Distributed Uplink-NOMA for Cloud Radio Access Networks

Koralia N. Pappi, *Member, IEEE*, Panagiotis D. Diamantoulakis, *Student Member, IEEE*, and George K. Karagiannidis, *Fellow, IEEE*

***Abstract*— We propose and study the application of distributed non-orthogonal multiple access (NOMA) for the uplink of cloud radio access networks. By considering that the remote radio heads are able to exchange digital information through high capacity links of the cloud, they can cooperate in performing interference cancellation, enhancing the users' achievable rate region. The achievable rates are presented in simple closed form, while simulations show that distributed NOMA can offer substantial improvement over benchmark schemes, by exploiting the cloud capabilities.**

***Index Terms*— NOMA, distributed systems, cloud radio access.**

I. Introduction

THE evolving fifth generation (5G) is envisioned to deal with an expected thousandfold increase in total mobile broadband data and a hundred fold increase in connected devices. These challenges are expected to be addressed by adopting a multitier heterogeneous architecture, which will leverage the integration of the cloud in future networks.

The limited spectral resources and the need of evolved interference management, have shifted focus on non-orthogonal multiple access (NOMA), which was proved to increase spectral efficiency, while it has been recognized as a promising multiple access technique for the 5G networks [1]–[3]. NOMA is substantially different from orthogonal multiple access schemes, since its basic principle is that the users can achieve multiple access by exploiting the power domain. Consequently, multi-user detection techniques are required to retrieve the users' signals at the receiver, such as joint decoding or successive interference cancellation (SIC). In [3], Al-Imari *et al.* investigate the use of NOMA for the uplink of a network, consisting of traditional nodes with fixed energy supplies. Uplink NOMA is able to achieve the system capacity upper bound, and it can also be used as a means to improve fairness among the users. However, decoding order optimization is a critical issue in fairness aware uplink NOMA systems [3], thus a set of different decoding orders can be used for corresponding fractions of time, by using *time-sharing* [4]. The application of multiple-input multiple-output (MIMO) techniques to NOMA has been considered in [5].

In future cloud radio access networks (CRANs) [6], a system of multiple distributed antennas is usually considered, which are connected to powerful processing units via high capacity links (e.g. optic fibers). This architecture enables the easier application of Coordinated Multi-Point (CoMP) [7], i.e., joint processing by distributed antennas. The main contribution of this paper is the application of distributed uplink NOMA with time-sharing (NOMA-TS), i.e., the introduction of SIC in C-RAN with distributed single antennas, which exploit the concept of cloud coordination, effectively acting as a distributed MIMO system.

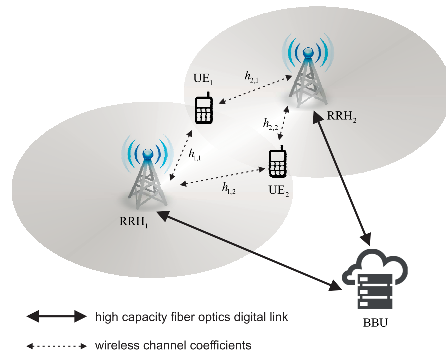


Fig. 1. System model of a cloud base station with two RRHs and two UEs.

II. System Model

We consider a network of M small cells, where in each cell, one remote radio head (RRH$_i$, $i = 1, 2, \ldots, M$) is situated at the center. All RRHs are connected with digital backhaul links to a baseband unit (BBU), and their capacity is assumed to be orders of magnitude higher than the capacity of the wireless links. We adopt a partially centralized cloud architecture [6], where the RRHs have decoding capabilities, and the BBU is able to perform complex calculations or centralized decisions. Therefore, we consider that the RRHs can exchange decoded messages with the BBU, without errors. The network serves N users with single antennas, which are referred to as user equipment (UE$_j$, $j = 1, 2, \ldots, N$). The UEs simultaneously send their messages during uplink, employing NOMA-TS. The communication process consists of time frames of unitary duration, during which the channel coefficient, $h_{i,j}$, from UE$_j$ to RRH$_i$, is constant. We consider that $h_{i,j}$ represents both small scale fading and path loss and it is perfectly estimated and known at the BBU, while $g_{i,j} = |h_{i,j}|^2$. The system model under investigation is depicted in Fig. 1, for the case of $M = N = 2$. Let P_j be the transmit power of UE$_j$, x_j the transmit signal of UE$_j$, and s_j the information-carrying symbol, assumed to belong to a zero-mean unit-variance Gaussian codebook. The received signal by RRH$_i$ is given by

$$y_i = \sum_{j=1}^{N} h_{i,j} x_j + n_i = \sum_{j=1}^{N} h_{i,j} \sqrt{P_j} s_j + n_i, \tag{1}$$

where n_i is additive white Gaussian noise (AWGN) at RRH$_i$.

