Design of Cooperative Non-Orthogonal Multicast Cognitive Multiple Access for 5G Systems: User Scheduling and Performance Analysis

Lu Lv, Jian Chen, *Member, IEEE,* Qiang Ni, *Senior Member, IEEE,* and Zhiguo Ding, *Senior Member, IEEE*

***Abstract*—Non-orthogonal multiple access (NOMA) is emerging as a promising, yet challenging, multiple access technology to improve spectrum utilization for the fifth generation (5G) wireless networks. In this paper, the application of NOMA to multicast cognitive radio networks (termed as MCR-NOMA) is investigated. A dynamic cooperative MCR-NOMA scheme is proposed, where the multicast secondary users serve as relays to improve the performance of both primary and secondary networks. Based on the available channel state information (CSI), three different secondary user scheduling strategies for the cooperative MCR-NOMA scheme are presented. To evaluate the system performance, we derive the closed-form expressions of the outage probability and diversity order for both networks. Furthermore, we introduce a new metric, referred to as mutual outage probability to characterize the cooperation benefit compared to non cooperative MCR-NOMA scheme. Simulation results demonstrate significant performance gains are obtained for both networks, thanks to the use of our proposed cooperative MCR-NOMA scheme. It is also demonstrated that higher spatial diversity order can be achieved by opportunistically utilizing the CSI available for the secondary user scheduling.**

***Index Terms*—Non-orthogonal multiple access (NOMA), cooperative NOMA, multicast cognitive radio networks, opportunistic user scheduling, 5G systems.**

I. Introduction

Recently, there has been a surge of increased interest in applying non-orthogonal multiple access (NOMA) to the next generation of wireless networks [1]–[3]. Empowered by the technology of superposition coding, multiple users' signals are superimposed at the transmitter with distinct power allocation factors, and successive interference cancellation (SIC) is implemented at the receiver, so that the negative impacts of inter-user interference can be readily mitigated. In this regard, all users can share the same time, frequency or code domain, leading to a boost in spectral efficiency. Particularly, downlink NOMA has been applied to 3rd generation partnership project long-term evolution (3GPP-LTE) systems [4], and it has also been designated as one of the key multiple access techniques for the fifth generation (5G) wireless networks [5]–[8].

This work was supported by the EU FP7 CROWN project under Grant PIRSES-GA-2013-610524, by the National Natural Foundation of China under Grant 61601347, and by the "111" project of China under Grant B38038. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. M. Abdallah. (*Corresponding author: Qiang Ni.*)

L. Lv and J. Chen are with the State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an, 710071, China (email: lulv@stu.xidian.edu.cn; jianchen@mail.xidian.edu.cn).

Q. Ni and Z. Ding are with the School of Computing and Communications, Lancaster University, Lancaster LA1 4WA, UK (email: q.ni@lancaster.ac.uk; z.ding@lancaster.ac.uk).

A. Related Works

In contrast to traditional water-filling power allocation strategy, NOMA allocates more power to the users with worse channel conditions, which results in a better tradeoff between the system throughput and user fairness. In this context, the work in [9] has studied the power allocation with max-min fairness criterion. An uplink NOMA-OFDM scheme with joint power and subcarrier allocations has been proposed in [10], where the performance of both link-level and system-level has been investigated. Since it is always challenging to recruit all the users to perform NOMA, and it is preferable to implement user grouping/pairing to reduce the system complexity. To this end, the impact of user pairing on downlink NOMA systems has been characterized in [11]. The benefits of cooperative communications have been well acknowledged in the literature, where currently the diversity and relaying techniques have been used for the robust and reliable NOMA transmissions [12]–[15]. In [12], a cooperation-based NOMA scheme for coordinated direct and relay transmissions has been introduced. The performance of transmit antenna selection for NOMA assisted multiple-input multiple-output (MIMO) relay networks has been examined in [13]. A diversity-oriented detection mechanism for cooperative relaying system using NOMA has been proposed in [14]. Inspired by user collaboration, a cooperative NOMA transmission scheme has been proposed in [15], with jointly considering simultaneous wireless information and power transfer (SWIPT).

Meanwhile, cognitive radio (CR) has been widely acknowledged as a promising solution for the spectrum scarcity of wireless applications [16]–[20]. The new technology allows secondary users (SUs) to opportunistically utilize the licensed spectrum occupied by primary users (PUs). As an efficient method of delivering the same content to multiple receivers while minimizing network resource usage, multicasting [21]–[24] is undoubtedly an attractive transmission technique to fully capitalize the potential of CR networks, especially urgent for the spectrum constrained SUs [25]–[27]. In [25], an energy-efficient chance-constrained resource allocation for multicast CR networks based on statistical channel state information (CSI) has been investigated. In [26], an optimal beamforming strategy for cooperative multicast in cognitive MIMO relay networks has been proposed. The exact outage performance of cooperative multicast in selective relaying CR networks has been analyzed in [27].

B. Contributions and Outcomes

In light of the above cited works, both NOMA and multicast CR have demonstrated their advantages in spectrum utilization. However, by appropriately combining NOMA with multicast CR networking (termed as MCR-NOMA), further performance improvement in terms of spectral efficiency can be achieved. One example is to consider a situation where the PU locates closely to the edge of the cell, e.g., its channel conditions are poor; while the multicast SUs seat closely to the base station (BS) and thus have better channel conditions. The use of NOMA admits both PU and SUs to transmit at the same frequency band simultaneously, which substantially enhances the spectrum utilization. However, due to the existence of inter-user interference at the PU and heterogeneity of wireless channels at the SUs, the reception reliability of both networks may be degraded. Thanks to the SIC at the SUs, the signals intended to the PU and SUs can be jointly decoded. In order to improve the reliability, it is wisely to recruit the SUs as candidate relays for both primary and secondary networks, where user scheduling can be well implemented at the SUs to exploit the cooperative diversity. This cooperation is a win-win strategy for both networks, in the sense that the SUs help relay the primary and secondary traffic to boost both PU's and SU's performance, and meantime SUs obtain the opportunity to access the licensed spectrum, therefore improving the connectivity of the SUs and the overall spectral efficiency. Most recently, the application of NOMA to underlay CR networks has been studied in [28], where multiple unicast SUs are served simultaneously via non cooperative NOMA transmissions. In [29], a novel NOMA assisted cooperative spectrum sharing scheme for unicast primary and secondary streaming has been proposed and analyzed.

Different from [28] and [29] which mainly focus on the performance of unicast CR-NOMA communications, in this paper, we investigate the design of cooperative mechanism for the MCR-NOMA, where the multicast SUs serve as relays and collaboratively retransmit the signals intended for PU and SUs, respectively. The considered scenario can be directly applied to current cellular networks where local SUs may have common packets for nearby receivers, for example, local marketers may send the same advertising messages to people who happen to be in the neighborhood, while the PU locates at the cell edge. The main contributions of this paper can be summarized as follows. First, we propose a dynamic cooperative MCR-NOMA scheme, whereby the multicast SUs, served as potential relays for the primary network, will also help improve the reliability of the secondary multicast transmissions if the signals for both networks are decoded correctly by the SUs. This cooperation is particularly preferred by the primary network when the PU's quality-of-service (QoS) cannot be met by the primary network itself. Second, based on the available CSI, we further propose three different secondary user scheduling strategies for the cooperative MCR-NOMA scheme. Third, we carry out analysis to provide the closed-form expressions to reveal the outage performance and the diversity order at both primary and secondary networks. Furthermore, we introduce a new metric, referred to as mutual outage probability, to characterize whether the primary and secondary networks can take advantage of this cooperation, e.g., achieving a mutual cooperation benefit. Results show that a win-win situation can be eventually achieved and a more sophisticated user scheduling strategy can promise better outage performance for the overall networks, but at a sacrifice of consuming more communication overhead to obtain the required CSI. Finally, we present simulation results to demonstrate the effectiveness of the analytical results and the availability of the proposed cooperative MCR-NOMA with secondary user scheduling.

C. Organization and Notations

The remainder of this paper is organized as follows. The system model is explained and the cooperative MCR-NOMA scheme is introduced in Section II. In Section III, the performance of three secondary user scheduling strategies based on different CSI assumptions is investigated. Simulation results and discussions are presented in Section IV. Finally, the main outcomes are summarized in Section V.

Notations: Throughout the paper, $\mathcal{P}(\cdot)$ symbolizes probability; $f_X(\cdot)$ and $F_X(\cdot)$ represent the probability density function (PDF) and the cumulative distribution function (CDF) of a random variable X, respectively; "$\simeq$" denotes asymptotically equal to, which holds in the high signal-to-noise ratio (SNR) regime.

II. System Model

Consider a downlink cooperative MCR-NOMA framework consisting of one BS, one unicast PU and N multicast SUs, as illustrated in Fig. 1, which is envisioned to be of crucial importance for emerging wireless networks.[1] Without loss of generality, we assume that the PU locates closely to the edge of the cell and its channel conditions are poor, while the multicast SUs seat closely to the BS such that they have better channel conditions. This provides an incentive for the PU granting the multicast SUs an access to the licensed spectrum, in exchange for the SUs acting as relays to improve PU's performance, due to the PU's QoS may not be satisfied by the primary network itself in this case. Such an assumption has been commonly used in the literature on cooperative CR networking (see, e.g., [18]–[20] and references therein). For each transmission from the BS, both unicast signal for the PU and multicast signal for the SUs are transmitted simultaneously by NOMA signaling. Each node is equipped with a single antenna and operates in a half-duplex mode. All the channels in the network experience independent but not necessarily identically distributed (i.n.i.d.) Rayleigh block fading, e.g., the channels remain constant within each transmission block,

[1]Note that the considered cooperative MCR-NOMA scenario can be treated as a special case of CR networks. Bearing in mind that CR is aimed at improved utilization of the radio spectrum [16], in this work, we propose to employ NOMA for the simultaneous primary (unicast signal) and secondary (multicast signal) transmissions, so as to maximally enhance the spectral efficiency. Therefore, the proposed model is in line with the basic principle of CR, and can be regarded as the NOMA assisted cooperative CR networks.

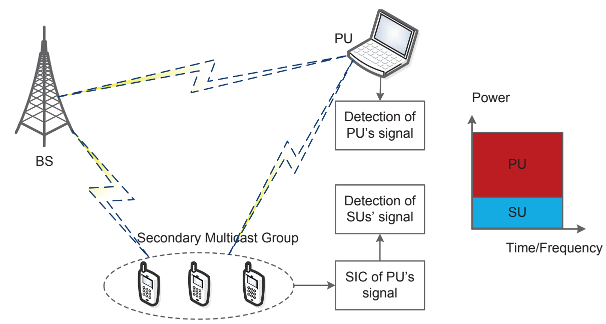


Fig. 1. Schematics of the downlink cooperative MCR-NOMA scenario.

but vary independently between different blocks. Denote $h_{b,p}$, $h_{b,n}$, $h_{n,p}$, $h_{n,n'}$ $(n, n' \in \{1, \cdots, N\})$ as the channel gains from the BS to the PU, the BS to SU n, SU n to the PU, and SU n to SU n', respectively. Thus, the channel power gains of these links follow an exponential distribution with means $d_{b,p}^{-\eta}$, $d_{b,n}^{-\eta}$, $d_{n,p}^{-\eta}$, and $d_{n,n'}^{-\eta}$, where $d_{i,j}$ denotes the Euclidean distance between nodes i and j, and η denotes the path loss exponent.

Note that in CR networks, the signal for the PU always has high priority to satisfy its real-time QoS constraint. While the SUs are to be served for low priority signal transmission, and are usually connected in an opportunistic manner [16]. Let x_p and x_s be the signals for the PU and SUs, with zero means and $E[x_p^* x_p] = E[x_s^* x_s] = 1$. The transmit power at each node is limited by P. The background noise is assumed to be white and Gaussian with variance of σ^2.

