Non-Orthogonal Multiple Access for High-Reliable and Low-Latency V2X Communications in 5G Systems

Boya Di, *Student Member, IEEE*, Lingyang Song, *Senior Member, IEEE*, Yonghui Li, *Senior Member, IEEE*, and Geoffrey Ye Li, *Fellow, IEEE*

***Abstract*— In this paper, we consider a dense vehicular communication network where each vehicle broadcasts its safety information to its neighborhood in each transmission period. Such applications require low latency and high reliability, and thus, we exploit non-orthogonal multiple access to reduce the access latency and to improve the packet reception probability. In the proposed two-fold scheme, the BS performs semi-persistent scheduling and allocates time-frequency resources in a non-orthogonal manner while the vehicles autonomously perform distributed power control with iterative signaling control. We formulate the centralized scheduling and resource allocation problem as equivalent to a multi-dimensional stable roommate matching problem, in which the users and time/frequency resources are considered as disjoint sets of objects to be matched with each other. We then develop a novel rotation matching algorithm, which converges to an L-rotation stable matching after a limited number of iterations. Simulation results show that the proposed scheme outperforms the traditional orthogonal multiple access scheme in terms of the access latency and reliability.**

***Index Terms*— Non-orthogonal multiple access, V2X broadcasting, resource allocation, scheduling problem, matching theory.**

I. Introduction

With the rapid development of intelligent transportation systems (ITS), a growing number of vehicular applications have emerged to provide a safer and more efficient driving experience for our daily life. Among various applications, safety critical services play a vital role in the blueprint of the future ITS, supported by the vehicle-to-everything (V2X) communications [1], including the vehicle-to-vehicle (V2V), vehicle-to-pedestrian (V2P), and vehicle-to-network/infrastructure (V2N/I). To achieve low-latency and high-reliability (LLHR) for the V2X services, the widely deployed Long Term Evolution (LTE) networks have been considered as a very promising solution to achieve large cell coverage, controllable latency, and high data rates even in a high-mobility scenario [2].

The LTE-based V2X services combine the ad hoc and cellular network architecture by exploiting the device-to-device (D2D) communications [3]. Hence, the end-to-end latency can be reduced compared to the cellular uplink/downlink (UL/DL) mode, and the quality of services can be guaranteed in contrast to the 802.11p [4]. However, unlike the traditional D2D communications, V2X applications require stringently low latency, which poses new challenges to the LTE-based vehicular network, especially in a dense environment causing severe data congestion [5]. One of the main reasons is that the existing LTE networks are based on the orthogonal multiple access (OMA), and the limited spectrum resource have not been fully and efficiently utilized [6], leading to the severe data congestion and low access efficiency in a dense network. Therefore, a more spectrally efficient radio access technology is required for the V2X services.

To handle the challenges of access collisions and massive connectivity, non-orthogonal multiple access (NOMA) schemes have been introduced as a potential solution for 5G wireless networks, which allow users to access the channel non-orthogonally by either power-domain [7] or code-domain multiplexing [8]. Multiple users with different types of traffic requests can transmit concurrently on the same channel to improve spectrum efficiency [9], [10] and alleviate the congestion of data traffic by supporting massive connectivity, thereby potentially reducing the data retransmissions as well as the access latency. To make the NOMA scheme more practical, various multi-user detection (MUD) techniques, such as successive interference cancellation (SIC) [11], have been applied at the end-user receivers to cope with co-channel interference caused by spectrum sharing among various users. Though the SIC receiver in NOMA system is slightly more complex than that in the conventional OMA system and requires a bit more processing time, with fast processing speed

Manuscript received January 28, 2017; revised May 15, 2017; accepted May 22, 2017. Date of publication July 11, 2017; date of current version September 15, 2017. This work was supported in part by the National 973 Project under Grant 2013CB336700, and in part by the National Nature Science Foundation of China under Grant U1301255 and Grant 61625101. The work of Y. Li was supported by the Australian Research Council under Grant DP150104019. The work of G. Y. Li was supported by the National Science Foundation under Grant 1405116 and Grant 1443894. *(Corresponding author: Lingyang Song.)*

B. Di and L. Song are with the State Key Laboratory of Advanced Optical Communication Systems and Networks, School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China (e-mail: diboya@pku.edu.cn; lingyang.song@pku.edu.cn).

Y. Li is with the School of Electrical and Information Engineering, The University of Sydney, Camperdown, NSW 2006, Australia (e-mail: yonghui.li@sydney.edu.au).

G. Y. Li is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: yl46@gatech.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSAC.2017.2726018

and powerful computation capability of current IC chips, the processing delay caused by SIC can be ignored compared to the access latency, caused by frequent retransmissions in a congested network. Thus in this paper, we will only focus on the access latency.

Capable of achieving high overloading transmission over limited resources, NOMA provides a new dimension for V2X services to alleviate the traffic congestion, thereby reducing the latency. In this paper, we consider the V2X broadcast scenario [12] where every vehicle needs to broadcast its safety information to the neighborhood[1] in each transmission period. Each period consists of multiple time slots and the transmitter-receiver (Tx-Rx) selection[2] needs to be determined for each slot such that all vehicles can update their safety information in at least one slot during this period. Sub-channel allocation in each time slot is performed to manage the co-channel interference caused by the non-orthogonal nature. Moreover, to perform joint decoding in an non-orthogonal manner, a different method of the real-time power control is applied by the users, in contrast to the traditional OMA-based case. New challenges are thus posed in the design of scheduling and resource allocation schemes. Meanwhile, the LLHR requirement and dense topology of vehicular networks need to be considered.

In this paper, we propose a NOMA-based mixed centralized/distributed (NOMA-MCD) scheme for V2X broadcasting. The centralized semi-persistent scheduling (SPS) [13] is performed by the BS where the Tx-Rx is selected and the time and frequency resources are allocated every few transmission periods. For the BS, Tx-Rx selection and resource allocation problem can be formulated as a non-linear integer programming problem to maximize the packet reception probability of the network. Afterwards, the autonomous distributed power control is performed by each user, and a control signaling scheme is designed, where the Tx users utilize the control signals to iteratively adjust their transmit power based on the feedback of the Rx users.

To tackle such a combinatorial optimization problem on Tx-Rx selection and resource allocation, we then decouple it into two multi-dimensional stable roommate (MD-SR) matching problems [14], in which the vehicles and time slots/sub-channels are considered as two disjoint sets of "students" and "rooms" such that multiple "students" can occupy the same "room". It is not trivial to find an efficient algorithm to solve such a NP-hard MD-SR problem since most existing matching algorithms [14]–[16] are designed for the two-dimensional SR problem in which at most two "students" can share one "room". Therefore, we develop a novel rotation matching algorithm for the MD-SR problem which converges to a stable matching.

Only limited works have discussed how to improve the performance of the safety critical applications from a NOMA-based perspective. The most related ones [17]–[19] focus on either the LTE-based D2D broadcasting systems or the cellular V2V unicast systems with the assumptions of fixed Tx-Rx selection or transmit power. In [17], a distributed scheme for the coordination of the D2D broadcast transmission has been presented in which the traffic collision has been avoided through the orthogonal use of the spectrum resources in the neighborhood of each Tx user. In [18], the resource management for D2D unicast safety-critical vehicular communications has been discussed and the sum rate of cellular mobile users has been guaranteed under the constraints of satisfying the vehicles' requirements on latency and reliability. In [19], a centralized sub-channel allocation scheme involving several D2D broadcast groups underlaying cellular network has been proposed by utilizing a greedy algorithm.

In this paper, we will investigate the LLHR NOMA-based cellular V2X broadcast communications. The main contributions of our work are listed below:

- We propose a novel NOMA-MCD scheme for the V2X broadcasting system combining the centralized SPS at the BS and the distributed power control of the vehicles. The access latency can be reduced and the reliability can be improved by NOMA in a dense network.
- We consider the centralized Tx-Rx selection and resource allocation problem as a packet reception probability maximization problem, and decouple it into two MD-SR matching problems, which can be solved by a novel rotation matching algorithm.
- An iterative signaling control scheme is proposed to solve the distributed power control problem of the Tx users, in which the Tx and Rx users exchange control messages for joint decoding and transmit power adjustment.

The rest of this paper is organized as follows. In Section II, we describe the system model of the NOMA-based cellular V2X broadcasting system and propose the NOMA-MCD scheme. In Section III, we formulate the centralized Tx-Rx selection and time-frequency resource allocation of the BS as a packet reception probability maximization problem and then solve it by utilizing the matching theory. Properties of the proposed rotation matching algorithms are analyzed. In Section IV, an iterative signaling control scheme is designed for the distributed power control problem of the Tx users. Simulation results are presented in Section V, and finally, we conclude the paper in Section VI.

II. System Model

In this section, we present the system model of the cellular V2X broadcasting system and propose the NOMA-based cellular V2X broadcasting scheme to reduce access collisions.

A. Scenario Description

Consider an urban V2X broadcast system as shown in Fig. 1. In every transmission period consisting of multiple time slots, each of N vehicles broadcasts safety-critical information[3] to its neighborhood, including other vehicles, pedestrians, and the road side units (RSU), in at least one time slot. If a vehicle broadcasts in one time slot, we say

[1] V2X broadcasting refers to V2V/I broadcasting for safety information.

[2] A vehicle is called a Tx user if it broadcasts in a time slot and a Rx user otherwise.

[3] For example, the cooperative awareness messages record periodic time-triggered position information with transmission interval of 100ms.

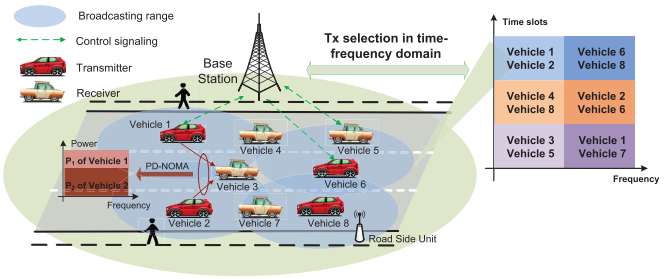


Fig. 1. System model of V2X transmission.

that it acts as a Tx user in this time slot, otherwise, it acts as a Rx user in this time slot.[4] For simplicity, we assume that all the pedestrians and RSUs always act as Rx users in each time slot. Direct data transmission between neighboring users is achieved by the D2D communications, i.e., users in proximity directly communicate with each other bypassing the BS. The available bandwidth is divided into K sub-channels for transmission.

Note that if all vehicles are scheduled to broadcast in the same time slots, then no one can successfully receive from its neighborhood due to the half-duplex nature, which also brings severe interference to those pedestrians due to limited frequency resources. To solve the above two issues, the vehicles need to be scheduled at the beginning of each transmission period such that the fewest vehicles broadcast simultaneously while all vehicles are guaranteed to broadcast in at least one time slot. Given the time scheduling, frequency resources also need to be allocated to the Tx users to improve the spectral efficiency.

