NOMA-based Improper Signaling for Multicell MISO RIS-assisted Broadcast Channels

Mohammad Soleymani* *Member, IEEE*, Ignacio Santamaria† *Senior Member, IEEE*, Eduard Jorswieck‡ *Fellow, IEEE*, and Sepehr Rezvani‡

Abstract—In this paper, we study the performance of reconfigurable intelligent surfaces (RISs) in a multicell broadcast channel (BC) that employs improper Gaussian signaling (IGS) jointly with non-orthogonal multiple access (NOMA) to optimize either the minimum-weighted rate or the energy efficiency (EE) of the network. We show that although the RIS can significantly improve the system performance, it cannot mitigate interference completely, so we have to employ other interference-management techniques to further improve performance. We show that the proposed NOMA-based IGS scheme can substantially outperform proper Gaussian signaling (PGS) and IGS schemes that treat interference as noise (TIN) in particular when the number of users per cell is larger than the number of base station (BS) antennas (referred to as overloaded networks). In other words, IGS and NOMA complement to each other as interference management techniques in multicell RIS-assisted BCs. Furthermore, we consider three different feasibility sets for the RIS components showing that even a RIS with a small number of elements provides considerable gains for all the feasibility sets.

Index Terms—Energy efficiency, improper Gaussian signaling, majorization minimization, MISO broadcast channels, NOMA, reflecting intelligent surface, spectral efficiency.

I. INTRODUCTION

Reconfigurable intelligent surfaces (RISs) are promising technologies for next generations of wireless communication systems [1]–[4]. RIS can modulate the channel coefficients, which provides some additional degrees of freedom in the system design. By optimizing the propagation channel, RIS can improve different performance metrics and/or can manage interference. In this work, we study the performance of RIS in a multicell broadcast channel (BC), which is an interference-limited system. It is known that RIS can improve the performance of such a system, however, it should be further investigated what is the main role of RIS in such improvements. For instance, are the RIS benefits primarily because of its ability to cancel interference in overloaded systems, or are they due to a coverage improvement? In this paper, we provide an answer for these important questions.

*Mohammad Soleymani is with the Signal and System Theory Group, Universität Paderborn, Germany, http://sst.upb.de (email: mohammad.soleymani@sst.upb.de).

†Ignacio Santamaria is with the Department of Communications Engineering, University of Cantabria (email: i.santamaria@unican.es).

‡ Eduard Jorswieck and Sepehr Rezvani are with Institute for Communications Technology, Technische Universität Braunschweig, 38106 Braunschweig, Germany (e-mails: {jorswieck,rezvani}@ifn.ing.tu-bs.de)

A. Related work

Interference has always been among the main performance restrictions for modern wireless communication systems, and therefore interference-management techniques play an essential role in improving the performance of interference-limited networks [5]. Treating interference as noise (TIN) is the optimal decoding strategy when the interference level is low [6]. Additionally, in the presence of strong interference, the optimal decoding strategy is successive interference cancellation (SIC) in which the signal of the interfering user is decoded at first, and then subtracted from the received signal [7]. However, for operational points between these two extreme cases the optimal strategy to handle interference is either unknown or involves very high complexities. In these cases, we usually have to employ effective, but not necessarily optimal, interference-management techniques such as improper Gaussian signaling (IGS) and/or non-orthogonal-multiple-access (NOMA)-based techniques.

IGS was employed as an interference-management technique for the first time in [8], where the authors showed that IGS can increase the degrees-of-freedom (DoF) of the 3-user IC. In the last decade, IGS has been applied to various interference-limited systems, showing that it can increase the spectral and energy efficiency of such systems with TIN [9]–[31]. For instance, the papers [11]–[14] showed that IGS can improve the rate and/or energy efficiency (EE) of the secondary user in cognitive radio systems. In [17]–[20], it is shown that IGS can enlarge the rate region of the 2-user IC. The papers [21], [22] showed that IGS can increase the spectral efficiency (SE) of the 2-user IC under realistic assumptions regarding the channel state information (CSI). These papers present a few representative examples of the superiority of IGS in interference-limited systems. For a more detailed overview of the applications of improper signaling in wireless communications, the reader is referred to [32].

TIN is a simple and practical decoding scheme, which is shown to be optimal from a generalized-DoF (GDoF) perspective when the interference level is low [33]. However, TIN can be highly suboptimal in the presence of strong interference. Under strong interference, the optimal decoding strategy is SIC, which is the basic technique in NOMA. As an example, [34] has shown that under strong interference NOMA can significantly improve the performance of TIN in a single-input single-output (SISO) BC. Of course, applying SIC among all users entails some challenges especially in large-scale multicell BCs with multiple-antenna BSs and many

users per cell [35]. One of the challenges to implement SIC is to obtain the optimal user ordering, which may require to solve an NP-hard problem [36]. Moreover, applying SIC with many users per cell requires complex devices to be able to decode and subtract the signal of several users. To address these issues, hybrid schemes have been proposed in which users are clustered so that SIC is only applied to the users in a given cluster.

How to solve the aforementioned challenges remains an open question in RIS-assisted systems. RIS is an emerging technology, which has been shown to be able to improve the spectral and energy efficiency of various systems [1]–[4], [31], [36]–[49]. For example, the papers [37]–[39], [46] considered various performance metrics such as global energy efficiency (GEE), sum rate, minimum SINR and showed that RIS can improve the spectral and energy efficiency of a single-cell BC with a multi-antenna BS and multiple single-antenna users. The paper [40] considered a multiple-input, multiple-output (MIMO) multicell BC and showed that RIS can improve the weighted-sum rate of the system. In [41], it was shown that RIS can improve the weighted-sum rate for secondary users in an underlay cognitive radio (CR) system. In addition, RIS can increase the achievable rate of an orthogonal-frequency-division-multiplexing (OFDM) scheme for a single-antenna point-to-point RIS-assisted system, as shown in [42]. The authors in [47] considered instantaneous, but noisy channel state information (CSI) at transmitters and proposed a framework for robust designs in MISO RIS-assisted systems.

Despite these promising results, there are still many research questions regarding the use of RIS in multicell BCs that must be thoroughly investigated. For example, RIS can modulate channels and cancel leakage and/or interference links [50]. It should be clarified whether and/or how we can exploit this feature in different interference-limited systems. For instance, in a single-cell SISO BC with two users, both the useful and interfering signals are received via the same channel, and canceling the interference link cancels the useful signal as well [51]. This means that there are some limitations on the use of RIS as an interference-management technique. As a result, other interference-management techniques may be needed to fully exploit benefits of RIS, especially in overloaded networks. Indeed, as we will show in this work, RIS cannot fully handle interference in overloaded multicell BCs, but it can greatly improve the system coverage and even increase the benefits of IGS and NOMA. Another important parameter for the performance of RIS is its position in the system. Due to a more severe large-scale path loss of RIS links, it is important to have a line-of-sight (LoS) link between BS and RIS as well as between RIS and users. Thus, RISs should be located close to users to reduce the path loss and increase the probability of having a LoS link. Otherwise, the benefit of RIS would be minor [52].

B. Motivation

In our previous study [52], we showed that IGS can improve the spectral and EE of multicell RIS-assisted BCs with TIN, so IGS and RIS provide additive benefits. As indicated, TIN is optimal only when the interference level is low. However, it is not the case in modern wireless communication systems, where the performance of such systems are mainly limited by interference [5]. Thus, it is important to investigate whether and how IGS can improve the performance of RIS-assisted systems when it is combined with another advanced interference-management technique, which employs SIC to cancel part of interference. This, indeed, motivates us to investigate the performance of more advanced interference-management techniques (a combination of IGS and NOMA) in a multicell RIS-assisted BC.

IGS has been shown to be a powerful interference-management technique; however, it also has its own limitations. Particularly, when we increase the number of resources in frequency (by OFDM [53]), in time (by time sharing [54]), or in space (by MIMO [10]) for a fixed number of users, the benefits of IGS decrease or may even vanish. Thus, one would expect that IGS might not be beneficial if NOMA is applied since a significant part of the interference would be canceled by SIC. However, we show that IGS can provide a considerable gain with NOMA. Indeed, our results show that NOMA and IGS can be employed as complementary and mutually beneficial interference-management techniques. Note that [24], [25] also showed that IGS with NOMA can improve the minimum rate of users in a multicell BC without RIS. In this paper, to the best of our knowledge, we apply a NOMA-based IGS scheme to RIS-assisted systems to improve rate and EE metrics for the first time in the literature. Additionally, we will provide a comprehensive analysis on the RIS performance in practical scenarios with realistic assumptions and try to answer the following research questions:

- When do we need interference-management techniques such as NOMA and IGS in RIS-assisted systems?
- How many RIS components do we need and how should they be configured/optimized to reap benefits from RIS?

These questions are very important for implementing RIS in practice since we should avoid adding unnecessary complexities introduced by considering additional interference-management techniques and/or adding more RIS components. We provide a brief comparison between our work and most related papers in the literature in Table I.

C. Contribution

In this paper, we propose NOMA-based IGS schemes to improve the spectral and energy efficiency of multicell RIS-assisted BCs. We make realistic assumptions regarding the channel model, large-scale path loss, and reflecting coefficients of RIS components by considering different feasibility sets for the reflecting coefficients based on the models in [2]. We consider a multicell BC in which multiple-antenna BSs serve several single-antenna users. Furthermore, we consider multiple RISs in the system such that each cell has at least one RIS. The reason is that, as we showed in [52], the position of RISs plays a key role, and distributed implementation of RIS locations can considerably outperform a co-located RIS implementation.

TABLE I: A brief comparison of the most related works.

	This paper	[52]	[24], [25]	[31], [46]	[9]–[12], [14]–[22]	[39]–[42]	[43]–[45]	[37]
IGS	√	√	√	√	√			
RIS	√	√		√		√	√	√
NOMA	√		√				√	
EE metrics	√	√						√

We aim at maximizing either the minimum-weighted rate or the EE of users, which are practical metrics, widely used in the literature [18], [23]–[26], [28], [55], [56]. Employing a rate/EE profile technique, the whole rate/EE region can be characterized by solving the maximization of the minimum-weighted rate/EE [18], [23]. We show that these metrics can be improved even by RIS with a small number of elements. However, a RIS alone is not able to fully cancel intracell interference when BSs are overloaded, which enforces us to employ additionally other interference-management techniques such as IGS and NOMA. In the proposed NOMA-based IGS schemes, we divide users into several clusters with two users each. We then apply NOMA to each 2-user cluster and employ IGS to handle the remaining interference. The proposed NOMA-based IGS scheme substantially outperforms the IGS schemes in [52] as well as NOMA with PGS from both spectral and energy efficiency points of view.

Three different feasibility sets for the RIS components are considered. In the first set, whose performance can be viewed as upper bound of more practical RIS-assisted systems, the amplitude and the phase of each RIS component can be optimized independently. Our proposed algorithms converge to a stationary point of the considered optimization problems for this feasibility set. In the second, a more practical feasibility set is considered in which only the phase of each RIS component can be optimized, while the amplitude is fixed. This feasibility set can be referred to as the reflecting RIS [3]. In the third, we employ the model in [57] in which the amplitude is not constant, but it is a function of the phase of RIS components. Our numerical results suggest that RIS can substantially improve the performance of the system with all these feasibility sets even with a small number of RIS components. Furthermore, reflecting RIS can perform close to the upper bound in some regimes. However, the performance gap increases with the transmission power of BSs.

We, furthermore, consider the impact of various parameters such as the number of BS antennas, the number of RIS components, and the number of users per cell on the performance of our proposed interference-management techniques. The goal is to find operational points that improve the system performance. Our numerical results show that the proposed NOMA-based IGS scheme can provide a significant gain when BSs are overloaded, i.e., the number of users per cell is larger than the number of transmit antennas at BSs. The benefits of the proposed schemes increase with the number of users per cell for a fixed number of BS antennas. However, the improvements are diminishing in the number of BSs antennas. Indeed, the more overloaded the BSs are, the more gains NOMA-based IGS provides. This finding is also in line with our previous studies in [10], [52]; which altogether corroborate that the higher the interference level is, the more benefits IGS
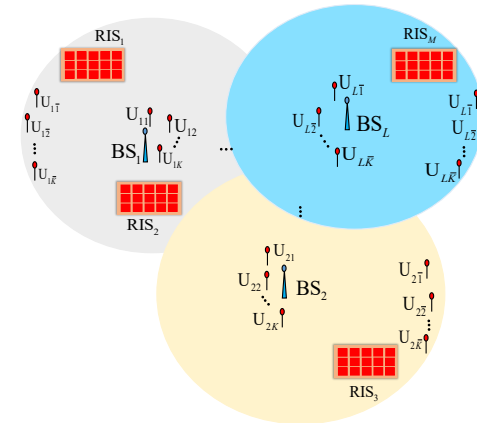

Fig. 1: A multicell broadcast channel with RIS.

provides.

The main contributions of this work can be summarized as in the following:

- We propose NOMA-based IGS schemes for multicell RIS-assisted BCs and show that RIS cannot fully mitigate interference even when the RIS components are properly optimized. Therefore, RIS has to be employed along with a suitable interference-management technique (such as IGS/NOMA) to fully reap its benefit.
- We show that the combination of NOMA with IGS techniques can substantially improve the performance of RIS-assisted multicell downlink channels whenever the BSs are overloaded.
- Our results show that RIS can provide significant gains even with a relatively small number of RIS components, suggesting that reconfigurable intelligent surfaces can be a promising and practical technology.

D. Paper outline

This paper is organized as follows. Section II presents the RIS and system model, states the assumptions for the proposed NOMA-based IGS scheme, and presents the rate/EE metrics. Section III and Section IV, respectively, present the proposed algorithms to solve the minimum-weighted rate and EE maximization problems. Section V provides numerical examples, and Section VI concludes the paper.

II. SYSTEM MODEL

We consider a multicell broadcast channel with L multi-antenna base stations (BSs) in which each BS has N_{BS} antennas and serves $2K$ single-antenna users. We assume that there are $M \geq L$ RISs with N_{RIS} reflecting elements in the system to assist the BSs. In [52], it was shown that a distributed implementation of RIS can outperform a centralized RIS in

which all RIS elements are co-located. The reason is that the link through RIS may suffer from a more severe large-scale path loss in comparison with the direct link. Thus, we assume that in each cell there is at least one RIS. Note that we consider symmetric BSs/cells/RISs for notational simplicity, but the analysis can be easily extended to an asymmetric scenario with different number of users associated to each BS, different number of BS antennas, or different number of reflecting elements in each RIS.

