Buffer-aided Relaying for Downlink NOMA Systems with Direct Links

Jianglong Li*, Xianfu Lei*, Panagiotis D. Diamantoulakis†, Panagiotis Sarigiannidis‡, and George K. Karagiannidis†
*School of Information Science and Technology, Southwest Jiaotong University, Chengdu, China
†Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece
‡Department of Informatics and Telecommunications Engineering, University of Western Macedonia, Kozani, Greece
e-mails: JLLi@my.swjtu.edu.cn, xflei81@gmail.com, padiaman@ieee.org, psarigiannidis@uowm.gr, geokarag@auth.gr

***Abstract*—Non-orthogonal multiple access (NOMA) has recently attracted the academic and industrial interest, due to offering higher spectral efficiency and connectivity compared to conventional orthogonal multiple access schemes. In this paper, buffer-aided relaying for a downlink NOMA system with direct links is proposed, while in order to take advantage the extra degrees of freedom, appropriate transmission modes are presented. Targeting at the throughput maximization, the corresponding optimization problem is formulated and solved and theoretical expressions are derived for the optimal mode selection policy and maximum throughput. Finally, simulation results illustrate the efficiency of the proposed scheme and its superiority compared with a previously presented baseline scheme.**

I. Introduction

Non-orthogonal multiple access (NOMA) has become a major paradigm for the design of radio access technologies (RATs) for the fifth generation (5G) wireless networks, since, compared to orthogonal multiple access (OMA), i.e., time/frequency/code division multiple access, it can increase spectral efficiency and connectivity [1]–[4]. Motivated by these advantages, NOMA has already been an approved study item of the 3-rd Generation Partnership Project (3GPP) in Release 15 [5]. The main principle of NOMA is that it can serve multiple users on the same orthogonal resource block. In particular, NOMA enables the simultaneous transmission of superimposed messages by using joint processing technique at the receivers, such as successive interference cancellation (SIC) [6]. Interestingly, NOMA is compatible with most RATs, e.g., OMA, multiple-input-multiple-output (MIMO), millimeter wave, full-duplex, etc [7]–[10].

Among others, NOMA can also be used in cooperative networks, in order to further extend the coverage and/or increase throughput. This can be achieved by using either a dedicated node or one of the users' nodes as a relay. Cooperative NOMA (C-NOMA) with a dedicated relay was proposed in [11], considering a two-user downlink scenario, where the information transmission to both users is assisted by the relay. Also, in [12] and [13], [14], C-NOMA with multi-antenna techniques and relay selection were proposed, respectively, in order to reduce the outage probability and increase the diversity gain. On the other hand, in [15], [16], C-NOMA was investigated in the context of a near-far user setup, where the weak user also receives its message from the near user, taking advantage of the SIC process performed at the strong user, according to which the strong user decodes both messages. Furthermore, in [17], full-duplex was used in order to increase the achievable rate and reduce the outage probability of C-NOMA systems with user cooperation.

The throughput of cooperative networks can also be improved by employing a buffer at the relay such that data can be queued until the relay-destination link is selected for transmission, taking advantage of channel ergodicity (see the pioneering work in [18] and references therein). This has motivated the investigation of buffer-aided relaying in the context of C-NOMA systems. More specifically, information transmission from one source to two users through a buffer-aided relay has been investigated in [19], with the aim to maximize the system throughput. In more detail, based on the channel conditions and by assuming, for simplicity (although with loss of generality), fixed target rates, the relay has either the option to perform NOMA and simultaneously serve both users or serve only one of them by using OMA. Also, in [20], a similar setup has been investigated, focusing though solely on NOMA and by assuming finite buffer size. In addition, a relay selection scheme for the downlink of buffer-aided C-NOMA systems has been reported in [21]. It is highlighted that in the above works, the exploitation of direct links between the source and the user has not been considered. However, in 5G scenarios with high-speed mobility and nodes' density, the exploitation of direct links in cooperative networks systems becomes of paramount importance. Motivated by this, the authors of [22] and [23] have investigated for the first time the combination of direct and buffer-aided relaying transmissions, to optimize system throughput in single-user and multi-users uplink systems, respectively. To the best of the authors' knowledge, there are is no work in existing literature that investigates C-NOMA for the downlink of buffer-aided systems with direct links. It should be highlighted that the concept of downlink is different from that of the uplink NOMA, since in the downlink all users receive the interfering messages from the same source. Consequently, in contrast to uplink NOMA that achieves the capacity of the multiple access channel for any decoding order of the users' messages, in downlink NOMA the system throughput heavily depends on the users' channels ordering [24].

978-1-5386-8088-9/19/$31.00 ©2019 IEEE

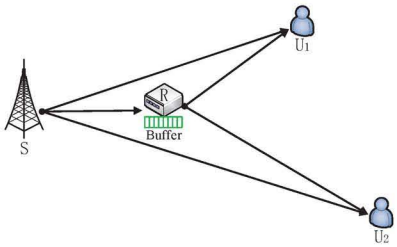

Fig. 1. Buffer-aided relaying for downlink NOMA system with direct links.