To meet the above requirements, the centralized control of the BS is applied to manage the frequency resource allocation and Tx-Rx selection in each time slot. In other words, at the beginning of each transmission period, the BS decides: 1) which role each vehicle plays (i.e., a Tx user or a Rx user) in each time slot of this transmission period; 2) the frequency resources each Tx user occupies for broadcasting given the time scheduling scheme.

Due to the dense topology of the network, when more than one Tx user (e.g., Tx user 1 and Tx user 2 in Fig. 1) are assigned the same time-frequency resources, severe collision may occur for those Rx users (e.g., Rx user 1) locating in the overlapping region of two adjacent Tx users' communication ranges. For the traditional OMA case, it is likely that a conflicting Rx user fails to decode the signals of multiple Tx users due to severe interference, leading to transmission delay caused by extra retransmissions. To reduce the probability of collision as well as the transmission delay, non-orthogonal allocation of radio resources is then considered such that one sub-channel can be occupied by multiple Tx users simultaneously [21]. The data rates of Tx-Rx links can be improved compared to the OMA case and each conflicting Rx user (such as Rx user 1 in Fig. 1 mentioned above) is more likely to decode the received superposed signals by utilizing the SIC technique. In this way, data collision can be reduced and the number of users supported by the system is also improved.

B. NOMA-Based Collision Avoidance

1) Channel Model: To better depict the Tx-Rx selection and sub-channel allocation in each time slot i, we introduce a binary variable $\gamma_j^{(i)}$ in which $\gamma_j^{(i)} = 1$ indicates that time slot i is assigned by the BS to user j, i.e., user j acts as a Tx user in time slot i. $\gamma_j^{(i)} = 0$ reflects that its acts a Rx user in time slot i. For any Tx user j (i.e., $\gamma_j^{(i)} = 1$), we define a binary variable $\mu_{j,k}^{(i)}$ in which $\mu_{j,k}^{(i)} = 1$ implies that subchannel k is allocated by the BS to Tx user j, otherwise, $\mu_{j,k}^{(i)} = 0$. When $\gamma_j^{(i)} = 0$, the value of $\mu_{j,k}^{(i)}$ makes no sense since $\gamma_j^{(i)} \cdot \mu_{j,k}^{(i)}$ always equals zero.

Based on the NOMA scheme, the received signal of Rx user m over subchannel k in time slot i (the ith time slot of a transmission period) can then be presented by

$$y_{m,k}^{(i)} = \sum_{j \in \mathcal{N}_m^{(i)}} \gamma_j^{(i)} \cdot \mu_{j,k}^{(i)} \sqrt{p_{j,k}^{(i)}} H_{j,m,k}^{(i)} s_j^{(i)} + n_m^{(i)}, \tag{1}$$

in which $\mathcal{N}_m^{(i)} = \left\{1 \le j \le N | d_{j,m}^{(i)} \le r\right\}$ denotes the index set of users within Rx user m's communication range of interest, i.e., a disk with the radius of r, $p_{j,k}^{(i)}$ is the transmit power of Tx user j, $H_{j,m,k}^{(i)}$ denotes the channel coefficient of subchannel k between Tx user j and Rx user m in time slot i, $s_j^{(i)}$ represents the transmitted symbol of Tx user j in time slot i, $n_m^{(i)} \sim \mathcal{CN}\left(0, {\sigma_n}^2\right)$ is the additive white Gaussian noise (AWGN) for Rx user m, and ${\sigma_n}^2$ is the noise variance.

In equation (1), the channel coefficient $H_{j,m,k}^{(i)}$ can be defined as $H_{j,m}^{(i)} = G \cdot h_{j,m,k}^{(i)} \cdot \beta_{j,m}^{(i)} \cdot \left(d_{j,m}^{(i)}\right)^{-\alpha}$, in which G is the constant power gain factor introduced by amplifier and antenna, $h_{j,m,k}^{(i)} \sim \mathcal{CN}(0,1)$ is a complex Gaussian variable representing Rayleigh fading, $\beta_{j,m}^{(i)}$ follows log-normal distribution representing shadowing fading, $d_{j,m}^{(i)}$ is the distance between users j and m in time slot i, and α represents the pathloss exponent. At the beginning of each transmission period, the BS can predict the distance between two users at any time slot of this period based on the periodic velocity and location information updated by the users.[5] Therefore, $d_{j,m}^{(i)}$ can be expressed as

$$d_{j,m}^{(i)} = i \cdot \frac{\left(\boldsymbol{v}_{j,m}^{(0)}\right)^T \left(\boldsymbol{x}_m^{(0)} - \boldsymbol{x}_j^{(0)}\right)}{\left\|\boldsymbol{x}_m^{(0)} - \boldsymbol{x}_j^{(0)}\right\|_2}, \quad i \ge 1, \tag{2}$$

in which $\boldsymbol{v}_{j,m}^{(0)}$ is the relative velocity vector between users j and m at the beginning of a transmission period, i.e., in time slot 0, and $\boldsymbol{x}_j^{(0)}, \boldsymbol{x}_m^{(0)} \in \mathbb{R}^2$ denote the positions of users j and m in time slot 0.

[4]Considering the implementation issues, we assume that both users and the BS perform in the half-duplex mode and the full-duplex enabled V2I case [20] is out of scope in this paper.

[5]Note that the prediction error is tolerable due to the density of the vehicular network and the short duration of each transmission period (no longer than 100ms).

2) SIC Decoding: The transmit power of the links corresponding to the conflicting Rx users will be carefully adjusted according to the NOMA principle. Each conflicting Rx user m decodes the received signals in a decreasing order of channel gains [11]. For example, consider two Tx users j and j' occupying subchannel k in time slot i transmitting to Rx user m satisfying $p_{j,k}^{(i)}\left|H_{j,m,k}^{(i)}\right|^2 < p_{j',k}^{(i)}\left|H_{j',m,k}^{(i)}\right|^2$. The conflicting Rx user m first detects the data signals of Tx user j'. After subtracting the decoded signals, it then decodes the signal from Tx user j while regarding the signals of those Tx users with lower channel gains than that of Tx user j as noise. For any conflicting Rx user m, the achievable rate obtained from Tx user j over subchannel k in time slot i can then be formally presented [7], [21]:

$$R_{j,m,k}^{(i)} = \log_2\left(1 + \frac{p_j^{(i)}\rho_{j,m,k}^{(i)}}{1 + \sum\limits_{j' \in \mathcal{S}_{j,m,k}^{(i)}} p_{j'}^{(i)}\rho_{j',m,k}^{(i)}}\right), \tag{3}$$

where $\rho_{j,m,k}^{(i)} = p_{j,k}^{(i)}\left|H_{j,m,k}^{(i)}\right|^2 / \left(n_m^{(i)}\right)^2$ represents the SNR of the Tx user j – Rx user m link, and $\mathcal{S}_{j,m,k}^{(i)}$ is the set of active Tx users with lower channel gains than Tx user j such that they cause interference to the Tx j – Rx m link when Rx user m decodes the signal of Tx user j.

3) Criteria for Successful Decoding: We assume that the rate threshold[6] for a Rx user to successfully decode the signal of a Tx user is $\bar{R}_{th}$. Therefore, for any Rx user m that receives from multiple Tx users over the sub-channel k in time slot i, two criteria that Rx user m should satisfy to successfully detect the signal $\boldsymbol{x}_j$ received from Tx user j are:

- The signals of other Tx users with channel gains higher than Tx user j are successfully decoded first;
- The rate threshold of Tx user j is satisfied, i.e., $R_{j,m,k}^{(i)} \geq \bar{R}_{th}$.

The above criteria can be mathematically expressed as

$$\prod_{\substack{j' \in \mathcal{N}_m^{(i)} \setminus \mathcal{S}_{j,m,k}^{(i)} \\ \gamma_{j'}^{(i)} \cdot \mu_{j',k}^{(i)} = 1}} \left(R_{j',m,k}^{(i)} - \bar{R}_{th}\right)^+ = 1, \tag{4}$$

in which $(\cdot)^+$ is a signed function. Conditions $j' \in \mathcal{N}_m^{(i)} \setminus \mathcal{S}_{j,m,k}^{(i)}$ and $\gamma_{j'}^{(i)} \cdot \mu_{j',k}^{(i)} = 1$ imply that the channel gain between active Tx user j' and Rx user m is no lower than that between Tx user j and Rx user m over subchannel k in time slot i.

C. NOMA-Based Mixed Centralized/Distributed Scheme

Based on the above non-orthogonal manner of reducing the collision, we aim to design a scheme in which each user can successfully broadcast the safety information to as many neighboring users as possible while satisfying the latency requirement. According to equations (3) and (4), system performance relies on Tx-Rx selection and frequency/power-domain resource allocation. Therefore, the following key problems need to be solved in the NOMA-based scheme design:

- How to determine the role of each user (i.e., Tx/Rx user) in each time slot of a transmission period;
- How to allocate the subchannels to the set of Tx users;
- How to perform the power control of the Tx users in each time slot.

To address these issues, we propose a two-fold NOMA-MCD scheme consisting of the following two phases: 1) centralized Tx-Rx selection and time-frequency resource allocation of the BS; 2) distributed power control of the users.

1) Centralized Tx-Rx Selection and Time-frequency Resource Allocation of the BS: In the traditional dynamic scheduling, resource allocation may cause significantly large delay since the users need to send resource request messages to the BS for every data packet. To reduce such UL latency, centralized SPS at the BS is considered in which a predefined sequence of resources are allocated to the users by the BS at the beginning of each SPS period consisting of one or more transmission periods. The resource allocation scheme keeps unchanged during each transmission period of the SPS period [13].

To take full advantage of the users' global position information obtained by the BS, the BS determines the Tx-Rx selection and subchannel allocation of each time slot at the beginning of each SPS period in the centralized NOMA-based SPS scheme. It also provides a more stable performance of latency and reliability compared to a distributed scheme in which the users compete with each other to access the channel.

2) Distributed Power Control of the Users: To benefit from the power domain multiplexing, the NOMA scheme requires prior knowledge of the users for SIC decoding such as the real-time CSI, which is hard for the BS to obtain due to the mobility of the vehicles. Distributed real-time power control of the users is then performed to achieve a better decoding effect after the centralized SPS at the BS. In each time slot, the transmit power of the Tx users is adjusted based on feedback sent from the conflicting Rx users via the control signaling.

Design of the above two phases will be presented in detail in Sections III and IV, respectively.