Manuscript received April 21, 2017; revised June 11, 2017; accepted June 12, 2017. Date of publication June 29, 2017; date of current version October 7, 2017. The associate editor coordinating the review of this paper and approving it for publication was T. Ngatched. *(Corresponding author: George K. Karagiannidis.)*

K. N. Pappi is with the Electrical and Computer Engineering Department, Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece, and also with Intracom S. A. Telecom Solutions, GR-57001 Thessaloniki, Greece (e-mail: kpappi@auth.gr).

P. D. Diamantoulakis and G. K. Karagiannidis are with the Electrical and Computer Engineering Department, Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece (e-mail: padiaman@auth.gr; geokarag@auth.gr).

Digital Object Identifier 10.1109/LCOMM.2017.2721366

III. Proposed Scheme

Uplink NOMA at RRH_i can be employed with fixed decoding order of the UEs' messages, or by applying *time-sharing*[1] [4], where multiple decoding orders are assumed. In this work, we assume a time-sharing scheme. Let $\mathbf{j}_\upsilon = [j_1, j_2, \ldots, j_N]$ be the decoding order vector, where $\upsilon = 1, \ldots N!$ enumerates all possible decoding orders and $j_n \in \{1, 2, \ldots, N\}$ is the index of the UE, whose message is decoded in the n^{th} order, e.g. $j_2 = 4$ means that the message of UE_4 is decoded 2^{nd}. Further, let $t(\mathbf{j}_\upsilon)$ be the time sharing fraction when the specific decoding order is applied, and $\sum_\upsilon t(\mathbf{j}_\upsilon) = 1$.

In contrast to an actual MIMO receiver where receive beamforming is feasible [8], distributed RRHs of a partially centralized CRAN architecture decode the received signals, not allowing their observations to be jointly processed through beamforming as e.g. in [9]. This is not possible in the partially centralized C-RAN setup; however, we can consider that the RRHs can exchange decoded messages. Note that this work focuses on the system capacity, which is a prerequisite for the investigation of practical systems with imperfect channel estimates, SIC faults, etc. Therefore, we assume that each RRH can use correctly decoded messages from other RRHs, in order to perform SIC. Regarding the number of exchanged messages between RRHs, the worst case scenario is that every UE message is decoded by a different RRH, in which case, a maximum of $\frac{N^2-N}{2}$ message exchanges may occur, which is not considered a bottleneck, since the backhaul capacity is assumed to be much higher than that of the wireless medium.

Let $\mathbf{i}_\omega = [i_1, i_2, \ldots, i_N]$ be the decoding RRH vector, where $\omega = 1, \ldots M^N$ enumerates the possible vectors, and $i_n \in \{1, 2, \ldots, M\}$ is the index of the RRH that decodes the message of UE_{j_n}. We denote by $\mathcal{M}(\mathbf{i}_\omega)$ the mode when the RRHs decode the UEs' messages according to the vector $\mathbf{i}_\omega$. Assuming specific vectors $\mathbf{i}_\omega$ and $\mathbf{j}_\upsilon$, the sum rate achieved by using the mode $\mathcal{M}(\mathbf{i}_\omega)$ during $t(\mathbf{j}_\upsilon)$ can be evaluated as

$$\mathcal{R}_{sum}|_{\mathbf{i}_\omega,\mathbf{j}_\upsilon} = \log_2\left(\left(1 + g_{i_N,j_N} p_{j_N}\right) \prod_{n=1}^{N-1} \left(1 + \frac{g_{i_n,j_n} p_{j_n}}{1+\sum\limits_{k=n+1}^{N} g_{i_n,j_k} p_{j_k}}\right)\right) \tag{2}$$

