A Novel NOMA Solution with RIS Partitioning

Aymen Khaleel, *Student Member, IEEE* and Ertugrul Basar, *Senior Member, IEEE*

***Abstract*—Reconfigurable intelligent surface (RIS) empowered communications with non-orthogonal multiple access (NOMA) has recently become an appealing research direction for next-generation wireless communications. In this paper, we propose a novel NOMA solution with RIS partitioning, where we aim to enhance the spectrum efficiency by improving the ergodic rate of all users, and to maximize the user fairness. In the proposed system, we distribute the physical resources among users such that the base station (BS) and RIS are dedicated to serve different clusters of users. Furthermore, we formulate an RIS partitioning optimization problem to slice the RIS elements between the users such that the user fairness is maximized. The formulated problem is shown to be a non-convex and non-linear integer programming (NLIP) problem with a combinatorial feasible set, which is challenging to solve. Therefore, we exploit the structure of the problem to bound its feasible set and obtain a sub-optimal solution by sequentially applying three efficient search algorithms. Furthermore, we derive exact and asymptotic expressions for the outage probability. Simulation results clearly indicate the superiority of the proposed system over the considered benchmark systems in terms of ergodic sum-rate, outage probability, and user fairness performance.**

***Index Terms*—Reconfigurable intelligent surface (RIS), non-orthogonal multiple access (NOMA), user fairness, sum-rate.**

I. Introduction

NON-ORTHOGONAL multiple access (NOMA) has been regarded as one of the promising technologies to address the increasing demand for high data rates, massive connectivity, and spectrum efficiency associated with fifth-generation (5G) and beyond networks. Due to the use of higher carrier frequencies and in order to fulfill the verticals' specific requirements, local service areas are among the main envisioned deployment models in 5G [1]. Within this context, NOMA can play an important role in order to efficiently exploit the spectrum and serve higher numbers of users in such networks [2]. This is because NOMA allows the sharing of the same time/frequency/code resources between users and thus, enhances the spectrum efficiency and decreases the latency by allowing more users to be connected in the same time-frequency slot [3]. In power domain (PD)-NOMA [4], the difference in channel gains of different users is exploited and, accordingly, different power levels are assigned to different users by the use of superposition coding (SC) at the base station (BS) side. At the receiver side, each user employs the successive interference cancellation (SIC) technique to recover its own message. Compared to the orthogonal multiple access (OMA), it has been shown that NOMA has a superior performance in terms of the outage probability and the ergodic sum-rate [5], [6], [7].

Recently, reconfigurable intelligent surface (RIS)-assisted communication has received growing attention as a potential next-generation technology due to its promising capabilities to control the wireless propagation environment. RIS is an array of low-cost and passive reflecting elements that can be used to re-engineer the electromagnetic waves by adjusting the reflection coefficient of each element [8], [9]. Due to their promising advantages, RISs have been integrated to many existing wireless technologies as a main or supportive module in the communications system. In [10], two RIS-assisted space shift keying (SSK) schemes are proposed to enhance the performance of the classical SSK system in terms of the bit error rate and system throughput. In [11], the authors proposed RIS-based receive quadrature reflecting modulation by partitioning the surface into two parts, where the in-phase and quadrature components are sent from each part separately. In [12], an RIS is integrated to a conventional single-input single-output (SISO) system to achieve an ultra-reliable communication system. In [13], an RIS is used as a part of the transmitter to perform different types of index modulation (IM) and thus, obviating the need for a multi-antenna BS. These techniques are extended to multiple-input multiple-output (MIMO) systems in [14], where an RIS is used to replace the radio frequency (RF) chains in Alamouti's scheme, and to boost the data rate for vertical Bell Labs layered space-time (VBLAST) system by using IM.

A. Related Works

Recently, the joint operation of RISs with NOMA has appeared as an appealing research direction and the combination of the hardware capabilities of RISs and the SC technique of PD-NOMA has been investigated. In [15] and [16], the rate performance and user fairness of an RIS-assisted NOMA system are optimized by maximizing the minimum decoding signal-to-interference-plus-noise ratio (SINR) of all users, which is achieved by the joint optimization of the active and passive beamforming at the BS and the RIS, respectively. The authors in [17] considered an RIS-assisted multiple-input single-output (MISO) NOMA system where the cell-center users are served by using spatial division multiple access (SDMA) and the cell-edge users are served by RISs. In [18], by considering a priority-oriented design, an RIS-assisted SISO NOMA network is proposed, where the passive beamforming weights are designed at the RIS side in order to enhance the spectrum efficiency. In [19], the energy efficiency is enhanced in a downlink RIS-assisted NOMA system by jointly optimizing the user clustering, passive beamforming, and power allocation. The user fairness

Manuscript received April 21, 2021; revised August 10, 2021 and October 13, 2021; accepted November 7, 2021. This work was supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under Grant 120E401. (Corresponding author: Ertugrul Basar.)

The authors are with the Communications Research and Innovation Laboratory (CoreLab), Department of Electrical and Electronics Engineering, Koç University, Sariyer 36050, Istanbul, Turkey. Email: akhaleel18@ku.edu.tr, ebasar@ku.edu.tr

is considered in [20], where the authors investigated the joint optimization of power allocation, decoding order, and the RIS phase shifts to maximize the minimum user rate considering a total power constraint. In [21], the deployment and passive beamforming design of an RIS are investigated for an RIS-assisted MISO NOMA system, in order to maximize the energy efficiency under the constraint of preserving the individual data rate requirements for users. Joint optimization for the active beamforming matrices at the BS and the reflection coefficient vector at the RIS is utilized in [22] in order to minimize the total transmit power for a multi-cluster MISO NOMA networks. In [23], a multi-cluster RIS-assisted MIMO NOMA network is considered, where by designing its passive beamforming weights, the RIS is employed in a signal cancellation mode to eliminate the inter-cluster interference. The authors in [24] and [25] used stochastic geometry to investigate the coverage probability and ergodic rate of an RIS-assisted multi-cell NOMA networks for outdoor scenarios by using Poisson cluster process (PCP) model. The authors in [26] considered the combination of joint transmission coordinated multipoint (JT-CoMP) with the RIS technology in order to enhance the cell-edge user ergodic rate performance without degrading the performance of the cell-center user. In [27], the authors investigated the impact of the coherent and the random discrete phase-shifting designs on an RIS-assisted NOMA system. Finally, the authors in [28] considered the resource allocation problem in an RIS-assisted NOMA system, and jointly optimized the channel assignments, power allocation, decoding order, and RIS reflection coefficients, in order to maximize the system throughput.

B. Motivation and Contributions

In light of the above discussion, it can be noted that a common physical resource (PR) allocation scheme is followed by all of the previous works, which can be summarized as follows. The BS and RIS are both used to serve all users by using a single SC message, where all users are assumed to be in the field-of-view (FoV) of the RIS. Furthermore, by adjusting its reflection coefficients, the RIS is used as a single unit to serve all users jointly. Considering this PR allocation scheme, the main factor that limits the users' performance becomes the mutual interference that underlies the SC technique, which can be seen clearly when the users are deployed randomly in and out of the FoV of the RIS. In such a users' deployment scenario, not all users share the same PR (BS and RIS) and therefore, the use of a single SC message for all users adversely affects user fairness and unnecessarily amplifies the mutual interference between the users' messages.

Against this background, we propose an RIS-assisted novel NOMA system, where a more efficient PR allocation scheme is proposed to enhance user fairness and effectively mitigate the impact of the mutual interference between users. In the proposed system, the users are grouped into two clusters, Cluster 1 (C_1) contains all the users out of the FoV of the RIS, and Cluster 2 (C_2) contains the ones inside it. The BS is dedicated to serve the users in C_1 and the RIS is portioned into sub-surfaces, where each sub-surface is exploited to serve a different user in C_2. The main contributions of this paper can be summarized as follows:

- To the best of the authors' knowledge, this study considers the PR problem when the users are deployed in and out of the FoV of the RIS, for the first time. To address this issue, we propose a novel PR scheme that employs RIS partitioning to mitigate the mutual interference between the users in and out of the FoV of the RIS. This leads to an effective enhancement in the performance of all users in terms of ergodic rate, outage probability, fair distribution of the PRs among users, and simplifies the detection process.
- We formulate an RIS partitioning optimization problem to find a proper RIS slicing that maximizes the fairness among C_2 users. Although the fact that the formulated problem is a non-convex and non-linear integer programming (NLIP) one, we exploit the structure of the problem, specifically the nature of transmission over the RIS, to provide a sub-optimal solution with marginal performance degradation. Note that, unlike the partitioning schemes adopted in [29] and [30], to obtain a scalable optimization framework for the phase-shift adjustment of large RISs, we partition the RIS so that each sub-surface works as a modulator and beamformer simultaneously. In this way, each sub-surface sends an independent data stream to the user assigned to it independently from the other sub-surfaces.
- We derive the exact and asymptotic outage probability expressions for users in C_2. Accordingly, under the uniform partitioning scenario, we obtain the required number of RIS elements that need to be allocated for each user in C_2 to obtain a given outage probability value for all users.
- With comprehensive computer simulations, we compare our proposed system with four different benchmark schemes and show the superiority of our proposed system in terms of the ergodic sum-rate, outage probability, and user fairness.

The rest of the paper is organized as follows. In Section II, we introduce the system model and describe the transmission mechanism in detail. The outage probability and its asymptotic behaviour are formulated in Section III. In Section IV, we introduce our proposed RIS partitioning approach and provide the system performance analysis of two special cases for the proposed system in different deployment scenarios. Computer simulations are provided in Section V followed by the conclusions in Section VI.[1]

II. RIS Partitioning and Rotating: System Model

Consider a downlink NOMA system where single-antenna users are served by a single-antenna BS[2] and an RIS of $N = N_H N_V$ elements, where N_H and N_V denote the number

[1] *Notation*: Matrices and column vectors are denoted by an upper and lower case boldface letters, respectively. $\mathbf{X} \in \mathbb{C}^{m\times k}$ denotes a complex-valued matrix $\mathbf{X}$ with $m \times k$ size, where $\mathbf{X}^T$ is the transpose and $[\mathbf{X}]_{n,\tilde{n}}$ is the $(n,\tilde{n})$-th entry. $\mathbf{0}_N$, $\mathbf{I}_N$, $\binom{m}{k}$, $\lfloor\cdot\rfloor$, $\lceil\cdot\rceil$, and $\mathrm{mod}(\cdot)$ are the N-dimensional all-zeros column vector, the $N \times N$ identity matrix, the binomial coefficient, the floor, ceiling, and modulus functions, respectively. $x \sim \mathcal{CN}(0,\sigma^2)$ stands for complex Gaussian distributed random variable (RV) with mean $\mathrm{E}[x] = 0$ and variance $\mathrm{VAR}[x] = \sigma^2$. $\mathbb{R}$, $\mathbb{Z}^+$, and $|\mathcal{S}|$ are the set of real numbers, the set of positive integer numbers, and the cardinality of the set $\mathcal{S}$, respectively.