The decode-and-forward protocol is applied at the SUs, and the proposed cooperative MCR-NOMA scheme consists of two contiguous time slots. During the first time slot, the BS broadcasts the superimposed signals $(\alpha_p x_p + \alpha_s x_s)$, where α_p and α_s denote the power allocation coefficients. Invoking the NOMA signalling,[2] we have the power allocation $\alpha_p > \alpha_s$ with $\alpha_p^2 + \alpha_s^2 = 1$. Note that fixed power allocation is used in this paper. Optimizing the power allocation coefficients can further improve the performance of cooperative MCR-NOMA, which is beyond the scope of this paper. The received signals at the PU during the first time slot are expressed as

$$y_{p,1} = \sqrt{P} h_{b,p}(\alpha_p x_p + \alpha_s x_s) + \omega_{p,1} \tag{1}$$

where $\omega_{p,1}$ denotes the additive noise at the PU. Since $\alpha_p > \alpha_s$, the received signal-to-interference-plus-noise ratio (SINR) at the PU during the first time slot is given by

$$\mathrm{SINR}_p = \frac{\alpha_p^2 |h_{b,p}|^2}{\alpha_s^2 |h_{b,p}|^2 + \frac{1}{\rho}} \tag{2}$$

where $\rho = P/\sigma^2$ represents the transmit SNR. Similarly, SU n $(1 \le n \le N)$ in the secondary multicast group observes

$$y_n = \sqrt{P} h_{b,n}(\alpha_p x_p + \alpha_s x_s) + \omega_n \tag{3}$$

where ω_n represents the additive noise at SU n. Using NOMA, SIC will be carried out at SU n to combat the negative effect of the inter-user interference. Specifically, SU n first decodes the signal for the PU, and then removes this component from the received signals to subtract its own information. Accordingly, the received SINR at SU n to detect x_p is given by

$$\mathrm{SINR}_{n,p} = \frac{\alpha_p^2 |h_{b,n}|^2}{\alpha_s^2 |h_{b,n}|^2 + \frac{1}{\rho}}. \tag{4}$$

After completely canceling out x_p, the received SNR at SU n to detect x_s is shown by

$$\mathrm{SNR}_{n,s} = \rho \alpha_s^2 |h_{b,n}|^2. \tag{5}$$

Recalling the natural ordering of SIC, we denote $\mathcal{S}_1$ as the set of SUs which can decode x_p correctly but cannot decode x_s, shown as follows

$$\mathcal{S}_1 = \Big\{ n : 1 \le n \le N, \mathrm{SINR}_{n,p} \ge \tau_p, \mathrm{SNR}_{n,s} < \tau_s \Big\} \tag{6}$$

and let $\mathcal{S}_2$ be the set of SUs which can decode both x_p and x_s correctly, given by

$$\mathcal{S}_2 = \Big\{ n : 1 \le n \le N, \mathrm{SINR}_{n,p} \ge \tau_p, \mathrm{SNR}_{n,s} \ge \tau_s \Big\} \tag{7}$$

where $\tau_p = 2^{2R_p} - 1$, $\tau_s = 2^{2R_s} - 1$, R_p and R_s denote the targeted data rates for x_p and x_s, respectively. Note that the sizes of $\mathcal{S}_1$ and $\mathcal{S}_2$ are represented by $|\mathcal{S}_1|$ and $|\mathcal{S}_2|$.

During the second time slot, according to the results of whether the signals x_p and x_s are decoded correctly by the scheduled SU, a dynamic cooperative MCR-NOMA scheme, which adaptively switches between the cooperative NOMA and the cooperative spatial division multiple access (SDMA), is described in the following three cases.

(i) *Event of* $1 \le |\mathcal{S}_2| \le N-1$: In this case, suppose that SU n is selected from $\mathcal{S}_2$ to help the BS with cooperative NOMA, where the selection criterion will be discussed in the following section. In particular, SU n re-encodes and forwards the signal mixture $(\alpha_p x_p + \alpha_s x_s)$ to the PU and other SUs. Therefore, during the second time slot, the PU receives

$$y_{p,2} = \sqrt{P} h_{n,p}(\alpha_p x_p + \alpha_s x_s) + \omega_{p,2} \tag{8}$$

where $\omega_{p,2}$ is the additive noise. Upon maximal ratio combining (MRC) over the two contiguous time slots, the resulting SINR at the PU is given by

$$\mathrm{SINR}_{p,end} = \frac{\alpha_p^2 |h_{b,p}|^2}{\alpha_s^2 |h_{b,p}|^2 + \frac{1}{\rho}} + \frac{\alpha_p^2 |h_{n,p}|^2}{\alpha_s^2 |h_{n,p}|^2 + \frac{1}{\rho}}. \tag{9}$$

After using SIC, SU n' belonging to $\mathcal{S}_2^c$ (e.g., $\mathcal{S}_2^c$ is the complementary set of $\mathcal{S}_2$, indicating the SUs that fail to decode signals x_p and x_s beforehand) tries to decode its own signal x_s with the following SNR:[3] $\mathrm{SNR}_{n',s} = \rho \alpha_s^2 |h_{n,n'}|^2$.

(ii) *Event of* $|\mathcal{S}_2| = 0$ *and* $|\mathcal{S}_1| \ne 0$*, or event of* $|\mathcal{S}_2| = N$: In this case, suppose that SU $\hat{n}$ is selected from $\mathcal{S}_1$ to help the

[2]Moreover, from the concept of CR that the PU is served with the high priority while the SUs are connected opportunistically, it is reasonable to allocate more power to the PU to satisfy its high QoS requirement [16].

[3]For mathematically tractable, SU n' in $\mathcal{S}_2^c$ does not combine the signals from the BS and SU n. Therefore, the results in this paper can be treated as a lower bound for system performance.

BS with cooperative SDMA.[4] Particularly, x_p is regenerated and forwarded by SU $\hat{n}$, and the received signal at the PU during the second time slot is

$$y_{\tilde{p},2} = \sqrt{P} h_{\hat{n},p} x_p + \omega_{p,2}. \tag{10}$$

Similarly, by using MRC the resulting SINR at the PU follows

$$\text{SINR}_{\tilde{p},end} = \frac{\alpha_p^2 |h_{b,p}|^2}{\alpha_s^2 |h_{b,p}|^2 + \frac{1}{\rho}} + \rho |h_{\hat{n},p}|^2. \tag{11}$$

Since there is no SU which can decode x_s correctly (e.g., $\mathcal{S}_2$ is empty), an outage will be always declared at the secondary network. Therefore, the transmission scheme essentially reduces to the conventional cooperative relaying by recruiting the SUs as pure relays, in order to guarantee the QoS requirement of the PU [20].

(iii) *Event of* $|\mathcal{S}_1| = |\mathcal{S}_2| = 0$: In this case, there exists no SU that can decode the signals x_p and x_s correctly, therefore, the PU receives signal x_p only from the direct link $h_{b,p}$ and the overall SINR can be computed by (2) straightforwardly.

A detailed summary of the proposed dynamic cooperative MCR-NOMA scheme is shown in **Algorithm 1**.

Algorithm 1 Dynamic cooperative MCR-NOMA scheme

1: During the first time slot, the BS broadcasts the signal mixture $(\alpha_p x_p + \alpha_s x_s)$ to the PU and multicast SUs;
2: With the detection results at the SUs, it follows:
(i) Event of $1 \le |\mathcal{S}_2| \le N-1$, the scheduled SU n retransmits the signal mixture $(\alpha_p x_p + \alpha_s x_s)$ with cooperative NOMA;
(ii) Event of $|\mathcal{S}_2| = 0$ and $|\mathcal{S}_1| \ne 0$; or event of $|\mathcal{S}_2| = N$: Being aware of the decoding results from other SUs, the scheduled SU n retransmits the signal x_p with cooperative SDMA;
(iii) Event of $|\mathcal{S}_1| = |\mathcal{S}_2| = 0$, it cancels the cooperative MCR-NOMA.
3: During the second time slot, by appropriately scheduling SU n under different available CSI, the proposed dynamic cooperative MCR-NOMA begins.

III. Secondary User Scheduling Strategies and Performance Analysis

It is worth pointing out that the scheduled SU plays a crucial role for the primary and secondary networks, therefore it is important to investigate the user scheduling strategies for the cooperative MCR-NOMA. Under different CSI assumptions, the performance of three secondary user scheduling strategies is analyzed in the following subsections, providing thus more insights for practical setups.

[4]Note that for the event of $|\mathcal{S}_2| = N$, all the SUs decode x_s correctly. Therefore, being aware of the decoding results from other SUs, the scheduled SU n will use all its power to relay x_p in the second time slot. Otherwise, it still forwards the signal mixture $(\alpha_p x_p + \alpha_s x_s)$, as will be detailed in Section III.

A. User Scheduling without Instantaneous CSI

Consider round robin scheduling, the BS will randomly select a SU served as a relay. This strategy does not require any CSI, which significantly reduces the communication overhead and is preferable to the limited-feedback systems. In this paper, two evaluation metrics, namely outage probability and diversity order, are used to characterize the system performance.

1) Outage Probability: Without loss of generality, we assume SU n is selected to help the BS. The outage probability for the primary network given the use of SU n is[5]

$$\begin{aligned}\mathcal{P}_p =& \mathcal{P}(\text{SINR}_p < \tau_p, \text{SINR}_{n,p} < \tau_p) \\ &+ \mathcal{P}(\text{SINR}_{\tilde{p},end} < \tau_p, \text{SINR}_{n,p} \ge \tau_p, \text{SNR}_{n,s} < \tau_s) \\ &+ \mathcal{P}(\text{SINR}_{p,end} < \tau_p, \text{SINR}_{n,p} \ge \tau_p, \text{SNR}_{n,s} \ge \tau_s).\end{aligned} \tag{12}$$

From (12), the first probability is for the event that neither the PU nor SU n can decode x_p. The second and third probabilities are for the events that SU n applies cooperative SDMA and cooperative NOMA but the overall SINR cannot support the targeted data rate.

On the other hand, the capacity of secondary multicasting is dominated by the SU with the weakest channel gain in order to minimize outage and retransmission [21]. Accordingly, the outage probability for the secondary network can be defined as

$$\begin{aligned}\mathcal{P}_s =& \big(1 - \mathcal{P}(\text{SINR}_{n,p} \ge \tau_p, \text{SNR}_{n,s} \ge \tau_s)\big) \\ &+ \mathcal{P}\big(\text{SINR}_{n,p} \ge \tau_p, \text{SNR}_{n,s} \ge \tau_s, \text{SINR}_{n,n'} < \tau_s\big)\end{aligned} \tag{13}$$

where n' denotes the SU with the worst relaying channel gain in $\mathcal{S}_2^c$ to SU n. From (13), the former part indicates the event that SU n fails in decoding x_s, and the latter part denotes the event that x_s can be decoded by SU n but cannot be decoded by SU n' (whose relaying channel gain is the weakest). The following theorem provides the outage probability for the primary and secondary networks achieved by the round robin scheduling strategy.

Theorem 1: For i.n.i.d. channels, the outage probability for the primary network with a randomly scheduled SU can be approximated as

$$\begin{aligned}\mathcal{P}_p \approx& \left(1 - e^{-\frac{\psi_p d_{b,p}^\eta}{\rho}}\right)\left(1 - e^{-\frac{\psi_p d_{b,n}^\eta}{\rho}}\right) \\ &+ \chi\left(e^{-\frac{\psi_p d_{b,n}^\eta}{\rho}} - e^{-\frac{\psi_s d_{b,n}^\eta}{\rho}}\right)\left(1 - e^{-\frac{(\tau_p - a_l) d_{n,p}^\eta}{\rho}}\right) \\ &+ \chi e^{-\frac{\psi_s d_{b,n}^\eta}{\rho}}\left(1 - e^{-\frac{(\tau_p - a_l) d_{n,p}^\eta}{\rho(\alpha_p^2 - \alpha_s^2(\tau_p - a_l))}}\right).\end{aligned} \tag{14}$$

In (14), χ can be computed by

$$\chi = \frac{\pi\varpi}{2L} \sum_{l=1}^{L} \frac{\sqrt{1-\theta_l^2}\, d_{b,p}^\eta}{\rho(\alpha_p^2 - \alpha_s^2 a_l)^2} e^{-\frac{a_l d_{b,p}^\eta}{\rho(\alpha_p^2 - \alpha_s^2 a_l)}} \tag{15}$$

where $\psi_p = \frac{\tau_p}{(\alpha_p^2 - \alpha_s^2 \tau_p)}$, $\psi_s = \frac{\tau_s}{\alpha_s^2}$, $\varpi = \frac{\alpha_p^2}{\alpha_s^2}$, $a_l = \frac{(\theta_l+1)\varpi}{2}$, $\theta_l = \cos(\frac{2l-1}{L}\pi)$, and L is a complexity-accuracy tradeoff

[5]Since there is no CSI exchange or feedback among the multicast SUs, so the randomly scheduled SU does not know whether the other SUs can detect x_s correctly, and it will always retransmit the signal mixture $(\alpha_p x_p + \alpha_s x_s)$ if both are decoded.

parameter. The outage probability for the secondary network is given by

$$\mathcal{P}_s = \left(1 - e^{-\frac{\psi_s d_{b,n}^{\eta}}{\rho}}\right) + \sum_{k=1}^{N-1} \sum_{\substack{\mathcal{S}_2^c \subseteq \{1,\cdots,N\} \\ |\mathcal{S}_2^c|=k}} e^{-\sum_{n\in\mathcal{S}_2} \frac{\psi_s d_{b,\tilde{n}}^{\eta}}{\rho}} \times \prod_{i=1}^{k} \left(1 - e^{\frac{\psi_s d_{b,n_i}^{\eta}}{\rho}}\right)\left(1 - e^{-\sum_{i=1}^{k} \frac{\psi_s d_{n,n_i}^{\eta}}{\rho}}\right) e^{-\frac{\psi_s d_{b,n}^{\eta}}{\rho}}. \tag{16}$$