III. NOMA-Based Centralized Semi-Persistent Scheduling

In this section, we formulate the centralized Tx-Rx selection and time-frequency resource allocation of the BS as a packet reception probability (PRP) maximization problem and then solve it by utilizing the matching theory.

A. Problem Formulation

Note that in the vehicular network, not only the full CSI is very costly for the BS to acquire, but also CSI can become outdated easily due to the rapid variation of the small scale fading caused by the mobility of vehicles [22]. Therefore, only partial CSI containing the path loss and shadowing,[7]

[6]The successful decoding rate threshold is obtained based on the SINR threshold for successful decoding.

[7]Note that the assumption of partial CSI is widely used in vehicular networks such as [18], [22].

i.e., $H_{j,m}^{(i)} = G \cdot \beta_{j,m}^{(i)} \cdot \left(d_{j,m}^{(i)}\right)^{-\alpha}$, can be used for the centralized resource allocation. Considering the practical issues, the BS assumes that each Tx user broadcasts with a fixed transmit power in the centralized SPS.

To improve the reliability of the network, we try to alleviate the probability of collision such that fewer retransmissions are required, which in turn reduces the latency. Since the collision reduction can improve the PRP of the network, we then aim at maximizing the PRP of each transmission period, which is proportional to the total number of signals successfully decoded by all Rx users.

Whether Rx user m can successfully decode the signal of Tx user j can be judged according to (4). The number of successfully decoded signals within one transmission period can then be given by,

$$\sum_{i=1}^{T_v}\sum_{k=1}^{K}\sum_{m=1}^{N}\left(1-\gamma_m^{(i)}\right)\sum_{j\in\mathcal{N}_m^{(i)}}\gamma_j^{(i)}\mu_{j,k}^{(i)}\prod_{\substack{j'\in\mathcal{N}_m^{(i)}\setminus s_{j,m,k}^{(i)}\\ \gamma_{j'}^{(i)}\cdot\mu_{j',k}^{(i)}=1}}\times\left(R_{j',m,k}^{(i)}-\bar{R}_{th}\right)^{+}, \quad (5)$$

in which T_v is the total number of time slots in each transmission period. However, we note that successful decoding using SIC is closely related to the gap between the signal strength of multiple Tx users in practice. In other words, a Rx user can decode two Tx users with distinctive signal strengths more easily according to the SIC principle [23], which cannot be accurately reflected just by a binary function. To be more general, we adopt the logistic function[8] instead of the indicator function to describe the probability that a signal is successfully decoded [24], [25]. When the data rate exceeds the rate threshold, the value of a logistic function approaches one as the rate grows, implying a higher probability of successful decoding. The probability that the signal of Tx user j is successfully decoded by Rx user m over subchannel k in time slot i is given as:

$$U_{j,m,k}^{(i)}=\gamma_j^{(i)}\mu_{j,k}^{(i)}\left(1-\gamma_m^{(i)}\right)\prod_{\substack{j'\in\mathcal{N}_m^{(i)}\setminus s_{j,m,k}^{(i)}\\ \gamma_{j'}^{(i)}\cdot\mu_{j',k}^{(i)}=1}}\frac{\gamma_{j',k}^{(i)}}{1+e^{-\eta\left[R_{j',m,k}^{(i)}-\bar{R}_{th}\right]}}, \quad (6)$$

where η is the slope parameter of the logistic function.

Based on the above discussion, the Tx-Rx selection and time-frequency resource allocation problem of the BS in one transmission period can then be formulated as,

$$\max_{\left\{\gamma_j^{(i)},\mu_{j,k}^{(i)}\right\}}\sum_{i=1}^{T_v}\sum_{k=1}^{K}\sum_{m=1}^{N}\sum_{j\in\mathcal{N}_m^{(i)}}U_{j,m,k}^{(i)} \quad (7)$$

$$s.t.\ \gamma_j^{(i)}+\gamma_{j'}^{(i)}\le 1,$$

$$\exists i\in[1,T_v],\forall\left\{j,j'\right\}\in\left\{1\le j,j'\le N|d_{j,j'}^{(i)}<r\right\}, \quad (7a)$$

$$\gamma_j^{(i)}\sum_{k=1}^{K}\mu_{j,k}^{(i)}\le K_{\max},1\le i\le T_v,1\le j\le N, \quad (7b)$$

$$1\le\sum_{i=1}^{T_v}\gamma_j^{(i)}\le T_{\max},1\le j\le N, \quad (7c)$$

$$\left(1-\gamma_m^{(i)}\right)\sum_{j\in\mathcal{N}_m^{(i)}}\gamma_j^{(i)}\cdot\mu_{j,k}^{(i)}\le K_u,$$

$$1\le m\le N,\quad 1\le k\le K,\ 1\le i\le T_v, \quad (7d)$$

$$\gamma_j^{(i)},\mu_{j,k}^{(i)}\in\{0,1\},\quad 1\le j\le N,\ 1\le k\le K,$$

$$1\le i\le T_v. \quad (7e)$$

Constraint (7*a*) denotes that any two users within each other's communication range cannot cannot act as Tx users simultaneously in at least one time slot. If not, they will never receive each other's safety information due to the half duplex nature. For the sake of user fairness, a Tx user can occupy no more than K_{max} subchannels within one time slot and at most T_{max} time slots are allocated to it in one transmission period as expressed in constraints (7*b*) and (7*c*). Considering the decoding complexity of SIC at the receiver, we assume that each subchannel can be assigned to at most K_u Tx users with overlapping communication disks simultaneously as shown in constraint (7*d*).

B. Matching Algorithm Design

The formulated problem in (7) is non-convex and NP-hard as discussed in Appendix A. To solve this problem, we develop a novel MD rotation matching algorithm. We first start with the traditional two-dimensional stable roommate (SR) problem [26] in which $2n$ students aim at selecting roommates given n identical rooms [14].

Based on the above SR problem, we consider the set of users and time slots/subchannels as two disjoint sets of "students" and "rooms". Note that one time slot cannot be assigned to two neighboring Tx users due to the half-duplex nature as expressed in constraint (7*a*). Therefore, time scheduling is considered prior to sub-channel allocation for the users and the time-frequency resources cannot be considered as a whole. Considering the above relation between the time slots and frequency resources, we formulate the optimization problem in (7) as two MD versions of the SR problem. Different from the 2D version, one "room" can be occupied by multiple "students" and the complicated interaction between multiple "students" is difficult to be captured via the traditional matching algorithms, thereby bringing new challenge to the problem.

To address the above challenge, we propose a rotation matching algorithm for the MD user-time matching problem and then extend it to the user-subchannel matching problem.

1) Matching Problem Formulation for the Tx-Rx Selection and Time Scheduling: We formulate the Tx-Rx selection and time scheduling problem as a two-sided *matching* between the set of users and time slots. If a user is matched with one time slot, we say that it acts as a Tx user in this time slot; otherwise it acts as a Rx user in this time slot. Denote the set of users as $\mathcal{N}$ and the set of time slots in one transmission period as $\mathcal{T}$.

[8]Here we take the logistic function as an example to depict the successful decoding probability. Other approximation methods can also be used.

Mathematically, a *matching* Ψ is defined as a mapping from the set $\mathcal{N} \cup \mathcal{T} \cup \varnothing$ into itself such that $\Psi(j) \subseteq \mathcal{T}$, $j \in \mathcal{N}$ and $\Psi(i) \subseteq \mathcal{N}$, $i \in \mathcal{T}$. The number of scheduled Tx users in time slot i is denoted as $|\psi(i)|$ and the number of time slots in which user j acts as a Tx user is denoted as $|\psi(j)|$ satisfying $|\psi(j)| \leq T_{max}$.

Any two users matching to the same time slot are defined as *matching peers*, which is an expansion of "roommates" in the SR problem. A pair of users $\left(j, j'\right)$ is defined as a *forbidden pair* in time slot i if $d_{j,j'}^{(i)} \leq r$, that is, they cannot be matched with this time slot simultaneously due to constraint $(7a)$ as mentioned above.[9]

a) Preference relation: To better describe the interaction between the users, we then investigate how each user selects its matching peers. The key idea is that each user tends to choose those far away from it as its matching peers so that the overlapping part of their communication disks will be small and the number of conflicting Rx users will decrease, thereby reducing the potential collision. For convenience, we call such potential collision as cross influence brought by multiple Tx users on the conflicting Rx users. Therefore, it is natural that the cross influence on the Rx users brought by any two Tx users j and j' in time slot i can be evaluated by the overlapping area of their communication disks, as shown below:

$$I_{j,j'}^{(i)} = \begin{cases} \left(2r - d_{j,j'}^{(i)}\right)^2, & \text{if } 2r > d_{j,j'}^{(i)}, \\ \varepsilon, & \text{otherwise}, \end{cases} \tag{8}$$

in which $-0.1 < \varepsilon < 0$ is small enough. Therefore, the average cross influence brought by user j in time slot i is then given by

$$Q_j^{(i)} = \begin{cases} \dfrac{1}{|\Psi(i)|+1} \displaystyle\sum_{j' \in \Psi(i)} I_{j,j'}^{(i)}, & \text{if } |\Psi(i)| > 1, \\ \varepsilon, & \text{if } |\Psi(i)| = 1, \end{cases} \tag{9}$$

User j *prefers* time slot i to time slot i' if user j causes smaller cross influence in time slot i and no forbidden pair exists when it matches with time slot i, i.e., $Q_j^{(i)} < Q_{j'}^{(i)}$ and $d_{j,j'}^{(i)} > r, \forall j' \in \Psi(i)$. For convenience, we denote the *preference relation* as $i \succ_j i'$. The total cross influence with respect to time slot i is the sum of the cross influence of all users matched with time slot i, which is omitted here.

b) Analysis on matching formulation: Based on equations (8) and (9), the matching between users and time slots is a MD geometric SR problem with forbidden pairs, which is different from the traditional 2D or 3D SR problem due to the extension of multiple dimension and forbidden pairs. In a traditional geometric SR problem, any user sets its preference based on the distance to other users and a time slot can only be occupied by two or three users. A stable matching can be established via a greedy algorithm such that no two/three users prefer each other to their current matches. However, in such a stable matching, the users who are closest to each other always match with the same time slot [15]. This may not be suitable for the case with forbidden pairs since the set of closest matching peers always contains forbidden pairs with the smallest distance, making the stable matching an infeasible solution. If there is no matched forbidden pair in the final matching, we then say that it is a *feasible solution*.