[2] The RIS is deployed in the direct line-of-sight (LoS) of the BS to compensate for the high path loss associated with the RIS. This leads to a rank-one BS-RIS channel, where no multiplexing gain can be achieved by using multiple transmit antennas. Furthermore, it has been shown that the array gain vanishes as the number of users increases irrespective of the number of transmit antennas and RIS size, and the number of users that can be efficiently served is one [31].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2021.3127725, IEEE Journal of Selected Topics in Signal Processing

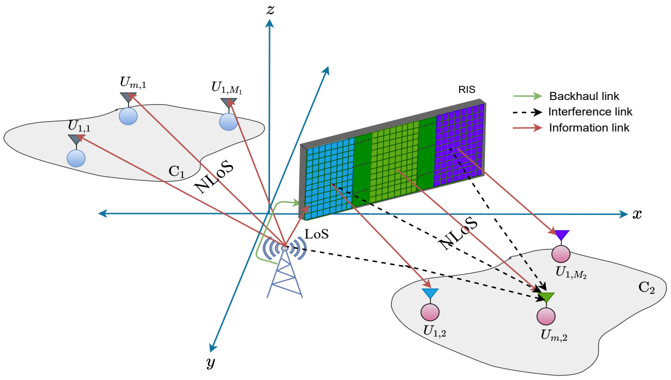


Fig. 1. RIS partitioning based NOMA system.

of elements per row and per column, respectively. The users are assumed to be grouped into two clusters[3], C_1 with M_1 users and C_2 with M_2 users, as shown in Fig. 1. The RIS is deployed close, yet in the far-field[4], to the BS in order to exploit the pure line-of-sight (LoS) channels and to make the use of a backhaul link from the BS to the RIS practical. Furthermore, denoting the m^{th} user in C_2 by $\mathrm{U}_{m,2}$, perfect channel state information (CSI) of the BS-RIS-$\mathrm{U}_{m,2}$ link for all users in C_2 and the channel gains of all users in C_1 need to be available at the BS side, under the quasi-static flat-fading channels assumption. Although the fact that the perfect CSI assumption is practically challenging, nevertheless, it is adopted by the vast majority of the RIS-assisted NOMA works in the literature [31], [34], [35]. Therefore, the CSI is assumed to be obtained by using one of the proposed methods in the literature [36], [37], and thus, the performance results provided in this study can serve as an upper bound to the ones achievable in practical implementation.

Instead of using a single SC message that combines all users' symbols, the C_1 and C_2 users' symbols are simultaneously transmitted over the BS direct link and the RIS reflection link, respectively, as follows.

A. The transmission of C_1 users' symbols

Let T_c denotes the duration of the channel coherence block, wherein all channels remain constant, while all channel coefficients are independent and identically distributed (i.i.d.) over different coherence blocks. Within T_c, the BS transmits x, which is the superposed signal of all symbols to be transmitted to M_1 users in C_1. As in conventional PD-NOMA, x is constructed as follows:

$$x = \sum_{m=1}^{M_1} \sqrt{P\zeta_m} x_m, \tag{1}$$

where P is the transmit power, ζ_m and x_m are the power allocation factor and the symbol to be transmitted to user m in C_1 ($\mathrm{U}_{m,1}$), respectively, where $\sum_{m=1}^{M_1} \zeta_m = 1$. Note that x can be represented in the form $x = ue^{-j\theta_{M_1}}$, where $u = |x|$ is the amplitude and $\theta_{M_1} = \arg(x)$ is the angle. From (1), it is clear that u and θ_{M_1} are both RVs, where $\theta_{M_1} \in [0, 2\pi)$ and $u \in \mathbb{R}$. Without loss of generality, in order to simplify our derivations, u is assumed to be a constant value that is equal to unity, which can be achieved by the proper design of the constellations used to obtain x_1 and x_2 for C_1 users [38], [39]. For example, consider the simple case of two users in C_1, where $x_1 \in \{\frac{t}{\sqrt{2}}(1+1j), \frac{t}{\sqrt{2}}(-1-1j)\}$ and $x_2 \in \{\frac{t}{\sqrt{2}}(1-1j), \frac{t}{\sqrt{2}}(-1+1j)\}$, then for any random values of ζ_1 and ζ_2 we obtain $u = |t|$, where $t \in \mathbb{R}$.

The signal received by each user m in C_1 is given by

$$\tilde{y}_m = \tilde{v}_m x + \tilde{z}_m, \tag{2}$$

where $\tilde{v}_m$ is the BS-$\mathrm{U}_{m,1}$ channel coefficient and $\tilde{z}_m$ is the additive white Gaussian noise (AWGN) sample at $\mathrm{U}_{m,1}$.

From (2), it can be noted that there is no interference from C_2 users' symbols received by C_1 users, due to the location of C_1 being behind the RIS. Although C_1 users still need to use SIC to recover their own signals, the SIC is performed with significantly less number of iterations due to the absence of C_2 users' interference. This simplifies the detection process and effectively reduces the error propagation associated with the SIC technique. In this way, the system performance for the M_1 users in C_1 follows the one of the conventional PD-NOMA, therefore, its analysis is omitted in this study and we focus on the system performance of C_2. Nevertheless, we still perform the numerical simulations in Section V for the users in C_1 along with C_2 users to show that the proposed system enhances the performance of both C_1 and C_2 users compared to the benchmark systems.

B. The transmission of C_2 users' symbols

Instead of using the SC technique to send the symbols of C_2 users in a single message from the BS, the symbol of each user m is sent over the RIS reflection link independent from the symbols belong to the other users. Therefore, the RIS is partitioned into M_2 sub-surfaces, where each sub-surface i has N_i elements and is allocated to serve a specific user m in C_2, for $i = m$, where $m = 1, 2, ...M_2$. Specifically, each sub-surface $i = m$ remodulates the same impinging signal x to reflect the $\mathrm{U}_{m,2}$ symbol over the RIS-$\mathrm{U}_{m,2}$ reflection link. Thus, by considering only a single reflection from the RIS elements [40] and within the same T_c, the signal received by each user m in C_2 can be obtained as follows:

$$y_m = \left[\mathbf{g}_{m,m}^T \mathbf{\Theta}_m \mathbf{h}_m + \sum_{\substack{i=1 \\ i \neq m}}^{M_2-1} \mathbf{g}_{i,m}^T \mathbf{\Theta}_i \mathbf{h}_i + v_m\right] x + z_m, \tag{3}$$

where $z_m \sim \mathcal{CN}(0, \sigma^2)$ is the AWGN sample at $\mathrm{U}_{m,2}$, v_m is the BS-$\mathrm{U}_{m,2}$ channel coefficient, $v_m \sim \mathcal{CN}(0, L_m^{\mathrm{BS}})$, under Rayleigh fading assumption [31], [41], where L_m^{BS} denotes the BS-$\mathrm{U}_{m,2}$ path gain. Assuming pure LoS links [41], $\mathbf{h}_i \in \mathbb{C}^{N_i \times 1}$ is the BS-(i^{th} sub-surface) LoS channel vector with $[\mathbf{h}_i]_n = \sqrt{L^{\mathrm{RISh}}} e^{-j\psi^{(n)}}$, where L^{RISh} denote the path gain, $\psi^{(n)} = \pi(n-1)\sin\psi_A \sin\psi_E$, where ψ_E and ψ_A denote the LoS elevation and azimuth angles of arrival (AoA)s at the RIS, respectively [31]. Since the RIS is deployed in the far-field of the BS, the path gain experienced by each element is assumed

[3] C_1 and C_2 are the main clusters that identify the users who are out of the FoV of the RIS and served by the BS, and the ones who are in the FoV of the RIS and served by the RIS, respectively. Therefore, it is possible to use a second level of clustering inside these two main clusters [32], however, the investigation of such a scenario is out of the scope of this study.

[4] This simplifies the system analysis, where for the the near-field assumption we need to consider the distances of the individual RIS elements from the BS and thus, the effective BS antenna area and the polarization mismatch associated with each RIS element [33].

to be the same, where the spacing between the elements is small compared to the BS-RIS distance. $\mathbf{g}_{i,m} \in \mathbb{C}^{N_i \times 1}$ is the RIS (i^{th} sub-surface)-U$_{m,2}$ channel vector, $[\mathbf{g}_{i,m}]_n = \sqrt{L_m^{\text{RISg}}} g_{i,m}^{(n)}$, where $g_{i,m}^{(n)}$ and L_m^{RISg} are the RIS (i^{th} sub-surface n^{th} element)-U$_{m,2}$ small-scale fading coefficient and path gain, respectively. $g_{i,m}^{(n)} = \beta_{i,m}^{(n)} e^{-j\phi_{i,m}^{(n)}}$, where $\beta_{i,m}^{(n)}$ and $\phi_{i,m}^{(n)}$ denote the channel amplitude and phase, respectively. Thus, the overall RIS-U$_{m,2}$ channel vector can be given as $\mathbf{g}_m = [\mathbf{g}_{1,m} \, ... \, \mathbf{g}_{i,m} \, ... \, \mathbf{g}_{M_2,m}]^T$, $\mathbf{g}_m \in \mathcal{CN}(\mathbf{0}_N, \mathbf{R}_{\text{RIS}})$, where $\mathbf{R}_{\text{RIS}} \in \mathbb{C}^{N \times N}$ is the RIS spatial correlation matrix. $\mathbf{\Theta}_i \in \mathbb{C}^{N_i \times N_i}$ is the matrix of reflection coefficients for the i^{th} sub-surface of the RIS, $[\mathbf{\Theta}_i]_n = \eta_i^{(n)} e^{j\Phi_i^{(n)}}$, where $\Phi_i^{(n)} \in [0, 2\pi)$ and $\eta_i^{(n)} = 1$, $\forall i, n$, assuming full reflection[5]. In what follows, we describe the process of remodulating the impinging signal x at the RIS sub-surfaces to reflect the symbols of the C$_2$ users.

By properly and independently adjusting its phase shifts, each sub-surface $i = m$ performs two roles, namely, maximizing the channel gain by making the BS-RIS-U$_{m,2}$ channel phases equal to zero (passive beamforming), and remodulating the impinging signal x to reflect the symbol to be transmitted to U$_{m,2}$. Note that since the transmission of the RIS to C$_2$ depends on the BS transmission to C$_1$, the symbol rate is the same for all users in both clusters. To serve user m in C$_2$, the phases of the $i = m$ sub-surface are adjusted such that the phase of each element is given as

$$\Phi_m^{(n)} = \phi_{m,m}^{(n)} + \psi^{(n)} + \theta_m + \theta_{M_1}, n = 1, ..., N_m, \quad (4)$$

where θ_m is the phase shift keying (PSK) symbol to be transmitted to user m in C$_2$, furthermore, $\Phi_m^{(n)}$ is assumed to be calculated at the BS and sent to the RIS controller over a backhaul link. By considering the close distance and the LoS link between the BS and the RIS, a wired/out-band wireless backhaul link can be used without affecting the useful bandwidth used in (1) [43], [44]. In this way, the BS and RIS perform a joint transmission synchronized by the backhaul link as in coordinated multi-point transmission [45], where the achievable capacity of the backhaul link is assumed to be much higher than the one of the RIS-C$_2$ link.