Proof: See Appendix A for details. ■

2) Mutual Outage Probability: It is worth noting that the proposed cooperative MCR-NOMA can lead to a win-win situation for both networks, e.g., the outage performance is enhanced for the PU while the SUs obtain the opportunity to access the licensed spectrum. In this regard, an outage event that either the primary or the secondary network is in outage, referred to as mutual outage probability, is considered in this paper. Since both networks should transmit reliably, the mutual outage probability can be formulated as

$$\mathcal{P}_{mop} = \mathcal{P}(\mathcal{O}_1) + \mathcal{P}(\mathcal{O}_2) \tag{17}$$

where $\mathcal{O}_1$ denotes the event that SU n cannot detect x_p and x_s, and $\mathcal{O}_2$ denotes the event that though x_p and x_s can be detected by SU n, either the PU or SU n' (n' represents the SU whose relaying channel is the weakest) is in outage. From (17), $\mathcal{P}(\mathcal{O}_1)$ can be further written as

$$\mathcal{P}(\mathcal{O}_1) = 1 - \mathcal{P}\big(\text{SINR}_{n,p} \geq \tau_p, \text{SNR}_{n,s} \geq \tau_s\big). \tag{18}$$

In addition, $\mathcal{P}(\mathcal{O}_2)$ can be defined as

$$\mathcal{P}(\mathcal{O}_2) = \mathcal{P}(|\mathcal{S}_2^c| = 0)\mathcal{P}(\text{SINR}_{\tilde{p},end} < \tau_p) + \sum_{k=1}^{N-1} \sum_{\substack{\mathcal{S}_2^c \subseteq \{1,\cdots,N\} \\ |\mathcal{S}_2^c|=k}} \mathcal{P}(|\mathcal{S}_2^c| = k) \times \Big(\mathcal{P}_{out,s}(|\mathcal{S}_2^c| = k) + \mathcal{P}(\text{SINR}_{p,end} < \tau_p)\Big). \tag{19}$$

As such, by substituting the results derived in Appendix A into (17), an explicit expression of $\mathcal{P}_{mop}$ can be obtained immediately.

3) Diversity Analysis: To obtain the robustness of the cooperative MCR-NOMA scheme in the high SNR regime, we provide a diversity analysis for the primary and secondary networks achieved by the round robin user scheduling strategy. In the following, Lemma 1 provides the asymptotic outage probability and the diversity order analysis.

Lemma 1: For a sufficiently large SNR region, e.g., $\rho \to \infty$, the asymptotic outage probability for the primary network can be expressed as

$$\mathcal{P}_p \simeq \frac{1}{\rho^2}\left(\psi_p^2 d_{b,p}^{\eta} d_{b,n}^{\eta} + \frac{\chi'(\tau_p - a_l) d_{n,p}^{\eta}}{\alpha_p^2 - \alpha_s^2(\tau_p - a_l)}\right) \tag{20}$$

where χ' is given by

$$\chi' = \frac{\pi\varpi}{2L} \sum_{l=1}^{L} \frac{\sqrt{1-\theta_l^2}\, d_{b,p}^{\eta}}{(\alpha_p^2 - \alpha_s^2 a_l)^2}. \tag{21}$$

The asymptotic outage probability for the secondary network can be expressed as

$$\mathcal{P}_s \simeq \frac{\psi_s d_{b,n}^{\eta}}{\rho} + \sum_{k=1}^{N-1} \sum_{\substack{\mathcal{S}_2^c \subseteq \{1,\cdots,N\} \\ |\mathcal{S}_2^c|=k}} \prod_{i=1}^{k} \frac{\psi_s d_{b,n_i}^{\eta}}{\rho} \sum_{i=1}^{k} \frac{\psi_s d_{n,n_i}^{\eta}}{\rho}. \tag{22}$$

The diversity order is defined as [28]: $D = -\lim_{\rho\to\infty} \frac{\log \mathcal{P}(\rho)}{\log \rho}$, and therefore we obtain

$$D_p = 2, \quad D_s = 1. \tag{23}$$

Furthermore, we define system diversity order as

$$D_{sys} = -\lim_{\rho\to\infty} \frac{\log \mathcal{P}_{mop}(\rho)}{\log \rho} = 1. \tag{24}$$

Proof: By using the simple fact that $1 - e^{-x} \simeq x$ when $x \to 0$, (20) and (22) can be obtained straightforwardly. Knowing the fact that the higher order of $\frac{1}{\rho}$ can be ignored in the high SNR regime, and thus it is readily to attain the diversity order shown in (23). After substituting (17) into (24), D_{sys} can be obtained similarly. This completes the proof. ■

Remark 1: As can be observed from (23), a diversity order of two is achieved for the primary network by using the round robin scheduling. This is due to the use of the proposed cooperative MCR-NOMA scheme. Surprisingly, the secondary network achieves only a diversity order of one, even if the randomly scheduled SU successfully decodes x_p and x_s, and then forwards them in the second time slot. Due to the performance of multicast transmission is dominated by the worst case (e.g., the randomly scheduled SU fails to decode x_p and x_s), and thus, it determines the diversity performance. However, the use of cooperative relaying can improve the outage performance when the randomly scheduled SU can fully decode x_p and x_s.

Furthermore, it is also observed that the system diversity order of the overall networks is only one, e.g., the worst case dominates the mutual outage probability. Therefore, from a system-level perspective, the transmission design for both primary and secondary networks should be jointly considered, in order to achieve the optimal performance.

Remark 2: Note that it is realistic to obtain the second order statistics of wireless channels without consuming much communication overhead. Such information can be determined by the distance between the transceivers and changes slowly compared to the instantaneous channel gains. In this context, intuition suggests that the optimal secondary user scheduling strategy for the overall network performance is to select the SU closest to the BS, which can be confirmed in Corollary 1.

Corollary 1: When the second order statistics of channel gains are available for user scheduling, selecting a SU that is closest to the BS minimizes the outage probability for both primary and secondary networks in the high SNR regime.

Proof: We refer to the high SNR asymptotic analysis to provide further insights, and then the proof can be completed by demonstrating that the outage probability is an increasing function of the distance. Suppose SU n is selected, a high SNR approximation for the outage probability of primary networks is given by (20). It is easy to verify that the derivatives of

$\mathcal{P}_p$ in terms of $d_{b,n}$ and $d_{n,p}$ are positive, e.g., $\frac{\partial \mathcal{P}_p}{\partial d_{b,n}} > 0$ and $\frac{\partial \mathcal{P}_p}{\partial d_{n,p}} > 0$, which indicates that the outage probability of primary networks is an increasing function in $d_{b,n}$ and $d_{n,p}$. Similarly, we can obtain that the outage probability of secondary networks is an increasing function in $d_{b,n}$ and $d_{n,n'}$. Therefore, choosing the SU closest to the BS, e.g., the smallest $d_{b,n}$, the outage probability for both primary and secondary networks can be minimized. This completes the proof. ■

It is important to point out that the use of the second order statistics of wireless channels for secondary user scheduling can improve the outage performance compared with the round robin scheduling. Surprisingly, the knowledge of average CSI is not helpful for improving the system diversity, since the diversity order remains unchanged with the round robin scheduling. This necessitates the need for designing user scheduling strategies aimed at diversity enhancement.

B. User Scheduling with Partial CSI

For many practical cooperative CR scenarios, it is common that the primary network is able to be aware of the number and identity of the SUs available in the band of interest. Often the CSI between the primary and secondary networks, e.g., $h_{b,n}$ and $h_{n,p}$, $n \in \{1, \cdots, N\}$, can be obtained via channel estimation [18]–[20]. Under such partial CSI, an opportunistically scheduled SU $n^\star$ can be determined by using the following criterion

$$n^\star = \arg \max_{n \in \mathcal{S}_2 \cup \mathcal{S}_1} \{|h_{n,p}|^2\}. \tag{25}$$

Note that whether applying NOMA or SDMA transmission for the cooperative MCR-NOMA scheme depends on the existence of $\mathcal{S}_2$ or $\mathcal{S}_1$, as aforementioned. The detailed operation process for this secondary user scheduling with partial CSI is introduced in **Algorithm 2**.[6]

Algorithm 2 User scheduling strategy with partial CSI

1: The BS broadcasts a training signal to the PU and multicast SUs. Based on the training symbol, each SU recovers $h_{b,n}$ and the PU estimates $h_{b,p}$;
2: The PU broadcasts training signal to the SUs, and each SU acquires $h_{n,p}$;
3: Each SU calculates (6) and (7) based on its local CSI, such that subsets $\mathcal{S}_1$ and $\mathcal{S}_2$ can be determined;
4: First, if $\mathcal{S}_2$ exists, each SU n ($n \in \mathcal{S}_2$) starts a virtual timer initiated by $t_n = t_0 \exp(-|h_{n,p}|^2)$, where t_0 is a constant. The timer of SU $n^\star$ with the best channel condition to the PU will expire first. Then, SU $n^\star$ broadcasts a flag message signaling its presence and other SUs back off;
5: Otherwise, if $\mathcal{S}_2$ does not exist while $\mathcal{S}_1$ exists, repeat procedure 4 until SU $n^\star$ can be determined;
6: Finally, the cooperative MCR-NOMA begins with the selected SU $n^\star$.

[6]Note that the forward and backward channels between the PU and SUs are reciprocal, because the transmissions of the training signals occur on the same frequency and coherence time.

It is worth pointing out that the opportunistic scheduling of the "best" SU $n^\star$ involves the discovery of the most appropriate SU in a distributed and "quick" fashion, well before the channels change again, and thus the execution time of the user scheduling is negligible.[7]

1) Outage Probability: With the proposed user scheduling strategy, the outage probability for the primary network can be formulated as[8]

$$\begin{aligned}\mathcal{P}_{p^\star} &= \mathcal{P}(|\mathcal{S}_1| = 0)\mathcal{P}(\mathrm{SINR}_p < \tau_p) \\ &+ \sum_{k_1=1}^{N} \sum_{\substack{\mathcal{S}_1 \subseteq \{1,\cdots,N\} \\ |\mathcal{S}_1| = k_1}} \mathcal{P}(|\mathcal{S}_1| = k_1)\mathcal{P}(\mathrm{SINR}_{\tilde{n},p} < \tau_p) \\ &+ \sum_{k_2=1}^{N} \sum_{\substack{\mathcal{S}_2 \subseteq \{1,\cdots,N\} \\ |\mathcal{S}_2| = k_2}} \mathcal{P}(|\mathcal{S}_2| = k_2)\mathcal{P}(\mathrm{SINR}_{n,p} < \tau_p).\end{aligned} \tag{26}$$

From (26), the first probability corresponds to the event that neither the PU nor SUs can decode x_p correctly. The second probability represents the event that $\mathcal{S}_2$ is empty while $\mathcal{S}_1$ exists, the combined SINR is smaller than the targeted threshold. The third probability is for the event that $\mathcal{S}_2$ is non-empty, but the combined SINR is below the targeted threshold.

As for the secondary network, if $|\mathcal{S}_2| = 0$, an outage event will be always declared. Therefore, the outage probability given the use of SU $n^\star$ can be expressed as

$$\begin{aligned}\mathcal{P}_{s^\star} &= \mathcal{P}(|\mathcal{S}_2| = 0) \\ &+ \sum_{k_2=1}^{N-1} \sum_{\substack{\mathcal{S}_2 \subseteq \{1,\cdots,N\} \\ |\mathcal{S}_2| = k_2}} \mathcal{P}(|\mathcal{S}_2| = k_2)\ \mathcal{P}_{out,s^\star}(|\mathcal{S}_2| = k_2).\end{aligned} \tag{27}$$

The closed-form expressions for the outage probability are given in the following theorem.

Theorem 2: For i.n.i.d. channels, the outage probability for the primary network achieved by the user scheduling under partial CSI can be approximated as (28), shown at the top of the next page. Next, the outage probability for the secondary network can be computed by (29), illustrated at the top of the next page.

Proof: See Appendix B for details. ■

2) Mutual Outage Probability: Under the partial CSI scenario, the mutual outage probability for the primary and secondary networks can be formulated as

$$\mathcal{P}_{mop^\star} = \mathcal{P}(\mathcal{O}_1^\star) + \mathcal{P}(\mathcal{O}_2^\star) \tag{30}$$

where $\mathcal{P}(\mathcal{O}_1^\star)$ is for the event that no SU can decode x_p and x_s correctly, and $\mathcal{P}(\mathcal{O}_2^\star)$ is for the event that the scheduled SU

[7]In the case of two or more timers expire simultaneously, all of the collided SUs back off and randomly generate their own timers. The timer that expires first will be designated the selected SU.