2) Rotation Matching Algorithm for the MD-Geometric SR Problem: A two-phase matching algorithm is designed in which a feasible solution is obtained in the first phase, and then the users rotate their matching peers to further reduce their average cross influence while avoiding the forbidden pairs in the second phase.

a) Phase 1 – obtaining a feasible solution: We utilize a greedy algorithm in which each user is assigned with one time slot if the matching problem is solvable. For any unmatched user j, if it can form a forbidden pair with user $j' \in \Psi(i)$, user j is not allowed to match with time slot i. Each user j searches for the set of available matched time slots obtained by

$$\mathcal{P}_j = \left\{ i \in \mathcal{T}_{matched} | \Psi(i) \cap \mathcal{H}_j^i = \varnothing \right\}, \tag{10}$$

in which $\mathcal{T}_{matched}$ is the set of all matched time slots in the current matching and $\mathcal{H}_j^i$ is the set of users that can form forbidden pairs with user j in time slot i. When user j is not allowed to match all of the matched time slots, i.e., $\mathcal{P}_j = \varnothing$, it then selects an unmatched time slot i', i.e., $\Psi\left(i'\right) = \varnothing$. Otherwise, it selects a matched time slot i^* such that it brings the smallest cross influence $Q_j^{(i)}$, i.e.,

$$T_{(i^*)} = \arg\min_i Q_j^{(i)}. \tag{11}$$

The whole process continues until all users or time slots are matched.

b) Phase 2 – rotation matching: We introduce the concept of *rotation* to better describe the interdependency of different users in the MD-geometric SR problem.

Consider a matching Ψ obtained from the first phase with a subset of users $\mathcal{N}_s \subseteq \mathcal{N}$ satisfying $|\mathcal{N}_s| = L \geq 2$. Each member of $\mathcal{N}_s$ is denoted as $N_s(j), 1 \leq j \leq L$, and a *matching pair* refers to a user $N_s(j)$ and a time slot that it is matched with, i.e., $(N_s(j), \Psi(N_s(j)))$, $1 \leq j \leq L$. A matching sequence with the normal order consists of L matching pairs, i.e.,

$$(N_s(1), \Psi(N_s(1))), \cdots, (N_s(L), \Psi(N_s(L))). \tag{12}$$

Based on the above notions, we then present the definition of rotation sequence as below:

Definition 1: Given a matching Ψ, a subset of users $\mathcal{N}_s \subseteq \mathcal{N}$ with size L, and a parameter $1 \leq l \leq L$, each user j forms a pair with the partner of the $[\mathrm{mod}_L(j+l)]$th user in $\mathcal{N}_s$, i.e., $\Psi(N_s(\mathrm{mod}_L(j+l)))$. Given the parameter l, there exist L such pairs since there are L users in $\mathcal{N}_s$. A rotation sequence with respect to l is then defined as the combination of L such pairs, i.e.,

$$\zeta_l = (N_s(1), \Psi(N_s(l+1))), (N_s(2), \Psi(N_s(l+2))), \cdots, (N_s(L), \Psi(N_s(l))), \tag{13}$$

where $1 \leq l \leq L-1$ implying that there exist $L-1$ possible rotation sequences given matching Ψ and user subset $\mathcal{N}_s$.

[9]Note that we have modified the definition here to better fit our scenario, which differs from the traditional definition of forbidden pair in [14].

When $l = L$, ζ_L refers to the original matching sequence with the normal order as shown in (12). The rotation sequence enables the matching relation of the members in $\mathcal{N}_s$ to be reorganized in a cyclic method.

Given a rotation sequence ζ_l, a *rotation matching* refers to

$$\Psi^{rot}_{\mathcal{N}_s,\zeta_l} = \Psi \backslash \{(N_s(1), \Psi(N_s(1))), \cdots, (N_s(L), \Psi(N_s(L)))\} \cup \zeta_l. \quad (14)$$

This means that each user in $\mathcal{N}_s$ is re-matched with the partner defined in the rotation sequence ζ_l. For example, user j is originally matched with time slot $\Psi(N_s(j))$. After the rotation matching, it does not match with its original partner any more but $\Psi(N_s(\mathrm{mod}_L(j+l)))$ instead. At least one user/time slot in the rotation sequence is allowed to be unmatched, i.e., it is matched with a dummy time slot or a dummy user.

Aiming at reducing the cross influence on the unmatched users (i.e., Rx users) brought by the matched users (i.e., Tx users), we evaluate the optimality of one rotation matching as well as its validity as shown below.

Definition 2: Consider a matching Ψ and a subset of users $\mathcal{N}_s \subseteq \mathcal{N}$ with size L. For $1 \le l \le L-1$, a rotation matching $\Psi^{rot}_{\mathcal{N}_s,\zeta_l}$ is *valid* if any user $j \in \mathcal{N}_s$ does not form forbidden pairs with any matching peers in $\Psi^{rot}_{\mathcal{N}_s,\zeta_l}\left[\Psi^{rot}_{\mathcal{N}_s,\zeta_l}(j)\right] \backslash \{j\}$. A rotation matching is *optimal* if it achieves the smallest average cross influence in (9) with no forbidden pairs among the matching peers, i.e., $l^* = \arg\min_{1\le l\le L} \sum_{j\in\mathcal{N}_s}\sum_{i\in\Psi^{rot}_{\mathcal{N}_s,\zeta_l}(j)} \mathcal{Q}^{(i)}_j$.

Note that in Definition 2, we have extended the optimality of a rotation matching to include the case of $l = L$ so as to compare the original matching sequence with the rotation sequences.

Remark 2: Two special cases of the rotation matching are discussed as below.

(1) When $L = 2$, the rotation matching is degraded to a swap matching as defined in [21] and [27], and the dummy nodes are valid in the rotation matching.

(2) For users $j, j' \in \mathcal{N}_s$, if $L > 2$, $\Psi(j)$ and $\Psi(j')$ are allowed to be the same; otherwise, there exists no rotation.

Note that the concept of rotation matching describes a more general case compared to the swap matching [27], where only two users exchange their matches for utility improvement. It provides a heuristic method to cope with the interaction of the players in the SR problem. To achieve a tolerable complexity of the algorithm, we set the length of each rotation sequence $L \le L_{max}$.

Based on the defined rotation matching, we then define the stability as shown below.

Definition 3: A MD-geometric SR matching Ψ is *L-rotation stable* if all rotation matchings in Ψ are optimal, i.e., no rotation matchings with $L \le L_{max}$ can further reduce the average cross influence of Ψ.

c) Description of the whole algorithm: Given the above definitions of the rotation and stability, we then propose a Tx-Rx selection and time scheduling algorithm (TRTSA) to obtain an L-rotation stable matching, as listed in detail in Table I.

TABLE I

TX-RX SELECTION AND TIME SLOT ALLOCATION ALGORITHM (TRTSA)

Input: Set of users $\mathcal{N}$; Set of time slots $\mathcal{T}$; locations of all the users.
Output: A L-rotation stable matching Ψ in which the set of Tx users and Rx users are determined for each time slot.
Initialization
Record current matching as Ψ, and construct the list $\mathcal{T}_{unmatched} = \{1, 2, \cdots, T_v\}$.
For each time slot i, each user j constructs the forbidden pair list $\mathcal{H}^{(i)}_j$.
Phase 1: ***Obtaining a feasible matching***
for $j = 1 : N$,
Obtain the available set of matched time slots as $\mathcal{P}_j$ according to (10).
if $\mathcal{P}_j = \varnothing$ and $|\mathcal{T}_{unmatched}| \neq \varnothing$
Randomly select a time slot i^* from $\mathcal{T}_{unmatched}$;
else
Obtain T_{i^*} according to (11).
end
end
Set $\Psi(j) = i^*$;
Update $\mathcal{T}_{unmatched} = \mathcal{T}_{unmatched} \backslash \{i^*\}$.
Phase 2: ***Rotation Matching***
Repeat:
Randomly select a user sequence $\mathcal{N}_s$ with length $L <= L_{max}$ which has not been searched before;
Find the optimal rotation matching $\Psi^{rot}_{\mathcal{N}_s,\zeta^*}$ with respect to $\mathcal{N}_s$ according to Definition 2;
Update $\Psi = \Psi^{rot}_{\mathcal{N}_s,\zeta^*}$;
until no rotation matching can further reduce the average cross influence to the network.
End of algorithm.

To start with, the BS collects each user's geometric information to construct the forbidden matching list of each user. In Phase 1, a feasible matching is obtained by the BS via a greedy algorithm in which each user is matched with exactly one time slot, i.e., each user acts as a Tx user in its matched time slot and acts as a Rx user in the other time slots. Phase 2 contains multiple iterations in each of which an optimal rotation matching is executed by the BS given a randomly selected rotation sequence. The iteration stops until no rotation sequence can further reduce the total cross influence of the network. The TRTSA is performed by the BS for centralized resource allocation.

3) Multi-Dimensional Stable Roommate Matching for the Sub-Channel Allocation: We now extend the above algorithm to solve the sub-channel allocation problem in each time slot. We re-construct the utility function based on (6) to formulate a non-convex integer programming problem, which is much more intractable than the time-user matching problem.

a) Definitions: We observe that the set of Tx users and subchannels can be considered as "students" and "rooms" in a MD SR problem, respectively. Denote the set of subchannels as $\mathcal{K}$. In each time slot i, a *user-subchannel matching* Φ_i is then defined as a mapping from the set $\Psi(i) \cup \mathcal{K} \cup \varnothing$ into itself such that $\Phi(j) \subseteq K$ and $\Phi(k) \subseteq \Psi(i)$, in which $j \in \Psi(i)$, $k \in \mathcal{K}$. Each Tx user can match with multiple sub-channels and each sub-channel can match with multiple Tx users. To be specific, we have $|\Phi(j)| \le K_{\max}$ based on constraint (7*b*). For constraint (7*d*), we present the following remark:

Remark 3: For sub-channel k, if there exists a user subset $\mathcal{N}_k \subseteq \Phi(k)$ satisfying that $\forall j, j' \in \mathcal{N}_k$, $d^{(i)}_{j,j'} \le 2r$ stands, then it is required that $|\mathcal{N}_k| \le K_u$.

Note that this is a tighter constraint compared to constraint (7d). However, due to the dense topology of the network, it is highly likely that there exists at least one Rx user in such an overlapping region, which makes Remark 3 a suitable approximation to constraint (7d).

b) Preference and utility functions: Different from the geometric utility function in (9), we further evaluate the co-channel cross influence based on the number of successfully decoded signals, which is our final goal of maximization. The utility functions of the Tx users and sub-channels need to be re-defined in this case.