The PSK modulation scheme is adopted for the users in C$_2$ due to the phase-dependent amplitude variation associated with the RIS reflection coefficient, where it is difficult to realize modulation schemes with a non-constant envelope [42].

According to the RIS phase adjustment in (4), (3) can be re-expressed as

$$\begin{aligned} y_m =& \sqrt{L_m^{\text{RIS}}} \left[e^{j\theta_m} \sum_{n=1}^{N_m} \beta_{m,m}^{(n)} + \sum\nolimits_{\substack{i=1 \\ i \neq m}}^{M_2 - 1} \left[\sum_{n=1}^{N_i} \beta_{i,m}^{(n)} e^{j\bar{\Phi}_i^{(n)}} \right] \right] \\ &+ e^{j\theta_{M_1}} v_m + z_m, \end{aligned} \quad (5)$$

where $L_m^{\text{RIS}} = L_m^{\text{RISh}} L_m^{\text{RISg}}$,$\bar{\Phi}_i^{(n)} = -\phi_{i,m}^{(n)} + \Phi_i^{(n)}$, $\Phi_i^{(n)}$ corresponds to the phase adjustment of the other sub-surface i for $i \in \{1, ..., M_2\} \setminus \{m\}$, and it is given by

$$\Phi_i^{(n)} = \phi_{i,i}^{(n)} + \psi^{(n)} + \theta_i + \theta_{M_1}, n = 1, ..., N_i. \quad (6)$$

The first three terms in (5) represent the amplitudes of the constructive combining, the intra-cluster, and inter-clusters interference, respectively. Furthermore, thanks to the remodulation process at the RIS, it can be noted that there is no interference from C$_1$ users' symbols received by C$_2$ users over the RIS link.

In order to detect its PSK symbol, each user m in C$_2$ is assumed to perform maximum likelihood (ML) detection in the presence of the interference coming from the other sub-surfaces ($i \neq m$) and the one comes from the BS, which are irremovable [46] and considered as noise. In this way, for the detection process, user m in C$_2$ needs to know the overall sum of the channel gains of the m^{th} sub-surface only. This, in turn, significantly reduces the training overhead and limits the spectrum efficiency degradation associated with it [36], [37]. Furthermore, by properly choosing the RIS size [47], each user m in C$_2$ relies on the amplification gain provided by its own sub-surface ($i = m$) to overcome the irremovable interference. Thus, the SINR for user m in C$_2$ to decoded its own message is given by

$$\text{SINR}_m = \frac{A_m}{I_m + \frac{1}{\rho}}, \quad (7)$$

where $\rho = \frac{P}{\sigma^2}$ denotes the transmit SNR, A_m and I_m denote the signal and total interference powers, respectively, and they are, from (5), given as

$$A_m = \left| \sqrt{L_m^{\text{RIS}}} \sum_{n=1}^{N_m} \beta_{m,m}^{(n)} \right|^2, \quad (8)$$

$$I_m = \left| I_{\text{RIS}} + v_m \right|^2, \quad (9)$$

and I_{RIS} is given by

$$I_{\text{RIS}} = \sqrt{L_m^{\text{RIS}}} \sum\nolimits_{\substack{i=1 \\ i \neq m}}^{M_2 - 1} \left[\sum_{n=1}^{N_i} \beta_{i,m}^{(n)} e^{j\bar{\Phi}_i^{(n)}} \right]. \quad (10)$$

From (7), the instantaneous transmission rate for U$_{m,2}$ and the sum-rate for the all M_2 users can be calculated, respectively, as follows:

$$R_m = \log_2(1 + \text{SINR}_m), \quad (11)$$

$$R = \sum_{m=1}^{M_2} R_m = \sum_{m=1}^{M_2} \log_2(1 + \text{SINR}_m). \quad (12)$$

In the following section we derive the outage probability for a given user m in C$_2$.

Remark 1. *It can be verified from* (2) *and* (5) *that the number of users in C_1 and C_2 has no effect on the performance of the users belong to the other cluster, where the BS and RIS independently serve C_1 and C_2, respectively. Furthermore, although the fact that the users in C_2 still receive interference from the users in C_1 through the BS-U_m link in addition to the sub-surfaces mutual interference, increasing the RIS size can effectively mitigate the impact of the overall interference on the users in C_2 [47]. Furthermore, the sum-rate gain provided by the proposed system does not require that the users in C_2 have different channel gains as is the case in conventional NOMA, which adds more flexibility to the system design. Finally, for the detection process, user m in C_2 needs to know the overall sum of the channel gains of the m^{th} sub-surface only, which effectively reduces the training overhead*

[5] Note that, practically, there are phase-dependent amplitude variations associated with the reflection coefficient of each RIS element [42]. However, we assume a constant reflection amplitude for the mathematical analysis simplification, as this assumption has no effect on the proposed system design.

in the channel estimation of the RIS channels.

III. Outage Probability Analysis

Denoting the data rate requirement for $\text{U}_{m,2}$ as γ_m^*, the outage probability for $\text{U}_{m,2}$ is given as [48]

$$P_m^{out} = P(R_m < \gamma_m^*) \tag{13}$$

By substituting (11) in (13), we obtain

$$\begin{aligned} P_m^{out} &= P\left(\log_2\left(1+\text{SINR}_m\right) < \gamma_m^*\right), \\ &= P\left(\text{SINR}_m < 2^{\gamma_m^*} - 1\right) \end{aligned} \tag{14}$$

Due to the spatial correlation between the RIS-U_m channels, and thus, the correlation through $\mathbf{\Theta}_i$, it is challenging to derive the distribution of SINR_m [49], [50][6]. Furthermore, by considering an inter-element separation of $\lambda/2$, where λ is the wavelength of the operating frequency, the spatially correlated channels become close to the i.i.d. case [51]. This approximation is verified through Monte Carlo simulations in Section V, where the ergodic-rates and outage probability curves with/without correlation are shown to be close to each other. In what follows the outage probability expression is given under the i.i.d Rayleigh fading channels assumption and thus, the obtained results can serve as an upper bound for the performance of the cases where the inter-element separation is less than $\lambda/2$ and the RIS channels are highly spatially-correlated.

Proposition 1. *The closed-form outage probability expression of user m in C_2, assuming uncorrelated channels $\mathbf{R}_{RIS} = \mathbf{I}_N$, is given by*

$$\begin{aligned} P_m^{out} = 1 - \mathcal{Q}_{\frac{1}{2}}\left(\frac{\mu_1}{s_1}, \sqrt{\frac{y}{s_1^2}}\right) + \left(\frac{s_2^2}{s_1^2+s_2^2}\right)^{\frac{1}{2}} exp\left(\frac{y}{2s_2^2}\right) \\ \times\, exp\left(-\frac{\mu_1^2}{2(s_1^2+s_2^2)}\right) \mathcal{Q}_{\frac{1}{2}}\left(\frac{\mu_1}{s_1}\sqrt{\frac{s_2^2}{s_1^2+s_2^2}}, \sqrt{\frac{y(s_1^2+s_2^2)}{s_1^2 s_2^2}}\right), \end{aligned} \tag{15}$$

where $\mathcal{Q}_k$ is the k^{th} order generalized Marcum Q-function [52], $y = \frac{2^{\gamma_m^*}-1}{\rho}$, $\mu_1 = \sqrt{L_m^{RIS}} N_m \frac{\sqrt{\pi}}{2}$, $s_1^2 = L_m^{RIS} N_m \frac{4-\pi}{4}$, and $s_2^2 = 0.5(2^{\gamma_m^*} - 1)(L_m^{RIS}(N - N_m) + L_m^{BS})$.

Proof. See Appendix A. ∎

Proposition 1 expresses the outage probability as a function of the transmit SNR and the ratio of $\sqrt{L_m^{\text{RIS}}}N_m$ to $\sqrt{L_m^{\text{RIS}}}(N-N_m)$ plus $\sqrt{L_m^{\text{BS}}}$, where these parameters reflect the amplification gain, the sub-surfaces interference, and the BS interference powers, respectively. To get more insight, we give the following corollaries that follow from Proposition 1.

Corollary 1. *The Asymptotic behaviour of the outage probability can be obtained, for $\rho \to \infty$ or $y \to 0$, as*

$$P_m^{\infty} = \left(\frac{s_2^2}{s_1^2+s_2^2}\right)^{\frac{1}{2}} exp\left(-\frac{\mu_1^2}{2(s_1^2+s_2^2)}\right), \tag{16}$$

Proof. The proof follows directly from Proposition 1 by letting $y = 0$, where $\mathcal{Q}_{\frac{1}{2}}(a, 0) = 1, \forall a$ [52]. ∎

From Corollary 1, at high SNR, the outage probability performance improves exponentially with the ratio of the amplification gain to the overall interference associated with the RIS sub-surfaces and the BS links.

Corollary 2. *Consider the case where the RIS is uniformly partitioned between M_2+1 users in C_2, where $N_m = N_i = N/(M_2+1), \forall i, m$, $\sqrt{L_m^{RIS}} M_2 N_m >> L_m^{BS}$, and $\rho \to \infty$, then, the outage probability is given by*

$$P_m^{\infty} = \left(\frac{2M_2}{2M_2+4-\pi}\right)^{\frac{1}{2}} exp\left(-\frac{\pi N_m}{2(2M_2+4-\pi)}\right), \tag{17}$$

for $M_2 >> 1$, (17) can be simplified to

$$P_m^{\infty} \approx exp\left(-\frac{\pi N_m}{4M_2}\right). \tag{18}$$

Thus, for a given P_m^{∞}, the required N_m for each user in C_2 can be obtained as

$$N_m \approx \lceil -\frac{4M_2}{\pi}\ln(P_m^{\infty})\rceil. \tag{19}$$

Proof. The proof follows directly from Corollary 1 by letting $s_2^2 = 0.5(\sqrt{L_m^{\text{RIS}}}(N-N_m) + L_m^{BS}) \approx 0.5\sqrt{L_m^{\text{RIS}}} M_2 N_m$, where, without loss of generality, $\gamma_m^* = 1$ bit per channel use (bpcu). ∎

Corollary 2 shows the outage probability performance at high SNR by considering only the mutual sub-surfaces interference impact. The BS link interference impact can be ignored by assuming large RIS size N and/or $L_m^{\text{BS}} >> L_m^{\text{RIS}}$. It can clearly be seen from (18) that the outage probability performance improves exponentially in proportion to the ratio of the sub-surface size allocated to user m to the number of the other served users (M_2). Although increasing N_m for each user, and hence increasing N, increases the number of the interferer signals received by each user, the outage probability still decreases exponentially for all users. This can be explained by the fact that the interference is of the incoherent type where the interferer signals are combined constructively/destructively in a random way.