To this end, in this work, we investigate a downlink C-NOMA system consisting of a single source, one buffer-aided relay, and two users, while, unlike the previously presented baseline scheme in [19], we also consider the direct link between the source and each user. Consequently, we assume that the users can receive information either directly from the source or the relay, proposing the utilization of NOMA in both cases. Also, in contrast to [19], the analysis does not depend on the assumption of fixed target rates for the information transmission to the users. Taking advantage of the extra degrees of freedom offered by buffer-aided relaying, three different transmission modes are introduced, which correspond to the three possible sets of nodes that simultaneously exchange information. Furthermore, the system throughput maximization problem is formulated and optimally solved, by deriving an expression for the optimal mode selection policy in closed form. The proposed mode selection policy enables the use of the optimal transmission mode, according to the instantaneous channel quality and under the buffer-stability constraint. Moreover, a theoretical formula for the maximum throughput is derived. Simulation results show the efficiency of the proposed scheme, especially in comparison with [19].

II. System Model

In this paper, we consider a downlink C-NOMA system consisting of four single-antenna nodes, namely a source (S), a relay (R) with an infinite buffer, and two users (U_1 and U_2), as shown in Fig.1. Furthermore, we assume that there is a direct link between S and each user. The node S either transmits information to two users through the direct links, or transmits the information to the relay. The relay node is capable to forward this information to the end users, by using the decode-and-forward (DF) protocol. We further assume that time-division duplex (TDD) is used and total time is divided into slots of equal length indexed by $i = 1, \cdots, N$. Moreover, the channel state information (CSI) of all links is available at S, which coordinates the data transmission scheduling at the beginning of each time slot. All channels are subject to independent flat Rayleigh fading, i.e., the channel coefficients remain constant in one time slot and independently change in the following time slots or in different links. We also assume additive white Gaussian noise (AWGN) with zero mean and unitary variance at both the relay and the users.

Hereinafer, $h^i_{s1}, h^i_{s2}, h^i_{sr}, h^i_{r1}, h^i_{r2}$ denote the channel fading gains of S-U_1 link, S-U_2, S-R , R-U_1 , R-U_2 links at the $i-$th time slot, respectively. Let $\Omega_{s1}, \Omega_{s2}, \Omega_{sr}, \Omega_{r1}, \Omega_{r2}$ denote the mean value of the exponentially distributed channel gains of S-U_1, S-U_2, S-R, R-U_1, and R-U_2 links, respectively, and $f_{xy}\left(\frac{1}{h_{xy}}\right) = \frac{1}{\Omega_{xy}} * e^{-\frac{h_{xy}}{\Omega_{xy}}}$ is the corresponding probability density function (PDF). The transmit power at S and R is denoted by P_s and P_r, respectively. Also, at the $i-$th time slot, $x^i_{jk}(j = s, r; k = r, 1, 2; j \neq k)$ is used to denote the transmitted signal from node j to k, while $y^i_k(k = r, 1, 2)$ and $z^i_k(k = r, 1, 2)$ are used to denote received signal and AWGN at node k, respectively. Meanwhile, $C^i_{jk} = \log_2(1+X^i_{jk})(j = s, r; k = r, 1, 2; j \neq k)$ denotes the capacity of each link, with X^i_{jk} being the corresponding instantaneous channel signal-to-interference-plus-noise ratio (SINR).

III. Problem Formulation

A. Transmission Modes

In the considered setup, the source needs to decide whether or not to use the relay, and under the condition that the relay is used, whether the latter should be operated in *receive* or *transmit* mode. Let $r_i \in \{0, 1\}$ be the decision variable that determines if the relay is utilized, with $r_i = 0$ implying that the relay is not utilized. Moreover, for $r_i = 1$, $q_i \in \{0, 1\}$ determines whether the relay is used to receive or transmit information, i.e., $q_i = 0$ and $q_i = 1$ imply that the relay is utilized for receiving and transmitting information, respectively. Based on the above, three different modes are identified, i.e., M_1, M_2, M_3, which are described in detail below.

M_1: The relay is not utilized. Since $r_i = 0$, S transmits information to the two users using the direct links and NOMA. The received signal at user k is

$$y^i_k = \sqrt{h^i_{sk}}\left(\sqrt{\alpha^i_1}x^i_{s1} + \sqrt{\alpha^i_2}x^i_{s2}\right) + z^i_k,\ k = 1, 2, \quad (1)$$

where α^i_1 and α^i_2 are the power allocation coefficients with $\alpha^i_1 + \alpha^i_2 = 1$. Considering ordering of the channel conditions, there are two possible rate pairs as follows

$$(C^i_{s1}, C^i_{s2}) = \begin{cases} \left(\log_2\left(1 + \alpha^i_1 h^i_{s1} P_s\right), \log_2\left(1 + \frac{\alpha^i_2 h^i_{s2} P_s}{1+\alpha^i_1 h^i_{s2} P_s}\right)\right), h^i_{s1} > h^i_{s2}, \\ \left(\log_2\left(1 + \frac{\alpha^i_1 h^i_{s1} P_s}{1+\alpha^i_2 h^i_{s1} P_s}\right), \log_2\left(1 + \alpha^i_2 h^i_{s2} P_s\right)\right), h^i_{s1} < h^i_{s2}. \end{cases} \quad (2)$$