For any Tx user j, its utility with respect to subchannel k represents the approximation of the number of successfully decoding signals in its communication disk, i.e., $\sum_{m\in\mathcal{N}_j^{(i)}} U_{j,m,k}^{(i)}$. If Tx user j obtains higher utility from subchannel k than from subchannel k', we have $k \succ_j k'$. Similarly, for any subchannel k, its preference $\succ_k$ over the subset of Tx users $\Psi(i)$ and $\Psi'(i)$ is defined based on the utility function as below. Given two matchings Φ, Φ', and $\Phi(k) = \Psi(i)$, $\Phi'(k') = \Psi'(i)$, we have

$$\{\Psi(i), \Phi\} \succ_k \left\{\Psi'(i), \Phi'\right\} \Leftrightarrow \sum_{j\in\Psi(i)} \sum_{m\in N_j^{(i)}} U_{j,m,k}^{(i)} > \sum_{j\in\Psi'(i)} \sum_{m\in N_j^{(i)}} U_{j,m,k'}^{(i)}. \quad (15)$$

Given a subset of Tx users $\mathcal{N}_s \subseteq \Psi(i)$, the index l^* for the optimal rotation sequence ζ_{l^*} can be obtained by

$$l^* = \arg\max_{1\le l\le L} \sum_{j\in\mathcal{N}_s} \sum_{m\in\mathcal{N}_j^{(i)}} \sum_{k\in\Phi^{rot}_{\mathcal{N}_s,\zeta_l}(j)} U_{j,m,k}^{(i)}, \quad (16)$$

and the optimal rotation matching is expressed as $\Phi^{rot}_{\mathcal{N}_s,\zeta_{l^*}}$. With the above definitions, the TRTSA proposed in Section III.B.2 can then be modified to a rotation matching algorithm for subchannel allocation (RMSA), as shown in Table II. The whole centralized Tx-Rx selection and time-frequency resource allocation scheme is then obtained after the BS performs the above two matching algorithms in Sections III.B.2 and III.B.3.

4) Analysis of the Proposed Matching Algorithms: The rotation matching is the core of both the TRTSA in Table I and the RMSA in Table II. Due to limited space, we use the TRTSA as an example to analyze the stability, convergence, complexity, and optimality.

a) Stability and convergence: As proved in Appendix C, the proposed TRTSA for the centralized Tx-Rx selection and time scheduling is guaranteed to converge to a final matching Ψ_{λ^*}. We then prove that Ψ_{λ^*} is an L-rotation stable matching in Appendix D.

The convergence of the RSMA can be guaranteed since the total utility increases after each rotation matching and there exists an upper bound of the total utility.

b) Complexity: Given the convergence of the TRTSA, we now analyze the complexity of its two phases in terms of the number of possible rotation matchings. As proved in Appendix E, the number of rotation matchings to be considered in each iteration of Phase 2 in the TRTSA is bounded by $N^{L_{\max}}$, with $L_{max} \ll N$. Given the total number of iterations I_{iter}, the complexity can then be expressed by $O\left(I_{iter} N^{L_{\max}}\right)$.

TABLE II

ROTATION MATCHING ALGORITHM FOR SUB-CHANNEL ALLOCATION (RMSA)

Initialization:
Set $\mathcal{L}_{record} = \varnothing$;
Each Tx user j randomly matches with a subset of subchannels with the size smaller than $K_{\max}$.
for each time slot i in the transmission period:
Repeat:
Select a user subset $\mathcal{N}_s \subseteq \Psi(i)$ and the corresponding rotation sequence $\zeta_L \notin \mathcal{L}_{record}$;
if ζ_L is not optimal
Set $\eta = L$;
while $\eta \ge 1$
Obtain l^{temp} based on equation (16);
if the updated matching $\Phi^{rot}_{\mathcal{N}_s,\zeta_{l^{temp}}}$ satisfies constraints (7*b*) and (7*d*)
Set $\zeta^* = \zeta_{l^{temp}}$;
Add ζ^* to $\mathcal{L}_{record}$;
break;
else
Remove l^{temp} from the candidate set $\{l\}$;
$\eta = \eta - 1$;
end
end
end
until no rotation matching can further improve the total utility of the network in equation (7).
end
End of algorithm.

Note that it is hard to get a closed form expression for I_{iter} since it is not trivial to know at which iteration the total cross influence stops decreasing, which is common for most heuristic algorithms [26], [28]. To evaluate the convergence, we will show the distribution of the total number of executed rotation matchings in both algorithms in Fig. 4.

For a special case of $L_{max} = N$, the influence of forbidden pairs is evaluated in the following corollary and proved in Appendix F.

Corollary 1: Given an initial matching Ψ and F forbidden pairs, at least $F\left[2^{N-1} - 2 + \frac{1}{\lfloor N/2\rfloor}\right]$ rotation matchings are not valid for the TRTSA with $L_{max} = N$.

Based on Corollary 1, we can prove the following corollary.

Corollary 2: Given $L_{max} = N$ and F forbidden pairs, the number of rotation matchings to be considered in each iteration is bounded by $(N-F)\cdot 2^{N-1} + 2F - \frac{F}{\lfloor N/2\rfloor}$ in each iteration of Phase 2 in the TRTSA.

Proof: When $L_{max} = N$, the number of all possible rotation sequences can be calculated by $\sum_{L=1}^{N} LC_N^L = N\cdot 2^{N-1}$. According to Corollary 1, the number of rotation matchings can then be evaluated as

$$N\cdot 2^{N-1} - F\left[2^{N-1} - 2 + \frac{1}{\lfloor N/2\rfloor}\right] = (N-F)\cdot 2^{N-1} + 2F - \frac{F}{\lfloor N/2\rfloor} \quad (17)$$

c) Optimality: After performing the algorithms proposed in Tables I and II, we can obtain a sub-optimal solution of the centralized Tx-Rx selection and time-frequency resource allocation problem, which is a natural result since the MD SR problem is NP-hard.

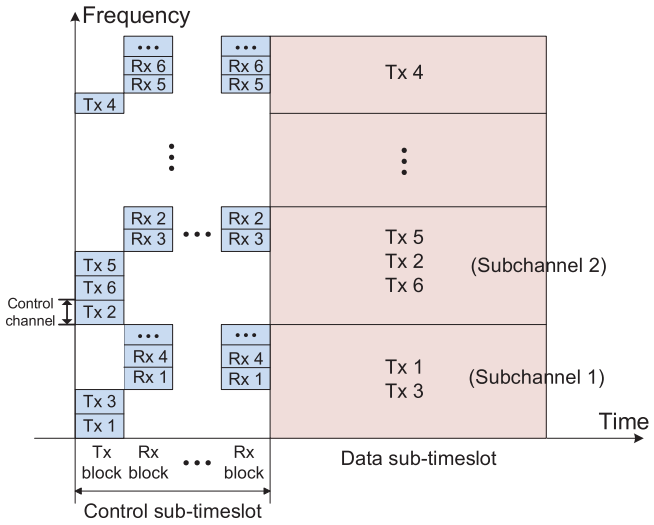


Fig. 2. The user association process within one time slot.

Proposition 1: All local minimum points of Ψ in the user-time matching problem and the local maximum point of Φ in the user-subchannel matching problem correspond to the L-rotation stable matchings.

Proof: We take the TRTSA for example. Suppose that the total cross influence of matching Ψ is a local minimum point. If it is not an L-rotation stable matching, then there must exist at least one rotation matching that can strictly reduce the cross influence of the network according to Definition 3, which is contradicted to the presupposition that Ψ is a local minimum. Therefore, matching Ψ must be L-rotation stable. ■

IV. NOMA-Based Distributed Power Control

In this section, we address the NOMA-based distributed power control scheme in which a power control problem is formulated for each Tx user and the control signaling between the Tx users and Rx users is considered.

To perform SIC decoding, necessary prior knowledge needs to be provided to the Rx users, e.g., the CSI of the intended Tx user – Rx user links, the number and transmit powers of the corresponding Tx users. Moreover, for a broadcasting system, the transmit power of each Tx user may influence multiple Rx users, and thus, power control is necessary. Furthermore, as mentioned in Section III, there is no centralized controller to provide real-time global information to the users. Therefore, the distributed power control requires information exchange between the Tx users and the Rx users, i.e., the *control signaling*. The function of the control signaling is two-fold: 1) the Rx users can obtain necessary information for SIC decoding; 2) the Tx users can obtain the local information from the Rx users for adjusting their own transmit powers. For such a distributed system, the above control signaling needs to updated iteratively to provide corresponding local information for each user.

A. Control Signaling

Note that control message may be transmitted separately with data in time within the same time slot to adapt to the changing topology of the vehicular networks [12]. We divide each time slot into two parts: the data control sub-timeslot and the data transmission sub-timeslot, as shown in Fig. 2. Since the control message is much smaller than the data message, we assume that the duration of control signaling is much shorter than that of the data transmission.

Since both Tx and Rx users need feedback from each other to obtain necessary information. Therefore, we divide each control sub-timeslot into multiple pairs of Tx blocks and Rx blocks,[10] i.e., multiple iterations of Tx-Rx control message exchange in the time domain. In each Tx block the Tx users broadcast their control message to its neighborhood and in each Rx block the Rx users broadcast their feedback to the neighboring Tx users. To be specific, each iteration works as below:

- *Tx block:* every Tx user j broadcasts a reference signal to its neighboring Rx users containing the transmit power and the frequency resources that it will occupy for the data transmission.[11] The transmit power, p_j, is determined based on the feedback sent by the Rx users in the previous Rx block. Each Rx user m obtains its neighboring Tx users' CSI via the received reference signals.
- *Rx block:* Based on the received message from previous Tx block, each Rx user m is informed of the subset of Tx users transmitting over the same sub-channel as well as their corresponding CSI. Therefore, Rx user m is able to predict whether it can successfully decode the coming data messages based on equation (4). To help the Tx users better adjust their transmit power, Rx user m calculates the potential co-channel interference to each Tx user j caused by other Tx users in the SIC decoding and broadcasts such information to neighboring Tx users for further processing in the next Tx block.

To constrain the total number of iterations such that the latency caused by signaling control is tolerable, we assume that there are T_c pairs of Tx and Rx blocks in each control sub-timeslot.

Similar to the data transmission, control message transmission also requires the frequency resources to be assigned. To obtain the precise CSI of the sub-channels for data transmission, the frequency resources allocated for the transmission of control message, including the reference signals, should be carefully assigned. To better distinguish different users, we adopt an OMA-based method for control message transmission. To be more specific, each user is assigned a unique frequency position in the control sub-timeslot such that the Rx (Tx) users can easily identify the Tx (Rx) users that broadcast to them. Through the pre-defined frequency resource for transmitting control message, the blind decoding complexity of the users can be saved [29]. For the Rx users, it is necessary to obtain the CSI of the sub-channels allocated to Tx users for data transmission so as to perform the SIC decoding in the coming data transmission. Therefore, we assume that the frequency positions assigned to each Tx user for control

[10] Each block is set for reference signal transmission with the duration of less than 200 μs.

[11] Note that before the distributed power control, the centralized time-frequency resource allocation has already been performed.