Remark 2. *Interestingly, the impact of the mutual sub-surfaces' interference between users can be effectively mitigated by increasing the overall RIS size for a given M_2 number of users in C_2. This is in contrast to the case of using SC, where increasing the transmit power has no effect on the mutual interference between the superposed symbols. Furthermore, it is worth noting that the outage probability in* (15) *can be generalized to the case where all the links, BS-RIS, BS-U_m, and RIS-U_m are Rician fading channels. In this case, by considering the same derivation steps provided in Appendix A, we obtain Y to be the difference of two independent and non-central chi-square random variables, which has the following characteristic function (CF) [53]*

$$\Psi_Y(w) = \frac{exp\left(\frac{jw\mu_A^2}{1-2jw\sigma_A^2}\right) exp\left(\frac{-jw\mu_I^2}{1+2jw\sigma_I^2}\right)}{(1-2jw\sigma_A^2)^{0.5}(1+2jw\sigma_I^2)}, \tag{20}$$

where (μ_A,σ_A) and (μ_I, σ_I) are the mean and standard deviation pairs of A_m and $\bar{I}_m$, respectively. Thus, by using Gil-Pelaez's inversion formula, P_m^{out} can be obtained as follows [54]

$$P(Y < y) = \frac{1}{2} - \int_0^{\infty} \frac{\Im\{e^{-jwy}\Psi_Y(w)\}}{w\pi} dw, \tag{21}$$

[6] In [49] and [50] a deterministic equivalent and moment matching approaches have been considered, respectively, to obtain the outage probability under spatially correlated channels. However, these works do not consider the intelligent phase shift adjustment.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2021.3127725, IEEE Journal of Selected Topics in Signal Processing

where the integration needs to be evaluated numerically with a suitable upper limit to avoid errors associated with the numerical calculations.

IV. RIS Partitioning Approach and Special Cases

In this section, we provide the approach we used to partition the RIS among C$_2$ users in order to maximize user fairness. Since each sub-surface of the RIS is allocated to serve a different user, each user m in C$_2$ receives a huge number of interferer signals from the other sub-surfaces allocated to the other users. Therefore, a proper number of reflecting elements (N_m) needs to be allocated for each user in order to guarantee maximum user fairness among them. In order to achieve this goal, we formulate an optimization problem where the Jain's fairness index [55] is the objective function to be maximized and $N_1, ..., N_m, ..., N_{M_2}$ are the decision variables to be determined, as follows:

$$\text{(P1)}: \max_{N_1,...,N_m,...,N_{M_2}} \quad \frac{(\frac{1}{M_2}\sum_{m=1}^{M_2}\bar{R}_m)^2}{\frac{1}{M_2}\sum_{m=1}^{M_2}\bar{R}_m^2}, \tag{20a}$$

$$\text{s.t.} \sum_{m=1}^{M_2} N_m = N,\ N_m \in \{1, ..., N-(M_2-1)\}, \forall m, \tag{20b}$$

where $\bar{R}_m = \mathbb{E}[R_m]$ is the ergodic rate of U$_{m,2}$, $\mathbb{E}[\cdot]$ is the statistical expectation operator over all the random channel realizations.

(P1) is NLIP problem with a non-convex feasible set represented by the constraint (20b) that has combinatorial growth as N increases. This makes (P1) a non-convex combinatorial optimization problem which is very challenging to solve if we consider the fact that a LIP is generally a non-deterministic polynomial-time (NP) hard problem [56]. IP optimization problems are usually solved by using branch-and-bound with the relaxation of the integrality constraint on the decision variables [57], which is, in this case, invalid for N_m, the number of RIS elements allocated to U$_{m,2}$. On the other side, considering the exhaustive search solution, the complexity lies in the explosively large size of the feasible set represented by the constraint (20b) which has $\binom{N-1}{M_2-1}$ feasible points. This implies that the required searching loops to scan all the feasible points are at a complexity level of $\mathcal{O}((N-1)^{\min(M_2-1,N-M_2)})$, or $\mathcal{O}((N-1)^{M_2-1})$ if we considered the fact that $N >> M_2$. Nevertheless, in order to shrink the size of the feasible set and thus, make the exhaustive search a practical method to solve (P1), we exploit the structure of the problem, specifically the nature of the transmission over the RIS, to bound the feasible set in (20b), as follows.

First, the number of RIS elements (N_m) allocated to any user m in C$_2$ cannot be randomly small, otherwise, the irremovable interference power from the other sub-surfaces overwhelms the amplification power from the m^{th} sub-surface. Motivated by the so-called interference temperature [58], a signal-to-interference power constraint (SIPC) [46] needs to be considered to protect each user U$_{m,2}$ from the interference belong to the other users. Based on the SIPC threshold, there is a minimum number of RIS elements N_{thr} that can be allocated for any user m in C$_2$ to protect it from the interference power associated with the remaining part of the RIS ($N - N_{thr}$). This means that the feasible set in (20b) needs to have N_{thr} (rather than unity) as the minimum element in the set, which in turn shrinks the size of the feasible set to $\binom{N-N_{thr}-1}{M_2-1}$ feasible points. Second, instead of considering a step size of one between any two successive elements in the feasible set of (20b), an adjustable step size b is considered, which reduces the feasible set further to $\binom{\frac{N-N_{thr}}{b}-1}{M_2-1}$ feasible points, where $\frac{N-N_{thr}}{b}$ needs to be an integer. Third, as in the power allocation of classical PD-NOMA, more RIS elements need to be allocated to the user with the weakest RIS-U$_{m,2}$ channel gain. Hence, the users need to be ordered according to their distances from the RIS, where U$_{m,2}$ is the m^{th} farthest user from the RIS and thus, the user with the m^{th} weakest RIS-U$_{m,2}$ channel gain. By considering the previously mentioned three bounding modifications on the feasible set in (20b), (P1) can be reformulated to obtain the following new optimization problem:

$$\text{(P1.1)}: \max_{N_1,...,N_m,...,N_{M_2}} \quad \frac{(\frac{1}{M_2}\sum_{m=1}^{M_2}\bar{R}_m)^2}{\frac{1}{M_2}\sum_{m=1}^{M_2}\bar{R}_m^2}, \tag{21a}$$

$$\begin{aligned}\text{s.t.} \sum_{m=1}^{M_2} N_m = N,\ N_m \in \mathcal{N},\ \forall m, \text{where } \mathcal{N} = \{N_{thr}, N_{thr}\\ + 1b, N_{thr}+2b, ..., N - N_{thr}(M_2-1)\}, \mathcal{N} \subset \mathbb{Z}^+,\end{aligned} \tag{21b}$$

$$N_1 \geq N_2 ... \geq N_{M_2}. \tag{21c}$$

By considering (P1.1), it can be observed that the combinatorial growth of the feasible set of (P1) is significantly limited by the three modifications considered on the constraint (20b), which in turn effectively reduces the search space. In what follows, we formulate two new optimization problems to find the proper N_{thr} and b for (P1.1):

$$\text{(P1.1.1)}: \min \quad N_{thr}, \tag{22a}$$

$$\text{s.t.}\ \ P(A_m|_{N_m=N_{thr}} \leq q\left|I_{\text{RIS}}|_{M_2=2,i\neq m}\right|^2) \leq \epsilon \tag{22b}$$

$$M_2 N_{thr} \leq N. \tag{22c}$$

Here, the motivation behind the constraint (22b) can be explained by the fact that from each user U$_{m,2}$ perspective, the RIS appears to be partitioned into two parts, the first part (with N_{thr} elements) that amplifies the signal of U$_{m,2}$, and the second part (with $N - N_{thr}$ elements) where the interference associated with the other users comes from. Thus, regardless of the noise and the BS-U$_{m,2}$ interference, the constraint (22b) ensures that for any user U$_{m,2}$ with N_{thr} RIS elements allocation, there is a low probability $\epsilon << 1$ that the interference power associated with the $N - N_{thr}$ part can be higher (by a factor of q) than the amplification power associated with the N_{thr} part. The two parameters ϵ and q need to be adjusted according to the size of the RIS and the number of users in C$_2$. Furthermore, the probability given in (22b) can be obtained from (16) by letting $L_m^{\text{BS}} = 0$ and $\gamma_m^* = 1$, as it is illustrated in Algorithm 1. The constraint (22c) ensures that for a given N, N_{thr} needs to have a small enough value such that all the M_2 users can have (at least) the same allocation of N_{thr} RIS elements, otherwise, N need to be increased. In what follows we formulate an optimization problem to find the proper step size b for (P1.1).

$$
\begin{aligned}
\text{(P1.1.2)}: \quad & \max \quad b, && \text{(23a)}\\
& \text{s.t.} \quad \bar{R}_m^{(N_{thr}+b)} - \bar{R}_m^{(N_{thr})} \leq \bar{r}, && \text{(23b)}\\
& \qquad\; b \leq N - M_2 N_{thr}, && \text{(23c)}
\end{aligned}
$$

where $\bar{R}_m^{(N_{thr}+b)}$ and $\bar{R}_m^{(N_{thr})}$ are obtained from (11) with simple modifications, as illustrated in Algorithm 2.

After determining N_{thr} in (P1.1.1), in (P1.1.2) we aim to find the maximum increment b that can be added to N_{thr} to get, accordingly, an ergodic rate increment upper-bounded by $\bar{r}$ b/s/Hz. Thus, b corresponds to the step size of the exhaustive search that ensures a maximum of $\bar{r}$ ergodic-rate resolution. In (23b), due to the fact that the users in C$_2$ experience different SNR values and different BS-U$_{m,2}$ interference power levels, the ergodic rate difference is calculated in the absence of the noise and the BS-U$_{m,2}$ interference effects. In this way, b determined from (P1.1.2) ensures that in the presence of noise and BS-U$_{m,2}$ interference, the search resolution to solve (P1.1) does not exceed $\bar{r}$, for all users. Hence, $\bar{r}$ needs to be adjusted according to the RIS size and number of users in C$_2$.

A. RIS Partitioning Algorithms

Note that Algorithm 3 is the main algorithm to obtain the RIS sub-optimum partition $N_1, N_2, ..., N_{M_2}$. Nevertheless, Algorithms 1, 2, and 3 need to be applied in order, where Algorithm 1 is used first to obtain N_{thr}, which is the input of Algorithm 2. In the same way, Algorithm 2 is used to obtain b, then, N_{thr} and b are used as inputs to Algorithm 3, which can be summarized as follows. First, the set $\mathcal{S}$ is constructed according to the constraints (21b) and (21c). The elements of $\mathcal{S}$ are the possible partitions the RIS can be partitioned into. When $b = 1$ is used in (21b), $\mathcal{S}$ contains all the possible partitions and the solution of the algorithm is a globally optimal solution, otherwise, for $b > 1$, the solution is a sub-optimal one. Second, for each partition s in $\mathcal{S}$, the ergodic transmission rate $\bar{R}_m$ for the all M_2 users are calculated and stored in the set $\mathcal{R}^{(s)}$. Third, for each partition s, Jain's index is calculated from $\mathcal{R}^{(s)}$ and stored in the set $\mathcal{J}$. Finally, the partition j^* associated with the maximum Jain's index is chosen as the optimum partition, and the set $\tilde{\mathcal{S}}^{(j^*)}$ contains the optimum number of RIS elements N_m^* needs to be allocated to each user m in C$_2$.