This is because only the user with the the stronger channel conditions can perform SIC. Since R is not utilized, the buffer state will not change, i.e., $Q^i = Q^{i-1}$, where Q^i denotes the total amount of information that is stored in the buffer during the i-th time slot. As a result, the instantaneous throughput in this mode is

$$\tau^i_1 = (1 - r_i)(C^i_{s1} + C^i_{s2}). \quad (3)$$

M_2: The relay is used to receive information. Since $r_i = 1$ and $q_i = 0$, S transmits the users' information to R with data rate

$$C^i_{sr} = \log_2(1 + h^i_{sr} P_s). \quad (4)$$

As in [19], the transmitted data flow is divided into two parts with rates ηC^i_{sr} and $(1 - \eta)C^i_{sr}$, each of which corresponds to U_1 and U_2, respectively, where the factor $0 < \eta < 1$ can be adjusted according to channels' statistical gains and buffer

state. As a result, if this mode is selected, the buffer state will change to $Q^i = Q^{i-1} + C^i_{sr}$. Since in this mode there is no data arrival at the users, the instantaneous throughput is $\tau^i_2 = 0$. It is worthy to mention that, due to the existence of the direct links, the users might also receive a copy of the transmitted messages by S that can be exploited using more complicated and less practical (especially in combination with NOMA) coding schemes [22], which are out of the scope of this work. Also, it needs to be noticed that if the direct links are stronger than the relaying links, there is a lack of motivation to use this mode instead of M_1.

M_3: The relay is utilized to transmit information to the users using NOMA. Since, $r_i = 1$ and $q_i = 1$, the received signal at each user is

$$y^i_k = \sqrt{h^i_{rk}}\left(\sqrt{\beta^i_1}x^i_{r1} + \sqrt{\beta^i_2}x^i_{r2}\right) + z^i_k,\ k = 1, 2, \quad (5)$$

where β^i_1 and β^i_2 are the power allocation coefficients and $\beta^i_1 + \beta^i_2 = 1$. Similarly to M_1, there are also two possible rate pairs, i.e.,

$$(C^i_{r1}, C^i_{r2}) = \begin{cases} \left(\log_2\left(1 + \beta^i_1 h^i_{r1} P_r\right), \log_2\left(1 + \frac{\beta^i_2 h^i_{r2} P_r}{1+\beta^i_1 h^i_{r2} P_r}\right)\right), h^i_{r1} > h^i_{r2}, \\ \left(\log_2\left(1 + \frac{\beta^i_1 h^i_{r1} P_r}{1+\beta^i_2 h^i_{r1} P_r}\right), \log_2\left(1 + \beta^i_2 h^i_{r2} P_r\right)\right), h^i_{r1} < h^i_{r2}. \end{cases} \quad (6)$$

Note that the buffer state constraints $C^i_{r1} < Q^i_1$ and $C^i_{r2} < Q^i_2$ are omitted, since buffers with infinite size have been assumed. In this mode, the buffer state will change to $Q^i_1 = Q^{i-1}_1 - C^i_{r1} - C^i_{r2}$, while the instantaneous throughput in this mode is

$$\tau^i_3 = r_i q_i (C^i_{r1} + C^i_{r2}). \quad (7)$$

B. System Throughput under Buffer-Stability Constraint

The system throughput, τ, is defined as the average sum throughput of two users over $N \to \infty$ time slots, i.e.,

$$\tau = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} (\tau^i_1 + \tau^i_2 + \tau^i_3) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \left((1 - r_i)(C^i_{s1} + C^i_{s2}) + r_i q_i (C^i_{r1} + C^i_{r2})\right), \quad (8)$$

Eq. (8) is subject to the requirement of equal arrival (A_1, A_2) and departure rates (D_1, D_2) at R for U_1 and U_2 [18], i.e,

$$\eta \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} r_i(1-q_i)C^i_{sr} + (1-\eta) \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} r_i(1-q_i)C^i_{sr} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} r_i q_i C^i_{r1} + \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} r_i q_i C^i_{r2}, \quad (9)$$

where the four aggregated terms correspond to A_1, A_2, D_1, D_2, respectively.

IV. System Throughput Maximization

A. Problem Formulation

As it has already been mentioned, in each time slot, it must hold that $r_i, q_i \in \{0, 1\}$, which is a set of combinatorial constraints. To facilitate the analysis, this can be replaced by the following equivalent set of constraints

$$\frac{1}{N} r_i(1 - r_i) = 0, \frac{1}{N} q_i(1 - q_i) = 0. \quad (10)$$

Based on above, the following optimization problem can be formulated which aims at the maximization of the system throughput:

$$\begin{aligned} \textbf{max}_{r_i, q_i} \quad & \lim_{N\to\infty} \tfrac{1}{N} \sum_{i=1}^{N} \left((1 - r_i)(C^i_{s1} + C^i_{s2}) + r_i q_i (C^i_{r1} + C^i_{r2})\right) \\ \textbf{s.t.} \quad & \mathrm{C}_1 : \lim_{N\to\infty} \tfrac{1}{N} \sum_{i=1}^{N} \left(r_i q_i (C^i_{r1} + C^i_{r2}) - r_i(1 - q_i)C^i_{sr}\right) = 0, \\ & \mathrm{C}_2 : \lim_{N\to\infty} \tfrac{1}{N} r_i(1 - r_i) = 0, \\ & \mathrm{C}_3 : \lim_{N\to\infty} \tfrac{1}{N} q_i(1 - q_i) = 0, \end{aligned} \quad (11)$$

where C_1 corresponds to (9). It is notable that due to the utilization of constraints C_2 and C_3 instead of $r_i, q_i \in \{0, 1\}$, the optimization problem in (11) is a non-combinatorial one and thus it can be efficiently solved by standard optimization methods.

B. Optimal Mode Selection

Theorem 1: The optimal mode selection policy that maximizes the system throughput is given by

$$(r^*_i, q^*_i) = \begin{cases} (1, 0), & \text{if } C^i_{s1} + C^i_{s2} < \frac{\rho}{1+\rho} C^i_{sr} \\ & \text{and } C^i_{r1} + C^i_{r2} < \rho C^i_{sr}, \\ (1, 1), & \text{if } C^i_{s1} + C^i_{s2} < \frac{1}{1+\rho}(C^i_{r1} + C^i_{r2}) \\ & \text{and } C^i_{sr} < \frac{1}{\rho}(C^i_{r1} + C^i_{r2}), \\ (0, \mathrm{NR}), & \text{otherwise}, \end{cases} \quad (12)$$

where ρ is the optimal decision threshold and its value depends on the Lagrange multiplier (LM) that corresponds to C_1, and NR stands for "not relevant".

Proof: The *Lagrangian* function of (11) is

$$\mathcal{L}(\tau, \lambda, \mu_i, \nu_i) = \frac{1}{N} \sum_{i=1}^{N} \big[(1 - r_i)(C^i_{s1} + C^i_{s2}) + \lambda r_i(1 - q_i) \times C^i_{sr} + (1 - \lambda) r_i q_i (C^i_{r1} + C^i_{r2}) - \nu_i r_i(1 - r_i) - \mu_i q_i(1 - q_i)\big] \quad (13)$$

where λ, μ_i, ν_i are the LMs associated the equality constraints. For the optimal values of q_i and r_i, it must hold that $\frac{\partial \mathcal{L}}{\partial q_i} = \frac{\partial \mathcal{L}}{\partial r_1} = 0$, from which, the following expressions are derived

$$q_i = \frac{\mu_i + \lambda r_i C^i_{sr} - (1 - \lambda) r_i (C^i_{r1} + C^i_{r2})}{2\mu_i} \quad (14)$$

$$r_i = \frac{\nu_i + C^i_{s1} + C^i_{s2} - \lambda(1 - q_i)C^i_{sr} - (1 - \lambda) q_i (C^i_{r1} + C^i_{r2})}{2\nu_i}. \quad (15)$$

Also, it should be noted that for the values of q_i, r_i that maximize $\mathcal{L}$, it must hold that $\frac{\partial^2 \mathcal{L}}{\partial q_i^2}, \frac{\partial^2 \mathcal{L}}{\partial r_i^2} < 0$, i.e.,

$$\mu_i, \nu_i < 0, \quad (16)$$

respectively. Hereinafter, we focus on the case that $r_i = 1$ is the optimal option, since $r_i = 0$ should be preferred otherwise.

In (14), for $r_i = 1$ and $q_i = 0$, we have

$$\mu_i = (1-\lambda)(C^i_{r1} + C^i_{r2}) - \lambda C^i_{sr}, \tag{17}$$

which, considering (16) leads to

$$C^i_{r1} + C^i_{r2} < \frac{\lambda}{1-\lambda} C^i_{sr}. \tag{18}$$

Similarly, for $r_i = 1$ and $q_i = 1$, from (14) and (16) it must hold that

$$C^i_{r1} + C^i_{r2} > \frac{\lambda}{1-\lambda} C^i_{sr}. \tag{19}$$

On the other hand, from (15), for $r_i = 1$ and $q_i = 0$ we have

$$\nu_i = C^i_{s1} + C^i_{s2} - \lambda C^i_{sr}, \tag{20}$$

which, by using (16) leads to

$$C^i_{s1} + C^i_{s2} < \lambda C^i_{sr}. \tag{21}$$

Similarly, from (15) and (16), for $r_i = 1$ and $q_i = 1$, the following condition is derived

$$C^i_{s1} + C^i_{s2} < (1-\lambda)(C^i_{r1} + C^i_{r2}). \tag{22}$$

By combining conditions (18), (19), (21), and (22) and letting $\rho = \frac{\lambda}{1-\lambda}$, we can derive (12). ■

C. Optimal Decision Threshold and Maximum Throughput

Considering (8), (9), and the optimal values of r_i, q_i for each time slot i, given by (12), the optimal decision threshold ρ^* and the maximum throughput τ^* can be derived by

$$\mathrm{E}\{r^*(1-q^*)C_{sr}\} = \mathrm{E}\{r^* q^* (C_{r1} + C_{r2})\}, \tag{23}$$

and

$$\tau^* = \mathrm{E}\{(1-r^*)(C_{s1} + C_{s2})\} + \mathrm{E}\{r^* q^* (C_{r1} + C_{r2})\}, \tag{24}$$

where $\mathrm{E}\{x\} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x_i$ denotes expectation. For the calculation of ρ^* and τ^*, a three-step approach is followed.