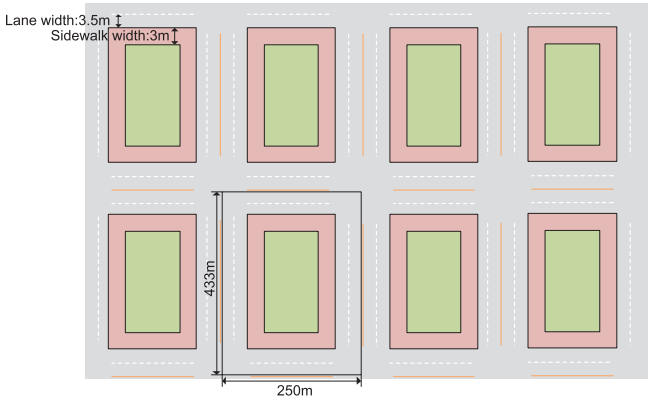


Fig. 3. Simulation setup for the road configuration in the urban scenario with the road width of $20m$.

message transmission are within the coherent bandwidth of the corresponding data sub-channels such that the precise CSI can be measured.

B. Power Control

We now describe how Tx user j decides its transmit power in each Tx block of time slot i. For convenience, we denote the set of Rx users sending feedback to Tx user j during the Rx block over subchannel k in time slot i as $\mathcal{B}_{j,k}^{(i)}$.

Note that in a dense V2X broadcast system, it is very hard that each Tx user j's signals can be successfully decoded by all neighboring Rx users. Therefore, we set a unique weighted value $w_j \in (0, 1]$ for each Tx user j such that at least $100w_j$ percent of the Rx users in the neighborhood should be able to decode the signal of Tx user j. The weighted value is updated once by the BS at the beginning of each SPS period. The power control problem of Tx user j over subchannel k in each Tx block of time slot i can then be formulated as below:

$$\min \quad p_{j,k}^{(i)} \tag{18}$$

$$s.t. \quad \sum_{m \in \mathcal{B}_{j,k}^{(i)}} \left[R_{j,m,k}^{(i)}(p_{j,k}^{(i)}) - \bar{R}_{th} \right]^{+} \geq w_j \left| \mathcal{B}_{j,k}^{(i)} \right|, \tag{18a}$$

$$p_{j,k}^{(i)} \leq P, \tag{18b}$$

in which P is the maximum transmit power of Tx user j over subchannel k. In problem (18), Tx user j aims to transmit with the minimum power such that at least $100w_j$ percent Rx users in $\mathcal{B}_{j,k}^{(i)}$ can decode their signals. In this way, the cross influence caused by multiple Tx users can be constrained to a tolerable level. Problem (18) can be solved by utilizing a bisection method [30]. The whole distributed power control and user association process is presented in detail in Table III.

TABLE III
CONTROL SUB-TIMESLOT IN THE DISTRIBUTED POWER CONTROL IN TIME SLOT i

Initialization: each user is assigned a unique frequency position for transmission in the control portion.
for $t = 1 : T_c$
Tx block:
if $t = 1$
Each Tx user broadcasts a reference signal with the initial transmit power $p_0^{(i)}$ over its allocated sub-channels;
else
Each Tx user j obtains the optimal transmit power $p_{j,k}^{(i)}$ over sub-channel k by solving problem (18);
It broadcasts a reference signal over sub-channel k with $p_{j,k}^{(i)}$.
end
Each Rx user m receives the control signals of neighboring Tx users;
After decoding, Rx user m knows the subset of sub-channels assigned to each Tx user in the coming data sub-timeslot;
Rx user m obtains the CSI and transmit power of neighboring Tx users.
Rx block:
for each Rx user m:
Set F as an empty matrix in which $F(l)$ ($l \geq 1$) represents the row vector.
Set $l = 1$.
for each Tx user j broadcasting to Rx user m:
Set the sub-channel occupied by Tx user j in the coming data sub-timeslot as k;
Calculate the potential interference item for Tx user j over sub-channel k as shown in (3) (i.e., equation (1) in this cover letter);
if (Rx user m predicts that it can decode the coming data message of Tx user j) **or** (Rx user m predicts that it cannot decode the coming data message of Tx user j but it can decode that of all the other Tx users with higher SIC decoding priority than Tx user j)
Add [id number of Tx user j, sub-channel k, interference item] to $F(l)$;
$l = l + 1$.
end
end
Rx user m broadcasts F to the neighboring Tx users.
end
end
End of the control portion.

V. SIMULATION RESULTS

In this section, we consider the urban scenario defined in [12], as shown in Fig. 3. The average inter-vehicle distance in the same lane is $2.5\text{s}\times v$, and the same density/speed in all lanes is adopted [12]. For simplicity, we assume that one SPS priod We evaluate the performance of the proposed NOMA-HCD scheme compared with the traditional OMA-based scheme and a NOMA-based geometric greedy algorithm (NOMA-GGA):

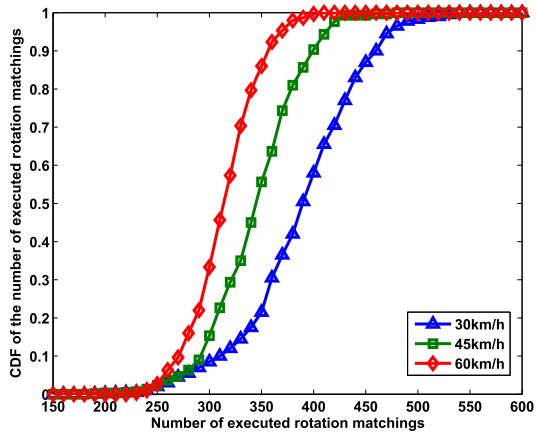
- **OMA-based scheme:** each Tx user transmits at maximum power. The Tx-Rx selection and time scheduling is solved by utilizing a greedy algorithm and the graph-based method in [19] is performed in the sub-channel allocation, in which the number of conflicting Rx users is minimized.
- **NOMA-GGA:** we adopt the geometric utility function (9) in the centralized SPS for both time and frequency resource allocation. A greedy algorithm proposed in Phase 1 of Table I is adopted.

Major simulation parameters are listed in Table IV. We consider the length of each scheduling period as 40ms for simplicity here, which can be easily extended to 100ms. The number of subchannels is set as 5 in the NOMA case[12]

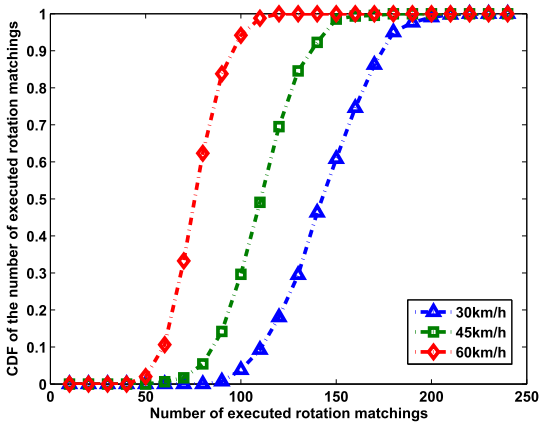
[12]The number of subchannels K in the NOMA case is set smaller than that in the OMA case considering the decoding complexity and signaling overhead caused by the increase of K.

TABLE IV
MAIN SIMULATION PARAMETERS

Parameter	Value
Peak power of each Tx user over each subchannel P	23 dBm
Noise power spectral density	-174 dBm/Hz
Carrier center frequency	2GHz
System bandwidth [12]	10 MHz
Decoding rate threshold	2 bits/Hz/s
Length of each time slot	1 ms
Rotation parameter L_{max}	4
Slope parameter η	4
Maximum number of Tx users transmitting to the same Rx users over the same subchannel simultaneously K_u	2
Maximum number of sub-channels that a Tx user occupies in one time slot K_{max}	2
Maximum number of time slots assigned to one user in a transmission period T_{max}	2
Number of Tx and Rx blocks in distributed power control T_c	6
Speed of vehicles	15 $\sim$ 60 km/h
Size of each packet	300 bytes
Ratio of control sub-timeslot in one timeslot	15%
Channel model	UMi model [31]



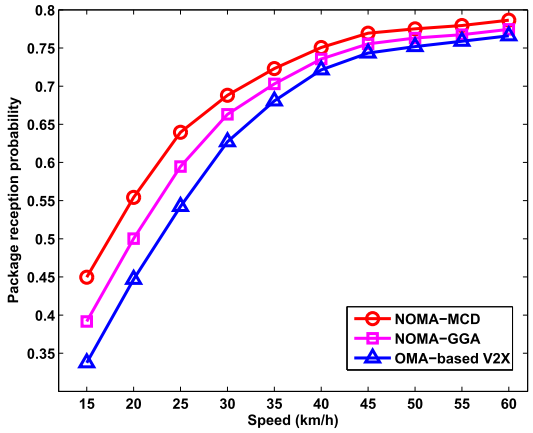
(a)



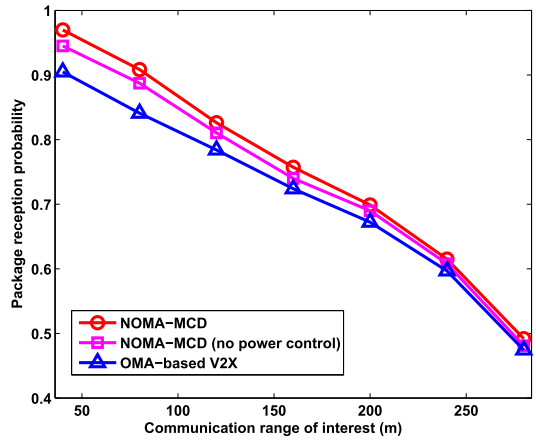
(b)

Fig. 4. (a) C.D.F. of the total number of executed rotation matchings in the TRTSA; (b) C.D.F. of the total number of executed rotation matchings in the RMSA.

and 10 in the OMA case, respectively. Each active Tx user occupies one time slot in the worst case in each transmission period, implying that it needs to successively



(a)



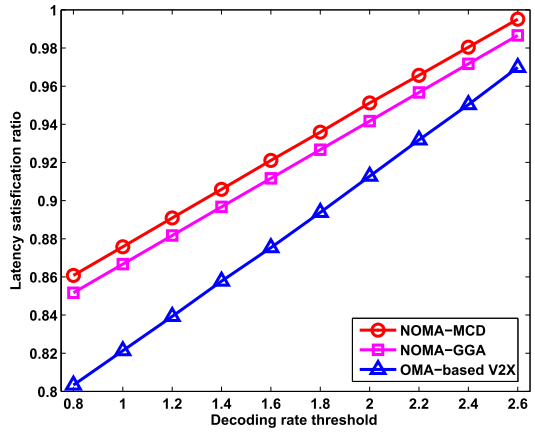
(b)

Fig. 5. (a) PRP v.s. speed of the vehicles; (b) PRP v.s. communication range of interest.