A. RIS model

In this paper, we assume perfect and global CSI similar to [36]–[41], [43]–[46], [50]. We employ the channel model in [40], [52] for RIS-assisted systems. For the sake of completeness, we review the model in this subsection and briefly discuss its features. In RIS-assisted systems, there can be two types of links between a transmitter and a receiver: a direct link and links through RISs. Thus, the channel between BS i and user k associated to BS l, i.e., u_{lk}, is [40], [52]

$$\mathbf{h}_{lk,i}\left(\{\boldsymbol{\Theta}\}\right) = \underbrace{\sum_{m=1}^{M} \mathbf{f}_{lk,m} \boldsymbol{\Theta}_m \mathbf{G}_{mi}}_{\text{Links through RIS}} + \underbrace{\mathbf{d}_{lk,i}}_{\text{Direct link}} \in \mathbb{C}^{1 \times N_{BS}}, \quad (1)$$

where $\mathbf{d}_{lk,i} \in \mathbb{C}^{1 \times N_{BS}}$ is the direct link between BS i and u_{lk}, $\mathbf{G}_{mi} \in \mathbb{C}^{N_{RIS} \times N_{BS}}$ is the channel matrix between BS i and RIS m, and $\mathbf{f}_{lk,m} \mathbb{C}^{1 \times N_{RIS}}$ is the channel vector between RIS m and u_{lk}. The matrix $\boldsymbol{\Theta}_m \in \mathbb{C}^{N_{RIS} \times N_{RIS}}$ contains the reflecting coefficients of RIS m as

$$\boldsymbol{\Theta}_m = \text{diag}\left(\theta_{m1}, \theta_{m2}, \cdots, \theta_{mN_{RIS}}\right), \quad (2)$$

where each θ_{mi} for all m, i is a complex-valued optimization parameter. We represent the set of reflecting coefficients $\{\boldsymbol{\Theta}_m\}_{m=1}^{M}$ as $\{\boldsymbol{\Theta}\}$. To simplify the notations, we drop the dependency of the channels with respect to $\{\boldsymbol{\Theta}\}$ unless it causes confusion. Note that in this paper, we consider a regular RIS, which can only reflect signals. Indeed, both transceivers should be in the reflection space of a RIS in order to have a link through the RIS. If BS i (or user u_{lk}) is not in the reflection space of RIS m, then we have $\mathbf{G}_{mi} = \mathbf{0}$ (or $\mathbf{f}_{lk,m} = \mathbf{0}$). Moreover, note that we can optimize the channels only through the reflecting coefficients θ_{mi} since the other parameters in (1) are given. The readers are referred to [40], [52] for a discussion on the large-scale and small-scale fading of the channels.

1) Feasibility sets for the RIS components: Here, we describe three common feasibility sets for the RIS components. In the ideal case, the amplitude and the phase of each reflecting coefficient can be independently optimized, which leads to the following feasibility set [2, Eq. (11)]

$$\mathcal{T}_U = \left\{\theta_{mi} : |\theta_{mi}|^2 \leq 1 \ \forall m, i\right\}. \quad (3)$$

This feasibility set provides a performance upper bound.

In a more realistic feasibility set, often found in the literature, we can tune the phase of each reflecting coefficient continuously between 0 and 2π, while the amplitude is fixed to 1 [2], [3]. This feasibility set can be written as

$$\mathcal{T}_I = \left\{\theta_{mi} : |\theta_{mi}| = 1 \ \forall m, i\right\}. \quad (4)$$

This case is also referred to as intelligent reflecting surface (IRS) to emphasize that there is only a passive phase-shifting beamforming at the intelligent surfaces [3]. This feasibility set has been widely used in the studies related to RIS [2], [3], [38]–[41], [46]. Note that, in some studies such as [39], an attenuation factor for the reflecting coefficients is considered such that $|\theta_{mi}| = \eta < 1$. In this paper, we do not consider this attenuation factor; however, our analysis can easily be adapted to the feasibility set in [39].

There is yet another feasibility set in which the amplitude is not constant, but instead it is a deterministic function of the phase as [57]

$$0 \leq |\theta|_{\min} \leq |\theta_{mi}| = f(\angle\theta_{mi}) \leq |\theta|_{\max} \leq 1 \ \forall m, i, \quad (5)$$

where $f(\cdot)$ is

$$f(\angle\theta_{mi}) = |\theta|_{\min} + (1 - |\theta|_{\min}) \left(\frac{\sin\left(\angle\theta_{mi} - \phi\right) + 1}{2}\right)^{\alpha}, \quad (6)$$

where $|\theta|_{\min}$, α and ϕ are non-negative constant values. According to this model, the amplitude and phase of RIS components are not independent optimization parameters. The feasibility set of θ_{mi} for this model is

$$\mathcal{T}_C = \left\{\theta_{mi} : |\theta_{mi}| = f(\angle\theta_{mi}), \angle\theta_{mi} \in [-\pi, \pi] \ \forall m, i\right\}. \quad (7)$$

Note that in [57], $|\theta|_{\max} = 1$. Moreover, when $\alpha = 0$, or equivalently $|\theta|_{\min} = 1$, this feasibility set is equivalent to $\mathcal{T}_I$. Note than $\mathcal{T}_I \subset \mathcal{T}_U$ and $\mathcal{T}_C \subset \mathcal{T}_U$, which implies that the solution for the feasibility set $\mathcal{T}_U$ is an upper bound for the performance of the other feasibility sets.

B. Signal model

The broadcast signal from BS l is

$$\mathbf{x}_l = \sum_{k=1}^{2K} \mathbf{x}_{lk} \in \mathbb{C}^{N_{BS} \times 1}, \quad (8)$$

which is a superposition of the signals intended for all users associated to BS l. We assume that $\mathbf{x}_{lk}$s for all l, k are uncorrelated zero-mean (and possibly improper) Gaussian signals. Let us remind the reader that the real and imaginary parts of improper signals are correlated and/or have unequal powers [58], [59]. To deal with impropriety, we employ the real-decomposition method, which is more suitable for studying this scenario. By the real-decomposition method, all parameters are written in a real domain by considering the real and imaginary parts of a signal as separate variables. We represent the real decomposition of $\mathbf{x}_{lk}$ by $\underline{\mathbf{x}}_{lk} = \left[\ \Re\{\mathbf{x}_{lk}\}^T \ \Im\{\mathbf{x}_{lk}\}^T\ \right]^T$, where $\Re\{x\}$ and $\Im\{x\}$ take, respectively, the real and imaginary part of x. Moreover, we represent the transmit covariance matrix of $\underline{\mathbf{x}}_{lk}$ by $\mathbf{P}_{lk} = \mathbb{E}\left\{\underline{\mathbf{x}}_{lk}\underline{\mathbf{x}}_{lk}^T\right\}$.

The received signal for the user u_{lk} is

$$y_{lk} = \sum_{i=1}^{L} \mathbf{h}_{lk,i} \sum_{j=1}^{2K} \mathbf{x}_{ij} + \mathbf{n}_{lk} \quad (9a)$$

$$= \underbrace{\mathbf{h}_{lk,l}\mathbf{x}_{lk}}_{\text{Desired signal}} + \underbrace{\mathbf{h}_{lk,l}\sum_{j=1,j\neq k}^{2K}\mathbf{x}_{lj}}_{\text{Intracell interference}} + \underbrace{\sum_{i=1,i\neq l}^{L}\mathbf{h}_{lk,i}\mathbf{x}_i}_{\text{Intercell interference}} + \underbrace{n_{lk}}_{\text{Noise}},$$

$$(9b)$$

where $\mathbf{h}_{lk,l} \in \mathbb{C}^{1\times N_{BS}}$ is the channel between BS l and u_{lk}, and $n_{lk} \in \mathbb{C}$ is a zero-mean additive white Gaussian noise with variance σ^2. Using the real-decomposition method to model IGS, (9) becomes

$$\underline{\mathbf{y}}_{lk} = \underbrace{\underline{\mathbf{H}}_{lk,l}\underline{\mathbf{x}}_{lk}}_{\text{Desired signal}} + \underbrace{\underline{\mathbf{H}}_{lk,l}\sum_{j=1,j\neq k}^{2K}\underline{\mathbf{x}}_{lj}}_{\text{Intracell interference}} + \underbrace{\sum_{i=1,i\neq l}^{L}\underline{\mathbf{H}}_{lk,i}\underline{\mathbf{x}}_i}_{\text{Intercell interference}} + \underbrace{\underline{\mathbf{n}}_{lk}}_{\text{Noise}},$$

$$(10)$$

where $\underline{\mathbf{y}}_{lk} = \begin{bmatrix} \Re\{y_{lk}\} & \Im\{y_{lk}\} \end{bmatrix}^T \in \mathbb{R}^{2\times1}$, $\underline{\mathbf{n}}_{lk} = \begin{bmatrix} \Re\{n_{lk}\} & \Im\{n_{lk}\} \end{bmatrix}^T \in \mathbb{R}^{2\times1}$, and the channels in the real-decomposition method, $\underline{\mathbf{H}}_{lk,i} \in \mathbb{R}^{2\times 2N_{BS}}$ for all i,l,k, are given by

$$\underline{\mathbf{H}}_{lk,i} = \begin{bmatrix} \Re\{\mathbf{h}_{lk,i}\} & -\Im\{\mathbf{h}_{lk,i}\} \\ \Im\{\mathbf{h}_{lk,i}\} & \Re\{\mathbf{h}_{lk,i}\} \end{bmatrix}. \quad (11)$$

C. NOMA

We assume that the intercell interference (ICI) is treated as noise. Note that treating ICI as noise is a practical assumption since it might be very inefficient to decode and cancel ICI in a multicell BC with many users per cell [36], [60]–[65]. The reason is that, due to the decodability constraint, the data transmission rates may be restricted by the weakest link between a BS and all the users (either inside or outside of the BS cell), which may significantly degrade the performance. Moreover, it may add unnecessary complexities to the system design such as finding the optimal decoding order of the ICI messages and highly increase the signaling overheads of the system. It should be noted that treating ICI as noise does not mean that ICI is overlooked in the system design. In this paper, we jointly employ IGS and RIS to manage ICI by a careful design of the transmit covariance matrices and RIS components.

We, additionally, employ a NOMA-based technique to handle a part of the intracell interference. Employing a full SIC scheme is very challenging even in a single-cell RIS-assisted broadcast channel [66], [67]. The optimal user ordering in a single-cell RIS-assisted BC is conducted based on the effective signal-to-interference-plus-noise ratio (SINR), which depends on channel gains. However, in RIS-assisted systems, the effective SINR and consequently the user ordering depend on the channel coefficients, which in turn depend on the reflecting coefficients [66]. Additionally, user ordering in a multicell BC is a difficult task since the users in a cell experience different levels of intercell interference, thus making it even more complicated to find the optimal user ordering in a multicell RIS-assisted BC, and to the best of our knowledge, the optimal decoding strategy for such systems is unknown. Moreover, assuming $2K$ users per cell, there are $L \times 2K!$ possible user orderings in the multicell BC, so an exhaustive

search is impossible from a practical point of view [66], [67]. In addition to the complexities involved with finding the optimal decoding order, there are also other challenges to apply NOMA to a large number of users. For instance, the receives should be able to decode and cancel the signals of multiple users, which may increase the complexities and cost of devices. Thus, it may not be realistic to employ a full NOMA scheme especially when the number of users per cell grows, which motivates employing hybrid schemes [34], [68]–[71]. Note that hybrid schemes can be very practical and efficient and may provide a suitable trade off between the system performance and complexity [34]. Therefore, in this paper, we employ a hybrid scheme in which the users in each cell are divided into clusters with two users each, and NOMA is applied only to the pairs of users. The optimal user ordering/pairing depends on the channel coefficients as well as the ICI level at the users for a multicell BC, and to the best of our knowledge, the optimal strategy for a multi-cell MISO RIS-assisted BC is unknown.

1) Proposed user ordering/pairing scheme: In this paper, we assume that the user pairing is given, and we do not tackle obtaining the optimal user pairing/ordering. Instead, we focus on different aspects of RISs as well as on the resulting optimization problems, and we leave the optimal user pairing/ordering scheme for a future study. Note that a suboptimal approach for user pairing can be to pair users based on their distance to the serving BS [24], [25]. For instance, in [24], it was assumed that users in each cell belong to two groups with the same number of members in each group. They can be called cell-centric users (CCUs) and cell-edge users (CEUs). Then each CCU is randomly paired with a CEU. Due to the large-scale fading, it is reasonable to assume that CEUs experience a higher ICI level than CCUs in systems without RIS. Additionally, the direct link between the BS and one of its cell-edge serving user is weaker than the direct link for a CCU associated to the BS. It should be noted that such assumptions may not hold in RIS-assisted systems since RISs can modulate the channels, which implies this pairing scheme is not necessarily optimal in a multi-cell MISO RIS-assisted BC. However, this pairing scheme can be practical and involves much lower complexities than the optimal strategy.