The convergence of the three algorithms is guaranteed as all algorithms have a predetermined finite number of iterations. Specifically, Algorithms 1 and 2 have $\bar{I}_1 = N/M_2$ and $\bar{I}_2 = N - M_2 N_{thr}$ maximum number of iterations, respectively. Likewise, Algorithm 3 has a maximum upper bound number of iterations $\bar{I}_3 = 2(\frac{\frac{N-N_{thr}}{b}-1}{M_2-1})$, which can be verified from the constraints (21b) and (21c) of (P1.1).

From the convergence analysis above, the computational complexity for Algorithm i, $i \in \{1, 2, 3\}$, can be given as $\mathcal{O}(\bar{C}_i \bar{I}_i)$, where $\bar{C}_i$ is the computational complexity of the functions inside the loops and $\bar{I}_i$ is the number of iterations for all loops, which is given above for each algorithm. By considering the required number of complex multiplications (CMs) as a metric, we obtain $N^2 + N$ as the number of CMs required to construct A_m and I_m/I_{RIS} in Algorithms 1

Algorithm 1 Solves (P1.1.1) to find N_{thr}.

```
Require: $L_m^{\text{RIS}}$, $M_2$, $N$, $q$, $\epsilon$.
Initialize $N_{thr} = 1, m = 1, i = 2$.
repeat
$\mu_m = \sqrt{L_m^{\text{RIS}}} N_{thr} \frac{\sqrt{\pi}}{2}, s_m^2 = L_m^{\text{RIS}} N_{thr} \frac{4-\pi}{4}$, $s_i^2 = 0.5 L_m^{\text{RIS}} q (N - N_{thr})$.
The probability in (22b) can be obtained from (16): $P_m^{\infty} = \left(\frac{s_i^2}{s_m^2+s_i^2}\right)^{\frac{1}{2}} \exp\left(-\frac{\mu_m^2}{2(s_m^2+s_i^2)}\right)$.
$N_{thr} = N_{thr} + 1$.
while $P_m^{\infty} <= \epsilon$ and $M_2 N_{thr} \leq N$.
return $N_{thr} = N_{thr} - 1$.
```

Algorithm 2 Solves (P1.1.2) to find the step size b.

```
Require: $N$, $N_{thr}$, $\bar{r}$, $\beta_{m,m}^{(n)}$, $\beta_{i,m}^{(n)}$, $\bar{\Phi}_i^n$, $\forall n$, and for any $m$ and $i$ such that $i \neq m$.
Initialize $b = 1$.
$\bar{R}_m^{(N_{thr})} = \mathbb{E}\left[\log_2\left(1 + \frac{A_m|_{N_m=N_{thr}}}{|I_{\text{RIS}}|_{N_i=N-N_{thr}}|^2}\right)\right]$. Perform the expectation over $10^4$ random channel realizations [59].
repeat
$\bar{R}_m^{(N_{thr}+b)} = \mathbb{E}\left[\log_2\left(1 + \frac{A_m|_{N_m=N_{thr}+b}}{|I_{\text{RIS}}|_{N_i=N-N_{thr}-b}|^2}\right)\right]$.
$\bar{r}_{diff} = \bar{R}_m^{(N_{thr}+b)} - \bar{R}_m^{(N_{thr})}$.
$b = b + 1$.
while $\bar{r}_{diff} \leq \bar{r}$ and $b \leq N - M_2 N_{thr}$.
return $b = b - 1$.
```

Algorithm 3 RIS partitioning algorithm to solve (P1.1)

```
Require: $b$, $N_{thr}$, $N$, $M_2$, $L_m^{\text{RIS}}$, $\rho$, $\upsilon_m$, $\beta_{i,m}^{(n)}$, $\phi_{i,m}^{(n)}$, $\bar{\Phi}_i^n$, $\forall i, n, m$.
Construct the set $\mathcal{N}$ in the constraint (21b).
According to the constraint (21b), construct the set $\mathcal{S}$ by finding all the solutions of $\sum_{m=1}^{M_2} N_m = N$, and excluding the ones that do not satisfy (21c). Thus, $\mathcal{S} = \{\tilde{\mathcal{S}}^{(1)}, ..., \tilde{\mathcal{S}}^{(s)}, ...\tilde{\mathcal{S}}^{(|\mathcal{S}|)}\}$, where $\tilde{\mathcal{S}}^{(s)} = \{N_1^{(s)}, ..., N_m^{(s)}, ...N_M^{(s)}\}$ corresponds to the RIS partition $s$, $\tilde{\mathcal{S}}^{(s)} \subset \mathcal{N}$, $\forall s$.
Construct the new sets $\mathcal{R}^{(1)}, ..., \mathcal{R}^{(s)}, ..., \mathcal{R}^{(|\mathcal{S}|)}$, where $\mathcal{R}^{(s)} = \{\emptyset\}$, $\forall s$.
for $s = 1 : |\mathcal{S}|$ do
    for $m = 1 : M_2$ do
        $R_m = \mathbb{E}\left[\log_2\left(1 + \frac{A_m}{I_m + \frac{1}{\rho}}\right)\right]$, where $N_i$, $N_m \in \tilde{\mathcal{S}}^{(s)}$, $\forall i$, $m$. Perform the expectation over $10^4$ random channel realizations.
        $\mathcal{R}^{(s)} = \mathcal{R}^{(s)} \cup \{R_m\}$.
    end for
end for
Construct a new set $\mathcal{J} = \{\emptyset\}$.
for $s = 1 : |\mathcal{S}|$ do
    $J = \frac{(\frac{1}{M_2}\sum_{m=1}^{M_2} R_m)^2}{\frac{1}{M_2}\sum_{m=1}^{M_2} R_m^2}$, where $R_m \in \mathcal{R}^{(s)}, \forall m$.
    $\mathcal{J} = \mathcal{J} \cup \{J\}$.
end for
$j^* = \underset{j=1:|\mathcal{J}|}{\arg\max}\, \mathcal{J}^{(j)}$
return $\tilde{\mathcal{S}}^{(j^*)} = \{N_1^*, N_2^*, ..., N_m^*, ..., N_{M_2}^*\}$.
```

and 2, where we considered the vector-matrix multiplication form given in (3). Consequently, for Algorithm 2, we obtain $\bar{I}_2(N^2+N)=(N-M_2N_{thr})(N^2+N)$ required number of CMs, hence, a complexity level of $\mathcal{O}_2(N^3)$. Likewise, for Algorithm 3, we have $0.5\bar{I}_3(N^2+N)=\big(\frac{\frac{N-N_{thr}}{b}-1}{M_2-1}\big)(N^2+N)$, hence, a complexity level of $\mathcal{O}_2(N^2\big(\frac{\frac{N-N_{thr}}{b}-1}{M_2-1}\big))$, where we considered only the first loop that requires CMs. For Algorithm 1, which has no CMs, the complexity level associated with $\bar{C}_1$ depends on the type of the used algorithm.

Remark 3. *Note that the partitioning process (Algorithms 1, 2, and 3) needs to be updated only when the RIS size N, transmit power P, number of users M_2 in C_2, or the distances of users from the RIS/BS changes. This significantly reduces the computational complexity cost associated with performing these algorithms as the mentioned parameters slowly change with time. Furthermore, a minimum data rate requirement for all or individual users can be straightforwardly included in the algorithms. Particularly, such a constraint can be included in Algorithm 1 to replace the constraint (22b).*

B. Special Cases with Different Number of Users in Clusters

Here, in addition to the general case introduced in Section II, we consider the system performance analysis for particular cases in order to shed some light on the performance of the proposed scheme under different settings, as follows.

i) *All users are located in* C$_2$: In this scenario $M_1=0$, and the same transmission mechanism described in Section II is used here with the following single modification. Since there are no users in C$_1$, the BS is assumed to transmit $x=\sqrt{P}\bar{x}$, where $\bar{x}$ is a predetermined symbol. Thus, assuming v_m is known at the receiver side, the signal $v_m\bar{x}$ received by each user m in C$_2$ over the BS-U$_{m,2}$ link can be removed readily. Furthermore, over the RIS-U$_{m,2}$ link, each sub-surface $i=m$ of the RIS remodulates $\bar{x}$ to reflect the PSK symbol to be transmitted to U$_{m,2}$, as described in Section II. Without loss of generality, an unmodulated carrier signal can be sent from the BS and, in this case, the BS can be compensated by a single RF signal generator (SG), which simplifies the transmitter architecture. Furthermore, R_m and P_m^{out} can be obtained from (11) and (15), respectively, by letting $v_m=0$ ($L_m^{\text{BS}}=0$). Likewise, for the asymptotic behaviour at high SNR, P_{out}^{∞} can be directly obtained from (17) and (18) by considering the uniform partitioning of the RIS elements to guarantee maximum user fairness.

ii) *Each cluster has a single user*: In this scenario, U$_{1,1}$ and U$_{1,2}$ are the only users in C$_1$ and C$_2$, respectively. Therefore, the BS transmits $x=\sqrt{P}x_1$, and U$_{1,1}$ experience the same performance in a SISO system. On the other side, the RIS (as a single unit) remodulates x and reflects the PSK symbol to be transmitted to U$_{1,2}$, as described in Section II. Since the whole surface is allocated for U$_{1,2}$, R_1, P_m^{out}, and P_m^{∞} can be obtained from (11), (15), and (16), respectively, by omitting the sub-surfaces interference term $I_{\text{RIS}}=0$ and letting $N_1=N$.

V. Simulation Results

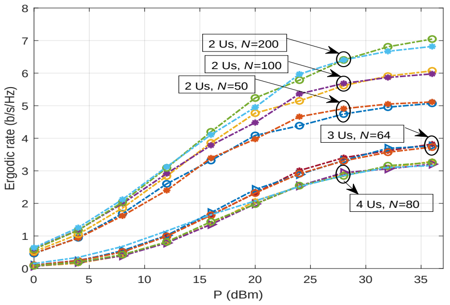


Fig. 2. The performance of the RIS partitioning algorithms with different N and M_2 values.

TABLE I
Users' distances for different deployment scenarios.

Users' deployment	$d_{1,1}$, $r_{1,1}$	$d_{2,1}$, $r_{2,1}$	$d_{1,2}$, $r_{1,2}$	$d_{2,2}$, $r_{2,2}$
$M_1=0$, $M_2=2$	-	-	150, 146	100, 104
$M_1=1$, $M_2=1$	150, 146	-	100, 104	-
$M_1=2$, $M_2=2$	150, 154	100, 104	250, 254	200, 204

TABLE II
RIS partitioning algoritms: input parameters and outcomes.