Step 1 : First, $I = \mathrm{E}\{(1-r^*)(C_{s1}+C_{s2})\}$ is calculated. To this end, it is important to find the integration domain in which it is non-zero, with respect to $r_i = 0$ in (12). Accordingly, I can be written as

$$I = I_1 + I_2 + I_3 + I_4 + I_5 + I_6. \tag{25}$$

In (25), I_1 is given by

$$I_1 = \int_{h_{sr}=0}^{\infty} \int_{h_{s1}=L_1}^{\infty} \int_{h_{s2=0}}^{h_{s1}} (C_{s1} + C_{s2}) f_{s2} dh_{s2} f_{s1} dh_{s1} f_{sr} dh_{sr}. \tag{26}$$

Since I_1 corresponds to the case $h_{s1} > h_{s2}$, $C_{s1} + C_{s2}$ is

$$C_{s1}+C_{s2} = \log_2(1 + \alpha_1 h_{s1} P_s) + \log_2\left(1 + \frac{\alpha_2 h_{s2} P_s}{1 + \alpha_1 h_{s2} P_s}\right). \tag{27}$$

Consequently, from $C_{s1} + C_{s2} > \frac{\rho}{1+\rho} C_{sr}$ in (12), we can get

$$L_1 = \frac{1}{\alpha_1 P_s}\left[\left[(1+h_{sr})^{\frac{\rho}{1+\rho}} \Big/ \left(1 + \frac{\alpha_2 h_{s1} P_s}{1 + \alpha_1 h_{s1} P_s}\right)\right] - 1\right]. \tag{28}$$

Also, I_2 that corresponds to the case $h_{s1} < h_{s2}$, can be calculated similarly to I_1. Regarding I_3, it can be derived by

$$I_3 = \int_{h_{r1}=0}^{\infty} \int_{h_{r2}=0}^{h_{r1}} \int_{h_{s1}=L_3}^{\infty} \int_{h_{s2=0}}^{h_{s1}} (C_{s1} + C_{s2}) \times f_{s2} dh_{s2} f_{s1} dh_{s1} f_{r2} dh_{r2} f_{r1} dh_{r1}. \tag{29}$$

Considering that I_3 corresponds to the case $h_{s1} > h_{s2}$, $h_{r1} > h_{r2}$, it holds that

$$C_{s1} + C_{s2} = \log_2(1 + \alpha_1 h_{s1} P_s) + \log_2\left(1 + \frac{\alpha_2 h_{s2} P_s}{1 + \alpha_1 h_{s2} P_s}\right) \tag{30}$$

and

$$C_{r1}+C_{r2} = \log_2(1 + \beta_1 h_{r1} P_r) + \log_2\left(1 + \frac{\beta_2 h_{r2} P_r}{1 + \beta_1 h_{r2} P_r}\right). \tag{31}$$

Thus, from $C_{s1} + C_{s2} > \frac{1}{1+\rho}(C_{r1} + C_{r2})$ in (12), we can get

$$L_3 = \frac{1}{\alpha_1 P_s}\left[\left[\frac{\left[(1+\beta_1 h_{r1} P_r)\left(1+\frac{\beta_2 h_{r2} P_r}{1+\beta_1 h_{r2} P_r}\right)\right]^{\frac{1}{1+\rho}}}{\left(1+\frac{\alpha_2 h_{s1} P_s}{1+\alpha_1 h_{s1} P_s}\right)}\right] - 1\right]. \tag{32}$$

Moreover, I_4, I_5, and I_6, which correspond to the cases $(h_{s1} > h_{s2}, h_{r1} < h_{r2})$ and $(h_{s1} < h_{s2}, h_{r1} > h_{r2})$, and $(h_{s1} < h_{s2}, h_{r1} < h_{r2})$, respectively, can be calculated similarly.

Step 2 : Similarly to the calculation of I, in order to derive $I' = \mathrm{E}\{r^* q^* (C_{r1} + C_{r2})\}$, with respect to $r_i = 1, q_i = 1$ in (12), the following expression can be used

$$I' = I'_1 + I'_2 + I'_3 + I'_4 + I'_5 + I'_6. \tag{33}$$

In (33), I'_1 can be expressed as

$$I'_1 = \int_{h_{sr}=0}^{\infty} \int_{h_{r1}=L'_1}^{\infty} \int_{h_{r2=0}}^{h_{r1}} (C_{r1} + C_{r2}) f_{r2} dh_{r2} f_{r1} dh_{r1} f_{sr} dh_{sr}. \tag{34}$$