2) Rate expressions for IGS: In the following, we consider the optimization problems for a given user pairing policy. We represent the user with a higher decoding order with k and its paired user with $\bar{k} = K + k$. According to the pairing scheme, user k associated to BS l, u_{lk}, first decodes the signal of user $\bar{k}$ associated to BS l, $\mathrm{u}_{l\bar{k}}$, and then subtract it from its received signal. In other words, the data intended for $\mathrm{u}_{l\bar{k}}$ should be transmitted in a data rate that is decodable by both u_{lk} and $\mathrm{u}_{l\bar{k}}$. Thus, the rate of user $\bar{k}$ associated to BS l, $\mathrm{u}_{l\bar{k}}$, is [24]

$$r_{l\bar{k}} = \min\left(\bar{r}_{l\bar{k}}, \bar{r}_{lk\to\bar{k}}\right), \quad (12)$$

where $\bar{r}_{l\bar{k}}$ is the rate at $\mathrm{u}_{l\bar{k}}$ when trying to decode its own signal:

$$\bar{r}_{l\bar{k}} = \frac{1}{2}\log_2\left|\mathbf{I} + \mathbf{D}_{l\bar{k}}^{-1}\underline{\mathbf{H}}_{l\bar{k},l}\mathbf{P}_{l\bar{k}}\underline{\mathbf{H}}_{l\bar{k},l}^T\right| \quad (13)$$

$$= \frac{1}{2} \log_2 \left| \mathbf{D}_{l\bar{k}} + \underline{\mathbf{H}}_{l\bar{k},l} \mathbf{P}_{l\bar{k}} \underline{\mathbf{H}}_{l\bar{k},l}^T \right| - \underbrace{\frac{1}{2} \log_2 |\mathbf{D}_{l\bar{k}}|}_{\bar{r}_{l\bar{k},2}}, \quad (14)$$

where $\mathbf{D}_{l\bar{k}}$ is the interference-plus-noise covariance matrix at $\mathrm{u}_{l\bar{k}}$

$$\mathbf{D}_{l\bar{k}} = \underbrace{\sum_{i=1,i\neq l}^{L} \underline{\mathbf{H}}_{l\bar{k},i} \mathbf{P}_i \underline{\mathbf{H}}_{l\bar{k},i}^T}_{\text{Intercell interference}} + \underbrace{\sum_{j=1,j\neq\bar{k}}^{2K} \underline{\mathbf{H}}_{l\bar{k},l} \mathbf{P}_{lj} \underline{\mathbf{H}}_{l\bar{k},l}^T}_{\text{Intracell interference}} + \underbrace{\frac{\sigma^2}{2}\mathbf{I}}_{\text{Noise}}, \quad (15)$$

where $\mathbf{P}_i = \sum_{j=1}^{2K} \mathbf{P}_{ij}$. Note that the intracell interference in (15) consists of two terms: intracell-intracluster and intracell-intercluster interference. The intracell-intracluster interference consists of the signal intended for the paired user k, i.e., $\underline{\mathbf{H}}_{l\bar{k},i} \mathbf{P}_i \underline{\mathbf{H}}_{l\bar{k},i}^T$, while the intracell-intercluster interference term consists of the signal intended for users of the other clusters in cell l. Moreover, the term $\bar{r}_{lk\to\bar{k}}$ is the rate at u_{lk} when trying to decode the signal intended for $\mathrm{u}_{l\bar{k}}$:

$$\bar{r}_{lk\to\bar{k}} = \frac{1}{2} \log_2 \left| \mathbf{I} + \left[\mathbf{D}_{lk} + \underline{\mathbf{H}}_{lk,l} \mathbf{P}_{lk} \underline{\mathbf{H}}_{lk,l}^T \right]^{-1} \underline{\mathbf{H}}_{lk,l} \mathbf{P}_{l\bar{k}} \underline{\mathbf{H}}_{lk,l}^T \right| \quad (16)$$

$$= \underbrace{\frac{1}{2} \log_2 \left| \mathbf{D}_{lk} + \underline{\mathbf{H}}_{lk,l} \mathbf{P}_{lk} \underline{\mathbf{H}}_{lk,l}^T + \underline{\mathbf{H}}_{lk,l} \mathbf{P}_{l\bar{k}} \underline{\mathbf{H}}_{lk,l}^T \right|}_{\bar{r}_{lk\to\bar{k},1}} - \underbrace{\frac{1}{2} \log_2 \left| \mathbf{D}_{lk} + \underline{\mathbf{H}}_{lk,l} \mathbf{P}_{lk} \underline{\mathbf{H}}_{lk,l}^T \right|}_{\bar{r}_{lk\to\bar{k},2}}, \quad (17)$$

where $\mathbf{D}_{lk}$ can similarly be written as

$$\mathbf{D}_{lk} = \underbrace{\sum_{i=1,i\neq l}^{L} \underline{\mathbf{H}}_{lk,i} \mathbf{P}_i \underline{\mathbf{H}}_{lk,i}^T}_{\text{Intercell interference}} + \underbrace{\sum_{j=1,j\neq k,\bar{k}}^{2K} \underline{\mathbf{H}}_{lk,l} \mathbf{P}_{lj} \underline{\mathbf{H}}_{lk,l}^T}_{\text{Intracell interference}} + \underbrace{\frac{\sigma^2}{2}\mathbf{I}}_{\text{Noise}}, \quad (18)$$

where the intracell interference consists of only the intracell-intercluster interference since the intracell-intracluster interference is canceled by employing SIC. Note that by transmitting data to $\mathrm{u}_{l\bar{k}}$ at rate $r_{l\bar{k}}$ in (12), it is ensured that the data is decodable at both u_{lk} and $\mathrm{u}_{l\bar{k}}$ users. Additionally, the rate for u_{lk} is

$$r_{lk} = \frac{1}{2} \log_2 \left| \mathbf{I} + \mathbf{D}_{lk}^{-1} \underline{\mathbf{H}}_{lk,l} \mathbf{P}_{lk} \underline{\mathbf{H}}_{lk,l}^T \right| \quad (19)$$

$$= \underbrace{\frac{1}{2} \log_2 \left| \mathbf{D}_{lk} + \underline{\mathbf{H}}_{lk,l} \mathbf{P}_{lk} \underline{\mathbf{H}}_{lk,l}^T \right|}_{\tilde{r}_{lk,1}} - \underbrace{\frac{1}{2} \log_2 |\mathbf{D}_{lk}|}_{\tilde{r}_{lk,2}}. \quad (20)$$

Note that we write the rates in two different formats so that we can use the more suitable format for each optimization problem. For instance, to optimize over transmit covariance matrices it is more convenient to write the rates as a difference of two concave functions in $\{\mathbf{P}\} = \{\mathbf{P}_{lk}, \forall l, k\}$ since it allows us to employ majorization-minimization (MM) algorithms. However, for optimizing over the reflecting coefficients, we employ the rate expressions in (14), (17) and (20), as will be discussed in the following sections.

3) Energy-efficiency: The EE of each user is defined as the ratio between its achievable rate and the power intended for the data transmission to the user, i.e., [72], [73]

$$EE_{lk} = \frac{r_{lk}}{P_c + \eta \mathrm{Tr}(\mathbf{P}_{lk})}, \quad (21)$$

where η^{-1} is the power transmission efficiency of each BS, and P_c is the constant power consumption for transmitting data to a user, given by [52, Eq. (27)].

III. MAXIMIZING THE MINIMUM-WEIGHTED RATE

As a figure of merit for the spectral efficiency and coverage, we maximize the minimum-weighted rate, which can be written as

$$\max_{r,\{\boldsymbol{\Theta}\}\in\mathcal{T},\{\mathbf{P}\}\in\mathcal{P}} r, \quad \text{s.t.} \quad \lambda_{lk} r_{lk} \geq r, \quad \forall l, k, \quad (22)$$

where λ_{lk} is the corresponding non-negative weight, $r_{l\bar{k}} = \min(\bar{r}_{l\bar{k}}, \bar{r}_{lk\to\bar{k}})$, and $\mathcal{P}$ denotes the feasibility set of the transmit covariance matrices, which for the IGS and PGS schemes are, respectively, [52]

$$\mathcal{P}_I = \left\{ \{\mathbf{P}\} : \sum_{k=1}^{2K} \mathrm{Tr}(\mathbf{P}_{lk}) \leq p_l, \mathbf{P}_{lk} \succeq \mathbf{0}, \forall l, k \right\}, \quad (23)$$

$$\mathcal{P}_P = \left\{ \{\mathbf{P}\} : \sum_{k=1}^{2K} \mathrm{Tr}(\mathbf{P}_{lk}) \leq p_l, \mathbf{P}_{lk} = \mathbf{P}_t, \mathbf{P}_{lk} \succeq \mathbf{0}, \forall l, k \right\}, \quad (24)$$

where p_l is the power budget of BS l, and $\mathbf{P}_t$ is a 2×2 block matrix that fulfills the structure in [52, Eq. (5)]. As can be verified through (23) and (24), PGS is a special case of IGS, and thus, the optimal IGS scheme never performs worse than the optimal PGS scheme. Readers can refer to [58], [59] for further details on improper signaling. Note that our proposed algorithms can be applied to both IGS and PGS schemes, and to write the equations for the most general case, we represent the feasibility set by $\mathcal{P}$. Moreover, note that employing the rate profile technique, we can characterize the rate region by solving (22) and varying the weights for all possible λ_{lk}, satisfying $\sum_{\forall l} \sum_{\forall k} \lambda_{lk}^{-1} = 1$ [52]. Additionally, the weights λ_{lk} can be chosen based on the priorities of the users [52].

Unfortunately, (22) is a non-convex difficult programming problem, and it is complicated to find its global optimal solution. Hence, we propose a suboptimal scheme by employing an iterative alternating optimization (AO) approach. That is, at iteration t, we first fix the reflecting coefficients to the solution of the previous step, $\{\boldsymbol{\Theta}^{(t-1)}\}$, and update the covariance matrices, $\{\mathbf{P}^{(t)}\}$. We then fix the covariance matrices $\{\mathbf{P}^{(t)}\}$ and optimize over reflecting coefficients to obtain $\{\boldsymbol{\Theta}^{(t)}\}$. We iterate this approach until a convergence criterion is met. Note that the convergence point of our algorithm may depend on the initial point since it is an iterative algorithm, based on MM. The initial point should be feasible and can be chosen either randomly or based on a heuristic approach [74]. To obtain a better performance, we can run the algorithms for several initial points and choose the best solution among them [74]. It should be noted that the proposed algorithms actively optimize the transmission parameters at BSs as well as the environment through RIS. Indeed, each iteration consists of two steps. The

first step is to optimize the transmit covariance matrices, which are transmission parameters at BSs. The second step is to optimize the RIS components, which modify the environment.

Unfortunately, the corresponding optimization problems are still complicated even when we fix covariance matrices or reflecting coefficients. To tackle these problems, we employ MM and propose suboptimal algorithms to solve the optimization problems.

A. Optimization of the transmit covariance matrices

In this subsection, we obtain transmit covariance matrices $\{\mathbf{P}^{(t)}\}$ for given reflecting coefficients $\{\mathbf{\Theta}^{(t-1)}\}$ by solving

$$\max_{r,\{\mathbf{P}\}\in\mathcal{P}} r, \qquad \text{s.t.} \quad \lambda_{lk}r_{lk} \geq r, \qquad \forall l,k. \qquad (25)$$

To solve this non-convex problem, we can apply difference of convex programming (DCP) to obtain a suboptimal solution for (25). The reason is that the rates can be written as a difference of two concave functions. Note that DCP falls into MM algorithms and obtains a stationary point of the original problem [75]. DCP is based on the first-order Taylor expansion that approximates the convex part of the rates with an affine (linear) function. To find a suitable surrogate function for the rates, we employ the results in the following lemma.

Lemma 1. Consider the function $f(\mathbf{P}) = \log\det(\mathbf{A} + \mathbf{B}\mathbf{P}\mathbf{B}^T)$, where $\mathbf{A} \in \mathbb{R}^{N\times N}$ and $\mathbf{B} \in \mathbb{R}^{N\times M}$ are constant matrices. Moreover, $\mathbf{A}$ and $\mathbf{P}$ are $M \times M$ real positive semi-definite matrices. Using the first-order Taylor expansion, an affine upper bound for $f(\mathbf{P})$ can be found as

$$f(\mathbf{P}) \leq f(\mathbf{P}^{(t)}) + \mathrm{Tr}\left(\mathbf{B}^T(\mathbf{A} + \mathbf{B}\mathbf{P}^{(t)}\mathbf{B}^T)^{-1}\mathbf{B}(\mathbf{P} - \mathbf{P}^{(t)})\right),$$

where $\mathbf{P}^{(t)}$ is any feasible fixed point.

Proof. Employing the first-order Taylor expansion, we have

$$f(\mathbf{P}) \leq f(\mathbf{P}^{(t)}) + \mathrm{Tr}\left(\left[\frac{\partial f(\mathbf{P})}{\partial \mathbf{P}}|_{\mathbf{P}^{(t)}}\right]^T (\mathbf{P} - \mathbf{P}^{(t)})\right),$$

where $\frac{\partial f(\mathbf{P})}{\partial \mathbf{P}}|_{\mathbf{P}^{(t)}}$ is the derivative of $f(\mathbf{P})$ with respect to $\mathbf{P}$ at $\mathbf{P}^{(t)}$, which is given by

$$\frac{\partial f(\mathbf{P})}{\partial \mathbf{P}}|_{\mathbf{P}^{(t)}} = \mathbf{B}^T(\mathbf{A} + \mathbf{B}\mathbf{P}^{(t)}\mathbf{B}^T)^{-1}\mathbf{B}.$$

Note that $\frac{\partial f(\mathbf{P})}{\partial \mathbf{P}}$ is symmetric, thus we have $\left[\frac{\partial f(\mathbf{P})}{\partial \mathbf{P}}|_{\mathbf{P}^{(t)}}\right]^T = \frac{\partial f(\mathbf{P})}{\partial \mathbf{P}}|_{\mathbf{P}^{(t)}}$. $\square$

A suitable concave lower bound for rates can be obtained by employing Lemma 1. That is, we keep the concave part of the rate functions and approximate the convex part by a linear lower-bound function using Lemma 1. Thus, the surrogate function for the rate of u_{lk}, i.e., the the user with a higher order, can be written as

$$r_{lk} \geq \tilde{r}_{lk} = r_{lk,1}\left(\{\mathbf{P}\},\{\mathbf{\Theta}^{(t-1)}\}\right) - r_{lk,2}^{(t-1)} -$$
$$\sum_{i=1}^{L}\sum_{j=1}^{2K}\mathrm{Tr}\left(\frac{\underline{\mathbf{H}}_{lk,i}^T(\mathbf{D}_{lk}^{(t-1)})^{-1}\underline{\mathbf{H}}_{lk,i}}{2\ln 2}\left(\mathbf{P}_{ij} - \mathbf{P}_{ij}^{(t-1)}\right)\right),$$
$$(26)$$

where $r_{lk,2}^{(t-1)} = r_{lk,2}\left(\{\mathbf{P}^{(t-1)}\},\{\mathbf{\Theta}^{(t-1)}\}\right)$, and $\mathbf{D}_{lk}^{(t-1)} = \mathbf{D}_{lk}\left(\{\mathbf{P}^{(t-1)}\},\{\mathbf{\Theta}^{(t-1)}\}\right)$. Similarly, a concave lower-bound for the rate of $u_{l\bar{k}}$, i.e., the user paired to u_{lk}, can be found as

$$\tilde{r}_{l\bar{k}} = \min\left(\tilde{r}_{l\bar{k}}^l,\tilde{r}_{lk\to\bar{k}}^l\right), \qquad (27)$$

where $\tilde{r}_{lk\to\bar{k}}^l$ is given by (28) on the top of the next page and

$$\tilde{r}_{l\bar{k}}^l = \bar{r}_{l\bar{k},1}\left(\{\mathbf{P}\},\{\mathbf{\Theta}^{(t-1)}\}\right) - \bar{r}_{l\bar{k},2}^{(t-1)}$$
$$- \sum_{i=1}^{L}\sum_{j=1}^{2K}\mathrm{Tr}\left(\frac{\underline{\mathbf{H}}_{l\bar{k},i}^T(\mathbf{D}_{l\bar{k}}^{(t-1)})^{-1}\underline{\mathbf{H}}_{l\bar{k},i}}{2\ln 2}\left(\mathbf{P}_{ij} - \mathbf{P}_{ij}^{(t-1)}\right)\right),$$
$$(29)$$

where $r_{l\bar{k},2}^{(t-1)} = r_{l\bar{k},2}\left(\{\mathbf{P}^{(t-1)}\},\{\mathbf{\Theta}^{(t-1)}\}\right)$, $r_{lk\to\bar{k},2}^{(t-1)} = r_{lk\to\bar{k},2}\left(\{\mathbf{P}^{(t-1)}\},\{\mathbf{\Theta}^{(t-1)}\}\right)$, and $\mathbf{D}_{l\bar{k}}^{(t-1)} = \mathbf{D}_{l\bar{k}}\left(\{\mathbf{P}^{(t-1)}\},\{\mathbf{\Theta}^{(t-1)}\}\right)$. Substituting the concave lower bounds in (25), we have

$$\max_{r,\{\mathbf{P}\}\in\mathcal{P}} r, \qquad \text{s.t.} \quad \lambda_{lk}\tilde{r}_{lk} \geq r, \qquad \forall l,k. \qquad (30)$$

This optimization problem is convex and can be efficiently solved to obtain a new set of transmit covariance matrices $\{\mathbf{P}^{(t)}\}$.