where $p_j = \frac{P_j}{N_0}$, and N_0 is the power spectral density of the additive noise. The expression in (2) can be easily extracted by summing the achievable rate for each UE, where the message of UE_{j_N} is decoded last by RRH_{i_N}, and the message of every other UE, e.g. UE_{j_n}, is decoded by RRH_{i_n}, considering the messages of all UE_{j_k} where $k > n$ as interference. The only difference of (2) from the sum rate of a single base station employing NOMA is that the receive antenna for the message of each UE_{j_n} is RRH_{i_n}, which leads to different channel coefficients according to which RRH decodes each message.

Special Case for $N = 2$, $M = 2$: Let $\mathbf{j}_1 = [1, 2]$ and $\mathbf{j}_2 = [2, 1]$. Thus, there are two possible decoding orders. During $t(\mathbf{j}_1)$, the message of UE_1 is decoded first, considering the message of UE_2 as interference. The message of UE_2 is decoded second and free from interference. For a given mode $\mathcal{M}(\mathbf{i}_\omega)$ where $\mathbf{i}_\omega = [i_1, i_2]$, the rates for UE_1 and UE_2 are calculated as

$$\mathcal{R}_1|_{\mathbf{i}_\omega,\mathbf{j}_1} = \log_2\left(1 + \frac{g_{i_1,1}p_1}{1 + g_{i_1,2}p_2}\right),$$
$$\mathcal{R}_2|_{\mathbf{i}_\omega,\mathbf{j}_1} = \log_2\left(1 + g_{i_2,2}p_2\right). \tag{3}$$

During $t(\mathbf{j}_2)$, the message of UE_2 is decoded first, considering the message of UE_1 as interference. The message of UE_1 is decoded free from interference. Similarly, the rates for UE_1 and UE_2 are calculated as

$$\mathcal{R}_1|_{\mathbf{i}_\omega,\mathbf{j}_2} = \log_2\left(1 + g_{i_1,1}p_1\right),$$
$$\mathcal{R}_2|_{\mathbf{i}_\omega,\mathbf{j}_2} = \log_2\left(1 + \frac{g_{i_2,2}p_2}{1 + g_{i_2,1}p_1}\right). \tag{4}$$

Observing (3) and (4), it can be easily concluded that the sum rate for each of the two decoding orders can be evaluated as

$$\mathcal{R}_{sum}|_{\mathbf{i}_\omega,\mathbf{j}_1} = \log_2\left(1 + g_{i_1,1}p_1\frac{1 + g_{i_2,2}p_2}{1 + g_{i_1,2}p_2} + g_{i_2,2}p_2\right),$$
$$\mathcal{R}_{sum}|_{\mathbf{i}_\omega,\mathbf{j}_2} = \log_2\left(1 + g_{i_1,1}p_1 + g_{i_2,2}p_2\frac{1 + g_{i_1,1}p_1}{1 + g_{i_2,1}p_1}\right). \tag{5}$$

Note that the resulting rates in (5) have the same form as the genie-aided outer bound in [10].

IV. Optimal Mode Selection and Achievable Sum Rate

During each time-sharing fraction $t(\mathbf{j}_\upsilon)$, only one mode maximizes the achievable sum rate. Therefore, the maximized sum rate for given time-sharing arguments is given by

$$\mathcal{R}_{sum} = \sum_{\upsilon=1}^{N!} t(\mathbf{j}_\upsilon) \max_\omega \left(\mathcal{R}_{sum}|_{\mathbf{i}_\omega,\mathbf{j}_\upsilon}\right), \tag{6}$$

while, if the objective is the maximization of the sum rate, (e.g. in contrast to the maximization of equal UE rates), then

$$R_{sum,\max} = \max_{t(\mathbf{j}_\upsilon),\forall\upsilon} \mathcal{R}_{sum}, \quad s.t. \sum_\upsilon t(\mathbf{j}_\upsilon) = 1. \tag{7}$$

Due to the expression in (6), if the objective is to maximize the sum rate, only one time fraction will be equal to 1, while all other time fractions will be zero. More specifically, (7) leads to

$$R_{sum,\max} = \max_\upsilon \left(\max_\omega \left(\mathcal{R}_{sum}|_{\mathbf{i}_\omega,\mathbf{j}_\upsilon}\right)\right). \tag{8}$$