[8]It is noteworthy that under partial CSI, there is no handshaking among the multicast SUs, so that the scheduled SU $n^\star$ always retransmits the signal mixture $(\alpha_p x_p + \alpha_s x_s)$ conditioned on $\mathcal{S}_2$ exists during the second time slot. Therefore, the reception reliability for the primary and secondary networks can be improved simultaneously.

$$\mathcal{P}_{p^\star} \approx \left(1-e^{-\frac{\psi_p d_{b,p}^\eta}{\rho}}\right)\prod_{n=1}^{N}\left(1-e^{-\frac{\psi_p d_{b,n}^\eta}{\rho}}\right)+\sum_{k_1=1}^{N}\sum_{\substack{\mathcal{S}_1\subseteq\{1,\cdots,N\}\\|\mathcal{S}_1|=k_1}}\chi\prod_{i=1}^{k_1}\left(e^{-\frac{\psi_p d_{b,n_i}^\eta}{\rho}}-e^{-\frac{\psi_s d_{b,n_i}^\eta}{\rho}}\right)\prod_{n\in\mathcal{S}_1^c}\left(1-e^{-\frac{\psi_p d_{b,n}^\eta}{\rho}}\right) \times\prod_{i=1}^{k_1}\left(1-e^{-\frac{(\tau_p-a_l)d_{n_i,p}^\eta}{\rho}}\right)+\sum_{k_2=1}^{N}\sum_{\substack{\mathcal{S}_2\subseteq\{1,\cdots,N\}\\|\mathcal{S}_2|=k_2}}\chi e^{-\sum_{j=1}^{k_2}\frac{\psi_s d_{b,n_j}^\eta}{\rho}}\prod_{n\in\mathcal{S}_2^c}\left(1-e^{-\frac{\psi_s d_{b,n}^\eta}{\rho}}\right)\prod_{j=1}^{k_2}\left(1-e^{-\frac{(\tau_p-a_l)d_{n_j,p}^\eta}{\rho(\alpha_p^2-\alpha_s^2(\tau_p-a_l))}}\right). \tag{28}$$

$$\mathcal{P}_{s^\star} = \prod_{n=1}^{N}\left(1-e^{-\frac{\psi_s d_{b,n}^\eta}{\rho}}\right)+\sum_{k_2=1}^{N-1}\sum_{\substack{\mathcal{S}_2\subseteq\{1,\cdots,N\}\\|\mathcal{S}_2|=k_2}}e^{-\sum_{j=1}^{k_2}\frac{\psi_s d_{b,n_j}^\eta}{\rho}}\sum_{j=1}^{k_2}\left(1+\sum_{i=1}^{k_2-1}\sum_{\substack{A_i=\{1,\cdots,j-1,j+1,\cdots,k_2\}\\|A_i|=i}}\frac{(-1)^i d_{n_j,p}^\eta}{d_{n_j,p}^\eta+\sum_{i\in A_i}d_{n_i,p}^\eta}\right) \times\left(1-e^{-\sum_{n\in\mathcal{S}_2^c}\frac{\psi_s d_{n_j,n}^\eta}{\rho}}\right)\prod_{n\in\mathcal{S}_2^c}\left(1-e^{-\frac{\psi_s d_{b,n}^\eta}{\rho}}\right). \tag{29}$$

$n^\star$ causes an outage either at the PU or SU n'. As a result, we can further express $\mathcal{P}_{mop^\star}$ as

$$\mathcal{P}_{mop^\star} = \mathcal{P}(|\mathcal{S}_2|=0)+\sum_{k_2=1}^{N-1}\sum_{\substack{\mathcal{S}_2\subseteq\{1,\cdots,N\}\\|\mathcal{S}_2|=k_2}}\mathcal{P}(|\mathcal{S}_2|=k_2) \times\Big(\mathcal{P}_{out,s^\star}(|\mathcal{S}_2|=k_2)+\mathcal{P}(\text{SINR}_{n,p}<\tau_p)\Big) +\mathcal{P}(|\mathcal{S}_2|=N)\mathcal{P}(\text{SINR}_{\tilde{n},p}<\tau_p). \tag{31}$$

Using the analytical results derived in Appendix B, one can readily obtain $\mathcal{P}_{mop^\star}$.

3) Diversity Analysis: For further insights, we provide diversity analysis for both the primary and secondary networks under the partial CSI scenario. In the following, Lemma 2 characterizes the asymptotic outage performance achieved by the proposed user scheduling.

Lemma 2: For the primary network, based on the analytical results, we derive the high SNR approximation for $\mathcal{P}_{p^\star}$ as

$$\mathcal{P}_{p^\star} \simeq \frac{1}{\rho^{N+1}}\left(\psi_p d_{b,p}^\eta\prod_{n=1}^{N}\psi_p d_{b,n}^\eta+\sum_{k_2=1}^{N}\sum_{\substack{\mathcal{S}_2\subseteq\{1,\cdots,N\}\\|\mathcal{S}_2|=k_2}} \times\prod_{n\in\mathcal{S}_2^c}\psi_s\chi' d_{b,n}^\eta\prod_{j=1}^{k_2}\frac{(\tau_p-a_l)d_{n_j,p}^\eta}{\left(\alpha_p^2-\alpha_s^2(\tau_p-a_l)\right)}\right). \tag{32}$$

For the secondary network, its asymptotic behavior follows

$$\mathcal{P}_{s^\star} \simeq \prod_{n=1}^{N}\frac{\psi_s d_{b,n}^\eta}{\rho}+\sum_{k_2=1}^{N-1}\sum_{\substack{\mathcal{S}_2\subseteq\{1,\cdots,N\}\\|\mathcal{S}_2|=k_2}}\prod_{n\in\mathcal{S}_2^c}\frac{\psi_s d_{b,n}^\eta}{\rho} \times\mathcal{P}(n^\star=n_j)\sum_{n\in\mathcal{S}_2^c}\frac{\psi_s d_{n_j,n}^\eta}{\rho} \tag{33}$$

where $\mathcal{P}(n^\star=n_j)$ is given by (68) in Appendix B. Based on the above equations, the diversity order can be derived as

$$D_{p^\star}=N+1,\quad D_{s^\star}=2,\quad D_{sys^\star}=2. \tag{34}$$

Proof: By applying the same rationale with (20) and (22), we can obtain the asymptotic outage probability straightforwardly. From (32), it is obvious that $\mathcal{P}_{p^\star}$ decays with ρ by a slope of $(N+1)$, so that we have $D_{p^\star}=N+1$. Knowing that the higher order terms of $\frac{1}{\rho}$ can be omitted when the transmit SNR is sufficiently large, we can obtain that the second summation in (33) has the term of $\frac{1}{\rho^2}$ as $k_2=N-1$ (since in this case, the dominant factor will be the term with $|\mathcal{S}_2^c|=1$). As a result, the remaining summations in (33), e.g., $\frac{1}{\rho^m}$, $m=3,\cdots,N$, can be ignored. Therefore, a diversity order of two is achieved. ■

Remark 3: It can be observed from (34), full diversity order (e.g., a diversity order of $(N+1)$ in the presence of $(N+1)$ individual fading paths) of the primary network is achieved by the proposed user scheduling strategy. Different from the round robin scheduling strategy, a diversity order of two is achieved at the secondary network. Such phenomenon is due to the fact that, the scheduled SU under partial CSI will always retransmit the signal mixture $(\alpha_p x_p+\alpha_s x_s)$ to the PU and other SUs (e.g., when $\mathcal{S}_2$ is a non-empty set), while the scheduled SU by the round robin strategy may be in outage (e.g., the worst case). As a result, the SUs that belong to $\mathcal{S}_2^c$ will receive the signal x_s twice, and only when these two signal copies failed in decoding results in an outage, and therefore achieving a diversity order of two at the secondary network. Therefore, compared to the round robin scheduling strategy, substantial outage performance can be achieved for both networks with a small amount of communications overhead.

Although the outage performance of the primary network is maximized, the system diversity order of the overall networks is still dominated by the secondary network, e.g., only a diversity order of two can be achieved, as indicated by (34). Therefore, to improve the overall outage performance, it is imperative to design the optimal secondary user scheduling strategy for both primary and secondary networks.

C. User Scheduling with Full CSI

In the CR networks with many SUs, it is more likely that the SUs will obtain information about the other SUs in their vicinity. Also, it is easy for SUs that are in the same vicinity to cooperate, since cooperating transmitters that are close to each other can exchange messages without transmitting significant power and creating much interference to the rest of the networks [16]. When full CSI is available for user scheduling, e.g., each SU knows not only the instantaneous CSI for the relaying link to the PU, but the CSI for the links between itself and other SUs as well [23]–[25]. An example of the scenario is the cellular networks, where the SUs are the device-to-device (D2D) multicast user equipments and the PU is the cellular user. Note that when cellular network coverage is available, the PU's and SUs' transmissions are under relatively tight network control, and thus the instantaneous CSI can be obtained by a simple manner [23]. In this case, the optimal user scheduling strategy is to jointly take the performance of both primary and secondary networks into consideration. To this end, we propose a new two-stage user scheduling strategy, e.g., the first stage is to ensure the successful detection of the PU, and the second stage is to maximize the worst-case secondary multicasting.

First, we construct a new subset $\mathcal{S}_3$ containing the SUs that aim to guarantee the successful detection of the PU, shown as

$$\mathcal{S}_3 = \Big\{n : 1 \le n \le N, \mathrm{SINR}_{n,p} \ge \tau_p, \\ \mathrm{SNR}_{n,s} \ge \tau_s, \mathrm{SINR}_{p,end} \ge \tau_p\Big\}. \quad (35)$$

According to (35), it is straightforward that $\mathcal{S}_3 \subseteq \mathcal{S}_2$.

Second, among the SUs in $\mathcal{S}_3$, the SU whose worst relaying gain is the strongest is selected to maximize the capacity of the multicast transmission, given by

$$n^* = \arg\max_{n\in\mathcal{S}_3} \big\{ \min_{n'\in\mathcal{S}_2^c} |h_{n,n'}|^2 \big\}. \quad (36)$$

The detailed operation process for the proposed user scheduling strategy proceeds in **Algorithm 3**.

Algorithm 3 User scheduling strategy with full CSI

1: The BS broadcasts a training signal to the PU and multicast SUs. Based on the training symbol, each SU recovers $h_{b,n}$ and the PU estimates $h_{b,p}$;
2: The PU broadcasts training signal containing a quantized version of $h_{b,p}$ to the SUs, and thus each SU acquires $h_{n,p}$ and $h_{b,p}$;
3: Each SU calculates (6), (7) and (35) based on its local CSI, so that subsets $\mathcal{S}_1$, $\mathcal{S}_2$ and $\mathcal{S}_3$ are determined;
4: Then, SU m ($m \in \{1, \cdots, N\}$) broadcasts a flag message "*Success*" or "*Failure*" in a sequential manner, according to whether it belongs to $\mathcal{S}_2$ or not;
5: If the flag message is "*Failure*", SU n ($n \in \mathcal{S}_2$) measures the CSI between SU m and itself, e.g., $h_{n,m}$, by receiving the flag message from SU m. As thus, the inter-user CSIs for the multicast SUs are obtained;
6: Each SU in $\mathcal{S}_3$ computes its worst relaying channel gain according to $T_n = \min_{n'\in\mathcal{S}_2^c} |h_{n,n'}|^2$;
7: By setting $t'_n = t_0 \exp(-T_n)$, a virtual timer process is started. SU n^* will expire first and other SUs in $\mathcal{S}_3$ will back off;
8: Finally, the cooperative MCR-NOMA begins with the selected SU n^*.

1) Outage Probability: Thanks to the handshaking among the SUs, the scheduled SU is able to know other SUs' decoding results to adapt its cooperative transmission strategy. Therefore, the outage probability for the primary network can be interpreted as

$$\mathcal{P}_{p^*} = \mathcal{P}(|\mathcal{S}_1| = 0)\mathcal{P}(\mathrm{SINR}_p < \tau_p) \\ + \sum_{k_1=1}^{N} \sum_{\substack{\mathcal{S}_1\subseteq\{1,\cdots,N\}\\ |\mathcal{S}_1|=k_1}} \mathcal{P}(|\mathcal{S}_1| = k_1)\mathcal{P}(\mathrm{SINR}_{\tilde{n},p} < \tau_p) \\ + \sum_{k_2=1}^{N-1} \sum_{\substack{\mathcal{S}_2\subseteq\{1,\cdots,N\}\\ |\mathcal{S}_2|=k_2}} \mathcal{P}(|\mathcal{S}_2| = k_2)\mathcal{P}(\mathrm{SINR}_{n,p} < \tau_p) \\ + \mathcal{P}(|\mathcal{S}_2| = N)\mathcal{P}(\mathrm{SINR}_{\hat{n},p} < \tau_p). \quad (37)$$

From (37), the former two probabilities are for the event that $\mathcal{S}_2$ is a null set, and the latter two probabilities are for the event that the SU in $\mathcal{S}_2$ which has the strongest relaying channel gain to cannot support the targeted data rate, e.g., $\mathcal{S}_3$ is an empty set.