B. Optimization of the reflecting coefficients

In this subsection, we update the reflecting coefficients while the covariance matrices are fixed to $\{\mathbf{P}^{(t)}\}$. That is, we solve the following optimization problem

$$\max_{r,\{\mathbf{\Theta}\}\in\mathcal{T}} r, \qquad \text{s.t.} \quad \lambda_{lk}r_{lk} \geq r, \qquad \forall l,k. \qquad (31)$$

Unfortunately, this problem is non-convex and difficult too. To solve it, we first find a suitable concave lower bound for the rates by employing the following lemma. Then, we consider each feasibility set separately.

Lemma 2 ([28], [46]). *The following inequality holds for all $N \times N$ positive definite matrices $\mathbf{Y}$ and $\bar{\mathbf{Y}}$, and any arbitrary $N \times M$ matrices $\mathbf{V}$ and $\bar{\mathbf{V}}$:*

$$\ln\left|\mathbf{I} + \mathbf{V}\mathbf{V}^H\mathbf{Y}^{-1}\right| \geq \ln\left|\mathbf{I} + \bar{\mathbf{V}}\bar{\mathbf{V}}^H\bar{\mathbf{Y}}^{-1}\right|$$
$$- \mathrm{Tr}\left(\bar{\mathbf{V}}\bar{\mathbf{V}}^H\bar{\mathbf{Y}}^{-1}\right) + 2\Re\left\{\mathrm{Tr}\left(\bar{\mathbf{V}}^H\bar{\mathbf{Y}}^{-1}\mathbf{V}\right)\right\}$$
$$- \mathrm{Tr}\left((\bar{\mathbf{Y}}^{-1} - (\bar{\mathbf{V}}\bar{\mathbf{V}}^H + \bar{\mathbf{Y}})^{-1})^H(\mathbf{V}\mathbf{V}^H + \mathbf{Y})\right). \quad (32)$$

Corollary 1. *A concave lower-bound $\hat{r}_{lk}$ and $\hat{r}_{l\bar{k}}$ for, respectively, r_{lk} and $r_{l\bar{k}}$ can be found as*

$$\hat{r}_{lk} = r_{lk}^{(t-1)} - \frac{1}{2\ln 2}\mathrm{Tr}\left(\bar{\mathbf{V}}_{lk}\bar{\mathbf{V}}_{lk}^T\bar{\mathbf{Y}}_{lk}^{-1}\right)$$
$$+ \frac{1}{\ln 2}\mathrm{Tr}\left(\bar{\mathbf{V}}_{lk}^T\bar{\mathbf{Y}}_{lk}^{-1}\underline{\mathbf{H}}_{lk,l}(\{\mathbf{\Theta}\})\mathbf{P}_{lk}^{1/2^{(t)}}\right)$$
$$- \frac{1}{2\ln 2}\mathrm{Tr}\left((\bar{\mathbf{Y}}_{lk}^{-1} - (\bar{\mathbf{V}}_{lk}\bar{\mathbf{V}}_{lk}^T + \bar{\mathbf{Y}}_{lk})^{-1})^T(\mathbf{Y}_{lk\to\bar{k}})\right)$$
$$\hat{r}_{l\bar{k}} = \min\left(\hat{r}_{l\bar{k}}^l,\hat{r}_{lk\to\bar{k}}^l\right),$$

where $r_{lk}^{(t-1)} = r_{lk}\left(\{\mathbf{P}^{(t)}\},\{\mathbf{\Theta}^{(t-1)}\}\right)$. Moreover, $\hat{r}_{l\bar{k}}^l$ and $\hat{r}_{lk\to\bar{k}}^l$ are, respectively, given by (33) and (34), on the top of this page, where $\bar{r}_{l\bar{k}}^{(t-1)} = \bar{r}_{l\bar{k}}\left(\{\mathbf{P}^{(t)}\},\{\mathbf{\Theta}^{(t-1)}\}\right)$ and

This article has been accepted for publication in IEEE Transactions on Signal Processing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSP.2023.3259145

8

$$\tilde{r}_{lk \to \bar{k}}^{l} = \bar{r}_{lk \to \bar{k},1}\left(\{\mathbf{P}\}, \{\boldsymbol{\Theta}^{(t-1)}\}\right) - \bar{r}_{lk \to \bar{k},2}^{(t-1)} - \sum_{i=1}^{L}\sum_{j=1}^{2K} \mathrm{Tr}\left(\frac{\underline{\mathbf{H}}_{lk,i}^{T}\left(\mathbf{D}_{lk}^{(t-1)} + \underline{\mathbf{H}}_{lk,i}\mathbf{P}_{lk}^{(t-1)}\underline{\mathbf{H}}_{lk,i}^{T}\right)^{-1}\underline{\mathbf{H}}_{lk,i}}{2\ln 2}\left(\mathbf{P}_{ij} - \mathbf{P}_{ij}^{(t-1)}\right)\right), \tag{28}$$

$$\hat{r}_{l\bar{k}}^{l} = \bar{r}_{l\bar{k}}^{(t-1)} - \frac{1}{2\ln 2}\,\mathrm{Tr}\left(\bar{\mathbf{V}}_{l\bar{k}}\bar{\mathbf{V}}_{l\bar{k}}^{T}\bar{\mathbf{Y}}_{l\bar{k}}^{-1}\right) + \frac{1}{\ln 2}\,\mathrm{Tr}\left(\bar{\mathbf{V}}_{l\bar{k}}^{T}\bar{\mathbf{Y}}_{l\bar{k}}^{-1}\underline{\mathbf{H}}_{l\bar{k},l}(\{\boldsymbol{\Theta}\})\mathbf{P}_{l\bar{k}}^{1/2^{(t)}}\right)$$
$$- \frac{1}{2\ln 2}\,\mathrm{Tr}\left((\bar{\mathbf{Y}}_{l\bar{k}}^{-1} - (\bar{\mathbf{V}}_{l\bar{k}}\bar{\mathbf{V}}_{l\bar{k}}^{T} + \bar{\mathbf{Y}}_{l\bar{k}})^{-1})^{T}\left(\mathbf{D}_{l\bar{k}} + \underline{\mathbf{H}}_{l\bar{k},l}(\{\boldsymbol{\Theta}\})\mathbf{P}_{l\bar{k}}^{(t)}\underline{\mathbf{H}}_{l\bar{k},l}^{T}(\{\boldsymbol{\Theta}\})\right)\right) \tag{33}$$

$$\hat{r}_{lk \to \bar{k}}^{l} = \bar{r}_{lk \to \bar{k}}^{(t-1)} - \frac{1}{2\ln 2}\,\mathrm{Tr}\left(\bar{\mathbf{V}}_{lk \to \bar{k}}\bar{\mathbf{V}}_{lk \to \bar{k}}^{T}\bar{\mathbf{Y}}_{lk \to \bar{k}}^{-1}\right) + \frac{1}{\ln 2}\,\mathrm{Tr}\left(\bar{\mathbf{V}}_{lk \to \bar{k}}^{T}\bar{\mathbf{Y}}_{lk \to \bar{k}}^{-1}\underline{\mathbf{H}}_{lk,l}(\{\boldsymbol{\Theta}\})\mathbf{P}_{l\bar{k}}^{1/2^{(t)}}\right)$$
$$- \frac{1}{2\ln 2}\,\mathrm{Tr}\left((\bar{\mathbf{Y}}_{lk \to \bar{k}}^{-1} - (\bar{\mathbf{V}}_{lk \to \bar{k}}\bar{\mathbf{V}}_{lk \to \bar{k}}^{T} + \bar{\mathbf{Y}}_{lk \to \bar{k}})^{-1})^{T}(\underline{\mathbf{H}}_{lk,l}(\{\boldsymbol{\Theta}\})\mathbf{P}_{l\bar{k}}^{(t)}\underline{\mathbf{H}}_{lk,l}^{T}(\{\boldsymbol{\Theta}\}) + \mathbf{Y}_{lk \to \bar{k}})\right), \tag{34}$$

$\bar{r}_{lk \to \bar{k}}^{(t-1)} = \bar{r}_{lk \to \bar{k}}\left(\{\mathbf{P}^{(t)}\}, \{\boldsymbol{\Theta}^{(t-1)}\}\right)$. *Furthermore, the other parameters are defined as* $\bar{\mathbf{V}}_{lk} = \underline{\mathbf{H}}_{lk,l}\left(\{\boldsymbol{\Theta}^{(t-1)}\}\right)\mathbf{P}_{lk}^{(t)1/2}$, $\bar{\mathbf{Y}}_{lk} = \mathbf{D}_{lk}\left(\{\mathbf{P}^{(t)}\}, \{\boldsymbol{\Theta}^{(t-1)}\}\right)$, *and*

$$\mathbf{V}_{lk \to \bar{k}} = \underline{\mathbf{H}}_{lk,l}\left(\{\boldsymbol{\Theta}^{(t-1)}\}\right)\mathbf{P}_{l\bar{k}}^{1/2},$$
$$\mathbf{Y}_{lk \to \bar{k}} = \mathbf{D}_{lk}\left(\{\mathbf{P}^{(t)}\}, \{\boldsymbol{\Theta}\}\right) + \underline{\mathbf{H}}_{lk,l}(\{\boldsymbol{\Theta}\})\mathbf{P}_{lk}^{(t)}\underline{\mathbf{H}}_{lk,l}^{T}(\{\boldsymbol{\Theta}\}),$$
$$\bar{\mathbf{Y}}_{lk \to \bar{k}} = \mathbf{D}_{lk}\left(\{\mathbf{P}^{(t)}\}, \{\boldsymbol{\Theta}^{(t-1)}\}\right)$$
$$+ \underline{\mathbf{H}}_{lk,l}(\{\boldsymbol{\Theta}^{(t-1)}\})\mathbf{P}_{lk}^{(t)}\underline{\mathbf{H}}_{lk,l}^{T}(\{\boldsymbol{\Theta}^{(t-1)}\}).$$

Proof. Substituting the corresponding parameters in Lemma 2 results in the lower bounds. Note that since we employ the real-decomposition method, we can substitute the Hermitian operator with the transpose operator. Additionally, we can remove operator $\mathfrak{R}$ because all the variables are real in the real-decomposition method. □

1) Feasibility set $\mathcal{T}_U$: Since $\mathcal{T}_U$ is a convex set, the following problem is convex:

$$\max_{r, \{\boldsymbol{\Theta}\} \in \mathcal{T}_U} r, \qquad \text{s.t.} \quad \lambda_{lk}\hat{r}_{lk} \geq r, \qquad \forall l, k. \tag{35}$$

Hence, we can solve (35) efficiently by existing convex optimization tools. Note that the algorithm for this feasibility set converges to a stationary point of (22) since the surrogate functions fulfill the three conditions in [10, Sec. III]. Our solution for $\mathcal{T}_U$ is summarized in Algorithm I.

Algorithm I NOMA-based IGS algorithm in Sec. III with $\mathcal{T}_U$.

Initialization

Set ϵ, $t = 1$, $\{\mathbf{P}\} = \{\mathbf{P}^{(0)}\}$, and $\{\boldsymbol{\Theta}\} = \{\boldsymbol{\Theta}^{(0)}\}$

While $\left(\min_{\forall l,k} \lambda_{lk}r_{lk}^{(t)} - \min_{\forall l,k} \lambda_{lk}r_{lk}^{(t-1)}\right)/\min_{\forall l,k} \lambda_{lk}r_{lk}^{(t-1)} \geq \epsilon$

   Optimizing over $\{\mathbf{P}\}$ by fixing $\{\boldsymbol{\Theta}^{(t-1)}\}$

      Obtain $\tilde{r}_{lk}^{(t-1)}$ based on (26)-(29)

      Compute $\{\mathbf{P}^{(t)}\}$ by solving (30)

   Optimizing over $\{\boldsymbol{\Theta}\}$ by fixing $\{\mathbf{P}^{(t-1)}\}$

      Obtain $\hat{r}_{lk}^{(t-1)}$ based on Corollary 1

      Compute $\{\boldsymbol{\Theta}^{(t)}\}$ by solving (35)

   $t = t + 1$

End (While)

Return $\{\mathbf{P}^\star\}$ and $\{\boldsymbol{\Theta}^\star\}$.