According to the partially centralized C-RAN architecture, the BBU is responsible for all complex centralized decisions [11], i.e. in this case accumulating all channel state information, calculating $\max_\omega \left(\mathcal{R}_{sum}|_{\mathcal{M}(\mathbf{i}_\omega),t(\mathbf{j}_\upsilon)}\right)$ for each $\mathbf{j}_\upsilon$, exchanging decoded messages between all RRHs, and - if applicable - optimizing $t(\mathbf{j}_\upsilon)$ $\forall\upsilon$. For M RRHs and N UEs, the BBU should calculate (2) for all possible $\mathbf{i}_\omega$ and $\mathbf{j}_\upsilon$, which are $M^N \times N!$ calculations. However, these calculations can be substantially minimized, due to the following observation. Based on the rates presented in the previous section, for a specific channel realization and for each decoding order, only one mode of operation achieves the best user rates, that is

$$\mathbf{i}_{\omega,best}|_{\mathbf{j}_\upsilon} = \arg\max_\omega \left(\mathcal{R}_{sum}|_{\mathbf{i}_\omega,\mathbf{j}_\upsilon}\right). \tag{9}$$

However, optimality of some $\mathbf{i}_\omega$ during one decoding order, restricts the candidate optimal modes for other decoding orders, since the values of (9) are not independent from each

[1] Not to be confused with *time-division*, which is orthogonal. Time-sharing allows multiple decoding orders for SIC within the same frame.

TABLE I
OPTIMAL MODE COMBINATIONS

$t(\mathbf{j}_1)$ (rows), $t(\mathbf{j}_2)$ (columns)	$\mathcal{M}(\mathbf{i}_1)$	$\mathcal{M}(\mathbf{i}_2)$	$\mathcal{M}(\mathbf{i}_3)$	$\mathcal{M}(\mathbf{i}_4)$
$\mathcal{M}(\mathbf{i}_1)$	✓	✓	✗	✗
$\mathcal{M}(\mathbf{i}_2)$	✗	✓	✗	✓
$\mathcal{M}(\mathbf{i}_3)$	✓	✗	✓	✗
$\mathcal{M}(\mathbf{i}_4)$	✗	✗	✓	✓

other, for each $\mathbf{j}_\upsilon$. For the special case of $M = N = 2$, the total number of calculations is $M^N \times N! = 8$. There are two decoding orders and four possible modes of operation, $\mathbf{i}_1 = [1, 1]$, $\mathbf{i}_2 = [1, 2]$, ,$\mathbf{i}_3 = [2, 1]$, and $\mathbf{i}_4 = [2, 2]$. During the decoding order $\mathbf{j}_1 = [1, 2]$, based on (3) it is

$$[i_1, i_2]_{best}|_{\mathbf{j}_1} = \arg\max_{i_1,i_2} \left(\left(1 + \frac{g_{i_1,1}p_1}{1 + g_{i_1,2}p_2}\right)\left(1 + g_{i_2,2}p_2\right)\right) \tag{10}$$

Similarly, for the decoding order $\mathbf{j}_2 = [2, 1]$, based on (4) it is

$$[i_1, i_2]_{best}|_{\mathbf{j}_2} = \arg\max_{i_1,i_2} \left(\left(1 + g_{i_1,1}p_1\right)\left(1 + \frac{g_{i_2,2}p_2}{1 + g_{i_2,1}p_1}\right)\right) \tag{11}$$

Based on (10) and (11), not all mode combinations for the two time fractions are optimal, because these maximal values cannot be satisfied simultaneously by any channel realization for some mode combinations. The achievable mode combinations are indicated in Table I. For example, the combination of $\mathcal{M}_{\mathbf{i}_1}$ during $t(\mathbf{j}_1)$, where $\mathbf{i}_1 = [1, 1]$ and $\mathcal{M}_{\mathbf{i}_4}$ during $t(\mathbf{j}_2)$, where $\mathbf{i}_4 = [2, 2]$, is not optimal; if $\mathcal{M}_{\mathbf{i}_1}$ is optimal during $t(\mathbf{j}_1)$, then according to (10),