The outage probability for the secondary network with the proposed user scheduling strategy can be formulated as

$$\mathcal{P}_{s^*} = \mathcal{P}(|\mathcal{S}_2| = 0) + \sum_{k_2=1}^{N-1} \sum_{k_3=0}^{k_2} \sum_{\substack{\mathcal{S}_2\subseteq\{1,\cdots,N\}\\ |\mathcal{S}_2|=k_2}} \sum_{\substack{\mathcal{S}_3\subseteq\mathcal{S}_2\\ |\mathcal{S}_3|=k_3}} \quad (38) \\ \mathcal{P}(|\mathcal{S}_2| = k_2, |\mathcal{S}_3| = k_3)\mathcal{P}_{out,s^*}(|\mathcal{S}_2| = k_2, |\mathcal{S}_3| = k_3).$$

The following theorem provides the closed-form expressions for the outage probability.

Theorem 3: With full CSI available, the outage probability for the primary network can be approximated as (39), shown at the top of the next page. In addition, the outage probability for the secondary network can be computed by (40), shown at the top of the next page.

Proof: See Appendix C for details. ■

2) Mutual Outage Probability: Under the full CSI scenario, the mutual outage probability achieved by the proposed user scheduling strategy is given by

$$\mathcal{P}_{mop^*} = \sum_{k_2=1}^{N-1} \sum_{\substack{\mathcal{S}_2\subseteq\{1,\cdots,N\}\\ |\mathcal{S}_2|=k_2}} \Big(\mathcal{P}(|\mathcal{S}_2| = k_2, |\mathcal{S}_3| = 0) \\ + \sum_{k_3=0}^{k_2} \sum_{\substack{\mathcal{S}_3\subseteq\mathcal{S}_2\\ |\mathcal{S}_3|=k_3}} \mathcal{P}(|\mathcal{S}_2| = k_2, |\mathcal{S}_3| = k_3) \\ \times \mathcal{P}_{out,s^*}(|\mathcal{S}_2| = k_2, |\mathcal{S}_3| = k_3)\Big) \\ + \mathcal{P}(|\mathcal{S}_2| = N, |\mathcal{S}_3| = 0) + \mathcal{P}(|\mathcal{S}_2| = 0). \quad (41)$$

Relying on the results derived in Appendix C, $\mathcal{P}_{mop^*}$ can be obtained straightforwardly.

$$\mathcal{P}_{p^*} \approx \left(1-e^{-\frac{\psi_p d_{b,p}^{\eta}}{\rho}}\right)\prod_{n=1}^{N}\left(1-e^{-\frac{\psi_p d_{b,n}^{\eta}}{\rho}}\right)+\sum_{k_1=1}^{N}\sum_{\substack{\mathcal{S}_1\subseteq\{1,\cdots,N\}\\|\mathcal{S}_1|=k_1}}\chi\prod_{i=1}^{k_1}\left(e^{-\frac{\psi_p d_{b,n_i}^{\eta}}{\rho}}-e^{-\frac{\psi_s d_{b,n_i}^{\eta}}{\rho}}\right)\prod_{n\in\mathcal{S}_1^c}\left(1-e^{-\frac{\psi_p d_{b,n}^{\eta}}{\rho}}\right)$$
$$\times\prod_{i=1}^{k_1}\left(1-e^{-\frac{(\tau_p-a_l)d_{n_i,p}^{\eta}}{\rho}}\right)+\sum_{k_2=1}^{N-1}\sum_{\substack{\mathcal{S}_2\subseteq\{1,\cdots,N\}\\|\mathcal{S}_2|=k_2}}\chi e^{-\sum_{j=1}^{k_2}\frac{\psi_s d_{b,n_j}^{\eta}}{\rho}}\prod_{n\in\mathcal{S}_2^c}\left(1-e^{-\frac{\psi_s d_{b,n}^{\eta}}{\rho}}\right)\prod_{n=1}^{N}\left(1-e^{-\frac{(\tau_p-a_l)d_{n,p}^{\eta}}{\rho}}\right)$$
$$\times\prod_{j=1}^{k_2}\left(1-e^{-\frac{(\tau_p-a_l)d_{n_j,p}^{\eta}}{\rho(\alpha_p^2-\alpha_s^2(\tau_p-a_l))}}\right)+\chi e^{-\sum_{n=1}^{N}\frac{\psi_s d_{b,n}^{\eta}}{\rho}}. \quad (39)$$

$$\mathcal{P}_{s^*} \approx \prod_{n=1}^{N}\left(1-e^{-\frac{\psi_s d_{b,n}^{\eta}}{\rho}}\right)+\sum_{k_2=1}^{N-1}\sum_{k_3=0}^{k_2}\sum_{\substack{\mathcal{S}_2\subseteq\{1,\cdots,N\}\\|\mathcal{S}_2|=k_2}}\sum_{\substack{\mathcal{S}_3\subseteq\mathcal{S}_2\\|\mathcal{S}_3|=k_3}}e^{-\sum_{j=1}^{k_2}\frac{\psi_s d_{b,n_j}^{\eta}}{\rho}}\prod_{n\in\mathcal{S}_2^c}\left(1-e^{-\frac{\psi_s d_{b,n}^{\eta}}{\rho}}\right)$$
$$\times\chi^{k_3}e^{-\sum_{r=1}^{k_3}\frac{(\tau_p-a_l)d_{n_r,p}^{\eta}}{\rho(\alpha_p^2-\alpha_s^2(\tau_p-a_l))}}\prod_{t=k_3+1}^{k_2-k_3}\chi\left(1-e^{-\frac{(\tau_p-a_l)d_{n_t,p}^{\eta}}{\rho(\alpha_p^2-\alpha_s^2(\tau_p-a_l))}}\right)\prod_{r=1}^{k_3}\left(1-e^{-\sum_{n\in\mathcal{S}_2^c}\frac{\psi_s d_{n_r,n}^{\eta}}{\rho}}\right). \quad (40)$$

Corollary 2: Under the full CSI scenario, the proposed user scheduling strategy is outage-optimal for the overall networks.

Proof: We prove this corollary using contradiction. Suppose that there exists another SU $n^\dagger$ ($n^\dagger \neq n^*$), which can achieve better outage performance than the use of SU n^*. Obviously, both SU $n^\dagger$ and SU n^* belong to $\mathcal{S}_3$, otherwise, an outage will be always declared at the primary network. However, according to the definition in (36), the selected SU n^* guarantees the worst-case outage performance of the secondary network, e.g., there exists no outage when the received SINR at the SU whose relaying channel is the worst is larger than the targeted threshold. Therefore, we have $n^\dagger = n^*$, which contradicts to the initial assumption that $n^\dagger \neq n^*$. As a result, we can conclude that SU n^* is the optimal solution under the full CSI scenario. This completes the proof. ∎

3) Diversity Analysis: In order to obtain further insights, the following lemma characterizes the asymptotic outage performance achieved by the secondary user scheduling under full CSI.

Lemma 3: Based on the analytical results given by (39) and (40), the high SNR approximation of the outage probability for the primary network can be computed by

$$\mathcal{P}_{p^*} \simeq \frac{1}{\rho^{N+1}}\Bigg(\psi_p d_{b,p}^{\eta}\prod_{n=1}^{N}\psi_p d_{b,n}^{\eta}+\sum_{k_2=1}^{N-1}\sum_{\substack{\mathcal{S}_2\subseteq\{1,\cdots,N\}\\|\mathcal{S}_2|=k_2}}\prod_{n\in\mathcal{S}_2^c}\psi_s\chi' d_{b,n}^{\eta}\prod_{j=1}^{k_2}\frac{(\tau_p-a_l)d_{n_j,p}^{\eta}}{\alpha_p^2-\alpha_s^2(\tau_p-a_l)}+\chi'\prod_{n=1}^{N}(\tau_p-a_l)d_{n,p}^{\eta}\Bigg). \quad (42)$$

For the secondary network, the outage probability asymptotically goes by

$$\mathcal{P}_{s^*} \simeq \prod_{n=1}^{N}\frac{\psi_s d_{b,n}^{\eta}}{\rho}+\sum_{k_2=1}^{N-1}\sum_{k_3=0}^{k_2}\sum_{\substack{\mathcal{S}_2\subseteq\{1,\cdots,N\}\\|\mathcal{S}_2|=k_2}}\sum_{\substack{\mathcal{S}_3\subseteq\mathcal{S}_2\\|\mathcal{S}_3|=k_3}}\left(\frac{\chi'}{\rho}\right)^{k_2}\times\prod_{n\in\mathcal{S}_2^c}\frac{\psi_s d_{b,n}^{\eta}}{\rho}\prod_{t=k_3+1}^{k_2-k_3}\frac{(\tau_p-a_l)d_{n_t,p}^{\eta}}{\rho\big(\alpha_p^2-\alpha_s^2(\tau_p-a_l)\big)}\times\prod_{r=1}^{k_3}\Bigg(\sum_{n\in\mathcal{S}_2^c}\frac{\psi_s d_{n_r,n}^{\eta}}{\rho}\Bigg). \quad (43)$$

Therefore, the diversity order can be readily found as

$$D_{p^*}=N+1,\quad D_{s^*}=N,\quad D_{sys^*}=N. \quad (44)$$

Proof: This can be proved similarly with Lemma 2. ∎

Remark 4: From Lemma 3, it is concluded that the proposed user scheduling strategy provides the full diversity order for both primary and secondary networks. In other words, by increasing the number of SUs, the outage probability for the overall networks reduces significantly. Therefore, under the full CSI available, the proposed user scheduling strategy fully exploiting the freedom offered by the cooperative MCR-NOMA scheme (e.g., maximizing the spatial diversity order), and it can be directly applied to the practical large-scale multicast CR networks.

Furthermore, a diversity order of N, e.g., the maximal spatial diversity gain under the full CSI scenario, can be achieved for the overall networks. Consequently, in order to maximize the mutual cooperation benefit, it is advisable to acquire the full CSI since it can achieve the most robust transmission for the considered cooperative MCR-NOMA scenario.

IV. Numerical Studies

To illustrate the performance of the secondary user scheduling strategies for the proposed cooperative MCR-NOMA

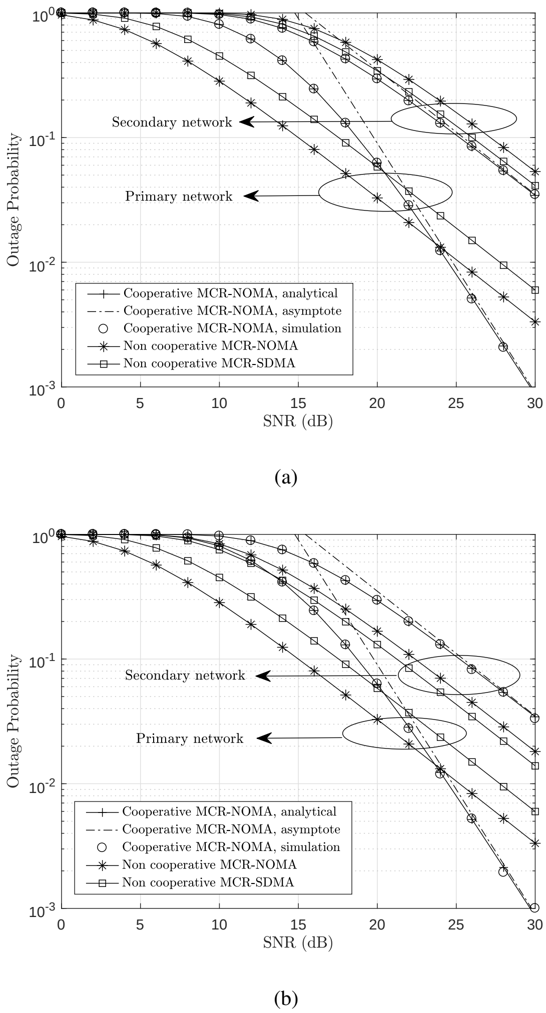


Fig. 2. Outage probability versus SNR with the round robin user scheduling strategy, $R_p = 1$ bits/Hz, $R_s = 1.5$ bits/Hz. (a) $N = 6$; (b) $N = 2$.

scheme, we provide numerical examples and evaluate the analytical results. Without loss of generality, we assume that the BS locates at the origin $(0, 0)$, the PU locates at $(0, 10)$, and all the SUs are randomly distributed inside a circle centered at $(0, 5)$ and with radius of 3. The path loss exponent is set to $\eta = 2.7$. Unless mentioned otherwise, the power allocation coefficients for the cooperative MCR-NOMA scheme are set as $\alpha_p^2 = 0.8$ and $\alpha_s^2 = 0.2$, respectively, and the Gaussian-Chebyshev parameter is chosen as $L = 10$.