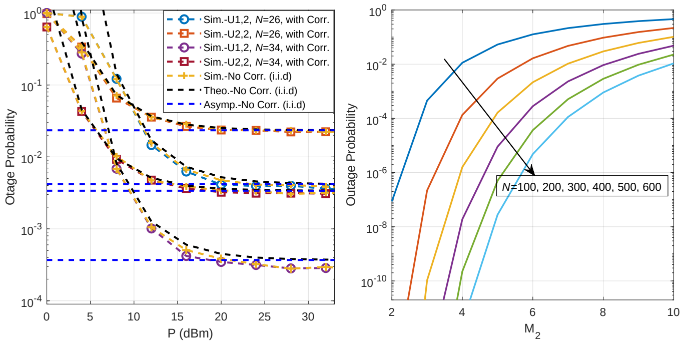
N, M_2	N_{thr}, q, ϵ	b, $\bar{r}$	$\lvert\mathcal{S}\rvert/\lvert\tilde{\mathcal{S}}\rvert$
$50, 2$	15, 1.5, 0.1	2, 0.3	$5/49$
$64, 3$	14, 1, 0.1	1, 0.1	$40/210$
$80, 4$	15, 1, 0.1	1, 0.1	$64/969$
$100, 2$	21, 1.5, 0.1	5, 0.5	$5/99$
$200, 2$	32, 1.5, 0.05	10, 0.7	$6/199$

In this section, we provide computer simulation results for the proposed scheme against the following four benchmark schemes: the time division multiple access (TDMA) as an OMA scheme, where the BS serves, with full power P, each user in its time slot with/without the RIS. PD-NOMA scheme, where the SC message is constructed as in (1). Finally, the RIS-assisted SISO NOMA scheme proposed in [15], [16][7]. For TDMA and classical NOMA schemes, in order to achieve maximum user fairness, we optimize the time and power allocation, respectively, at the transmitter side using exhaustive search. In order to guarantee a fair comparison, we consider the maximum fairness between users as the common threshold for all schemes, while the comparison lies in the ergodic transmission rates and outage probability performance. In what follows, we describe the path loss models and other simulation parameters.

For the BS-U$_{m,c}$ link, where c denotes the cluster number, we obtain the path loss as $(L_{m,c}^{\text{BS}})^{-1}=C_0\left(d_{m,c}/D_0\right)^{\alpha}$ [40], where $C_0=-30$ dB is the path gain at the reference distance $D_0=1$ m, $d_{m,c}$ is the BS-U$_{m,c}$ distance, and $\alpha=-3.5$ is the path gain exponent. For the BS-RIS-U$_{m,2}$ link, we consider that the RIS is located in the far-field with respect to the BS by ensuring that $r_s=\lceil\frac{N\lambda}{2}\rceil$ [60], thus, we obtain the path loss (with maximum gain RIS elements) as $(L_m^{\text{RIS}})^{-1}=\lambda^4/(256\pi^2r_s^2r_m^2)$ [61], where r_s and r_m are the BS-RIS (the center point) and the RIS (the center point)-U$_{m,2}$ distances, all in meters, respectively. Due to the small spacing distances between the RIS elements compared to r_s and r_m, we consider r_s and r_m for the path loss calculations, consequently, the path losses for all the RIS elements are

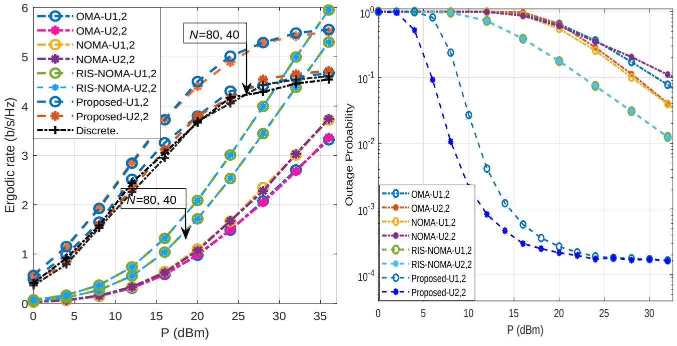
[7] This work does not consider the spatial correlation for the RIS channels in the obtained power allocation solution. Nevertheless, spatially correlated channels are used in our simulations of this benchmark scheme.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2021.3127725, IEEE Journal of Selected Topics in Signal Processing



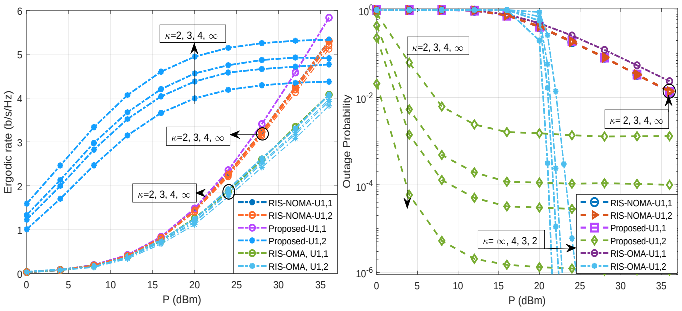
(a) (b)

Fig. 3. Outage probability performance with different N values, for (a) the general case, $M_2 = 2$, $M_1 \geq 1$, and a targeted transmission rate of 0.75 bpcu for all users, (b) different numbers of users in C$_2$ with uniform partitioning, obtained from (18).



(a) (b)

Fig. 4. The comparison of the proposed and benchmark schemes with $M_1 = 0$ and $M_2 = 2$, in terms of (a) ergodic rate with N =40, (b) the outage probability, with a targeted transmission rate of 1.2 bpcu for all users, and $N = 40$.



(a) (b)

Fig. 5. The comparison of the proposed scheme with the RIS-OMA (TDMA) and RIS-NOMA benchmark schemes, with $M_1 = M_2 = 1$ and $N = 40$, in terms of (a) ergodic rate, (b) outage probability, with a targeted transmission rate of 1.2 bpcu for all users.

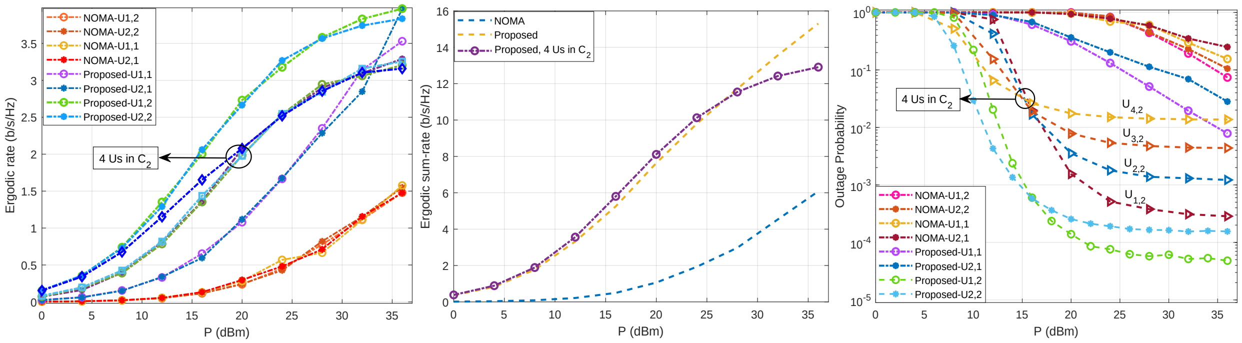
considered the same. In Table I, we give the users' distances from the BS and RIS, for different deployment scenarios, which are the used distances in our simulations, unless stated otherwise. The entries of the spatial correlation matrix are given as $[\mathbf{R}_{\text{RIS}}]_{n,\tilde{n}} = \text{sinc}\left(2 \left\|\mathbf{u}_n - \mathbf{u}_{\tilde{n}}\right\| / \lambda\right),\ n, \tilde{n} = 1, ..., N$ [51], where λ is calculated for a 1.8 Ghz operating frequency, $\text{sinc}(a) = \sin(\pi a)/(\pi a)$ is the sinc function, $\mathbf{u}_n = [0,\ i(n)d_l, \tilde{i}(n)d_w\ ]^T$, where $i(n)$, $\tilde{i}(n)$, d_l, and d_w are the horizontal index, vertical index, length, and width of the element n, respectively, $i(n) = \text{mod}(n-1, N_H)$, $\tilde{i}(n) = \lfloor (n-1)/N_H \rfloor$. Furthermore, the noise power is assumed to be fixed and identical for all users in C$_1$ and C$_2$, $\sigma^2 = -90$ dBm, $\forall m$, and the simulations are performed with 10^4 random channel realizations.

Fig. 2 illustrates the performance of the sub-optimal solution to (P1), which is represented by Algorithms 1, 2, and 3 to solve (P1.1) (the reformulated version of (P1)). Here, for each given P value, we consider multiple users in C$_2$ where the RIS needs to be partitioned among them to maximize the user fairness in terms of the ergodic-rate. It can be seen that the proposed algorithms achieve a result close to the perfect user fairness between the users with different N values. Furthermore, Table II shows that Algorithms 1 and 2 reduce the search space for Algorithm 3 significantly, where $|\mathcal{S}|/|\tilde{\mathcal{S}}|$ denote the number of possible partitions provided by the sub-optimal/optimal solution of Algorithm 3. This makes the exhaustive search a practical solution to solve (P1) and find the proper RIS partition.

Fig. 3(a) shows the outage probability performance for the general case, with different N values. Note that, as the performance improves with the increase of N, the saturation happens at the high SNR region due to the interference coming from BS and sub-surfaces. It can be also seen that the theoretical and simulation curves are in perfect agreement with each other. Furthermore, it can be noted that the i.i.d. approximation of the RIS-U$_{m,2}$ channels is accurate, where the curves with/without spatial correlation are perfectly matched. Fig. 3(b) shows the outage probability performance of users in C$_2$ versus the number of users M_2, where the RIS is partitioned uniformly between all users as stated in Corollary 2. It is worth noting that, although the ratio of the amplification to the interferer signals is the same, increasing the RIS size enhances the performance of all users due to the nature of the incoherent interference. In what follows we compare the proposed scheme with different benchmark schemes by considering the two special cases introduced in Section IV-B and the general case introduced in Section II.

In Figs. 4 (a) and (b), we consider the first case with $N = 40$, $M_1 = 0$, and $M_2 = 2$. As it is shown in Fig. 4(a), all the schemes achieve almost the same user fairness, nevertheless, the proposed scheme shines out with an improvement in the required P of 8 dB compared to the RIS-NOMA scheme and 14 dB compared to TDMA-OMA and PD-NOMA schemes. However, the performance gain achieved by the proposed scheme is bounded by a saturation point due to the mutual sub-surfaces interference. It can be also noted that the discrete phase shift adjustment with only 8 phase shift levels achieves almost the same performance as in the continuous case, where, assuming uniform quantization for the interval $[0, 2\pi)$, 3 bits are used to select the phase shift from $Z = 2^3$ levels in the finite set $\mathcal{F} = \{0, \Delta\Phi, ..., \Delta\Phi(Z-1)\}$, where $\Delta\Phi = \frac{2\pi}{Z}$ [62]. Fig 4 (b) shows the outage probability comparison with a targeted transmission rate of 1.2 bpcu for all users. It can be seen that the proposed scheme outperforms the benchmark schemes with a around 20 dB in the required P for both users before the saturation region.