$$\frac{g_{1,1}p_1}{1 + g_{1,2}p_2} > \frac{g_{2,1}p_1}{1 + g_{2,2}p_2}, \quad g_{1,2}p_2 > g_{2,2}p_2, \tag{12}$$

which leads to the conclusion that it must also hold that $g_{1,1}p_1 > g_{2,1}p_1$. However, if $\mathcal{M}_{\mathbf{i}_4}$ was optimal during $t(\mathbf{j}_2)$, it should hold from (11), that $g_{2,1}p_1 > g_{1,1}p_1$, which contradicts with the above. The rest of the combinations can be excluded in a similar manner. This dependency between optimal modes for each decoding order can significantly reduce the search space for the BBU operation. In the case of $N = M = 2$, with the use of Table I, only 6 out of 8 calculations are needed.

V. ENHANCED CAPACITY REGIONS

In this section, we will present the enhancement of the capacity region based on the proposed method for message exchange. In contrast to the capacity region of a single antenna, where the sum rate is constant regardless of the time-sharing arguments, depending on the maximal values in (6), it is obvious that the sum rate may depend on the selection of $t(\mathbf{j}_\upsilon)$. However, we will prove that the achieved capacity region is always greater or at least equal to the one achieved by only one RRH.

Theorem 1: The sum rate given in (6), for any selection of time sharing arguments $t(\mathbf{j}_\upsilon)$, is greater than or equal to the sum rate achieved by the best of the all RRHs, for the same time sharing arguments.

Proof: In order to prove the above, it is enough to express the sum rate, when only one RRH is selected to decode all messages, as

$$\mathcal{R}_{sum,RRHbest} = \max_{w:i_1=i_2=\cdots=i_N} \left(\sum_{\upsilon=1}^{N!} t(\mathbf{j}_\upsilon)\mathcal{R}_{sum}|_{\mathbf{i}_w,\mathbf{j}_\upsilon}\right), \tag{13}$$

since, when only one RRH decodes all messages, then all indices in $\mathbf{i}_w$ must be equal to the index of the best RRH. It is obvious that

$$\begin{aligned}\sum_{\upsilon=1}^{N!} t(\mathbf{j}_\upsilon) \max_{w} \mathcal{R}_{sum}|_{\mathbf{i}_w,\mathbf{j}_\upsilon} &\geq \max_{w} \sum_{\upsilon=1}^{N!} t(\mathbf{j}_\upsilon)\mathcal{R}_{sum}|_{\mathbf{i}_w,\mathbf{j}_\upsilon} \\ &\geq \max_{w:i_1=i_2=\cdots=i_N} \left(\sum_{\upsilon=1}^{N!} t(\mathbf{j}_\upsilon)\mathcal{R}_{sum}|_{\mathbf{i}_w,\mathbf{j}_\upsilon}\right)\end{aligned} \tag{14}$$

since $w : i_1 = i_2 = \cdots = i_N$ is a subset of all possible values for w, and therefore $\mathcal{R}_{sum,RRHbest} \leq \mathcal{R}_{sum}$. ■

As an example, we will examine the possible extension of the capacity region for the case of $M = N = 2$. Based on the various mode combinations, as shown in Table I, we can discern three cases, regarding the combined capacity region of the system: i) it remains unchanged and equal to the capacity region of the best RRH, ii) the rate is improved for one of the two UEs, iii) both UEs rates are improved. We consider three examples, where we set $p_1 = p_2 = 10$. For all schemes, the enhanced capacity region is compared to the capacity regions of the two RRHs, but also to the capacity region achieved by TDMA, when each UE's message is decoded by the RRHs with the best channel state.

A. Case 1 No Enhancement

Let $g_{1,1} = 0.6$, $g_{1,2} = 0.2$, $g_{2,1} = 0.3$, and $g_{2,2} = 0.1$. Based on (9), the best mode for both $t(\mathbf{j}_1)$ and $t(\mathbf{j}_1)$ is $\mathcal{M}(\mathbf{i}_1)$. This means that both UEs' messages are decoded by RRH_1. As seen in Fig. 2(a), the combined capacity region coincides with the capacity region of RRH_1. For the specific channel realization, using RRH_2 will provide no further benefit.