For illustration purpose, the performance of the non cooperative MCR-NOMA/SDMA schemes is also plotted. Specifically, the non cooperative MCR-NOMA/SDMA schemes can be revised as follows. For the non cooperative MCR-NOMA scheme, the BS serves the PU and multicast SUs simultaneously via direct NOMA transmissions. While for the non cooperative MCR-SDMA scheme, orthogonal time or frequency resources are allocated to the PU and multicast SUs, so as to create parallel spatial pipes for the primary and secondary transmissions.

In Fig. 2, the outage performance of the cooperative MCR-NOMA achieved by the round robin user scheduling is illustrated. It can be observed from Fig. 2(a) that the derived analytical results match perfectly the simulations, and the asymptotic curves always converge to the analytical one fast in the high SNR regions. As a benchmark, the performance of the non cooperative MCR-NOMA and non cooperative MCR-SDMA schemes is also presented. Several observations can be drawn as follows: 1) For the primary network, the proposed cooperative MCR-NOMA achieved by the round robin scheduling strategy gives better outage performance than its non cooperative MCR-SDMA when the SNR is larger than 20 dB, e.g., a tradeoff exists between the two schemes, and the outage probability gap becomes even larger when SNR increases. This phenomenon can be explained by using the simple fact that, full transmit power P is applied for the transmission of signal x_p by the non cooperative MCR-SDMA, while only a fraction of P (e.g., $\alpha_p^2 = 0.8$) for signal x_p is applied by the cooperative MCR-NOMA, therefore yielding larger received SNR in the low to medium SNR regions. On the other hand, in the absence of CSI acquisition/exchange, the round robin scheduling can still ensure a diversity order of two at the primary network, higher than its non cooperative counterpart (e.g., a diversity order of one). Therefore, lower outage probability can be achieved by the cooperative MCR-NOMA at high SNRs due to the increased diversity gains. 2) The proposed cooperative MCR-NOMA outperforms the non cooperative MCR-NOMA when the SNR is larger than 24 dB, which can be understood by the fact that the $1/2$ spectral efficiency loss caused by the two time-slot cooperation dominates the outage performance in the low to medium SNR regions. 3) The non cooperative MCR-NOMA always behaves superior outage performance than the non cooperative MCR-SDMA, since the use of NOMA allows the PU and SUs to be served simultaneously, without consuming additional time or frequency resource. As a result, the overall spectrum utilization can be significantly improved. Similar performance gains in terms of the outage probability can also be seen from Fig. 2(b).

Then attention is shifted to the performance of the secondary network. As shown in Fig. 2(a), interestingly, note that when $N = 6$ the outage performance achieved by a randomly scheduled SU obtains 3 dB SNR gain over the non cooperative MCR-NOMA, while the situation is opposite when $N = 2$, as is observed from Fig. 2(b). This is because a large value of N means that the probability that not all the SUs can be eventually successful is larger. However, when N is small, e.g., $N = 2$, the randomly scheduled SU is more likely to be the worst-case SU which is in outage, thus leading to further outage events for the cooperative transmission. Moreover, similar outage performance for the comparison between the cooperative MCR-NOMA and non cooperative MCR-SDMA can also be observed from Fig. 2(b). Due to the fact that for the non cooperative MCR-SDMA, when more SUs participate in the communication, the probability of the existence of the worst-case SU becomes larger, which increases the outage probability for the secondary multicasting. In addition, the slope of the outage curves are the same for the cooperative and non cooperative MCR-NOMA/SDMA schemes, which confirms Lemma 1.

Next, we assess the impact of the proposed user scheduling

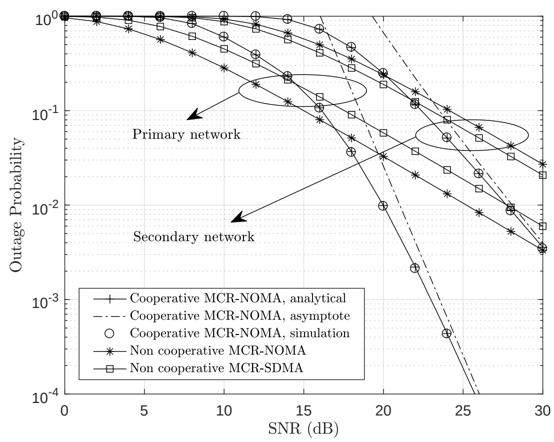


Fig. 3. Outage probability achieved by the proposed strategy with partial CSI, $R_p = 1$ bits/Hz, $R_s = 1.5$ bits/Hz, $N = 3$.

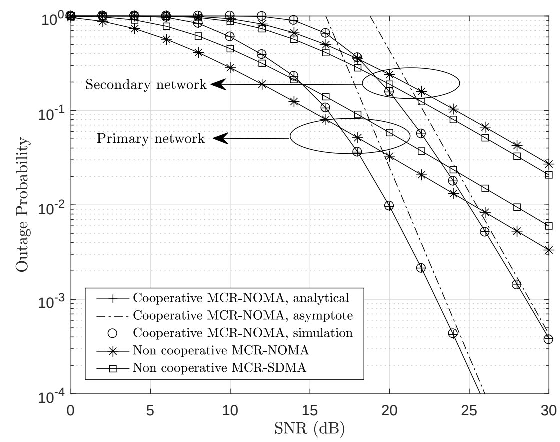


Fig. 5. Outage probability achieved by the proposed strategy with full CSI, $R_p = 1$ bits/Hz, $R_s = 1.5$ bits/Hz, $N = 3$.

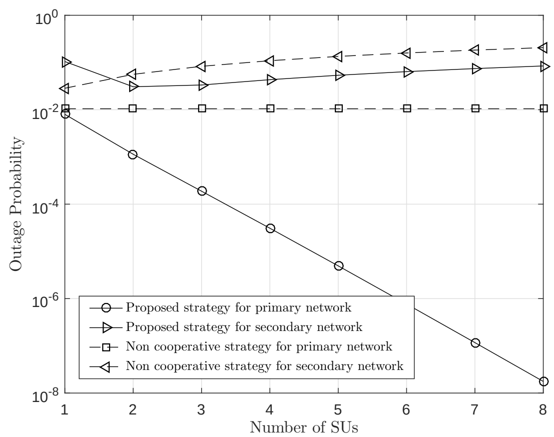


Fig. 4. Outage probability versus number of SUs N with partial CSI, $R_p = 1$ bits/Hz, $R_s = 1.5$ bits/Hz, $\rho = 25$ dB.

strategy using partial CSI on the outage performance. In Fig. 3, we first look at the case for the primary network. As illustrated from the figure, the slope of the outage curve is much steeper than that achieved by the round robin scheduling, because a higher diversity order (e.g., full diversity order) is achieved by the proposed strategy, as indicated by Lemma 2. Notably, in the medium to high SNR regions (e.g., 16–30 dB), the proposed strategy achieves a sufficient lower outage probability compared with the non cooperative MCR-NOMA/SDMA. On the other hand, as a convenient by-product of the NOMA signaling at the scheduled SU, a diversity order of two can be achieved at the secondary network, as shown in Fig. 3. Furthermore, in contrast to the phenomenon in Fig. 2, the outage performance of the secondary network achieved by the proposed strategy always outperforms the non cooperative MCR-NOMA when SNR is larger than 20 dB.

Fig. 4 shows the outage probability as a function of N for different MCR-NOMA schemes. One important observation is that the outage probability for the primary network reduces significantly with an increase in the number of SUs, e.g., the value of N, due to its monotonically decreasing property in N. It is also worth noting that the outage performance for the secondary network meliorates when N varies from 1 to 2, because of the enhanced diversity order. However, it becomes worse with further increase in N, a similar phenomenon can be observed from the non cooperative scheme for the secondary network. This can be explained using the simple fact that with a large number of SUs, the diversity order remains the same but the probability that not all the SUs are eventually decodable becomes larger. Therefore, there exists a tradeoff of the reception reliability between the primary and secondary networks under the partial CSI scenario, e.g., increasing the number of SUs is beneficial to the primary network but not to the secondary network.

In what follows, we investigate the performance achieved by the proposed strategy under full CSI scenario. Fig. 5 plots the outage probability versus SNR for the primary and secondary networks. One can observe that the proposed strategy under full CSI maximizes the spatial diversity gain for both networks, e.g., the achievable diversity orders for the primary and secondary networks are proportional to $N+1$ and N, respectively. Such an observation confirms the analytical results developed in Lemma 3. Another interesting observation is that the intersection between the cooperative and non cooperative strategies for the secondary network moves left, e.g., yielding nearly 3 dB SNR gain compared with the strategy under partial CSI. By further increasing N, such SNR gain becomes more evident, which can be validated by Lemma 3.

The impact of the number of SUs on the outage performance achieved by the proposed user scheduling strategy under full CSI is characterized by Fig. 6. As expected, with an increase in N, substantial outage performance improvement can be obtained at both the primary and secondary networks. Particularly, by fixing $\rho = 25$ dB, the proposed strategy always outperforms the non cooperative strategy for the primary network with any N and for the secondary network with $N \geq 2$, thus validating the advantage of the proposed cooperative MCR-NOMA under full CSI. With the proposed user scheduling strategy, it is suggested to recruit more SUs in the cooperation, and is preferable for the large-scale multicast

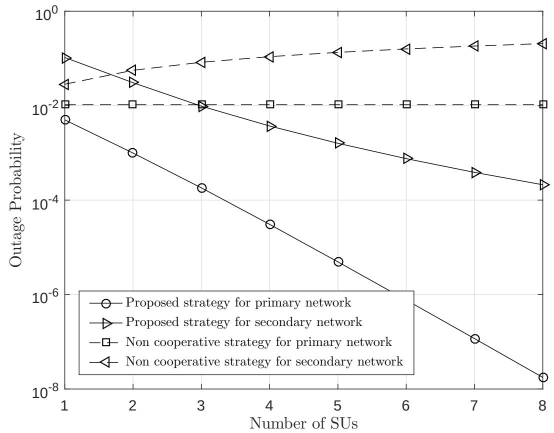


Fig. 6. Outage probability versus number of SUs N with full CSI, $R_p = 1$ bits/Hz, $R_s = 1.5$ bits/Hz, $\rho = 25$ dB.

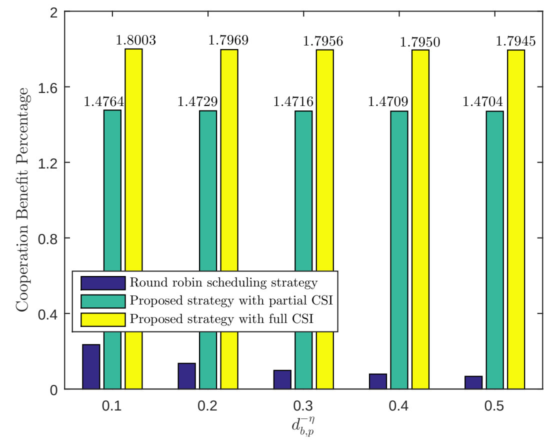


Fig. 8. Cooperation benefit percentage Λ as a function of $d_{b,p}^{-\eta}$ achieved by different secondary user scheduling strategies, $\rho = 25$ dB, $N = 3$.

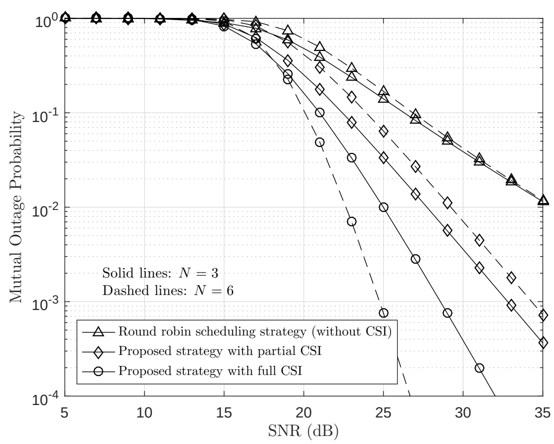


Fig. 7. Mutual outage probability achieved by different secondary user scheduling strategies, $R_p = 1$ bits/Hz, $R_s = 1.5$ bits/Hz.

CR networks.

Recall that the mutual outage probability is a good metric to judge the reliability of the PU-SU transmissions and evaluate whether the PU-SU collaboration can achieve a win-win situation or not. In Fig. 7, the mutual outage probability achieved by different user scheduling strategies is presented. It is clear from the figure that the mutual outage probability decreases as a function of SNR, which guarantees an acceptable level of outage for both networks. Furthermore, among the three strategies, the proposed strategy under full CSI achieves the lowest outage probability, since it ensures the maximal diversity gain offered by the PU-SU cooperation. From the figure, it also demonstrates that different from other two strategies, the performance of the proposed strategy under full CSI can be further improved with an increase in N, and thus it is the outage-optimal strategy for the considered cooperative MCR-NOMA as mentioned in Corollary 2.