---

2) Feasibility set $\mathcal{T}_I$: Unfortunately, $\mathcal{T}_I$ is not a convex set, which makes the resulting surrogate problem non-convex. To address this issue, we rewrite the constraint $|\theta_{m_n}| = 1$ as the two following constraints

$$|\theta_{mn}|^2 \leq 1 \tag{36}$$
$$|\theta_{mn}|^2 \geq 1, \tag{37}$$

for all m, n. The constraint (36) is a convex constraint, but we have to approximate the constraint (37). To this end, we approximate (37) by the following affine lower bound

$$|\theta_{mn}|^2 \geq |\theta_{mn}^{(t-1)}|^2 + 2\mathfrak{R}\left(\theta_{mn}^{(t-1)}(\theta_{mn} - \theta_{mn}^{(t-1)})^*\right). \tag{38}$$

Thus, a surrogate function for the constraint (37) can be written as

$$|\theta_{mn}^{(t-1)}|^2 + 2\mathfrak{R}\left(\theta_{mn}^{(t-1)}(\theta_{mn} - \theta_{mn}^{(t-1)})^*\right) \geq 1. \tag{39}$$

To speed up convergence, we introduce a positive parameter, ϵ, and rewrite (39) as

$$|\theta_{mn}^{(t-1)}|^2 + 2\mathfrak{R}\left(\theta_{mn}^{(t-1)}(\theta_{mn} - \theta_{mn}^{(t-1)})^*\right) \geq 1 - \epsilon. \tag{40}$$

The inequality in (40) is a linear constraint, which makes the following problem convex:

$$\max_{r, \{\boldsymbol{\Theta}\}} r, \qquad \text{s.t.} \quad \lambda_{lk}\hat{r}_{lk} \geq r, \qquad \forall l, k, \tag{41a}$$
$$(40), (36). \tag{41b}$$

The optimization problem (41) can be efficiently solved. Let us represent the normalized solution of (41) by $\{\hat{\boldsymbol{\Theta}}\}$. We update the reflecting coefficients for the feasibility set $\mathcal{T}_I$ as

$$\{\boldsymbol{\Theta}^{(t)}\} = \begin{cases} \{\hat{\boldsymbol{\Theta}}\} & \text{if} \quad \min_{\forall l,k}\left\{r_{lk}\left(\{\mathbf{P}^{(t)}\}, \{\hat{\boldsymbol{\Theta}}\}\right)\right\} \geq \\ & \quad \min_{\forall l,k}\left\{r_{lk}\left(\{\mathbf{P}^{(t)}\}, \{\boldsymbol{\Theta}^{(t-1)}\}\right)\right\} \\ \{\boldsymbol{\Theta}^{(t-1)}\} & \text{Otherwise.} \end{cases} \tag{42}$$

This updating rule ensures that the algorithm generates a sequence of non-decreasing minimum weighted rates, which guarantees its convergence. Note that although our proposed algorithm converges, we do not make any claim on its optimality.

3) Feasibility set $\mathcal{T}_C$: The feasibility set $\mathcal{T}_C$ is not convex. To convexify this feasibility set, we relax the relationship between the phase and amplitude of reflecting components and approximate (5) as

$$|\theta|_{\min} \leq |\theta_{mi}| \leq 1 \ \forall m, i. \tag{43}$$

The constraint $|\theta_{mi}| \leq 1$ is convex since it can be written as $|\theta_{mi}|^2 \leq 1$. However, we also have the constraint $|\theta_{mi}| \geq |\theta|_{\min}$, or equivalently $|\theta_{mi}|^2 \geq |\theta|_{\min}^2$, which is similar to the constraint (37). Thus, we can apply a similar procedure and employ the lower bound in (38), which yields

$$|\theta_{mn}^{(t-1)}|^2 + 2\Re\left(\theta_{mn}^{(t-1)}(\theta_{mn} - \theta_{mn}^{(t-1)})^*\right) \geq |\theta|_{\min}^2. \tag{44}$$

Hence, the surrogate optimization problem is

$$\max_{r,\{\mathbf{\Theta}\}} \ r, \qquad \text{s.t.} \quad \lambda_{lk}\hat{r}_{lk} \geq r, \qquad \forall l,k, \tag{45a}$$

$$(36),(44). \tag{45b}$$

Let us call the solution of (45) $\{\mathbf{\Theta}^\star\}$. We take the phases of $\{\mathbf{\Theta}^\star\}$ to generate feasible reflecting coefficients by choosing

$$\{\hat{\mathbf{\Theta}}\} = f(\angle\{\mathbf{\Theta}^\star\}). \tag{46}$$

To make the solution of the algorithm non-decreasing, we update the reflecting coefficients as (42). Since the algorithm generates a sequence of non-decreasing minimum-weighted rates, the convergence is guaranteed. However, the proposed algorithm for $\mathcal{T}_C$ does not necessarily converge to a stationary point of (22).

C. Discussion on computational complexity

Since our proposed algorithm to solve the minimum-weighted-rate maximization (MWRM) problem is iterative, its computational complexity depends on the total number of iterations and the complexities involved with obtaining a solution in each iteration. Note that such computational complexities highly depend on the implementation of our algorithms. In this subsection, we provide a brief discussion on the computational complexity of the proposed algorithm and provide an approximation for the upper bound of the number of multiplications to obtain a solution of our algorithm.

Our proposed algorithm is iterative, and each iteration consists of two steps. In the first step, we obtain the transmit covariance matrices by solving (30), which is a convex optimization problem. According to [76, Chapter 11], the total number of Newton steps to solve a convex optimization problem grows with the square root of the number of inequalities in the problem. There are $2KL$ inequalities in (30). To compute each lower bound in (26), there are approximately $8 + 24LN_{BS}^2$ multiplications. Moreover, there are $2LK$ lower bounds for the rates since the total number of users is $2LK$. Thus, the complexities involved with solving (30) can be approximated as $\mathcal{O}\left(\sqrt{KL}\left(L^2KN_{BS}^2\right)\right)$. In the second step, we update the RIS components. We consider three different feasibility sets for the RIS components. Here, we present the computational complexities to run our proposed algorithm only for the feasibility $\mathcal{T}_I$ due to a space restriction. It is very straightforward to repeat the analysis for the other feasibility

sets. To update the RIS components for $\mathcal{T}_I$, we solve the convex optimization problem (41). The number of constraints in (41) is $2LK + 2MN_{RIS}$. Additionally, the approximate number of multiplications to obtain each lower bound $\hat{r}_{lk}$ can be approximated as $\mathcal{O}\left(N_{BS}^2L + MN_{RIS}N_{BS}\right)$. Thus, the computational complexity to solve (41) can be approximated as $\mathcal{O}\left(LK\sqrt{KL + MN_{RIS}}\left(N_{BS}^2L + MN_{RIS}N_{BS}\right)\right)$. We set the maximum number of iterations to N. Hence, the total computational complexity of our proposed algorithm can be approximated as

$$\begin{aligned}\mathcal{O}\left(N\left(\sqrt{KL}\left(L^2KN_{BS}^2\right)\right.\right.\\ \left.\left.+ LK\sqrt{KL + MN_{RIS}}\left(N_{BS}^2L + MN_{RIS}N_{BS}\right)\right)\right).\end{aligned}$$

IV. MAXIMIZING THE MINIMUM WEIGHTED EE

In this section, we solve the minimum-weighted-EE maximization (MWEEM) problem, i.e.,

$$\max_{\{\mathbf{\Theta}\}\in\mathcal{T},\{\mathbf{P}\}\in\mathcal{P},e} \ e \quad \text{s.t.} \quad \lambda_{lk}EE_{lk} \geq e, \quad \forall l,k \tag{47a}$$

$$r_{lk} \geq r_{lk}^{th}, \qquad \forall l,k, \tag{47b}$$

where the constraint (47b) is related to the rate constraint for all users, and r_{lk}^{th} is the corresponding minimum required rate for the user. Similar to the MWRM problem, we can characterize the EE region by employing the EE profile technique. That is, we solve (47) and vary the weights for all possible λ_{lk}, satisfying $\sum_{\forall l}\sum_{\forall k}\lambda_{lk}^{-1} = 1$ [23]. Moreover, similar to the MWRM problem, the non-negative weights λ_{lk} can be chosen based on the priorities of the users [52]. This problem is not convex, and is even more complicated than the MWRM problem since EE functions have fractional structures in $\{\mathbf{P}\}$, and (47) is actually a multiple-fractional problem. To solve (47), we employ an AO approach similar to the proposed scheme in Section III. That is, we first fix the reflecting coefficient matrices and solve (47) by optimizing over the covariance matrices. To this end, we employ MM alongside with the generalized Dinkelbach algorithm (GDA). Note that GDA is an iterative algorithm to solve fractional programming problems. After updating the covariance matrices, we fix them and solve (47) over the reflecting-coefficient matrices. The resulting problem has a structure similar to (31), and hence we can employ a similar approach to solve it. However, due to the rate constraint and the format of EE functions, we have to modify the constraints.

A. Optimization of the transmit covariance matrices

In this subsection, we solve (47) for a fixed $\{\mathbf{\Theta}^{(t-1)}\}$, i.e.,

$$\max_{\{\mathbf{P}\}\in\mathcal{P},e} \ e \quad \text{s.t.} \quad \lambda_{lk}EE_{lk} \geq e, \qquad \forall l,k \tag{48a}$$

$$r_{lk} \geq r_{lk}^{th}, \qquad \forall l,k. \tag{48b}$$

This problem is not convex, and we employ MM to solve it. That is, we first approximate the rates by the concave lower bounds in (26)-(29), which results in

$$\max_{\{\mathbf{P}\}\in\mathcal{P},e} \ e \quad \text{s.t.} \quad \frac{\lambda_{lk}\tilde{r}_{lk}}{P_c + \eta\text{Tr}\left(\mathbf{P}_{lk}\right)} \geq e, \qquad \forall l,k \tag{49a}$$

This article has been accepted for publication in IEEE Transactions on Signal Processing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSP.2023.3259145
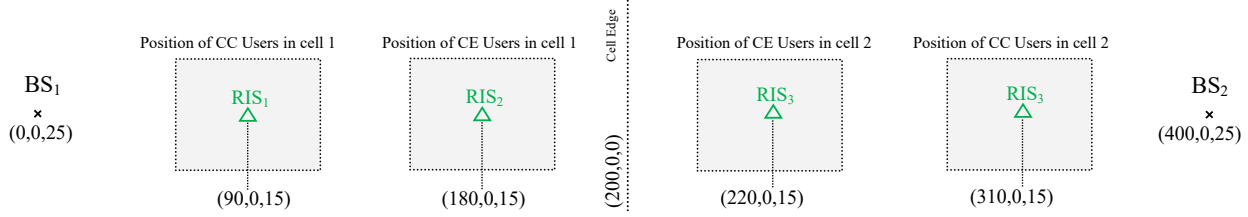
10

Fig. 2: System topology.

$$\tilde{r}_{lk} \geq r_{lk}^{th}, \qquad \forall l, k. \qquad (49b)$$

Although (49) is not convex, we can obtain its global optimal solution by the GDA. The GDA is an extension of the Dinkelbach algorithm to solve multiple-ratio fractional programming problems [72]. The optimum solution of (49) is obtained by iteratively solving

$$\max_{\{\mathbf{P}\} \in \mathcal{P}, e} e \qquad (50a)$$

$$\text{s.t.} \quad \lambda_{lk} \tilde{r}_{lk} - \mu^{(t,n)} \left(P_c + \eta \text{Tr} \left(\mathbf{P}_{lk} \right) \right) \geq e, \quad \forall l, k, \quad (50b)$$

$$\tilde{r}_{lk} \geq r_{lk}^{th}, \qquad \forall l, k, \quad (50c)$$

and updating $\mu^{(t,n)}$ as

$$\mu^{(t,n)} = \min \left\{ \min_{\forall l,k} \left\{ \frac{\lambda_{lk} \tilde{r}_{lk} \left(\{\mathbf{P}^{(t,n-1)}\}, \{\mathbf{\Theta}^{(t-1)}\} \right)}{P_c + \eta \text{Tr} \left(\mathbf{P}_{lk}^{(t,n-1)} \right)} \right\}, \right.$$
$$\left. \min_{\forall l,\bar{k}} \left\{ \frac{\lambda_{lk} r_{l\bar{k}}^l \left(\{\mathbf{P}^{(t,n-1)}\}, \{\mathbf{\Theta}^{(t-1)}\} \right)}{P_c + \eta \text{Tr} \left(\mathbf{P}_{l\bar{k}}^{(t,n-1)} \right)} \right\} \right\}. \quad (51)$$

The GDA linearly converges to the optimum solution of (49).

B. Optimization of the reflecting coefficients

In this subsection, we consider the optimization of the weighted-minimum-EE over reflecting coefficients for a fixed set of covariance matrices. The MWEEM problem for a fixed $\{\mathbf{P}^{(t)}\}$ can be simplified to

$$\max_{\{\mathbf{\Theta}\} \in \mathcal{T}, e} e \quad \text{s.t.} \quad r_{lk} \geq \min \left(e \tilde{\lambda}_{lk}, r_{lk}^{th} \right), \quad \forall l, k, \quad (52)$$

where $\tilde{\lambda}_{lk} = \left(P_c + \eta \text{Tr} \left(\mathbf{P}_{lk}^{(t)} \right) \right) / \lambda_{lk}$. The structure of this problem is similar to the MWRM problem in Section III-B. Hence, we can easily employ the algorithms in Section III-B to solve (52). To this end, we first substitute the rates by the lower bounds in Theorem 1, which yields

$$\max_{\{\mathbf{\Theta}\} \in \mathcal{T}, e} e \quad \text{s.t.} \quad \hat{r}_{lk} \geq \min \left(\tilde{\lambda}_{lk} e, r_{lk}^{th} \right), \quad \forall l, k. \quad (53)$$

This optimization problem is very similar for the different feasibility sets and can be solved similarly, so we skip the details due to limited space. We summarize our proposed algorithm for the MWEEM in Algorithm II.

Algorithm II NOMA-based IGS algorithm in Sec. IV with $\mathcal{T}$.

Initialization

Set ϵ, $t = 1$, $\{\mathbf{P}\} = \{\mathbf{P}^{(0)}\}$, and $\{\mathbf{\Theta}\} = \{\mathbf{\Theta}^{(0)}\}$

While $\left(\min_{\forall l,k} \lambda_{lk} EE_{lk}^{(t)} - \min_{\forall l,k} \lambda_{lk} EE_{lk}^{(t-1)} \right) / \min_{\forall l,k} \lambda_{lk} EE_{lk}^{(t-1)} \geq \epsilon$

    Optimizing over $\{\mathbf{P}\}$ **by fixing** $\{\mathbf{\Theta}^{(t-1)}\}$

      Obtain $\tilde{r}_{lk}^{(t-1)}$ based on (26)-(29)

      Compute $\{\mathbf{P}^{(t)}\}$ by solving (49)

    Optimizing over $\{\mathbf{\Theta}\}$ **by fixing** $\{\mathbf{P}^{(t-1)}\}$

      Obtain $\hat{r}_{lk}^{(t-1)}$ based on Corollary 1

      Compute $\{\mathbf{\Theta}^{(t)}\}$ by solving (53)

  $t = t + 1$

End (While)

Return $\{\mathbf{P}^\star\}$ and $\{\mathbf{\Theta}^\star\}$.