In Figs. 5(a) and 5(b), we consider the second case with $N = 40$ and $M_1 = M_2 = 1$, as follows. Fig. 5(a) shows



(a) (b) (c)

Fig. 6. The comparison of the proposed and the NOMA benchmark schemes with $M_1 = M_2 = 2$ and $M_1 = 0, M_2 = 4$ with $N = 40, 80$, respectively, in terms of (a) ergodic rate, (b) ergodic sum-rate, (c) outage probability, with a targeted transmission rate of 0.75 bpcu for all users.

that $\text{U}_{1,1}$ in the proposed scheme and $\text{U}_{1,1}$ and $\text{U}_{1,2}$ in the RIS-NOMA scheme, achieve almost the same ergodic rate performance, with a 2 dB improvement in the required P for the proposed scheme at the high SNR region. On the other hand, $\text{U}_{1,2}$ in the proposed scheme achieves a 12-24 dB improvement compared to the other users when there is no phase estimation errors, which can be explained by the asymptotic squared power gain of $\mathcal{O}(N^2)$ the RIS provides to $\text{U}_{1,2}$ [40]. However, there is a saturation point for the performance of $\text{U}_{1,2}$ due to the BS-$\text{U}_{1,2}$ interference link. Considering the user fairness, contrary to might be concluded for the first glance from Fig. 5(a), the proposed scheme achieves the maximum user fairness, which can be explained as follows. Since the communication link over the RIS is blocked for $\text{U}_{1,1}$ in both schemes, the most efficient option, in terms of the sum-rate performance and user fairness, is to fully allocate the RIS to serve $\text{U}_{1,2}$ and the BS to serve $\text{U}_{1,1}$ in both schemes, which is what the proposed scheme basically does. On the other side, the use of SC in the benchmark scheme makes the RIS amplifies the interference from $\text{U}_{1,1}$ to $\text{U}_{1,2}$ without any benefit for $\text{U}_{1,1}$, which is unfair to $\text{U}_{1,2}$. In Fig. 5(b), $\text{U}_{1,2}$ in the proposed scheme outperforms the other users in both schemes in terms of the outage probability with a remarkable improvement of 20-35 dB in the required P, while the other users achieve the same performance, which agrees with the ergodic rate results shown in Fig. 5(a). For the phase estimation errors, we consider the von Mises distribution, where κ is the concentration parameter. It can be seen from Figs. 5(a) and 5(b) that the benchmark schemes are considerably less sensitive to the phase estimation errors compared to the proposed scheme, which can be explained as follows. The benchmark schemes, unlike the proposed one, have a direct information link which, with this small RIS size, dominates the RIS link and therefore, the performance loss due to the phase estimation errors does not appear clearly. With a larger RIS size that makes the BS-RIS-$\text{U}_{m,2}$ reflection link dominate, the benchmark schemes are also expected to experience a similar performance behavior, with the phase estimation errors.

Finally, in Figs. 6(a)-(c), we consider the general case of four users for the proposed and NOMA schemes. Fig. 6(a) shows the ergodic rate performance for all users, where almost maximum user fairness is achieved by the NOMA scheme. On the other side, the proposed scheme provides a close to the maximum fairness performance when all users are located in C_2, $M_1 = 0, M_2 = 4$, with $N = 80$. A better user fairness performance is noted for $M_1 = 2, M_2 = 2$, and $N = 40$ between the two users in each cluster, which corresponds to the maximum fairness between all the four users due to the same reasons explained in the discussion of Fig. 5(a). Furthermore, when $M_1 = M_2 = 2, N = 40$, the performance of the proposed scheme outperforms the one of NOMA scheme for each user significantly, with more than 12 dB and 22 dB improvement for the two users in C_1 and the two users in C_2, respectively. On the other side, when $M_1 = 0, M_2 = 4, N = 80$, 18 dB gain is observed compared to NOMA scheme, with a 4 dB less gain compared to the previous case due to the mutual sub-surfaces interference. Overall, for both users' deployment scenarios, an improvement of 18 dB in the ergodic sum-rate is achieved by the proposed scheme compared to the NOMA scheme, as shown in Fig. 6(b). In Fig. 6(c), when $M_1 = M_2 = 2, N = 40$, the proposed scheme outperforms NOMA scheme in the outage probability performance with around 22 dB in the required P for the first and second users in C_2, and around 5 dB and 10 dB for the first and second users in C_1, respectively. For the case where $M_1 = 0, M_2 = 4, N = 80$, the proposed scheme outperforms the NOMA scheme with 22 dB until the saturation region.

VI. Conclusion

In this paper, we have introduced a novel downlink NOMA solution with RIS partitioning in order to mitigate the mutual interference between users in a local beyond 5G network. In the proposed system, the ergodic rates and outage probabilities of all users are enhanced and the user fairness is maximized by the fair and efficient distribution of the PR among users. The potential of the proposed PR distribution is perfectly illustrated in Fig. 5, where the BS and RIS are used in a very efficient way compared to the classical use of the SC technique. Furthermore, we have proposed three efficient searching algorithms to, sequentially, obtain a sub-optimal solution for the RIS partitioning optimization problem, with insignificant performance degradation. By considering different users' deployment scenarios, it was shown that the proposed system provides remarkable performance gain in all of the considered different environment settings. We have derived the exact and asymptotic outage probability expressions for the proposed system in all the cases including the effect of the number of users in C_2 and the RIS size. The computer simulations show that the proposed system outperforms the OMA, RIS-OMA, NOMA, and RIS-NOMA benchmark systems significantly, in terms of outage probability, ergodic rates of all users, and user fairness. Finally, removing the sub-surfaces and/or the BS interference for C_2 users appears as a future research direction that is worth investigating.

Appendix A
Proof of Proposition 1

From (14), we obtain

$$\begin{aligned} P_{\text{out}} &= \left(\frac{A_m}{I_m + \frac{1}{\rho}} < 2^{\gamma_m^*} - 1 \right) \\ &= P\left(Y < y\right) \end{aligned} \tag{24}$$

where $Y = A_m - \bar{I}_m$, $y = \frac{2^{\gamma_m^*}-1}{\rho}$, and $\bar{I}_m$ is given by

$$\bar{I}_m = (2^{\gamma_m^*} - 1) I_m = |\sqrt{(2^{\gamma_m^*} - 1)}(I_{\text{RIS}} + v_m)|^2. \tag{25}$$

From (24), we note that the outage probability is equivalent to the cumulative distribution function (CDF) of Y. In order to find the CDF of Y, we first find the distribution of the RVs A_m and $\bar{I}_m$, as follows. By considering (8), we note that $\beta_{m,m}^{(n)}$ is a Rayleigh distributed RV with a mean $\mathrm{E}[\beta_{m,m}^{(n)}] = \sqrt{\pi}/2$, and a variance $\mathrm{VAR}[\beta_{m,m}^{(n)}] = (4 - \pi)/4$. According to the central limit theorem (CLT), for $N_m >> 1$, the term inside the squared parenthesis is a Gaussian RV, $\sim \mathcal{N}(\sqrt{L_m^{\text{RIS}}} N_m \frac{\sqrt{\pi}}{2}, L_m^{\text{RIS}} N_m \frac{4-\pi}{4})$. Thus, A_m is a non-central chi-square (χ^2) RV with one degree of freedom. Likewise, by considering (10), for $N - N_m >> 1$, I_{RIS} is a Gaussian RV, $I_{\text{RIS}} \sim \mathcal{CN}(0, L_m^{\text{RIS}}(N - N_m))$. Consequently, the constant-scaled sum of the two independent Gaussian RVs in (25), $\sqrt{(2^{\gamma_m^*} - 1)}(I_{\text{RIS}} + v_m)$, is a Gaussian RV, $\sim \mathcal{CN}(0, (2^{\gamma_m^*} - 1)(L_m^{\text{RIS}}(N - N_m) + L_m^{\text{BS}})$. Thus, $\bar{I}_m$ is a central χ^2 RV with two degrees of freedom, and consequently, Y is the difference of a non-central and central independent χ^2 RVs, where its CDF is given in (15) [63]. This completes the proof of Proposition 1. ■

References

[1] "Feasibility study on temporary spectrum access for local high-quality wireless networks," ETSI, Sophia Antipolis, France, Rep. TR 103 588, p. 30. Feb. 2018.

[2] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE Communications Surveys Tutorials*, vol. 20, no. 3, pp. 2294–2323, 2018.

[3] Y. Saito *et al.*, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *2013 IEEE 77th Veh. Technol. Conf. (VTC Spring)*, June 2013, pp. 1–5.

[4] 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, Oct. 2017.

[5] 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 Processing Lett.*, vol. 21, no. 12, pp. 1501–1505, Jul. 2014.

[6] X. Yue, Z. Qin, Y. Liu, S. Kang, and Y. Chen, "A unified framework for non-orthogonal multiple access," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5346–5359, May 2018.

[7] Z. Wei, L. Yang, D. W. K. Ng, J. Yuan, and L. Hanzo, "On the performance gain of noma over oma in uplink communication systems," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 536–568, 2020.

[8] E. Basar *et al.*, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116 753–116 773, Aug. 2019.

[9] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106–112, Nov. 2019.

[10] Q. Li, M. Wen, S. Wang, G. C. Alexandropoulos, and Y.-C. Wu, "Space shift keying with reconfigurable intelligent surfaces: Phase configuration designs and performance analysis," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 322–333, 2021.

[11] J. Yuan, M. Wen, Q. Li, E. Basar, G. C. Alexandropoulos, and G. Chen, "Receive quadrature reflecting modulation for RIS-empowered wireless communications," *IEEE Trans. on Vehicular Technology*, vol. 70, no. 5, pp. 5121–5125, 2021.

[12] E. Basar, "Transmission through large intelligent surfaces: A new frontier in wireless communications," in *2019 European Conf. Netw. Commun. (EuCNC)*, June. 2019, pp. 112–117.

[13] ——, "Reconfigurable intelligent surface-based index modulation: A new beyond MIMO paradigm for 6G," *IEEE Trans. Commun.*, vol. 68, no. 5, pp. 3187–3196, Feb. 2020.

[14] A. Khaleel and E. Basar, "Reconfigurable intelligent surface-empowered MIMO systems," *IEEE Systems Journal*, pp. 1–9, Aug. 2020.

[15] G. Yang, X. Xu, Y.-C. Liang, "Intelligent reflecting surface assisted non-orthogonal multiple access," Dec. 2019. [Online]. Available: arXiv:1907.03133.

[16] G. Yang, X. Xu, and Y. -C. Liang, "Intelligent reflecting surface assisted non-orthogonal multiple access," in *2020 IEEE Wireless Commun. Netw. Conf. (WCNC)*, May 2020, pp. 1–6.

[17] Z. Ding and H. V. Poor, "A simple design of IRS-NOMA transmission," *IEEE Commun. Lett.*, vol. 24, no. 5, pp. 1119–1123, Feb. 2020.

[18] T. Hou *et al.*, "Reconfigurable intelligent surface aided NOMA networks," *IEEE J. Sel. Areas in Commun.*, vol. 38, no. 11, pp. 2575–2588, Jul. 2020.