Finally, in order to explicitly characterize the cooperation benefits of the proposed cooperative MCR-NOMA scheme, based on the analytical results in Theorems 1–3, we define the cooperation benefit percentage as the percentage reduction for the overall network. That is

$$\Lambda = \left(\frac{\mathcal{P}_{non,p} - \mathcal{P}_p}{\mathcal{P}_{non,p}} + \frac{\mathcal{P}_{non,s} - \mathcal{P}_s}{\mathcal{P}_{non,s}} \right) \quad (45)$$

where $\mathcal{P}_{non,p}$ and $\mathcal{P}_{non,s}$ are the outage probabilities of non cooperative direct transmission for the primary and secondary networks. This metric is presented by Fig. 8, without loss of generality and for the illustration purpose, independent and identically distributed fading channel is assumed. It is worth noting that the cooperation-based results outperform the non cooperative MCR-NOMA scheme, especially preferred by the primary network when $h_{b,p}$ is relatively weak ($d_{b,p}^{-\eta}$ is small) and the QoS requirement of the PU may not be satisfied by the primary network itself. On the other hand, there always exists a minimum benefit gain by the three secondary user scheduling strategies, and as expected, more available CSI assisted secondary user scheduling could produce a more considerable cooperation benefit percentage.

V. Conclusion

In this paper, we have investigated the application of NOMA to multicast CR networks, where a dynamic cooperative MCR-NOMA scheme has been proposed. The rationale behind is that the use of NOMA allows the primary and secondary networks to access the licensed spectrum simultaneously, while the coexistence of inter-user interference and heterogeneity of wireless channels degrades the reception probability. In order to realize more robust transmission, the secondary network provides the cooperation for the compensation of accessing the spectrum. As a result, the reception probability of both networks can be improved. Based on the cooperative MCR-NOMA scheme, we have further proposed three different secondary user scheduling strategies to exploit the inherent spatial diversity, and have investigated their impacts on the outage probability, diversity order and mutual outage probability. Numerical results have been presented to validate the analysis. Substantial performance gains have been shown using the proposed cooperative MCR-NOMA with secondary user scheduling strategies than the non cooperative MCR-NOMA/SDMA, and a mutual cooperation benefit can be eventually achieved. It is expected that our proposed cooperative MCR-NOMA scheme can be deemed as a promising solution for future large-scale multicast CR networks.

Appendix A: Proof of Theorem 1

First, substituting (2) and (9) into (12), after some algebraic manipulations, the outage probability for the primary network can be rewritten as

$$\begin{aligned}\mathcal{P}_p \overset{(a)}{=} & \underbrace{\mathcal{P}\bigg(|h_{b,p}|^2 < \frac{\psi_p}{\rho}, |h_{b,n}|^2 < \frac{\psi_p}{\rho}\bigg)}_{Q_1} \\ & + \underbrace{\mathcal{P}\bigg(\frac{\alpha_p^2|h_{b,p}|^2}{\alpha_s^2|h_{b,p}|^2+\frac{1}{\rho}} + \rho|h_{n,p}|^2 < \tau_p, \frac{\psi_p}{\rho} \le |h_{b,n}|^2 < \frac{\psi_s}{\rho}\bigg)}_{Q_2} \\ & + \underbrace{\mathcal{P}\bigg(\frac{\alpha_p^2|h_{b,p}|^2}{\alpha_s^2|h_{b,p}|^2+\frac{1}{\rho}} + \frac{\alpha_p^2|h_{n,p}|^2}{\alpha_s^2|h_{n,p}|^2+\frac{1}{\rho}} < \tau_p, |h_{b,n}|^2 \ge \frac{\psi_s}{\rho}\bigg)}_{Q_3}\end{aligned} \tag{46}$$

where ψ_p and ψ_s can be readily found in Theorem 1, and step (a) can be obtained by assuming the condition $\psi_p < \psi_s$, otherwise, Q_2 is always equal to zero. Because all channels are independently distributed, Q_1 in (46) can be easily derived as

$$Q_1 = \left(1 - e^{-\frac{\psi_p d_{b,p}^\eta}{\rho}}\right)\left(1 - e^{-\frac{\psi_p d_{b,n}^\eta}{\rho}}\right). \tag{47}$$

Now, we focus on the calculation of Q_2, which can be rewritten as

$$\begin{aligned}Q_2 = & \underbrace{\int_0^{\varpi} \mathcal{P}\Big(|h_{n,p}|^2 < \frac{\tau_p - x}{\rho}\Big) f_X(x)dx}_{Q_{21}} \\ & \times \underbrace{\mathcal{P}\Big(\frac{\psi_p}{\rho} \le |h_{b,n}|^2 < \frac{\psi_s}{\rho}\Big)}_{Q_{22}}.\end{aligned} \tag{48}$$

In (48), we define $X = \frac{\alpha_p^2|h_{b,p}|^2}{\alpha_s^2|h_{b,p}|^2+\frac{1}{\rho}}$, so that its PDF is given by

$$f_X(x) = \frac{d_{b,p}^\eta}{\rho(\alpha_p^2 - \alpha_s^2 x)^2} e^{-\frac{x d_{b,p}^\eta}{\rho(\alpha_p^2 - \alpha_s^2 x)}} \tag{49}$$

where $0 \le x \le \varpi$. However, it is challenging to obtain a closed-form expression for Q_{21}. In this case, we will use Gaussian-Chebyshev quadrature [31] to find an approximation of Q_{21} as follows.

$$Q_{21} \approx \chi\left(1 - e^{-\frac{(\tau_p - a_l)d_{n,p}^\eta}{\rho}}\right) \tag{50}$$

where χ, ϖ, a_l, θ_l, and L can be found in Theorem 1. In addition, Q_{22} is computed by

$$Q_{22} = e^{-\frac{\psi_p d_{b,n}^\eta}{\rho}} - e^{-\frac{\psi_s d_{b,n}^\eta}{\rho}}. \tag{51}$$

Combining (50) and (51) with (48), a closed-form expression for Q_2 is attained. Using the same rationale, an approximation of Q_3 can be expressed as

$$Q_3 \approx \chi e^{-\frac{\psi_s d_{b,n}^\eta}{\rho}}\left(1 - e^{-\frac{(\tau_p - a_l)d_{n,p}^\eta}{\rho(\alpha_p^2 - \alpha_s^2(\tau_p - a_l))}}\right). \tag{52}$$

Combining the foregoing results with (12), the first part of Theorem 1 is proved.

Next, attention is shifted to the calculation of $\mathcal{P}_s$. Denote the first part in (13) as Q_4, which can be readily computed by

$$Q_4 = 1 - e^{-\frac{\psi_s d_{b,n}^\eta}{\rho}}. \tag{53}$$

The second part of (13), which is denoted by Q_5, can be rewritten as

$$Q_5 = \mathcal{P}(\mathrm{SINR}_{n,p} \ge \tau_p, \mathrm{SNR}_{n,s} \ge \tau_s)\mathcal{P}(\mathrm{SINR}_{n,n'} < \tau_s). \tag{54}$$

According to total probability theorem, $\mathcal{P}(\mathrm{SINR}_{n,n'} < \tau_s)$ can be solved as

$$\begin{aligned}&\mathcal{P}(\mathrm{SINR}_{n,n'} < \tau_s) \\ &= \sum_{k=1}^{N-1} \sum_{\substack{\mathcal{S}_2^c \subseteq \{1,\cdots,N\} \\ |\mathcal{S}_2^c| = k}} \mathcal{P}(|\mathcal{S}_2^c| = k)\mathcal{P}_{out,s}(|\mathcal{S}_2^c| = k)\end{aligned} \tag{55}$$

where $\mathcal{P}(|\mathcal{S}_2^c| = k)$ is calculated as

$$\begin{aligned}&\mathcal{P}(|\mathcal{S}_2^c| = k) \\ &= \prod_{i=1}^{k} \mathcal{P}\Big(|h_{b,n_i}|^2 < \frac{\psi_s}{\rho}\Big) \prod_{\tilde{n} \in \mathcal{S}_2} \mathcal{P}\Big(|h_{b,\tilde{n}}|^2 \ge \frac{\psi_s}{\rho}\Big) \\ &= \prod_{i=1}^{k}\left(1 - e^{-\frac{\psi_s d_{b,n_i}^\eta}{\rho}}\right) e^{-\sum_{\tilde{n} \in \mathcal{S}_2} \frac{\psi_s d_{b,\tilde{n}}^\eta}{\rho}}\end{aligned} \tag{56}$$

and $\mathcal{P}_{out,s}(|\mathcal{S}_2^c| = k)$ is computed by

$$\begin{aligned}\mathcal{P}_{out,s}(|\mathcal{S}_2^c| = k) &= \mathcal{P}\Big(\min_{i=1,\cdots,k}\{|h_{n,n_i}|\} < \frac{\psi_s}{\rho}\Big) \\ &= 1 - e^{-\sum_{i=1}^{k} \frac{\psi_s d_{n,n_i}^\eta}{\rho}}.\end{aligned} \tag{57}$$

Combining the results in (53)–(57), the second part of Theorem 1 is proved.

Appendix B: Proof of Theorem 2

The probabilities for $|\mathcal{S}_1| = 0$ and $|\mathcal{S}_2| = 0$ are given by

$$\mathcal{P}(|\mathcal{S}_1| = 0) = \prod_{n=1}^{N}\left(1 - e^{-\frac{\psi_p d_{b,n}^\eta}{\rho}}\right) \tag{58}$$

$$\mathcal{P}(|\mathcal{S}_2| = 0) = \prod_{n=1}^{N}\left(1 - e^{-\frac{\psi_s d_{b,n}^\eta}{\rho}}\right). \tag{59}$$

Furthermore, the probabilities for $|\mathcal{S}_1| = k_1$ and $|\mathcal{S}_2| = k_2$ are calculated, respectively, as

$$\mathcal{P}(|\mathcal{S}_1| = k_1) = \prod_{i=1}^{k_1} \left(e^{-\frac{\psi_p d_{b,n_i}^{\eta}}{\rho}} - e^{-\frac{\psi_s d_{b,n_i}^{\eta}}{\rho}} \right) \times \prod_{n \in \mathcal{S}_1^c} \left(1 - e^{-\frac{\psi_p d_{b,n}^{\eta}}{\rho}} \right) \quad (60)$$

$$\mathcal{P}(|\mathcal{S}_2| = k_2) = e^{-\sum_{j=1}^{k_2} \frac{\psi_s d_{b,n_j}^{\eta}}{\rho}} \prod_{n \in \mathcal{S}_2^c} \left(1 - e^{-\frac{\psi_s d_{b,n}^{\eta}}{\rho}} \right). \quad (61)$$

In addition, we can rewrite $\mathcal{P}(\mathrm{SINR}_{\tilde{n},p} < \tau_p)$ as

$$\mathcal{P}(\mathrm{SINR}_{\tilde{n},p} < \tau_p) = \mathcal{P}\left(\frac{\alpha_p^2 |h_{b,p}|^2}{\alpha_s^2 |h_{b,p}|^2 + \frac{1}{\rho}} + \rho |h_{n^\star,p}|^2 < \tau_p \right) \quad (62)$$

where the CDF of $\rho |h_{n^\star,p}|^2$ is given by

$$F_Y(y) = \prod_{i=1}^{k_1} \left(1 - e^{-\frac{y d_{n_i,p}^{\eta}}{\rho}} \right). \quad (63)$$

Substituting (63) into (62), and with the help of the Gaussian-Chebyshev quadrature [31], the probability $\mathcal{P}(\mathrm{SINR}_{\tilde{n},p} < \tau_p)$ can be approximated as

$$\mathcal{P}(\mathrm{SINR}_{\tilde{n},p} < \tau_p) \approx \chi \prod_{i=1}^{k_1} \left(1 - e^{-\frac{(\tau_p - a_l) d_{n_i,p}^{\eta}}{\rho}} \right). \quad (64)$$

Similarly, we can approximate $\mathcal{P}(\mathrm{SINR}_{n,p} < \tau_p)$ as

$$\mathcal{P}(\mathrm{SINR}_{n,p} < \tau_p) \approx \chi \prod_{j=1}^{k_2} \left(1 - e^{-\frac{(\tau_p - a_l) d_{n_j,p}^{\eta}}{\rho(\alpha_p^2 - \alpha_s^2(\tau_p - a_l))}} \right). \quad (65)$$

Substituting the above equations into (26), the first part of Theorem 2 is proved.