C. Discussion on computational complexity

The computation complexity of our proposed algorithm for WMEEM can be computed similar to the approach in Section III-C. The computational complexities involved with updating the RIS components is similar to Section III-C since the optimization problems (35) and (53) have a similar structure, and thus, their computational complexities are in the same order. However, updating the transmit covariance matrices has higher computational complexities comparing to the WMRM problem since (49) has a fractional structure, and we employ the GDA to solve it. Indeed, we have to solve a sequence of convex optimization problems ((50)) to update the transmit covariance matrices. We set the maximum number of the GDA iterations N_I. The maximum number of Newton iterations to solve (50) is related to $\sqrt{4LK}$ since the number of constraints in (50) is $4LK$. The main complexity in solving each Newton iteration is to compute the lower bounds for the user rates, which approximately has $8 + 24LN_{BS}^2$ multiplications. Thus, the overall computational complexity of solving the WMEEM problem can be approximated as

$$\mathcal{O} \left(N \left(N_I \sqrt{KL} \left(L^2 K N_{BS}^2 \right) \right. \right.$$
$$\left. \left. + LK \sqrt{KL + MN_{RIS}} \left(N_{BS}^2 L + MN_{RIS} N_{BS} \right) \right) \right).$$

V. NUMERICAL RESULTS

We consider a two-cell broadcast channel with $2K$ users as well as two RIS locations in each cell as shown in Fig. 2, unless it is explicitly mentioned otherwise. The dimensions in Fig. 2 are in meters. We assume that the users are uniformly located into two different clusters/positions in each cell: a cell centric and a cell edge region [24]. In each cluster, the users are located in a $20m \times 20m$ square with a RIS at its center. Note that the position of each user is modeled by a random

variable with a uniform distribution and varies in each iteration of the Monte Carlo simulations. The heights of BSs, RISs and users are, respectively, 25m, 15m and 1.5m. It is worth emphasizing that we consider a specific system topology to accurately model the large-scale fading, which may have a great impact on the performance of RIS [37], [39], [40], [46]. We further assume that the power budget of each BS is equal to P watts. Additionally, we assume that the weights of all users are equal to 1, i.e., $\lambda_{l,k} = 1$ for all l, k. The parameters for the feasibility set $\mathcal{T}_C$ are $|\theta|_{\min} = 0.2$, $\alpha = 1.6$, and $\phi = 0.43\pi$ [57]. We employ a random feasible initial point for running the proposed algorithms. We assume that the direct link between a BS and a user is non-line-of-sight (NLoS), which implies that the small-scale fading for a direct link is Rayleigh. It means that the channel coefficients for direct links are drawn from a zero-mean complex proper Gaussian distribution with unit variance times an attenuation coefficient, which models the large-scale fading. Furthermore, we assume that the links related to RISs are line-of-sight (LoS), which implies that their small-scale fading is Rician. Thus, the direct link between BS i and RIS m, denoted by $\mathbf{G}_{mi}$, is modeled as [40]

$$\mathbf{G}_{mi} = \sqrt{\beta_{mi}} \left(\sqrt{\frac{\gamma}{1+\gamma}} \mathbf{G}_{mi}^{\mathrm{LoS}} + \sqrt{\frac{1}{1+\gamma}} \mathbf{G}_{mi}^{\mathrm{NLoS}} \right),$$

where β_{mi} is the channel-attenuation coefficient due to the large-scale fading, γ is the Rician factor, $\mathbf{G}_{mi}^{\mathrm{LoS}}$ is the LoS component, and $\mathbf{G}_{mi}^{\mathrm{NLoS}}$ is the NLoS Rayleigh component. We consider $\gamma = 3$ for all links. Although the NLoS component $\mathbf{G}_{mi}^{\mathrm{NLoS}}$ is a random variable, the LoS component $\mathbf{G}_{mi}^{\mathrm{LoS}}$ is a deterministic variable modeled as [40]

$$\mathbf{G}_{mi}^{\mathrm{LoS}} = \mathbf{a}_{r_{mi}} \left(\phi_{mi}^A \right) \mathbf{a}_{t_{mi}}^H \left(\phi_{mi}^D \right),$$

where $\phi_{mi}^A/\phi_{mi}^D \sim \mathrm{Unif}[0, 2\pi]$ is the angle of arrival/departure for the link between BS i and RIS m, and [40]

$$\mathbf{a}_{r_{mi}} = \left[1, e^{j\frac{2\pi d \sin(\phi_{mi}^A)}{\lambda}}, \cdots, e^{j\frac{2(N_{RIS}-1)\pi d \sin(\phi_{mi}^A)}{\lambda}} \right],$$
$$\mathbf{a}_{t_{mi}} = \left[1, e^{j\frac{2\pi d \sin(\phi_{mi}^D)}{\lambda}}, \cdots, e^{j\frac{2(N_{BS}-1)\pi d \sin(\phi_{mi}^D)}{\lambda}} \right].$$

Note that the difference between LoS and NLoS links is not only in the small-scale fading, but also they have different large-scale fading. The path loss in dB for each link can be modeled as [46], [52]

$$\mathrm{PL} = \mathrm{PL}_0 + G_0 - 10\alpha \log_{10}\left(\frac{d}{d_0}\right),$$

where PL_0 is the path loss at the reference distance d_0, G_0 is the transmitter antenna gain, α is the path-loss exponent, and d is the distance between the transmitter and receiver. The effective channel attenuation coefficient is $\beta = 10^{\mathrm{PL}/10}$ [52]. Note that the path-loss exponent α depends to a large extent on whether a link is LoS or NLoS, where NLoS links experience a larger attenuation. The simulation parameters related to the large-scale and small-scale fading are the same as in [52]. Due to the space restriction, we refer the readers to [52, Table II] for additional details on the rest of the parameters. Employing Monte Carlo simulations, we average the results over 100 channel realizations.

In this paper, we compare the proposed scheme with the algorithms in [52], which considers TIN. We also compare the proposed algorithms with the traditional systems, i.e., systems with no RIS. To summarize, the considered algorithms in the simulations are as follows:

- $\mathbf{IR}_X\mathbf{N}$ (or $\mathbf{PR}_X\mathbf{N}$) refers to the proposed NOMA-based IGS (or PGS) scheme for $\mathcal{T}_X$, where X can be equal to U, I and C.
- $\mathbf{IR}_R\mathbf{N}$ (or $\mathbf{PR}_R\mathbf{N}$) refers to the NOMA-based IGS (or PGS) scheme with random reflecting coefficients.
- $\mathbf{IR}_U\mathbf{T}$ (or $\mathbf{PR}_U\mathbf{T}$) refers to the proposed IGS (or PGS) scheme in [52] for the feasibility set $\mathcal{T}_U$ with TIN. Note that this scheme can be seen as an upper bound for the IGS (or PGS) performance with TIN.
- $\mathbf{IN}$ (or $\mathbf{PN}$) refers to the proposed NOMA-based IGS (or PGS) scheme, which is applied to traditional multicell BCs (i.e., without RIS).

A. Fairness rate

In this subsection, we consider the MWRM problem. We refer to the maximum of the minimum rate as a fairness rate since it is usually the case that all users receive an equal achievable rate when the minimum rate is maximized [30], [52]. To fully evaluate the system performance, we show the effect of different parameters on the performance of RIS-assisted systems. These parameters are power budget, transmit antennas at the BSs, number of users and number of RIS elements. Finally, we briefly illustrate the convergence of our proposed algorithms.

1) Impact of power budget: In Fig. 3, we show the effect of the BS power budget on the minimum user rate. As can be observed, the RIS-assisted systems (both IGS and PGS) with NOMA outperform the RIS-assisted systems with TIN as well as the systems without RIS. However, the IGS (and PGS) schemes with random RIS phases perform even worse than the IGS (and PGS) schemes with no-RIS especially at high SNR regimes, which indicates the importance of optimizing RIS coefficients. This is rather a surprising result, which shows that employing RIS without a proper optimization may even make the system performance worse. In Fig. 3c, we compare the IGS and PGS schemes. As can be observed, the IGS scheme can significantly outperform the PGS schemes, and the benefits of IGS increase with P. The reason is that the interference level increases with P, so more improvements are expected by IGS as an interference-management technique. Another interesting behavior in Fig. 3 is that IGS increases the benefits of NOMA. Indeed, it suggests that these interference-management techniques can be seen as a complementary of each other since NOMA handles the interference from the paired user (intracell), while IGS mainly handles ICI as well as the remaining intracell interference. Furthermore, we can observe that RIS increases the benefits of IGS.

In Fig. 3, we also consider the impact of the feasibility set for the RIS components. As can be observed, RIS can improve the system performance with the three considered feasibility sets. Moreover, we observe that the amplitude of RIS components can have a considerable impact on the
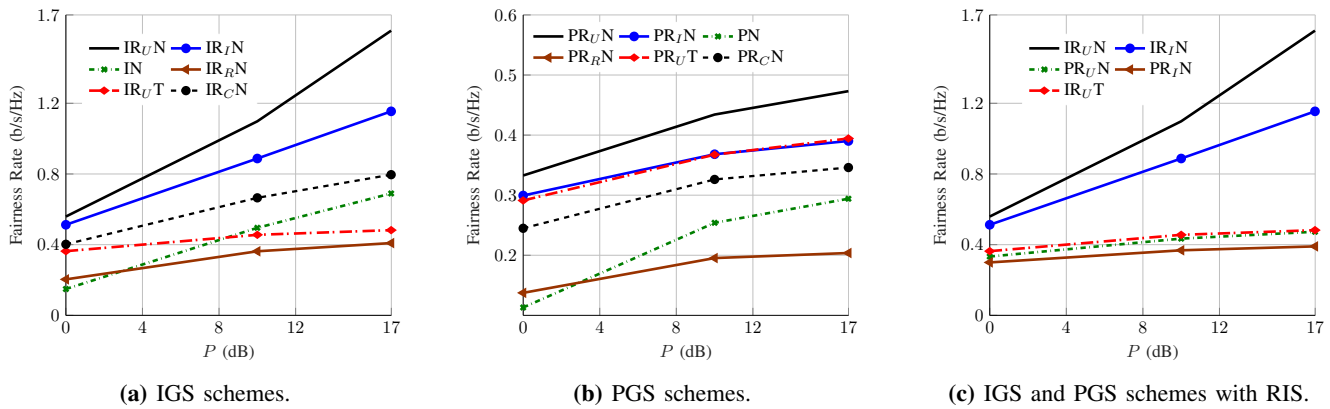
(a) IGS schemes.

(b) PGS schemes.

(c) IGS and PGS schemes with RIS.

Fig. 3: The average fairness rate versus P for $N_{BS} = N_u = 1$, $N_{RIS} = 25$, $L = 2$, $K = 2$, $M = 4$.

(a) IGS schemes.

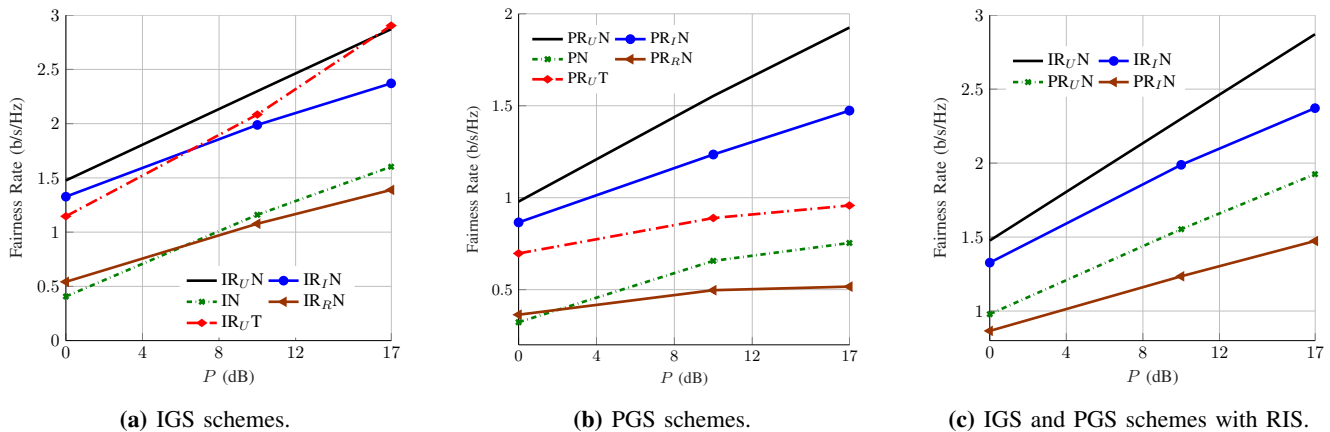**(b)** PGS schemes.

(c) IGS and PGS schemes with RIS.

Fig. 4: The average fairness rate versus P for $N_{BS} = 2$, $N_{RIS} = 25$, $L = 2$, $K = 2$, $M = 4$.

performance. As expected, RIS with the feasibility set $\mathcal{T}_U$ outperforms the other feasibility sets, which are a subset of $\mathcal{T}_U$.

In Fig. 4, we show the average fairness rate versus P for $N_{BS} = 2$, $N_{RIS} = 25$, $L = 2$, $K = 2$, $M = 4$. As can be observed, the overall performance is similar to the single-antenna systems. For instance, as shown in Fig. 4c, the IGS scheme can still significantly outperform the PGS scheme. The main difference between Fig. 4 and Fig. 3 is in the performance gap between the proposed NOMA scheme and the TIN scheme proposed in [52], which decreases by increasing the number of transmit antennas. The reason is that by increasing the number of transmit antennas, we actually increase the number of spatial resources, which consequently, reduces the interference level. It is expected that the benefits of interference-management techniques decreases whenever there is a low interference level, as will be corroborated in the next example.

In Fig. 4, we also observe an interesting behavior in the performance of NOMA. That is, IGS with TIN performs very close to IGS with NOMA at high SNR. This happens since the channels in IGS are not degraded, which means that NOMA can be suboptimal especially at high SNRs. This in turn implies that we should carefully choose the signaling scheme based on the operational regime. We elaborate more on this issue in this section.
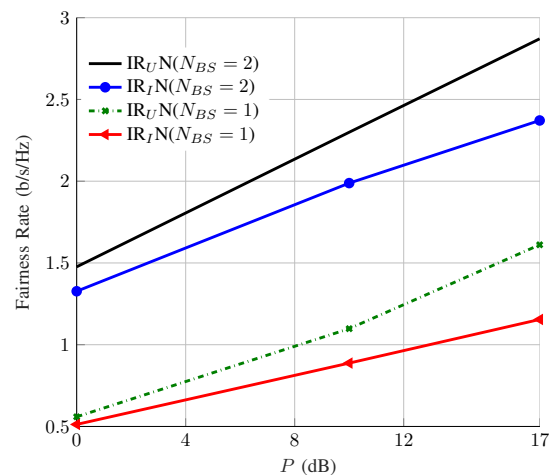

Fig. 5: The average fairness rate versus P for $N_u = 1$, $N_{RIS} = 25$, $L = 2$, $K = 2$, $M = 4$.

