal processing for wireless communications, in 2005. He has received the 2010 IEEE ComSoc Stephen O. Rice Prize Paper Award, the 2013 IEEE VTS James Evans Avant Garde Award, the 2014 IEEE VTS Jack Neubauer Memorial Award, the 2017 IEEE ComSoc Award for Advances in Communication, and the 2017 IEEE SPS Donald G. Fink Overview Paper Award. He has also received the 2015 Distinguished Faculty Achievement Award from the School of Electrical and Computer Engineering, Georgia Tech. He has organized and chaired many international conferences including the Technical Program Vice-Chair of the IEEE ICC 2003, the Technical Program Co-Chair of the IEEE SPAWC 2011, the General Chair of the IEEE GlobalSIP 2014, the Technical Program Co-Chair of the IEEE VTC 2016 (Spring), and the General Co-Chair of the IEEE VTC 2019 (Fall). He has been involved in editorial activities for over 20 technical journals for the IEEE, including the Founding Editor-in-Chief of the IEEE 5G Tech Focus.

SHENGJIE GUO received the B.Eng. degree from the University of Electronic Science and Technology of China, Chengdu, China, and the Ph.D. degree with Louisiana State University, Baton Rouge, LA, USA, in 2012 and 2018, respectively. His research interests include spectrum- and energy-efficient communications, simultaneous wireless information and power transfer, device-to-device communications, and vehicular communications.

SHAOQIAN LI (M'02–SM'12–F'16) received the B.E. degree in communication technology from the Northwest Institute of Telecommunication Engineering (currently, Xidian University), Xi'an, China, in 1981, and the M.E. degree in information and communication systems from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 1984.

In 1984, he joined the UESTC, as an Academic Member, where he has been a Professor of information and communication systems, since 1997, and the Ph.D. Supervisor, since 2000. He is currently the Director of the National Key Laboratory of Science and Technology on Communications, UESTC. He holds over 60 granted and filed patents. In his research areas, he has published over 100 journal papers, 100 conference papers, and two edited books. His general interests include the areas of wireless and mobile communications, anti-jamming technologies, and signal processing for communications subjects. His current research topics focus on multiple-antenna signal processing technologies for mobile communications, cognitive radios, and coding and modulation for next-generation mobile broadband communications systems.

Prof. Li has been a member of the Communication Expert Group, National 863 Plan, since 1998, and the FuTURE Project, since 2005. He is currently a member of the Board of Communications and Information Systems, the Academic Degrees Committee, the State Council of China, and the Expert Group of Key Special-Project on Next-Generation Broadband Wireless Mobile Communications of China (approved by the State Council, since 2007). He has served in various IEEE conferences as a Technical Program Committee (TPC) Member. He was the TPC Co-Chair of the IEEE International Conference on Communications, Circuits, and Systems, in 2005, 2006, and 2008. He is also a Member of the Editorial Board of the Chinese *Science Bulletin* and the *Chinese journal of Radio Science*.