Considering that I'_1 corresponds to the case $h_{r1} > h_{r2}$, it holds that

$$C_{r1}+C_{r2} = \log_2(1 + \beta_1 h_{r1} P_r) + \log_2\left(1 + \frac{\beta_2 h_{r2} P_r}{1 + \beta_1 h_{r2} P_r}\right), \tag{35}$$

and, thus, from $C_{r1} + C_{r2} > \rho C_{sr}$ in (12), L'_1 can be written as

$$L'_1 = \frac{1}{\beta_1 P_r}\left[\left[(1+h_{sr})^{\rho} \Big/ \left(1 + \frac{\beta_2 h_{r1} P_r}{1 + \beta_1 h_{r1} P_r}\right)\right] - 1\right]. \tag{36}$$

Considering that I'_2 corresponds to the case $h_{r1} < h_{r2}$, it can be calculated similarly to I'_1. Moreover, I'_3 is

$$I'_3 = \int_{h_{s1}=0}^{\infty} \int_{h_{s2}=0}^{h_{s1}} \int_{h_{r1}=L'_3}^{\infty} \int_{h_{r2=0}}^{h_{r1}} (C_{r1} + C_{r2}) \times f_{r2} dh_{r2} f_{r1} dh_{r1} f_{s2} dh_{s2} f_{s1} dh_{s1}. \tag{37}$$

Taking into account that I_3' corresponds to $h_{r1} > h_{r2}, h_{s1} > h_{s2}$, it holds that

$$C_{s1} + C_{s2} = \log_2\left(1 + \alpha_1 h_{s1} P_s\right) + \log_2\left(1 + \frac{\alpha_2 h_{s2} P_s}{1 + \alpha_1 h_{s2} P_s}\right) \tag{38}$$

and

$$C_{r1} + C_{r2} = \log_2\left(1 + \beta_1 h_{r1} P_r\right) + \log_2\left(1 + \frac{\beta_2 h_{r2} P_r}{1 + \beta_1 h_{r2} P_r}\right), \tag{39}$$

thus, from $C_{r1} + C_{r2} > (1+\rho)(C_{s1} + C_{s2})$ in (12), it is derived that

$$L_3' = \frac{1}{\beta_1 P_r}\left[\left[\frac{\left[(1+\alpha_1 h_{s1} P_s)\left(1+\frac{\alpha_2 h_{s2} P_s}{1+\alpha_1 h_{s2} P_s}\right)\right]^{1+\rho}}{\left(1+\frac{\beta_2 h_{r1} P_r}{1+\beta_1 h_{r1} P_r}\right)}\right] - 1\right]. \tag{40}$$

Moreover, I_4', I_5', and I_6', which correspond to the cases $(h_{s1} > h_{s2}, h_{r1} < h_{r2})$, $(h_{s1} < h_{s2}, h_{r1} > h_{r2})$, and $(h_{s1} < h_{s2}, h_{r1} < h_{r2})$, respectively, can be derived similarly.

Step 3 : Finally, by considering the case of $r_i = 1, q_i = 0$ in (12), $I'' = \mathrm{E}\{r^*(1-q^*)C_{sr}\}$ can be written as

$$I'' = I_1'' + I_2'' + I_3'' + I_4'', \tag{41}$$

where

$$I_1'' = \int_{h_{s1}=0}^{\infty}\int_{h_{s2}=0}^{h_{s1}}\int_{h_{sr}=L_1''}^{\infty} C_{sr} f_{sr} dh_{sr} f_{s2} dh_{s2} f_{s1} dh_{s1}.$$

Considering that I_1'' corresponds to the case $h_{s1} > h_{s2}$, it holds that

$$C_{s1} + C_{s2} = \log_2\left(1 + \alpha_1 h_{s1} P_s\right) + \log_2\left(1 + \frac{\alpha_2 h_{s2} P_s}{1 + \alpha_1 h_{s2} P_s}\right), \tag{42}$$

which, in combination with $C_{sr} > \frac{\rho}{1+\rho}(C_{s1} + C_{s2})$ form (12), L_1'' can be written as

$$L_1'' = \frac{1}{P_s}\left[\left[(1+\alpha_1 h_{s1} P_s)\left(1 + \frac{\alpha_2 h_{s2} P_s}{1+\alpha_1 h_{s2} P_s}\right)\right]^{\frac{\rho}{1+\rho}} - 1\right]. \tag{43}$$

Also, in (41), I_2'' corresponds to the case $h_{s1} < h_{s2}$ and can be calculated similarly. Moreover, I_3'' corresponds to the case $h_{r1} > h_{r2}$ and can be given by

$$I_3'' = \int_{h_{r1}=0}^{\infty}\int_{h_{r2}=0}^{h_{r1}}\int_{h_{sr}=L_3''}^{\infty} C_{sr} f_{sr} dh_{sr} f_{h_{r2}} d_{h_{r2}} f_{r1} dh_{r1}. \tag{44}$$

Considering that $h_{r1} > h_{r2}$, $C_{r1} + C_{r2}$ is given by

$$C_{r1} + C_{r2} = \log_2\left(1 + \beta_1 h_{r1} P_r\right) + \log_2\left(1 + \frac{\beta_2 h_{r2} P_r}{1+\beta_1 h_{r2} P_r}\right), \tag{45}$$

which, from $C_{sr} > \frac{1}{\rho}(C_{r1} + C_{r2})$ in (12), leads to

$$L_3'' = \frac{1}{P_s}\left[\left[(1+\beta_1 h_{r1} P_r)\left(1 + \frac{\beta_2 h_{r2} P_r}{1+\beta_1 h_{r2} P_r}\right)\right]^{\frac{1}{\rho}} - 1\right]. \tag{46}$$

Furthermore, I_4'' that corresponds to the case $h_{r1} < h_{r2}$ can be derived similarly.