2) Impact of transmit antennas: Fig. 5 shows the average fairness rate for IGS schemes versus P for $N_{RIS} = 25$, $L = 2$, $K = 2$, $M = 4$ and different number of transmit antennas at the BSs. As can be observed, the average fairness rate significantly increases with the number of transmit antennas. Moreover, we observe a similar behavior for the IGS schemes with different feasibility sets. Hence, it can be expected that
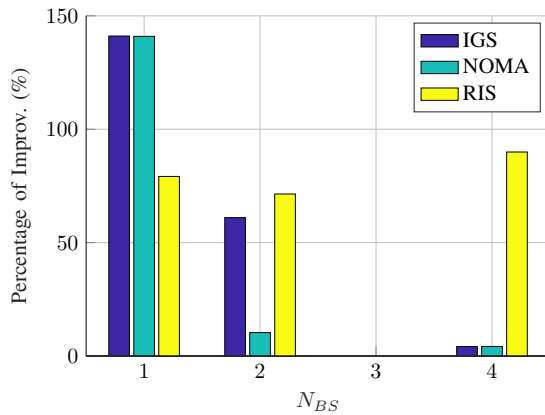
Fig. 6: The performance improvement by IGS/NOMA/RIS versus N_{BS} for $P = 10$W, $N_{RIS} = 25$, $L = 2$, $K = 2$, $M = 4$.

(a) Average fairness rate.   (b) Benefits by IGS.

Fig. 7: The average fairness rate versus K for $N_{BS} = 2$, $P = 10$, $N_{RIS} = 25$, $L = 2$, $M = 4$.

Fig. 8: The average fairness rate versus N_{RIS} for $N_{BS} = 2$, $P = 10$, $M = 2$.

the performance gap between the upper bound performance of RIS remains relatively constant if we increase the number of transmit antennas. Furthermore, this gap increases with the power budget (or equivalently the transmission power) of the BS.

In Fig. 6, we show the relative performance improvements in the average fairness rate by each technique versus N_{BS} for $P = 10$W, $N_{RIS} = 25$, $L = 2$, $K = 2$, $M = 4$. We obtain the relative improvements of IGS by comparing the average fairness rate of IR_IN and PR_IN schemes. Furthermore, the benefits of NOMA are obtained by comparing the performance of IR_IN and IR_IT schemes. Finally, the RIS benefits are evaluated by comparing the performance of IR_IN and IN schemes. As can be observed, the benefits of IGS as well as NOMA decrease with the number of transmit antennas. The reason is that the number of (spatial) resources per users increases with the number of antennas, which in turn decreases the interference level. Thus, the benefits of IGS as an interference-management technique are expected to be reduced. When the number of transmit antennas is lower than the number of users per cell, $N_{BS} < 2K$, BSs are overloaded, and the interference level is very high. As a result, we should employ interference-management techniques such as IGS and/or NOMA. However, when $N_{BS} > 2K$, BSs are underloaded, and the interference-level is low. Thus, we can simply treat the interference as noise and employ PGS without a considerable performance degradation. We also observe in Fig. 6 that the benefits of RIS remain relatively constant when we switch to underloaded systems. This interesting result can show that the main role of RIS in this set up is to improve the coverage and channel condition of users. Thus, RIS can significantly improve the system performance in both underloaded and overloaded regimes. However, in overloaded regimes, we have to employ interference-management techniques to reap the benefits of RIS and achieve a better performance.

3) Impact of number of users per cell: Fig. 7 shows the average fairness rate versus K for $N_{BS} = 2$, $N_{RIS} = 25$, $L = 2$, $M = 4$. As can be observed, the benefits of employing IGS increase with the number of users. Indeed, there are only minor benefits by IGS when the number of users per cell is low.
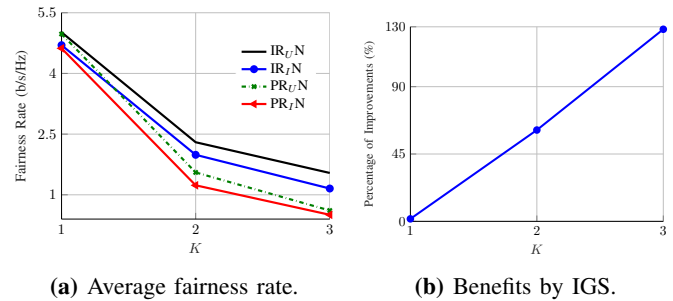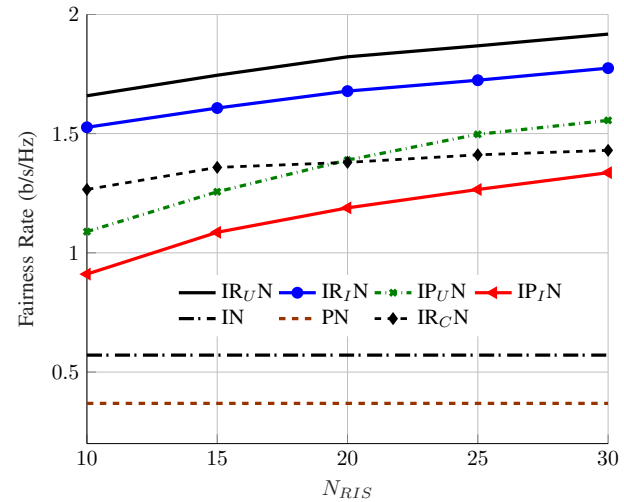
Particularly, when $K = 1$, there is no intracell interference for the CCUs, and hence, the interference level is significantly reduced, which in turn, considerably decreases the benefits of IGS. On the contrary, the interference level increases with the number of users per cell for a fixed number of antennas. Thus, the benefits of IGS significantly increase with the number of users per cell as illustrated in Fig. 7b. Through Figs. 6 and 7, we observe that we can get a very high gain by interference-management techniques whenever the BSs are overloaded, i.e., when the number of users per cell is larger than the number of transmit antennas $(2K > N_{BS})$. However, we can employ a simple PGS scheme with TIN when BSs are underloaded $(2K \leq N_{BS})$ without a considerable performance loss.

4) Impact of number of RIS components: Figure 8 shows the average fairness rate versus N_{RIS} for $N_{BS} = 2$, $P = 10$, $M = 2$. In this figure, there is only one RIS per cell, which is located close to the CEUs. As expected, the average minimum rate increases with N_{RIS}; however, the improvements in the rate are not significant, and even with a relatively low number of RIS components $(N_{RIS} = 10)$, RIS can provide a considerable gain. This means that we do not necessarily need a high number of RIS components to get benefits, which can be interesting from practical point of view.

5) Convergence: Figure 9 shows the average fairness rate versus the number of iterations. As expected, the algorithms
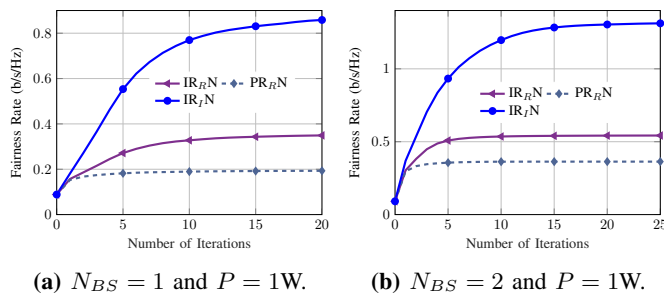
(a) $N_{BS} = 1$ and $P = 1$W. **(b)** $N_{BS} = 2$ and $P = 1$W.

Fig. 9: The average fairness rate versus number of iterations for $N_{RIS} = 25$, $L = 2$, $K = 2$, and $M = 4$.

for RIS-assisted systems require more iterations for convergence and are slower than the algorithms that do not optimize over RIS components. However, the RIS-assisted algorithms outperform the other algorithms after only a few iterations. For instance, in Fig. 9b, the IGS algorithm provides a better solution than the final value of the I-R algorithm after just 3 iterations.

B. EE maximization

In this subsection, we provide the results for the maximization of the minimum EE of users. We expect that we observe the same behavior in the performance of the proposed algorithms by varying different parameters such as K, N_{BS}, N_{RIS}. Hence, due to the space restriction, we provide only a figure for the EE maximization.

In Fig. 10, we show the minimum EE of users versus P_c for $P = 1$w, $N_{BS} = N_u = 1$, $N_{RIS} = 25$, $L = 2$, $K = 2$, $M = 2$. As can be observed, IGS can significantly outperform PGS. Moreover, the difference between the IGS scheme and the upper bound (I-M scheme) is not significant, which is because of a lower transmission power.

In Fig. 11, we show the average fairness EE achieved by the NOMA-based IGS schemes proposed in Section III and in Section IV for $P_c = 10$dB, $N_{BS} = 1$, $N_{RIS} = 15$, $L = 2$, $K = 2$, $M = 2$. Indeed, we compare the performance of our proposed algorithms, which consider two different utility functions, from a fairness EE point of view. In other words, this figure shows how much EE is lost if the minimum rate of users is maximized. The EE loss increases with the power budget. The reason is that if we consider a point-to-point system, the rate strictly increases with the transmission power. However, the EE is maximized by a specific transmission power and is strictly decreasing for transmission powers higher than the optimum transmission power (see [73, Fig. 2]). When the optimum transmission power for the MWEEM problem is lower than the power budget, the solution of the MWEEM problem does not change by increasing the power budget. That is why the fairness EE is almost constant for $P \geq 10$ dB. On the contrary, the transmission power of the solution of the MWRM problem is expected to be increased by the power budget, which results in a lower energy efficient operational point. Additionally, note that the MWEEM problem is equivalent to MWRM for a very large P_c [73]. Thus, the algorithms for the MWRM problem perform close to the algorithms for

the MWEEM problem when power budget is much lower than P_c.

C. Summary of numerical results

Our main findings in the numerical results can be summarized as in the following:

- RIS always improves the performance of the system; however, we have to properly optimize their elements to reap the full benefits of reconfigurable intelligent surfaces. In some cases, a RIS with random phases may even worsen performance, which shows the importance of optimizing the RIS components.
- The combination of NOMA with appropriate user pairing plus IGS can highly improve the system performance from both spectral and energy efficiency perspectives when the BSs are overloaded, i.e., when $N_{BS} < 2K$. In this case, the interference level is high, which significantly degrades the system performance if interference is not properly managed.
- Our results show that, even with a relatively low number of RIS components, RIS can provide a considerable performance gain. This suggests that RIS can be a promising technology from a practical point of view.

VI. CONCLUSION

In this paper, we proposed NOMA-based IGS schemes to improve the spectral and energy efficiency of multicell RIS-assisted BCs. We showed that RIS can significantly improve the system performance in both overloaded ($N_{BS} < 2K$) and underloaded ($N_{BS} \geq 2K$) regimes. In overloaded regimes, RIS is unable to completely mitigate undesired consequences of interference, and we have to employ interference-management techniques to fully exploit the benefits of RIS. In other words, RIS cannot be used alone as an interference-management technique when $N_{BS} < 2K$. However, in underloaded regimes, the interference level is low, and there are only minor benefits (if there exist any) by employing IGS/NOMA. We also showed that, even with a relatively low number of RIS components, RIS can considerably improve the system performance, which indicates that RIS can be efficiently implemented in practice. Moreover, we showed that RIS components should be carefully optimized. Otherwise, RIS may only provide minor benefits or even may decrease the system performance, especially in SISO systems.

As a future work, it is interesting to investigate the performance of IGS and/or NOMA in the presence of imperfect CSI, employing a robust design similar to [47]. Moreover, analyzing the performance gap between our proposed algorithms and the global optimal solution can be also another challenging direction for a future study.

REFERENCES

[1] C. Huang, *et al.*, "Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends," *IEEE Wireless Commun.*, vol. 27, no. 5, pp. 118–125, 2020.
[2] Q. Wu *et al.*, "Intelligent reflecting surface aided wireless communications: A tutorial," *IEEE Trans. Commun.*, 2021.
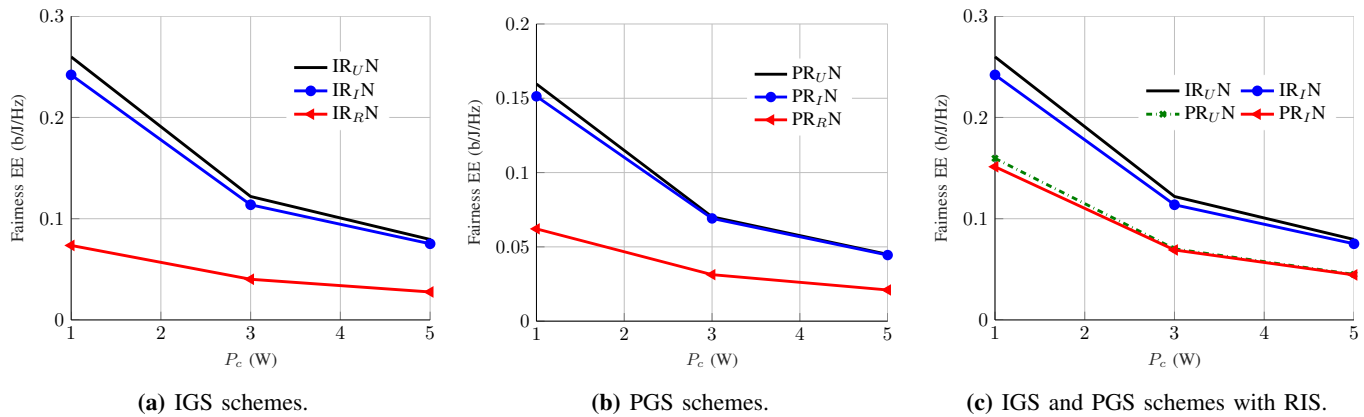
This article has been accepted for publication in IEEE Transactions on Signal Processing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSP.2023.3259145

15

(a) IGS schemes.

(b) PGS schemes.

(c) IGS and PGS schemes with RIS.

Fig. 10: The average fairness EE versus P_c for $P = 1$w, $N_{BS} = N_u = 1$, $N_{RIS} = 25$, $L = 2$, $K = 2$, $M = 2$.
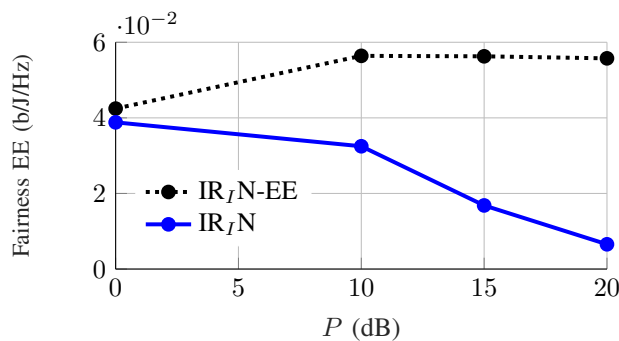

Fig. 11: The average fairness rate versus P_c for SNR= 0dB, $N_{BS} = 1$, $N_{RIS} = 15$, $L = 2$, $K = 2$, $M = 2$.