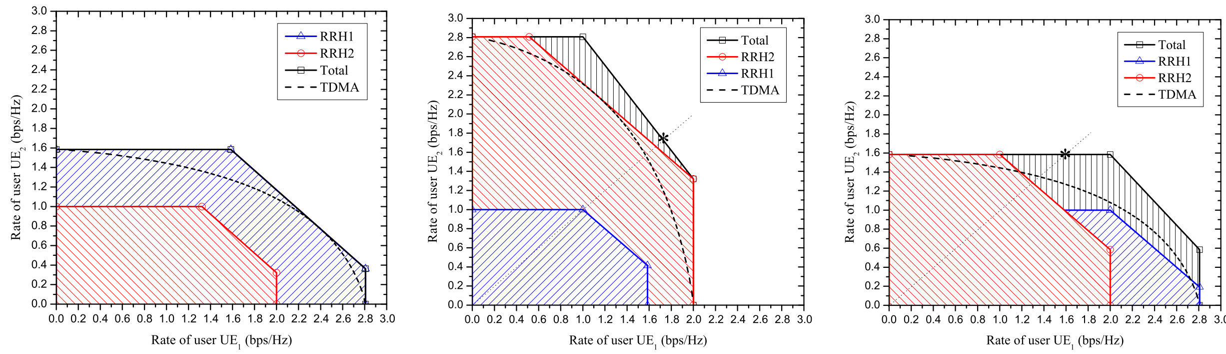
B. Case 2 One-Side Enhancement

Let $g_{1,1} = 0.2$, $g_{1,2} = 0.1$, $g_{2,1} = 0.3$, and $g_{2,2} = 0.6$. The best mode for $t(\mathbf{j}_1)$ is $\mathcal{M}(\mathbf{i}_2)$ and for $t(\mathbf{j}_2)$ is $\mathcal{M}(\mathbf{i}_4)$. This means that all messages are decoded by RRH_2, apart from the message of UE_1 during $t(\mathbf{j}_1)$, when it is decoded by RRH_1 and then forwarded to RRH_2. As seen in Fig. 2(b), the combined capacity region is enhanced regarding the rate of UE_1.

C. Case 3 Two-Side Enhancement

Let $g_{1,1} = 0.6$, $g_{1,2} = 0.1$, $g_{2,1} = 0.3$, and $g_{2,2} = 0.2$. The best mode for time fractions is $\mathcal{M}(\mathbf{i}_2)$. This means that the message of UE_1 is decoded by RRH_1 while the message of UE_2 is decoded by RRH_2 for both time fractions. However, in contrast to plain cell association, the two RRHs exchange decoded messages for the implementation of SIC. As it can be seen in Fig. 2(c), the combined capacity region is enhanced regarding the rates of both messages.

It is evident from the above that the most beneficial capacity region enhancement occurs when the individual capacity regions of the RRHs differ the most from each other. Also note that, with this capacity region enhancement, apart from the maximum sum rate improvement, an improvement is also achieved when users are required to operate with common rate. The points of the capacity region that maximize the equal rate for Cases 2 and 3 are marked with asterisk, while they can be achieved by selecting proper time-sharing factors.



(a) Case 1 - No enhancement (b) Case 2 - One-side enhancement (c) Case 3 - Two-side enhancement

Fig. 2. Capacity region examples with RRH coordination.

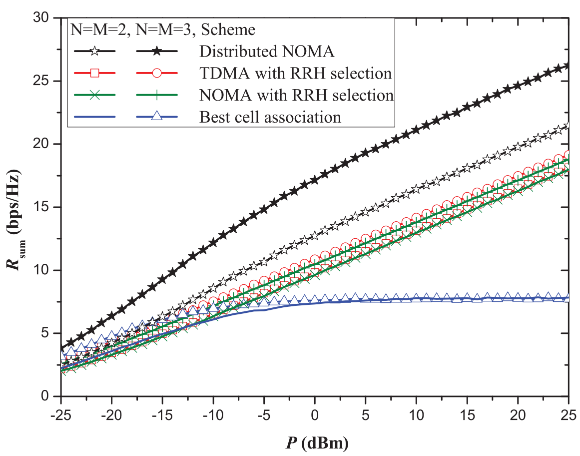


Fig. 3. Average sum rate improvement offered by distributed NOMA for $N = M = 2$ and $N = M = 3$.