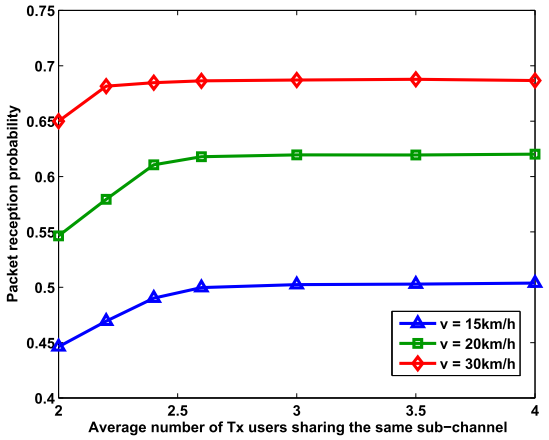
transmitting 300 bytes in this slot. Given the above parameters in Table IV, the required transmission rates for meeting the low latency requirement is 1.41 bits/Hz/s in the NOMA case and 2.4 bits/Hz/s in the OMA case, respectively. Simulation results are obtained as shown below.

Let the random variable $\tilde{Y}$ denote the total number of executed rotation matchings λ^* required for the proposed matching algorithms to converge, which is proportional to the computational complexity of our proposed algorithms. Fig. 4 shows the cumulative distribution function (C.D.F.) of $\tilde{Y}$, $\Pr\left(\tilde{Y} \leq \tilde{y}\right)$, versus $\tilde{y}$ for different speeds of vehicles with the communication range of interest $r = 150$. From Fig. 4, $\tilde{Y}$ in the TRTSA is much smaller than the total number of potential rotation matchings as expressed in Appendix E due to the existence of the forbidden pairs. Moreover, $\tilde{Y}$ in the RMSA is lower than that in the TRTSA due to a smaller scale of the subchannel allocation problem.

Fig. 5(a) illustrates the PRP versus the speed of the vehicles, with the communication range of interest as 150m. In the simulation, we consider a packet successfully received only when both the latency requirement and the criteria for successful decoding in equation (4) are satisfied. We observe



(a)



(b)

Fig. 6. (a) Latency satisfaction ratio v.s. decoding rate threshold; (b) PRP v.s. the average number of users sharing the same sub-channel.

that the PRP increases with the speed of the vehicles, and the PRP growth becomes slower as the speed grows. When the speed increases, the density of users decreases in the network, and thus, there is less potential collision, leading to an increased PRP. Our proposed NOMA-MCD scheme performs better than the NOMA-GAA scheme and the OMA-based scheme due to a smart utilization of the time and frequency resources. Furthermore, the gap between the OMA-based scheme and the NOMA-MCD scheme becomes smaller as the speed grows, indicating that the non-orthogonal manner of resource utilization works better in a dense network.

Fig. 5(b) shows the PRP versus the communication range of interest r. The communication range of interest r is set such that each Tx user only cares whether the Rx users within this range can receive its messages. It influences the formation of forbidden pairs in the TRTSA, i.e., any two vehicles cannot act as Tx users simultaneously within the distance of r. As the range r becomes larger, the number of active Tx users in one time slot decreases, i.e., the number of the accessed Tx users is smaller, leading to the decrease of the average PRP of the network. Moreover, the gap between the NOMA-MCD scheme and the OMA-based scheme becomes smaller as r increases since the possibility of orthogonal use of the frequency resources is higher. Fig. 5(b) also shows that our proposed NOMA-MCD scheme performs better than the case without power control.

Fig. 6(a) presents the latency satisfaction ratio versus the decoding rate threshold $\bar{R}_{th}$ with the speed $\upsilon = 30km/h$. The latency satisfaction ratio is defined as (the number of successfully transmitted packets satisfying both reliability and latency requirements) / (the number of successfully decoded Tx users given the decoding rate (or SNR) threshold). The latency satisfaction ratio increases with the decoding rate threshold. As the decoding rate threshold grows, the data rate of those successfully decoded Tx users increases. Therefore, it is more likely for them to satisfy the latency requirement. NOMA outperforms OMA in this case since the spectrum efficiency of the OMA case is lower than that of the NOMA case, leading to higher transmission delay.

Fig. 6(b) depicts the PRP versus the average number of Tx users sharing the same sub-channel simultaneously with the communication range of interest $r = 150$m. With different settings of the speed, the PRP grows to a stable value as K_u increases because each Tx user has more opportunity to access the channels. Since the computational complexity of SIC grows with K_u, the balance between the spectrum efficiency and decoding complexity can be reached when fewer than 3 Tx users share the same sub-channel.

VI. Conclusions

In this paper, we have studied the scheduling and resource allocation problem in a cellular V2X broadcasting system. To reduce the access collision and improve the reliability of the network, we have proposed a NOMA-based mixed centralized/distributed scheme, in which the time-frequency resources can be fully employed in a non-orthogonal manner by the BS while the dynamic power control along with the iterative control signaling is autonomously performed by the vehicles to achieve a better performance of power domain multiplexing. To solve the centralized Tx-Rx selection and resource allocation problem, we have re-formulated it as two MD-SR problems and proposed a novel rotation matching algorithm. Simulation results have showed that the NOMA-MCD scheme outperforms the OMA-based scheme especially in a dense network. In the future work, more practical issues such as the selection of modulation and coding schemes will be discussed.

Appendix A
Proof of NP-Hardness of Problem (7)

We prove that even the reduced version of (7) is a NP-hard problem, implying the original problem is also NP-hard [32].

Given the Tx-Rx selection and time slot allocation, the original problem is then reduced to a sub-channel allocation problem in which the BS allocates sub-channels to the Tx users in each time slot such that the number of successfully decoded packets is maximized. We construct an instance of the reduced problem in which $K_u = 2$, and establish the equivalence between this instance and a vertex coloring problem [33]. Since the vertex coloring problem has been proved to

be NP-hard, the instance of the reduced version of (7) is also NP-hard.

Consider the set of Tx users as nodes in a undirected conflicting graph. An edge exists between two nodes j and j' denoting they cannot be allocated the same sub-channel if either of the following conditions is satisfied: 1) Tx j and j' lie in each other's communication disk; 2) $R_{j,m,k} < \bar{R}_j$ or $R_{j',m,k} < \bar{R}_{j'}$, $\forall m \in \mathcal{N}_j^n \cap \mathcal{N}_{j'}^n$. Considering the set of sub-channels as colors, we aim at coloring as many nodes as possible such that no two adjacent vertices share the same color. For those dyed nodes, we say the corresponding Tx users can successfully send messages to the Rx users in their communication disks. Therefore, the equivalence between the vertex coloring problem and the instance is established.

A special case of problem (7) is NP-hard, and thus, the original problem in (7) is also NP-hard.

APPENDIX B
PROOF OF CONVERGENCE OF TRTSA

The TRTSA consists of two phases in which N iterations are required for the first phase. Therefore, the convergence depends on the second phase, i.e., rotation matching. Given an initial matching structure Ψ_0 obtained from Phase 1, the structure changes as below: $\Psi_0 \to \Psi_1 \to \Psi_2 \to \cdots$. Without loss of generality, we assume that a specific rotation sequence $\mathcal{N}_\mathcal{S}$ is considered during the structure change from Ψ_λ to $\Psi_{\lambda+1}$, i.e., $\Psi_{\lambda+1} = \Psi^{rot}_{\mathcal{N}_\mathcal{S},\zeta_{l^*}}$. Since ζ_{l^*} is a valid and optimal rotation matching, the relation of the total cross influence satisfies

$$\sum_{j\in\mathcal{N}_\mathcal{S}} \sum_{i\in\Psi_{\lambda+1}} \left(\sum_{j\in\mathcal{N}_\mathcal{S}} \sum_{i\in\Psi_\lambda(j)} Q_j^{(i)}\right. \tag{19}$$

Therefore, the total cross influence reduces after each rotation matching. The lower bound of the cross influence exists when all the users lie outside each other's communication disks, which can be expressed by $T_v\varepsilon$. Therefore, we can always find a rotation matching ζ_{l^*} after which there are no permitted rotation matchings in the current matching structure Ψ_{λ^*}.

APPENDIX C
PROOF OF STABILITY OF TRTSA

We know that the proposed TRTSA converges to a final matching Ψ_{λ^*}. Based on the definition of the rotation matching, the corresponding rotation sequence ζ_l in Ψ_{λ^*} is either the optimal one or the only valid one given any subset of users $\mathcal{N}_\mathcal{S} \subseteq \mathcal{N}$. Thus, the matches of $\mathcal{N}_\mathcal{S}$ with respect to ζ_l must be the best choice for users in $\mathcal{N}_\mathcal{S}$, with the strategies of other users in $\mathcal{N}\backslash\mathcal{N}_\mathcal{S}$ unchanged. Hence, the final matching is an L-rotation stable matching.

APPENDIX D
PROOF OF COMPLEXITY OF TRTSA

The complexity of TRTSA is determined by its two phases. For the first phase, the complexity is $O(N)$ since there are N iterations in total. The second phase consists of multiple iterations in each of which all valid rotation matchings are examined. We first calculate the upper bound of the number of all possible rotation matchings in each iteration, i.e., $\sum_{L=1}^{L_{\max}} LC_N^L = O\left(N^{L_{\max}}\right)$. In practice, one iteration involves a smaller number of rotations due to forbidden pairs.

APPENDIX E
PROOF OF COROLLARY 1

A forbidden pair $\{j, j'\}$ makes a rotation matching ζ_l invalid when $j \in \mathcal{N}^s$, $j' \notin \mathcal{N}^s$, and $\Psi\left(j'\right) \cap \Psi^{rot}_{\mathcal{N}^s,\zeta_l}(j) \neq \varnothing$, i.e., Tx users j and j' match with the same time slot after the rotation.

Given a forbidden pair $\{j, j'\}$, we consider a subset of users $\mathcal{N}^s$ in which $p \in \mathcal{N}^s$ and $p \in \Psi\left(\Psi\left(j'\right)\right)$. If $|\mathcal{N}^s| = 3$, then there are $(N-2)$ potential invalid rotation matchings. If $|\mathcal{N}^s| = 4$, then $C(K-2,2)$ rotation matchings are invalid. Note that when two forbidden pairs are involved in this subset of users, at least $C(K-2,2)/2$ rotation matchings are invalid. Following this methods, the lower bound of the number of invalid rotation matching can be calculated as

$$\begin{aligned} &F\left[C_{N-2}^1 + \frac{C_{N-2}^2}{\lfloor (2+2)/2 \rfloor} + \cdots + \frac{C_{N-2}^{N-2}}{\lfloor (N-2+2)/2 \rfloor}\right] \\ &= F\left[C_{N-2}^1 + \sum_{m=1}^{\lfloor (N-1)/2 \rfloor - 1} \frac{2}{2m+2} C_{N-1}^{2m+1} + \frac{1}{\lfloor N/2 \rfloor} C_{N-2}^{N-2}\right] \\ &= F\left[2^{N-1} - 2 + \frac{1}{\lfloor N/2 \rfloor}\right] \end{aligned} \tag{20}$$

REFERENCES

[1] G. Karagiannis *et al.*, "Vehicular networking: A survey and tutorial on requirements, architectures, challenges, standards and solutions," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 4, pp. 584–615, Sep. 2011.