Based on previous analysis, the optimal decision threshold ρ^* can be easily calculated from

$$\sum_{k=1}^{4} I_k'' = \sum_{k=1}^{6} I_k' \tag{47}$$

with the aid of one-dimensional search algorithm, while the maximum throughput is given by

$$\tau^* = \sum_{k=1}^{6}(I_k + I_k'). \tag{48}$$

V. Simulation Results and Discussion

In this section, the performance of the proposed buffer-aided relaying scheme with NOMA and direct links is evaluated and compared with the baseline scheme in [19]. Note that this baseline scheme maximizes the system throughput when R transmits information to two users by using buffer-aided relaying in a C-NOMA system without direct links. Also, note that in [19], fixed data rates (R_1 and R_2) for the communication between the relay and the users have been assumed, and when these rates cannot be achieved the relay adaptively chooses to transmit information to a single user. Moreover, we illustrate the performance of the proposed scheme when the direct links are not utilized, which is easily obtained by setting $h_{s1}^i = h_{s2}^i = 0, \forall i$ in the proposed analysis. We further assume that $\Omega_{xy} = 1/d_{xy}^2$, where d_{xy} is the distance between nodes x and y and that $d_{s1} = 0.9$, $d_{s2} = 1.1$, $d_{sr} = d$ and $d_{ri} = d_{si} - d, \forall i \in \{1, 2\}$, while $h_{sj}^i > h_{sk}^i \Rightarrow \alpha_j^i = 0.3$ and $h_{rj}^i > h_{rk}^i \Rightarrow \beta_j^i = 0.3$. Furthermore, we consider two cases for the distance d and fixed data rates of the baseline scheme, i.e., $d = 0.2, 0.3$ and $(R1, R2) = (1, 2), (2, 4)$ bps/Hz.

In Fig. 2, the achievable throughput of all considered schemes is illustrated versus the relay transmit power P_r, for $P_s = 2$ dB. As it can be observed, the proposed scheme with direct links performs better than both the baseline and the proposed one without direct links, for the whole range of P_r and for both values of d. The reason for this is twofold: i) the adaptive selection between the direct and relaying links increases the average gain of the utilized channels, ii) the performance is less prone to buffer-stability constraint, since the source can also directly transmit information to the users. Moreover, it is notable that the performance of the baseline scheme is also limited by the constraint of fixed data rates from the relay to users. For example, in the region of low P_r, higher fixed rates reduce throughput, since the system has less opportunities to perform NOMA. This also justifies the higher performance achieved by the proposed scheme, even when the direct links are not used. Furthermore, the performance of all schemes is affected by the relay placement, while the preferred value of d reduces with the increase of P_r. This is because as P_r increases, the bottleneck is due to the S-R link.

Similar observations can be made in Fig. 3, where the throughput is plotted versus P_s, assuming $P_s = P_r$. However, due to the increase of the transmit power of both S and R, in contrast to Fig. 2, the proposed scheme with direct links does

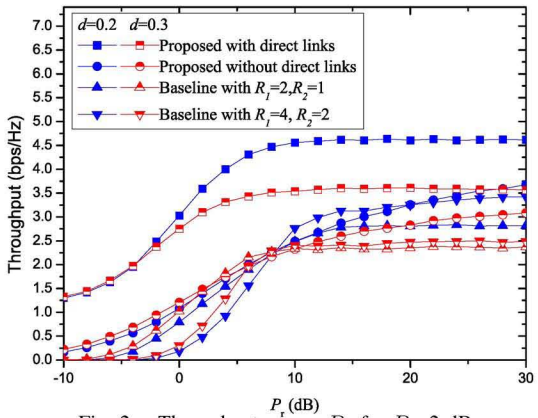

Fig. 2. Throughput versus P_r for P_s=2 dB

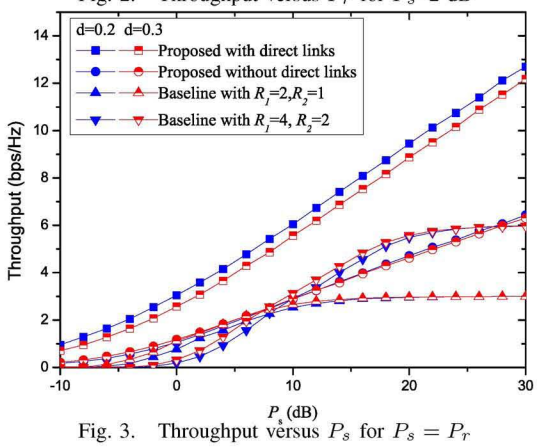

Fig. 3. Throughput versus P_s for $P_s = P_r$

not meet a floor in the range of $[-10, 30]$ dB. Also, in this case, the impact of d on the system throughput is less severe.