[3] M. Di Renzo *et al.*, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2450–2525, 2020.

[4] M. A. ElMossallamy *et al.*, "Reconfigurable intelligent surfaces for wireless communications: Principles, challenges, and opportunities," *IEEE Trans. Cogn. Commun. Netw.*, vol. 6, no. 3, pp. 990–1002, 2020.

[5] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, 2014.

[6] V. S. Annapureddy and V. V. Veeravalli, "Gaussian interference networks: Sum capacity in the low-interference regime and new outer bounds on the capacity region," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3032–3050, 2009.

[7] H. Sato, "The capacity of the Gaussian interference channel under strong interference (corresp.)," *IEEE Trans. Inf. Theory*, vol. 27, no. 6, pp. 786–788, 1981.

[8] V. R. Cadambe, S. A. Jafar, and C. Wang, "Interference alignment with asymmetric complex signaling–Settling the Høst-Madsen-Nosratinia conjecture," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4552–4565, 2010.

[9] S. Javed, O. Amin, B. Shihada, and M.-S. Alouini, "Improper Gaussian signaling for hardware impaired multihop full-duplex relaying systems," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1858–1871, 2019.

[10] M. Soleymani, I. Santamaria, and P. J. Schreier, "Improper Gaussian signaling for the K-user MIMO interference channels with hardware impairments," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 11 632–11 645, 2020.

[11] M. Gaafar, O. Amin, W. Abediseid, and M.-S. Alouini, "Underlay spectrum sharing techniques with in-band full-duplex systems using improper Gaussian signaling," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 235–249, 2017.

[12] O. Amin, W. Abediseid, and M.-S. Alouini, "Overlay spectrum sharing using improper Gaussian signaling," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 1, pp. 50–62, 2017.

[13] M. Soleymani, C. Lameiro, P. J. Schreier, and I. Santamaria, "Energy-efficient design for underlay cognitive radio using improper signaling," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Processing (ICASSP)*, 2019, pp. 4769–4773.

[14] C. Lameiro, I. Santamaría, and P. J. Schreier, "Benefits of improper signaling for underlay cognitive radio," *IEEE Wireless Commun. Lett.*, vol. 4, no. 1, pp. 22–25, 2015.

[15] C. Lameiro, I. Santamaría, and P. J. Schreier, "Rate region boundary of the SISO Z-interference channel with improper signaling," *IEEE Trans. Commun.*, vol. 65, no. 3, pp. 1022–1034, 2017.

[16] S. Lagen, A. Agustin, and J. Vidal, "On the superiority of improper Gaussian signaling in wireless interference MIMO scenarios," *IEEE Trans. Commun.*, vol. 64, no. 8, pp. 3350–3368, 2016.

[17] S. Lagen, A. Agustin, and J. Vidal, "Coexisting linear and widely linear transceivers in the MIMO interference channel," *IEEE Trans. Signal Process.*, vol. 64, no. 3, pp. 652–664, 2016.

[18] Y. Zeng, C. M. Yetis, E. Gunawan, Y. L. Guan, and R. Zhang, "Transmit optimization with improper Gaussian signaling for interference channels," *IEEE Trans. Signal Process.*, vol. 61, no. 11, pp. 2899–2913, 2013.

[19] Z. K. Ho and E. Jorswieck, "Improper Gaussian signaling on the two-user SISO interference channel," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3194–3203, 2012.

[20] M. Soleymani, C. Lameiro, I. Santamaria, and P. J. Schreier, "Improper signaling for SISO two-user interference channels with additive asymmetric hardware distortion," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8624–8638, 2019.

[21] M. Soleymani, C. Lameiro, I. Santamaria, and P. J. Schreier, "Robust improper signaling for two-user SISO interference channels," *IEEE Trans. Commun.*, vol. 67, no. 7, pp. 4709–4723, 2019.

[22] M. Soleymani, I. Santamaria, C. Lameiro, and P. J. Schreier, "Ergodic rate for fading interference channels with proper and improper Gaussian signaling," *Entropy*, vol. 21, no. 10, p. 922, 2019.

[23] M. Soleymani, C. Lameiro, I. Santamaria, and P. J. Schreier, "Energy-Efficient improper signaling for K-User interference channels," in *Proc. IEEE Eu. Signal Process. Conf. (EUSIPCO)*, 2019, pp. 1–5.

[24] H. D. Tuan, A. A. Nasir, H. H. Nguyen, T. Q. Duong, and H. V. Poor, "Non-orthogonal multiple access with improper Gaussian signaling," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 3, pp. 496–507, 2019.

[25] A. A. Nasir, H. D. Tuan, H. H. Nguyen, T. Q. Duong, and H. V. Poor, "Signal superposition in NOMA with proper and improper Gaussian signaling," *IEEE Trans. Commun.*, vol. 68, no. 10, pp. 6537–6551, 2020.

[26] A. A. Nasir, H. D. Tuan, T. Q. Duong, and H. V. Poor, "Improper Gaussian signaling for broadcast interference networks," *IEEE Signal Process. Lett.*, vol. 26, no. 6, pp. 808–812, 2019.

[27] H. Nguyen, H. Tuan, D. Niyato, D. Kim, and H. Poor, "Improper Gaussian signaling for D2D communication coexisting MISO cellular networks," *IEEE Trans. Wireless Commun.*, 2021.

[28] H. Yu, T. D. Hoang, A. A. Nasir, T. Q. Duong, and L. Hanzo, "Improper Gaussian signaling for computationally tractable energy and information beamforming," *IEEE Trans. Veh. Technol.*, 2020.

[29] H. Yu, H. D. Tuan, T. Q. Duong, Y. Fang, and L. Hanzo, "Improper Gaussian signaling for integrated data and energy networking," *IEEE Trans. Commun.*, vol. 68, no. 6, pp. 3922–3934, 2020.

This article has been accepted for publication in IEEE Transactions on Signal Processing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSP.2023.3259145

16

[30] H. Park, S.-H. Park, J.-S. Kim, and I. Lee, "SINR balancing techniques in coordinated multi-cell downlink systems," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 626–635, 2013.

[31] H. Yu, H. D. Tuan, E. Dutkiewicz, H. V. Poor, and L. Hanzo, "Maximizing the geometric mean of user-rates to improve rate-fairness: Proper vs. improper Gaussian signaling," *IEEE Trans. Wireless Commun.*, 2021.

[32] S. Javed, O. Amin, B. Shihada, and M.-S. Alouini, "A journey from improper Gaussian signaling to asymmetric signaling," *IEEE Commun. Surv. Tutor.*, vol. 22, no. 3, pp. 1539–1591, 2020.

[33] C. Geng, N. Naderializadeh, A. S. Avestimehr, and S. A. Jafar, "On the optimality of treating interference as noise," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1753–1767, 2015.

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

[35] B. Clerckx, Y. Mao, R. Schober, E. Jorswieck, D. J. Love, J. Yuan, L. Hanzo, G. Y. Li, E. G. Larsson, and G. Caire, "Is NOMA efficient in multi-antenna networks? a critical look at next generation multiple access techniques," *IEEE open j. Commun. Soc.*, 2021.

[36] W. Ni, X. Liu, Y. Liu, H. Tian, and Y. Chen, "Resource allocation for multi-cell IRS-aided NOMA networks," *IEEE Trans. Wireless Commun.*, 2021.

[37] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, 2019.

[38] 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, 2019.

[39] Q.-U.-A. Nadeem, *et al.*, "Asymptotic max-min SINR analysis of reconfigurable intelligent surface assisted MISO systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 7748–7764, 2020.

[40] C. Pan, H. Ren, K. Wang, W. Xu, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Multicell MIMO communications relying on intelligent reflecting surfaces," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5218–5233, 2020.

[41] L. Zhang, Y. Wang, W. Tao, Z. Jia, T. Song, and C. Pan, "Intelligent reflecting surface aided MIMO cognitive radio systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 11 445–11 457, 2020.

[42] Y. Yang, B. Zheng, S. Zhang, and R. Zhang, "Intelligent reflecting surface meets OFDM: Protocol design and rate maximization," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4522–4535, 2020.

[43] J. Zuo, Y. Liu, Z. Qin, and N. Al-Dhahir, "Resource allocation in intelligent reflecting surface assisted NOMA systems," *IEEE Trans. Commun.*, vol. 68, no. 11, pp. 7170–7183, 2020.

[44] X. Mu, Y. Liu, L. Guo, J. Lin, and N. Al-Dhahir, "Exploiting intelligent reflecting surfaces in NOMA networks: Joint beamforming optimization," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6884–6898, 2020.

[45] G. Yang, X. Xu, Y.-C. Liang, and M. Di Renzo, "Reconfigurable intelligent surface-assisted non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 20, no. 5, pp. 3137–3151, 2021.

[46] H. Yu *et al.*, "Joint design of reconfigurable intelligent surfaces and transmit beamforming under proper and improper Gaussian signaling," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2589–2603, 2020.

[47] G. Zhou, C. Pan, H. Ren, K. Wang, and A. Nallanathan, "A framework of robust transmission design for IRS-aided MISO communications with imperfect cascaded channels," *IEEE Trans. Signal Process.*, vol. 68, pp. 5092–5106, 2020.

[48] M. H. N. Shaikh, V. A. Bohara, A. Srivastava, and G. Ghatak, "A downlink RIS-aided NOMA system with hardware impairments: Performance characterization and analysis," *IEEE Open J. Signal Process.*, vol. 3, pp. 288–305, 2022.

[49] M. Soleymani, I. Santamaria, and E. Jorswieck, "Spectral and energy efficiency maximization of MISO STAR-RIS-assisted URLLC systems," *arXiv preprint arXiv:2301.00583*, 2023.

[50] T. Jiang and W. Yu, "Interference nulling using reconfigurable intelligent surface," *IEEE J. Sel. Areas Commun.*, 2022.

[51] Z. K. Ho and E. A. Jorswieck, "Instantaneous relaying: Optimal strategies and interference neutralization," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6655–6668, 2012.

[52] M. Soleymani, I. Santamaria, and P. J. Schreier, "Improper signaling for multicell MIMO RIS-assisted broadcast channels with I/Q imbalance," *IEEE Trans. Green Commun. Netw.*, 2022.

[53] M. Soleymani, C. Lameiro, P. J. Schreier, and I. Santamaria, "Improper signaling for OFDM underlay cognitive radio systems," in *2018 IEEE Statistical Signal Process. Workshop (SSP)*. IEEE, 2018, pp. 722–726.

[54] C. Hellings and W. Utschick, "Improper signaling versus time-sharing in the two-user Gaussian interference channel with TIN," *IEEE Trans. Inf. Theory*, vol. 66, no. 5, pp. 2988–2999, 2020.

[55] M. Soleymani, I. Santamaria, and P. J. Schreier, "Distributed algorithms for spectral and energy-efficiency maximization of K-user interference channels," *IEEE Access*, vol. 9, pp. 96 948–96 963, 2021.

[56] M. Soleymani, I. Santamaria, B. Maham, and P. J. Schreier, "Rate region of the K-user MIMO interference channel with imperfect transmitters," in *Proc. IEEE Eu. Signal Process. Conf. (EUSIPCO)*, 2020, pp. 1–5.

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

[58] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data: the Theory of Improper and Noncircular Signals*. Cambridge University Press, 2010.

[59] T. Adali, P. J. Schreier, and L. L. Scharf, "Complex-valued signal processing: The proper way to deal with impropriety," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5101–5125, 2011.

[60] J. Cui, Y. Liu, Z. Ding, P. Fan, and A. Nallanathan, "QoE-based resource allocation for multi-cell NOMA networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 9, pp. 6160–6176, 2018.

[61] L. Lei, L. You, Y. Yang, D. Yuan, S. Chatzinotas, and B. Ottersten, "Load coupling and energy optimization in multi-cell and multi-carrier NOMA networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 11 323–11 337, 2019.

[62] F. Guo, H. Lu, X. Jiang, M. Zhang, J. Wu, and C. W. Chen, "QoS-aware user grouping strategy for downlink multi-cell NOMA systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 12, pp. 7871–7887, 2021.

[63] L. You and D. Yuan, "A note on decoding order in user grouping and power optimization for multi-cell NOMA with load coupling," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 495–505, 2020.

[64] K. Wang, Y. Liu, Z. Ding, A. Nallanathan, and M. Peng, "User association and power allocation for multi-cell non-orthogonal multiple access networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5284–5298, 2019.

[65] L. You, D. Yuan, L. Lei, S. Sun, S. Chatzinotas, and B. Ottersten, "Resource optimization with load coupling in multi-cell NOMA," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4735–4749, 2018.

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

[67] M. Zeng, X. Li, G. Li, W. Hao, and O. A. Dobre, "Sum rate maximization for IRS-assisted uplink NOMA," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 234–238, 2021.

[68] Y. Liu, S. Zhang, X. Mu, Z. Ding, R. Schober, N. Al-Dhahir, E. Hossain, and X. Shen, "Evolution of NOMA toward next generation multiple access (NGMA) for 6G," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 4, pp. 1037–1071, 2022.

[69] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions," *IEEE Trans. Veh.Technol.*, vol. 65, no. 8, pp. 6010–6023, 2015.

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

[71] Z. Ding, R. Schober, and H. V. Poor, "Unveiling the importance of SIC in NOMA systemsPart II: New results and future directions," *IEEE Commun. Lett.*, vol. 24, no. 11, pp. 2378–2382, 2020.

[72] A. Zappone and E. Jorswieck, "Energy efficiency in wireless networks via fractional programming theory," *Found Trends in Commun. Inf. Theory*, vol. 11, no. 3-4, pp. 185–396, 2015.

[73] S. Buzzi, I. Chih-Lin, T. E. Klein, H. V. Poor, C. Yang, and A. Zappone, "A survey of energy-efficient techniques for 5G networks and challenges ahead," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 697–709, 2016.

[74] T. Lipp and S. Boyd, "Variations and extension of the convex–concave procedure," *Optim. Eng.*, vol. 17, no. 2, pp. 263–287, 2016.

[75] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, 2017.

[76] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.