On the other hand, relying on the total probability theorem, the probability of $\mathcal{P}_{out,s^\star}(|\mathcal{S}_2| = k_2)$ in (27) can be rewritten as

$$\mathcal{P}_{out,s^\star}(|\mathcal{S}_2| = k_2) = \sum_{j=1}^{k_2} \mathcal{P}(n^\star = n_j) \big(1 - \mathcal{P}(\mathrm{SINR}_{n',p} \geq \tau_p, \mathrm{SNR}_{n',s} \geq \tau_s) \big). \quad (66)$$

From (66), the probability of $\mathcal{P}(n^\star = n_j)$ can be formulated as

$$\mathcal{P}(n^\star = n_j) = \mathcal{P}\Big(\bigcap_{\substack{i=1 \\ i \neq j}}^{k_2} \big(|h_{n_j,p}|^2 > |h_{n_i,p}|^2 \big) \Big). \quad (67)$$

Note that the events $\big(|h_{n_j,p}|^2 > |h_{n_i,p}|^2\big)$ $(i \neq j)$ are not mutually exclusive, which is cumbersome to proceed forward. To do this, we define $Z = |h_{n_j,p}|^2$ and use the law of conditional probability, then (67) can be rewritten as

$$\mathcal{P}(n^\star = n_j) = 1 + \sum_{i=1}^{k_2 - 1} \sum_{\substack{A_i = \{1, \cdots, j-1, j+1, \cdots, k_2\} \\ |A_i| = i}} (-1)^i \times \frac{d_{n_j,p}^{\eta}}{d_{n_j,p}^{\eta} + \sum_{i \in A_i} d_{n_i,p}^{\eta}}. \quad (68)$$

which is obtained by using the multinomial expansion identity [30, eq. (33)]. Additionally, due to the multicast transmission in the secondary network, the outage performance is dominated by the worst-case scenario, and then $\mathcal{P}(\mathrm{SINR}_{n_j,p} \geq \tau_p, \mathrm{SNR}_{n_j,s} \geq \tau_s)$ is shown as

$$\mathcal{P}(\mathrm{SINR}_{n_j,p} \geq \tau_p, \mathrm{SNR}_{n_j,s} \geq \tau_s) = e^{-\sum_{n \in \mathcal{S}_2^c} \frac{\psi_s d_{n_j,n}^{\eta}}{\rho}}. \quad (69)$$

Therefore, combining results in (66)–(69) with (27), the rest of Theorem 2 is obtained.

Appendix C: Proof of Theorem 3

When $|\mathcal{S}_2| = N$, all the SUs can decode x_p and x_s correctly, and the optimal scheduled SU will transmit only x_p in this case. By using (59)–(65), a closed-form approximation for (39) is obtained immediately.

With the help of conditional probability, the probability of $\mathcal{P}(|\mathcal{S}_2| = k_2, |\mathcal{S}_3| = k_3)$ can be rewritten as

$$\mathcal{P}(|\mathcal{S}_2| = k_2, |\mathcal{S}_3| = k_3) = \mathcal{P}(|\mathcal{S}_2| = k_2) \mathcal{P}(|\mathcal{S}_3| = k_3 \mid |\mathcal{S}_2| = k_2). \quad (70)$$

From (70), $\mathcal{P}(|\mathcal{S}_2| = k_2)$ can be obtained by (61), and the remaining term can be computed by

$$\mathcal{P}(|\mathcal{S}_3| = k_3 \mid |\mathcal{S}_2| = k_2) \approx \chi^{k_3} e^{-\sum_{r=1}^{k_3} \frac{(\tau_p - a_l) d_{n_r,p}^{\eta}}{\rho(\alpha_p^2 - \alpha_s^2(\tau_p - a_l))}} \times \prod_{t=k_3+1}^{k_2 - k_3} \chi \left(1 - e^{-\frac{(\tau_p - a_l) d_{n_t,p}^{\eta}}{\rho(\alpha_p^2 - \alpha_s^2(\tau_p - a_l))}} \right) \quad (71)$$

where we have used the Gaussian-Chebyshev quadrature [31]. For the event of $k_3 = 0$, we have $\mathcal{P}_{out,s^*}(|\mathcal{S}_2| = k_2, |\mathcal{S}_3| = 0) = 1$, otherwise, $\mathcal{P}_{out,s^*}(|\mathcal{S}_2| = k_2, |\mathcal{S}_3| = k_3)$ is given by

$$\mathcal{P}_{out,s^*}(|\mathcal{S}_2| = k_2, |\mathcal{S}_3| = k_3) = \prod_{r=1}^{k_3} \mathcal{P}\left(\min_{n \in \mathcal{S}_2^c} \left\{ |h_{n_r,n}|^2 \right\} < \frac{\psi_s}{\rho} \right) = \prod_{r=1}^{k_3} \left(1 - e^{-\sum_{n \in \mathcal{S}_2^c} \frac{\psi_s d_{n_r,n}^{\eta}}{\rho}} \right). \quad (72)$$

Finally, substituting (59), (70)–(72) into (38), the second part of Theorem 3 is proved.

References

[1] L. Dai, B. Wang, Y. Yuan, S. Han, C.-L. I, and Z. Wang, "Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, Sep. 2015.

[2] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, and H. V. Poor, "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 185–191, Feb. 2017.

[3] S. M. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentails and challenges," *IEEE Commun. Surveys & Tutorials*, to be published.

[4] 3rd Generation Partnership Project (3GPP), "Study on downlink multiuser superposition transmission for LTE," Shanghai, China, Mar. 2015.

[5] "5G radio access: Requirements, concept and technologies," *NTT DOCOMO, Inc., Tokyo, Japan, 5G Whitepaper*, Jul. 2014.

[6] METIS, "Proposed solutions for new radio access," *Mobile and wireless communications enablers for the 2020 information society (METIS)*, Deliverable D.2.4, Feb. 2015.

[7] 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.

[8] 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.

[9] S. Timotheou and I. Krikidis, "Fairness for non-orthogonal multiple access in 5G systems," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1647–1651, Oct. 2015.

[10] M. Al-Imari, P. Xiao, M. Imran, and R. Tafazolli, "Uplink non-orthogonal multiple access for 5G wireless networks," in *Proc. 11th Int. Symp. Wireless Commun. Syst. (ISWCS)*, Aug. 2014, pp. 781–785.

[11] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G non-orthogonal multiple access downlink transmissions," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6010–6023, Aug. 2016.

[12] J.-B. Kim and I.-H. Lee, "Non-orthogonal multiple access in coordinated direct and relay transmission," *IEEE Commun. Lett.*, vol. 19, no. 11, pp. 2037–2040, Nov. 2015.

[13] J. Men and J. Ge, "Non-orthogonal multiple access for multiple antenna relaying networks," *IEEE Commun. Lett.*, vol. 19, no. 10, pp. 1686–1689, Oct. 2015.

[14] 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.

[15] 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.

[16] A. Goldsmith, S. A. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proc. IEEE*, vol. 97, no. 5, pp. 894–914, May 2009.

[17] J. Chen, L. Lv, Y. Liu, Y. Kuo, and C. Ren, "Energy efficient relay selection and power allocation for cooperative cognitive radio networks," *IET Commun.*, vol. 9, no. 13, pp. 1661–1668, Aug. 2015.

[18] G. Zheng, Z. Ho, E. A. Jorswieck, and B. Ottersten, "Information and energy cooperation in cognitive radio networks," *IEEE Trans. Signal Process.*, vol. 69, no. 2, pp. 2290–2303, May 2014.

[19] T. Elkourdi and O. Simeone, "Spectrum leasing via cooperation with multiple primary users," *IEEE Trans. Veh. Technol.*, vol. 61, no. 2, pp. 820–825, Feb. 2012.

[20] O. Simeone, I. Stanojev, S. Savazzi, Y. Bar-Ness, U. Spagnolini, and R. Pickholtz, "Spectrum leasing to cooperating secondary ad-hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 203–213, Jan. 2008.

[21] B. Niu, H. Jiang, and H. V. Zhao, "A cooperative multicast strategy in wireless networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 6, pp. 3136–3143, Jul. 2010.

[22] B. Hu, H. V. Zhao, and H. Jiang, "Wireless multicast using relays: Incentive mechanism and analysis," *IEEE Trans. Veh. Technol.*, vol. 62, no. 5, pp. 2204–2219, Jun. 2013.

[23] X. Lin, R. Ratasuk, A. Ghosh, and J. G. Andrews, "Modeling, analysis, and optimization of multicast device-to-device transmissions," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4346–4359, Aug. 2014.

[24] L. Yang, J. Chen, H. Zhang, H. Jiang, S. A. Vorobyov, and D. T. Ngo, "Cooperative wireless multicast: Performance analysis and time allocation," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5810–5819, Jul. 2016.

[25] L. Xu and A. Nallanathan, "Energy-efficient chance-constrained resource allocation for multicast cognitive OFDM network," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1298–1306, May 2016.

[26] S. Li, W. Xu, K. Yang, K. Niu, and J. Lin, "Distributed cooperative multicast in cognitive multi-relay multi-antenna systems," *IEEE Signal Process. Lett.*, vol. 22, no. 3, pp. 288–292, Mar. 2015.

[27] L. Yang, J. Chen, Y. Kuo, and H. Zhang, "Outage performance of DF-based cooperative multicast in spectrum-sharing cognitive relay networks," *IEEE Commun. Lett.*, vol. 18, no. 7, pp. 1250–1253, Jul. 2014.

[28] Y. Liu, Z. Ding, M. Elkashlan, and J. Yuan, "Non-orthogonal multiple access in large-scale underlay cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 10152–10157, Dec. 2016.

[29] L. Lv, Q. Ni, Z. Ding, and J. Chen, "Application of non-orthogonal multiple access in cooperative spectrum-sharing networks over Nakagami-m fading channels," *IEEE Trans. Veh. Technol.*, DoI: 10.1109/TVT.2016.2627559.

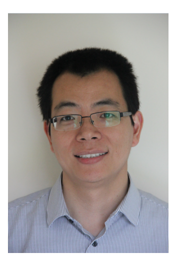
[30] A. Bletsas, A. G. Dimitriou, and J. N. Sahalos, "Interference-limited opportunistic relaying with reactive sensing," *IEEE Trans. Wireless Commun.*, vol. 9, no. 1, pp. 14–20, Jan. 2010.

[31] E. Hildebrand, *Introduction to Numerical Analysis*. New York, NY, USA: Dover, 1987.

Lu Lv received the B.Eng degree in Telecommunications Engineering from Hebei Normal University in 2013. He is currently working toward the Ph.D. degree in Telecommunications Engineering at Xidian University. From March 2016 to September 2016, he was a visiting Ph.D. student at Lancaster University, UK. Since December 2016, he has been with the University of Alberta, Canada, as a joint Ph.D. student sponsored by the China Scholarship Council (CSC). His current research interests focus on cooperative relaying, physical layer security, non-orthogonal multiple access, and mmWave communications.

Jian Chen (M'14) received the B.Eng degree from Xi'an Jiaotong University, China, in 1989, the M.Eng degree from Xi'an Institute of Optics and Precision Mechanics of Chinese Academy of Sciences in 1992, and the Ph.D. degree in Telecommunications Engineering from Xidian University, China, in 2005. He is currently a Full Professor at the School of Telecommunications Engineering, Xidian University. He was a Visiting Scholar with the University of Manchester from 2007 to 2008. His research interests are cognitive radio, physical layer security, wireless sensor networks, compress sensing and signal processing.

Qiang Ni (M'04–SM'08) received the Ph.D. degree in engineering from Huazhong University of Science and Technology (HUST), Wuhan, China. He is currently a Full Professor and the Head of Communication Systems Research Group at the School of Computing and Communications, Lancaster University, Lancaster, UK. His research interests lie in the area of future generation communications and networking systems, including energy and spectrum efficient green wireless communications, non-orthogonal multiple access (NOMA), 5G, software defined network (SDN), game theory, heterogeneous networks, cognitive radio network systems, energy harvesting, etc. He had published more than 180 research papers in international journals and conferences. He was an IEEE 802.11 Wireless Standard Working Group Voting Member and a Contributor to the IEEE Wireless Standards.



Zhiguo Ding (S'03–M'05–SM'15) received the B.Eng in Electrical Engineering from the Beijing University of Posts and Telecommunications in 2000, and the Ph.D degree in Electrical Engineering from Imperial College London in 2005. From July 2005 to August 2014, he was working in Queen's University Belfast, Imperial College and Newcastle University. Since September 2014, he has been with Lancaster University as a Chair Professor. From October 2012 to September 2016, he has been also with Princeton University as a Visiting Research Collaborator. His research interests are 5G networks, game theory, cooperative and energy harvesting networks and statistical signal processing. He is serving as an Editor for *IEEE Transactions on Communications*, *IEEE Transactions on Vehicular Networks*, *IEEE Wireless Communication Letters*, *IEEE Communication Letters*, and *Journal of Wireless Communications and Mobile Computing*. He received the best paper award in IET Comm. Conf. on Wireless, Mobile and Computing, 2009, IEEE Communication Letter Exemplary Reviewer 2012, and the EU Marie Curie Fellowship 2012-2014.