VI. Conclusion

In this paper, we have investigated buffer-aided relaying for a downlink NOMA system with direct links. In order to take advantage of the extra degrees of freedom, three different transmission modes have been introduced. Targeting at the throughput maximization, the corresponding optimization problem has been formulated and optimally solved, deriving appropriate theoretical expressions for the modes selection and maximum throughput. Finally, simulation results have illustrated the superiority of the proposed scheme compared to the baseline scheme [19].

VII. Acknowledgment

This work was supported by the Sichuan Science and Technology Program under Grant 2017HH0035, the NSFC project under Grant 61501382, and the Fundamental Research Funds for the Central Universities under Grant 2682018CX27. (Xianfu Lei is the corresponding author of this paper.)

References

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

[2] S. M. R. Islam, N. Avazov, O. A. Dobre, and K. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Commun. Surveys Tuts*, vol. 19, no. 2, pp. 721–742, Secondquarter 2017.

[3] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, C. I, 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.

[4] X. Chen, Z. Zhang, C. Zhong, R. Jia, and D. W. K. Ng, "Fully non-orthogonal communication for massive access," *IEEE Trans. Commun.*, vol. 66, no. 4, pp. 1717–1731, Apr. 2018.

[5] *Study on Non-Orthogonal Multiple Access (NOMA) for NR*, 3GPP, 2018, Rel. 15.

[6] 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.

[7] H. Zhang, B. Wang, C. Jiang, K. Long, A. Nallanathan, V. C. M. Leung, and H. V. Poor, "Energy efficient dynamic resource optimization in noma system," *IEEE Trans. Wireless Commun.*, vol. 17, no. 9, pp. 5671–5683, Sep. 2018.

[8] Z. Ding, R. Schober, and H. V. Poor, "A general mimo framework for NOMA downlink and uplink transmission based on signal alignment," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4438–4454, Jun. 2016.

[9] Z. Zhang, Z. Ma, Y. Xiao, M. Xiao, G. K. Karagiannidis, and P. Fan, "Non-orthogonal multiple access for cooperative multicast millimeter wave wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 8, pp. 1794–1808, Aug. 2017.

[10] M. Mohammadi, H. A. Suraweera, and C. Tellambura, "Uplink/downlink rate analysis and impact of power allocation for full-duplex cloud-rans," *IEEE Trans. Wireless Commun.*, vol. 17, no. 9, pp. 5774–5788, Sep. 2018.

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

[12] 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.

[13] Z. Ding, H. Dai, and H. V. Poor, "Relay selection for cooperative NOMA," *IEEE Wireless Commun. Lett.*, vol. 5, no. 4, pp. 416–419, Aug. 2016.

[14] Z. Yang, Z. Ding, Y. Wu, and P. Fan, "Novel relay selection strategies for cooperative NOMA," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 10 114–10 123, Nov. 2017.

[15] 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.

[16] T. N. Do, D. B. da Costa, T. Q. Duong, and B. An, "Improving the performance of cell-edge users in NOMA systems using cooperative relaying," *IEEE Trans. Commun.*, vol. 66, no. 5, pp. 1883–1901, May 2018.

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

[18] N. Zlatanov, R. Schober, and P. Popovski, "Buffer-aided relaying with adaptive link selection," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 8, pp. 1530–1542, Aug. 2013.

[19] S. Luo and K. C. Teh, "Adaptive transmission for cooperative NOMA system with buffer-aided relaying," *IEEE Commun. Lett.*, vol. 21, no. 4, pp. 937–940, Apr. 2017.

[20] Q. Zhang, Z. Liang, Q. Li, and J. Qin, "Buffer-aided non-orthogonal multiple access relaying systems in rayleigh fading channels," *IEEE Trans. Commun.*, vol. 65, no. 1, pp. 95–106, Jan. 2017.

[21] N. Nomikos, T. Charalambous, D. Vouyioukas, G. K. Karagiannidis, and R. Wichman, "Relay selection for buffer-aided non-orthogonal multiple access networks," in *Proc. IEEE Globecom Workshops*, Dec. 2017, pp. 1–6.

[22] N. Zlatanov, R. Schober, and L. Lampe, "Buffer-aided relaying in a three node network," in *Proc. IEEE International Symposium on Information Theory Proceedings*, Jul. 2012, pp. 781–785.

[23] R. Liu, P. Popovski, and G. Wang, "On buffer-aided multiple-access relay channel," *IEEE Commun. Lett.*, vol. 20, no. 10, pp. 2051–2054, Oct. 2016.

[24] P. D. Diamantoulakis, K. N. Pappi, G. K. Karagiannidis, H. Xing, and A. Nallanathan, "Joint downlink/uplink design for wireless powered networks with interference," *IEEE Access*, vol. 5, pp. 1534–1547, 2017.