[2] L. Song and J. Shen, *Evolved Network Planning and Optimization for UMTS and LTE*. Boca Raton, FL, USA: CRC Press, 2010.

[3] L. Song, D. Niyato, Z. Han, and E. Hossain, *Wireless Device-to-Device Communications and Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2015.

[4] G. Araniti, C. Campolo, M. Condoluci, A. Iera, and A. Molinaro, "LTE for vehicular networking: A survey," *IEEE Commun. Mag.*, vol. 51, no. 5, pp. 148–157, May 2015.

[5] S. Sun, J. Hu, Y. Peng, X. Pan, L. Zhao, and J. Fang, "Support for vehicle-to-everything services based on LTE," *IEEE Wireless Commun.*, vol. 23, no. 3, pp. 4–8, Jun. 2016.

[6] L. Song, Y. Li, Z. Ding, and H. Poor. (Oct. 2016). "Resource management in non-orthogonal multiple access systems: State of the art and research challenges." [Online]. Available: https://arxiv.org/abs/1610.09465

[7] 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 Veh. Technol. Conf.*, Dresden, Germany, Jun. 2013, pp. 1–5.

[8] K. Au *et al.*, "Uplink contention based SCMA for 5G radio access," in *Proc. IEEE Global Commun. Conf. Workshop (Globecom Workshop)*, Austin, TX, USA, Dec. 2014, pp. 900–905.

[9] 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.

[10] F. Liu, P. Mahonen, and M. Petrova, "Proportional fairness-based power allocation and user set selection for downlink NOMA systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[11] D. Tse and P. Viswanath, *Fundamentals Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[12] *LTE-Based V2X Services*, document RP-161894, 3GPP, Sep. 2016.

[13] L.-C. Tung and M. Gerla, "LTE resource scheduling for vehicular safety applications," in *Proc. 10th Annu. Conf. Wireless On-Demand Netw. Syst. Services (WONS)*, Mar. 2013, pp. 116–118.

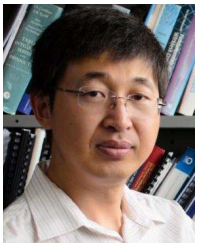
[14] D. Gusfield and R. Irving, *The Stable Marriage Problem: Structure and Algorithms*. Cambridge, MA, USA: MIT Press, 1989.
[15] E. Arkiny, S. Baez, A. Efratx, K. Okamoto, J. Mitchell, and V. Polishchukk, "Geometric stable roommates," *Inf. Process. Lett.*, vol. 109, no. 4, pp. 219–224, Jan. 2009.
[16] D. Manlove, *Algorithmics of Matching Under Preferences*. Singapore: World Scientific, 2013.
[17] Z. Wu, V. Park, and J. Li, "Enabling device to device broadcast for LTE cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 1, pp. 58–70, Jan. 2016.
[18] W. Sun, D. Yuan, E. Ström, and F. Brännström, "Cluster-based radio resource management for D2D-supported safety-critical V2X communications," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2756–2769, Apr. 2016.
[19] C. W. Yeh, G. Y. Lin, M. J. Shih, and H. Y. Wei, "Centralized interference-aware resource allocation for device-to-device broadcast communications," in *Proc. IEEE Int. Conf. Internet Things*, Sep. 2014, pp. 304–307.
[20] Y. Liao, K. Bian, L. Song, and Z. Han, "Full-duplex MAC protocol design and analysis," *IEEE Commun. Lett.*, vol. 19, no. 7, pp. 1185–1188, Jul. 2015.
[21] 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.
[22] X. Cheng, L. Yang, and X. Shen, "D2D for intelligent transportation systems: A feasibility study," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 1784–1793, Aug. 2015.
[23] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6010–6023, Aug. 2016.
[24] M. Proebster, M. Kaschub, T. Werthmann, and S. Valentin, "Context-aware resource allocation for cellular wireless networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2012, no. 1, p. 216, Jul. 2012.
[25] F. Blanquez-Casado, G. Gomez, M. del Carmen Aguayo-Torres, and J. T. Entrambasaguas, "eOLLA: An enhanced outer loop link adaptation for cellular networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2016, no. 1, p. 20, Jan. 2016.
[26] Y. Gu, W. Saad, and M. Bennis, "Matching theory for future wireless networks: Fundamentals and applications," *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 52–59, Nov. 2015.
[27] E. Baron, C. Lee, A. Chong, B. Hassibi, and A. Wierman, "Peer effects and stability in matching markets," in *Proc. Int. Symp. Algorithmic Game Theory (SAGT)*, Amalfi, Italy, Oct. 2011, pp. 117–129.
[28] B. Di, T. Wang, L. Song, and Z. Han, "Collaborative smartphone sensing using overlapping coalition formation games," *IEEE Trans. Mobile Comput.*, vol. 16, no. 1, pp. 30–43, Jan. 2017.
[29] *Discussion on Resource Pool Structure and Control Signaling for PC5-Based V2V*, document 3GPP TSG RAN WG1 Meeting#83, LG Electronics, Anaheim, CA, USA, Nov. 2015.
[30] J. Dennis and R. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1983.
[31] "Guidelines for evaluation of radio interface technologies for IMT-advanced," Int. Telecommun. Union (ITU), Geneva, Switzerland, Tech. Rep. ITU-R M.2135-1, 2009.
[32] L. Wolsey and G. Nemhauser, *Integer and Combinatorial Optimization*. Hoboken, NJ, USA: Wiley, 2014.
[33] T. Jensen and B. Toft, *Graph Coloring Problems*. New York, NY, USA: Wiley, 1995.

Boya Di (S'17) received the B.S. degree in electronic engineering from Peking University in 2014, where she is currently pursuing the Ph.D. degree with the School of Electrical Engineering and Computer Science. Her current research interests include matching theory, non-orthogonal multiple access, and 5G V2X networks.

Lingyang Song (S'03–M'06–SM'12) received the Ph.D. degree from the University of York, U.K., in 2007. He was a Research Fellow with the University of Oslo, Norway, until rejoining Philips Research U.K., in 2008. In 2009, he joined the School of Electronics Engineering and Computer Science, Peking University, China, as a Full Professor. His main research interests include MIMO, cognitive and cooperative communications, physical layer security, and wireless ad hoc/sensor networks. He received the K. M. Stott Prize for excellent research from the University of York.

He has authored two text books, *Wireless Device-to-Device Communications and Networks* and *Full-Duplex Communications and Networks* (U.K.: Cambridge University Press) and the co-editor of two other books, *Orthogonal Frequency Division Multiple Access (OFDMA)-Fundamentals and Applications* and *Evolved Network Planning and Optimization for UMTS and LTE (IEEE ComSoc Best Readings)* (USA: Auerbach Publications, CRC Press). He was a recipient of the 2012 IEEE Asia Pacific Young Researcher Award, and received eight best paper awards, including the IEEE WCNM 2007, the IEEE ICCC 2012, the ICST Chinacom 2012, the IEEE WCNC2012, the IEEE WCSP 2012, the IEEE ICC 2014, the IEEE Globecom 2014, and the IEEE ICC 2015, and one Best Demo Award in the ACM Mobihoc 2015.

Dr. Song served as the TPC Co-Chair of the International Conference on Ubiquitous and Future Networks in 2011 and 2012 and the Registration Co-Chair of the First IEEE International Conference on Communications in China in 2012. He served as the Symposium Co-Chair of the International Wireless Communications and Mobile Computing Conference in 2009 and 2010, the IEEE International Conference on Communication Technology in 2011, the IEEE International Conference on Communications in 2014 and 2016, the IEEE Vehicular Technology Conference in 2016, and the IEEE Global Communication Conference in 2016. He is currently on the Editorial Board of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the *China Communications*, and the *Journal of Network and Computer Applications*.



Yonghui Li (M'04–SM'09) received the Ph.D. degree from the Beijing University of Aeronautics and Astronautics in 2002. From 1999 to 2003, he was affiliated with Linkair Communication Inc., where he held a position of project manager with responsibility for the design of physical layer solutions for the LAS-CDMA system. Since 2003, he has been with the Centre of Excellence in Telecommunications, The University of Sydney, Australia, where he is currently a Professor with the School of Electrical and Information Engineering. He was a recipient of the Australian Queen Elizabeth II Fellowship in 2008 and the Australian Future Fellowship in 2012.

His current research interests are in the area of wireless communications, with a particular focus on MIMO, millimeter wave communications, machine to machine communications, coding techniques, and cooperative communications. He holds a number of patents granted and pending in these fields. He received the best paper awards from the IEEE International Conference on Communications 2014 and the IEEE Wireless Days Conferences 2014. He is currently an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and an Executive Editor of the *European Transactions on Telecommunications*.

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, USA, 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 Post-Doctoral Research Associate with the University of Maryland at College Park, College Park, MD, USA, from 1994 to 1996. He was with AT&T Labs–Research at Red Bank, NJ, USA, 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. He has also been holding a Cheung Kong Scholar Title at the University of Electronic Science and Technology of China since 2006. His general research interests include statistical signal processing and communications, with emphasis on cross-layer optimization for spectral- and energy-efficient networks, cognitive radios and opportunistic spectrum access, and practical issues in LTE systems. In these areas, he has authored around 200 journal papers in addition to around 40 granted patents and numerous conference papers. He was an IEEE Fellow for his contributions to signal processing for wireless communications in 2005. His publications have been cited over 28 000 times and has been recognized as the World's Most Influential Scientific Mind, also known as a Highly-Cited Researcher by Thomson Reuters. He received the 2010 Stephen O. Rice Prize Paper Award, the 2013 WTC Wireless Recognition Award, the 2017 Award for Advances in Communication from the IEEE Communications Society, the 2013 James Evans Avant Garde Award, and the 2014 Jack Neubauer Memorial Award from the IEEE Vehicular Technology Society. He 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'03, the Technical Program Co-Chair of the IEEE SPAWC'11, the General Chair of the IEEE GlobalSIP'14, and the Technical Program Co-Chair of the IEEE VTC'16 (Spring). 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.