[19] M. Zhang, M. Chen, Z. Yang, H. Asgari, and M. Shikh-Bahaei, "Joint user clustering and passive beamforming for downlink NOMA system with reconfigurable intelligent surface," in *2020 IEEE 31st Annual Int. Symposium on Personal, Indoor and Mobile Radio Commun.*, Aug. 2020, pp. 1–6.

[20] Y. Xu *et al.*, "Fair non-orthogonal multiple access communication systems with reconfigurable intelligent surface," in *2020 IEEE 31st Annual Int. Symposium on Personal, Indoor and Mobile Radio Communications*, Aug. 2020, pp. 1–6.

[21] X. Liu, Y. Liu, Y. Chen, and H. V. Poor, "RIS enhanced massive non-orthogonal multiple access networks: Deployment and passive beamforming design," *IEEE J. Sel. Areas Commun.*, pp. 1–1, Aug. 2020.

[22] Y. Li, M. Jiang, Q. Zhang, and J. Qin, "Joint beamforming design in multi-cluster MISO NOMA reconfigurable intelligent surface-aided downlink communication networks," *IEEE Trans. Commun.*, pp. 1–1, Oct. 2020.

[23] T. Hou, Y. Liu, Z. Song, X. Sun, and Y. Chen, "MIMO-NOMA networks relying on reconfigurable intelligent surface: A signal cancellation based design," *IEEE Trans. Commun.*, pp. 1–1, Aug. 2020.

[24] C. Zhang, W. Yi, Y. Liu, "Reconfigurable intelligent surfaces aided multi-cell NOMA networks: A stochastic geometry model," Aug. 2020. [Online]. Available: arXiv:2008.08457.

[25] C. Zhang, W. Yi, Y. Liu, Z. Qin, K.K. Chai, "Downlink analysis for reconfigurable intelligent surfaces aided NOMA networks," Jun. 2020. [Online]. Available: arXiv:2006.13260.

[26] M. Elhattab, M. A. Arfaoui, C. Assi, and A. Ghrayeb, "Reconfigurable intelligent surface assisted coordinated multipoint in downlink NOMA networks," *IEEE Commun. Lett.*, pp. 1–1, Oct. 2020.

[27] Z. Ding, R. Schober, and H. V. Poor, "On the impact of phase shifting designs on IRS-NOMA," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1596–1600, Apr. 2020.

[28] J. Zuo, Y. Liu, Z. Qin, and N. Al-Dhahir, "Resource allocation in intelligent reflecting surface assisted NOMA systems," *IEEE Trans. Commun.*, pp. 1–1, Aug. 2020.

[29] M. Najafi, V. Jamali, R. Schober, and H. V. Poor, "Physics-based modeling and scalable optimization of large intelligent reflecting surfaces," Apr. 2020. [Online]. Available: arXiv:2004.12957.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2021.3127725, IEEE Journal of Selected Topics in Signal Processing

[30] D. Xu, V. Jamali, X. Yu, D. W. K. Ng, and R. Schober, "Optimal resource allocation design for large IRS-assisted SWIPT systems: A scalable optimization framework," Apr. 2021. [Online]. Available: arXiv:2104.03346.

[31] Q. U. A. Nadeem, A. Kammoun, A. Chaaban, M. Debbah, and M. S. Alouini, "Asymptotic max-min sinr analysis of reconfigurable intelligent surface assisted miso systems," *IEEE Trans. on Wireless Commun.*, vol. 19, no. 12, pp. 7748–7764, 2020.

[32] M. S. Ali, H. Tabassum, and E. Hossain, "Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (NOMA) systems," *IEEE Access*, vol. 4, pp. 6325–6343, 2016.

[33] A. de Jesus Torres, L. Sanguinetti, and E. Björnson, "Near- and far-field communications with large intelligent surfaces," Nov. 2020. [Online]. Available: arXiv:2011.13835.

[34] H. Yu, H. D. Tuan, A. A. Nasir, T. Q. Duong, and H. V. Poor, "Joint design of reconfigurable intelligent surfaces and transmit beamforming under proper and improper gaussian signaling," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2589–2603, 2020.

[35] C. Pan, H. Ren, K. Wang, M. Elkashlan, A. Nallanathan, J. Wang, and L. Hanzo, "Intelligent reflecting surface aided mimo broadcasting for simultaneous wireless information and power transfer," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1719–1734, 2020.

[36] Q. U. A. Nadeem, H. Alwazani, A. Kammoun, A. Chaaban, M. Debbah, and M. S. Alouini, "Intelligent reflecting surface-assisted multi-user miso communication: Channel estimation and beamforming design," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 661–680, 2020.

[37] H. Alwazani, Q. U. A. Nadeem, and A. Chaaban, "Channel estimation for distributed intelligent reflecting surfaces assisted multi-user miso systems," in *2020 IEEE Globecom Workshops (GC Wkshps)*, Dec. 2020, pp. 1–6.

[38] W. Xia, Y. Zhou, G. Yang, and R. T. Chen, "Optimal minimum euclidean distance-based precoder for NOMA with finite-alphabet inputs," *IEEE Access*, vol. 7, pp. 45 123–45 136, Apr. 2019.

[39] J. Zhang, X. Wang, T. Hasegawa, and T. Kubo, "Downlink non-orthogonal multiple access (NOMA) constellation rotation," in *2016 IEEE 84th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2016, pp. 1–5.

[40] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, Aug. 2019.

[41] Q.-U.-A. Nadeem, A. Zappone, and A. Chaaban, "Intelligent reflecting surface enabled random rotations scheme for the MISO broadcast channel," Mar. 2021. [Online]. Available: arXiv:2103.09898.

[42] S. Abeywickrama, R. Zhang, Q. Wu, and C. Yuen, "Intelligent reflecting surface: Practical phase shift model and beamforming optimization," *IEEE Trans. on Commun.*, vol. 68, no. 9, pp. 5849–5863, 2020.

[43] U. Siddique, H. Tabassum, and E. Hossain, "Downlink spectrum allocation for in-band and out-band wireless backhauling of full-duplex small cells," *IEEE Trans. on Commun.*, vol. 65, no. 8, pp. 3538–3554, 2017.

[44] S. Atapattu, R. Fan, P. Dharmawansa, G. Wang, J. Evans, and T. A. Tsiftsis, "Reconfigurable intelligent surface assisted two–way communications: Performance analysis and optimization," *IEEE Trans. Commun.*, vol. 68, no. 10, pp. 6552–6567, 2020.

[45] A. Osseiran, J. Monserrat, and W. Mohr, "*Mobile and Wireless Communications for IMT-Advanced and Beyond*," Hoboken, NJ, USA: Wiley, 2011.

[46] X. Kang, H. F. Chong, Y. Chia, and S. Sun, "Ergodic sum-rate maximization for fading cognitive multiple-access channels without successive interference cancelation," *IEEE Trans. Veh. Techno.*, vol. 64, no. 9, pp. 4009–4018, Oct. 2014.

[47] T. Hou *et al.*, "Reconfigurable intelligent surface aided NOMA networks," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2575–2588, Jul. 2020.

[48] J. G. Proakis, "*Digital Communications*," 5th ed. New York: McGrawHill, 2008.

[49] A. Papazafeiropoulos, C. Pan, A. Elbir, P. Kourtessis, S. Chatzinotas, and J. M. Senior, "Coverage probability of distributed IRS systems under spatially correlated channels," Feb. 2021. [Online]. Available: arXiv:2102.09416.

[50] T. V. Chien, A. K. Papazafeiropoulos, L. T. Tu, R. Chopra, S. Chatzinotas, and B. Ottersten, "Outage probability analysis of IRS-assisted systems under spatially correlated channels," Feb. 2021. [Online]. Available: arXiv:2102.11408.

[51] E. Björnson and L. Sanguinetti, "Rayleigh fading modeling and channel hardening for reconfigurable intelligent surfaces," *IEEE Wireless Commun. Lett.*, pp. 1–1, 2020.

[52] A. Annamalai, C. Tellambura, and J. Matyjas, "A new twist on the generalized marcum q-function qm(a, b) with fractional-order m and its applications," in *2009 6th IEEE Consumer Commun. Netw. Conf.*, Jan. 2009, pp. 1–5.

[53] M. Simon, "*Probability Distributions Involving Gaussian Random Variables*," New York, NY, USA: Springer, 2002.

[54] A. Mathai and S. B. Provost, " *Quadratic Forms in Random Variables: Theory and Applications*," New York, NY, USA: Marcel Dekker, 1992.

[55] R. Jain, D.-M. Chiu, W. R. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer system," Eastern Research Laboratory, Digital Equipment Corporation Hudson, MA, 1984, vol. 38.

[56] L. A. Wolsey and G. L. Nemhauser, *Integer and Combinatorial Optimization*. New York: John Wiley & Sons, 1988.

[57] S. Burer and A. N. Letchford, "Non-convex mixed-integer nonlinear programming: A survey," *Surv. Oper. Res. Manage. Sci.*, vol. 17, no. 2, pp. 97–106, Jun. 2012.

[58] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[59] R. Zhang, S. Cui, and Y. Liang, "On ergodic sum capacity of fading cognitive multiple-access and broadcast channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5161–5178, Oct. 2009.

[60] W. Tang *et al.*, "Wireless communications with reconfigurable intelligent surface: Path loss modeling and experimental measurement," *IEEE Trans. Wireless Commun.*, pp. 1–1, Sep. 2020.

[61] S.W. Ellingson, "Path loss in reconfigurable intelligent surface-enabled channels," Dec. 2019. [Online]. Available: arXiv:1912.06759.

[62] Q. Wu R. Zhang, "Beamforming optimization for intelligent reflecting surface with discrete phase shifts," *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, United Kingdom, May 2019.

[63] M. Simon, "*Probability Distributions Involving Gaussian Random Variables*," New York, NY, USA: Springer, 2002.

Aymen Khaleel received the B.Sc. degree from the University of Anbar, Al Anbar, Iraq, in 2013, and the M.Sc. degree from Turkish Aeronautical Association University, Ankara, Turkey, in 2017. He is currently pursuing his Ph.D. in Electrical and Electronics Engineering at Koç University, Istanbul, Turkey, where he is currently a Teaching Assistant. His research interests include MIMO systems, index modulation, intelligent surfaces-based systems. He serves as a Reviewer for *IEEE Communications Magazine*, *IEEE Transactions on Vehicular Technology*, and *IEEE communications letters*.

Ertugrul Basar received his Ph.D. degree from Istanbul Technical University in 2013. He is currently an Associate Professor with the Department of Electrical and Electronics Engineering, Koç University, Istanbul, Turkey and the director of Communications Research and Innovation Laboratory (CoreLab). His primary research interests include beyond 5G systems, index modulation, intelligent surfaces, waveform design, and signal processing for communications. Dr. Basar currently serves as a Senior Editor of *IEEE Communications Letters* and an Editor of *IEEE Transactions on Communications* and *Frontiers in Communications and Networks*. He is a Young Member of Turkish Academy of Sciences and a Senior Member of IEEE.