VI. Simulation Results and Conclusion

We consider a system with i) $M = N = 2$ and ii) $M = N = 3$. The relative distances between the UEs and the RRHs are $d_{i,j} = 30$ m, when $i = j$, and $d_{i,j} = 60$ m, when $i \neq j$. $d_{1,2} = 8$ m, where $d_{i,j}$ is the distance of UE$_j$ from RRH$_i$. We consider that $h_{i,j}$ is the combination of Rayleigh fast fading and path loss, where the 3GPP pico cell path loss model in [12] has been adopted, considering $W = 1$ MHz bandwidth, noise spectral density -174 dBm/Hz, and noise figure 6 dB.

In Fig. 3 we present the maximum sum rate, $\mathcal{R}_{sum,\max}$, which can be achieved by the proposed distributed NOMA scheme. Its performance is compared against three schemes: i) the selection of the best RRH, which receives all UEs' messages using NOMA, ii) the best cell association, where each UE is assigned to a RRH, and each RRH decodes only one message, regarding the rest as interference, iii) TDMA with RRH selection, where each UE is decoded by its best RRH. For TDMA with RRH selection, we also consider that the sum rate is maximized through optimal degrees of freedom (DOF) allocation [13]. It can be easily seen that the proposed scheme offers substantial performance improvement, compared to all three benchmark schemes. What is more interesting is the fact that this gain is greater, as more UEs and RRHs are added. Regarding the benchmark schemes, it is observed that cell association, which does not allow any coordination or a multiple access scheme, reaches a floor for high SNR values, due to the prevailing interference. Accordingly, as the value of SNR or the number of users grows, the RRH selection schemes for NOMA and TDMA cannot offer such gains, because of the lack of cooperation between receiving RRHs. Finally, simulations for the case of $N = M = 2$ showed that, for all values of transmit power above $p = 0$ dBm, 66% of channel realizations led to two-side enhancement, 32% led to one-side enhancement, while less than 1% of channel realizations led to no enhancement of the capacity region.

The proposed scheme offers an efficient solution for implementing coordinated distributed multiple access, by exploiting the benefits offered by C-RAN. It can be effectively combined with user grouping and frequency selection schemes, while it can also be extended to accommodate MIMO RRH units.

References

[1] "5G radio access: Requirements, concept and techologies," NTT DoCoMo, Tokyo, Japan, White Paper, Jul. 2014.

[2] Z. Ding *et al.*, "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.

[3] M. Al-Imari *et al.*, "Uplink non-orthogonal multiple access for 5G wireless networks," in *Proc. 11th Int. Symp. Wireless Commun. Syst. (ISWCS)*, Aug. 2014, pp. 781–785.

[4] P. D. Diamantoulakis *et al.*, "Wireless-powered communications with non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8422–8436, Dec. 2016.

[5] Z. Ding *et al.*, "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, Jan. 2016.

[6] "C-RAN: The road towards green RAN," China Mobile Res. Inst., Beijing, China, White Paper, Oct. 2011.

[7] M. Sawahashi *et al.*, "Coordinated multipoint transmission/reception techniques for LTE-advanced coordinated and distributed MIMO," *IEEE Wireless Commun.*, vol. 17, no. 3, pp. 26–34, Jun. 2010.

[8] Y. Yuan and Z. Ding, "The application of non-orthogonal multiple access in wireless powered communication networks," in *Proc. IEEE 17th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2016, pp. 1–5.

[9] Z. Ding and H. V. Poor, "The use of spatially random base stations in cloud radio access networks," *IEEE Signal Process. Lett.*, vol. 20, no. 11, pp. 1138–1141, Nov. 2013.

[10] G. Kramer, "Outer bounds on the capacity of Gaussian interference channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 3, pp. 581–586, Mar. 2004.

[11] K. N. Pappi *et al.*, "Cloud compute-and-forward with relay cooperation," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3415–3428, Jun. 2015.

[12] *European Telecoms Standards Institute (ETSI) Technical Report v9.0.0*, Standard ETSI TR 136 931, May 2011.

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