NOMA-based Improper Signaling for MIMO STAR-RIS-assisted Broadcast Channels with Hardware Impairments

Mohammad Soleymani*, Ignacio Santamaria†, and Eduard Jorswieck‡
*Signal & System Theory Group, Universität Paderborn, Germany
†Dept. Communications Engineering, Universidad de Cantabria, Spain
‡ Institute for Communications Technology, Technische Universität Braunschweig, Germany
Email: mohammad.soleymani@sst.upb.de, i.santamaria@unican.es, jorswieck@ifn.ing.tu-bs.de

Abstract—This paper proposes schemes to improve the spectral efficiency of a multiple-input multiple-output (MIMO) broadcast channel (BC) with I/Q imbalance (IQI) at transceivers by employing a combination of improper Gaussian signaling (IGS), non-orthogonal multiple access (NOMA) and simultaneously transmit and reflect (STAR) reconfigurable intelligent surface (RIS). When there exists IQI, the output RF signal is a widely linear transformation of the input signal, which may make the output signal improper. To compensate for IQI, we employ IGS, thus generating a transmit improper signal. We show that IGS alongside with NOMA can highly increase the minimum rate of the users. Moreover, we propose schemes for different operational modes of STAR-RIS and show that STAR-RIS can significantly improve the system performance. Additionally, we show that IQI can highly degrade the performance especially if it is overlooked in the design.

Index Terms—Improper Gaussian signaling, I/Q imbalance, MIMO broadcast channels, non-orthogonal multiple access, reflecting intelligent surface, spectral efficiency.

I. INTRODUCTION

A main target of the sixth generation of wireless communication systems (6G) is to improve spectral efficiency (SE) of wireless networks by around 10 times higher than the SE of 5G networks [?]. To this end, we have to encompass some challenges such as interference and hardware impairments (HWI), which may highly affect the system performance. Interference has been always among the main performance limitations of modern wireless communication systems [?]. Hence, interference-management techniques such as non-orthogonal multiple access (NOMA) and improper Gaussian signaling (IGS) are expected to continue playing a key role in such networks. Moreover, devices are never ideal in practice, and thus, HWI-aware techniques should be employed to compensate for the non-idealities [?], [?], [?], [?], [?].

A source for HWI is an imbalance between the in-phase and quadrature signals, which is known as I/Q imbalance (IQI) [?], [?]. In the presence of IQI, the output transmitted/received signal is a widely linear transformation (WLT) of the input signal, which means that output digital baseband signal can be improper even if the input signal is proper. Note that in a zero-mean improper signal, the real and imaginary parts of the signal are not independent and identically distributed [?].

To compensate for IQI, we employ IGS, which is shown to be a powerful interference-management technique to improve the SE and/or energy efficiency of various systems [?], [?], [?], [?], [?].

In addition to interference-management and/or HWI-aware techniques, 6G should employ some emerging technologies like reconfigurable intelligent surface (RIS) to fulfill the ever increasing SE demands. RIS has been shown to be a promising technology to improve the performance of various interference-limited and/or interference-free systems [?], [?], [?], [?], [?], [?], [?]. A regular passive RIS can provide an additional degree of freedom in the system design by controlling the reflections, which can be used to improve the coverage and/or neutralize interference. To employ a regular RIS, both the transmitter and receiver should be in the reflection space of the regular RIS, which may restrict its applicability. To address this issue, simultaneously-reflect-and-transmit (STAR-) RISs can be employed, which provide a 360° coverage [?], [?], [?], [?], [?], [?]. In a STAR-RIS, each component can not only transmit, but also can reflect at the same time. Thus, STAR-RIS can cover a wider area and is expected to improve the system performance especially when a regular RIS cannot cover all the users.

In this paper, we propose NOMA-based IGS schemes to maximize the minimum-weighted rate of users in a MIMO STAR-RIS-assisted BC with IQI. To this end, we extend the proposed scheme in [?] to MIMO systems with IQI. Note that [?] proposed schemes to enhance the SE and EE of a multi-cell MISO RIS-assisted BC by considering regular RISs and perfect devices. However, in this work, we study STAR-RIS and focus on different modes of STAR-RIS with different assumptions regarding the feasibility sets for optimizing the RIS components. We consider two operational modes for STAR-RIS. First, we assume that all the STAR-RIS components can simultaneously operate in both the transmission and reflections modes, which is known as energy splitting (ES). Second, we divide the STAR-RIS components into two groups and operate the components in a group only in the transmission mode, while the components for the other group operate only in the reflection mode. This operational mode is known as the mode
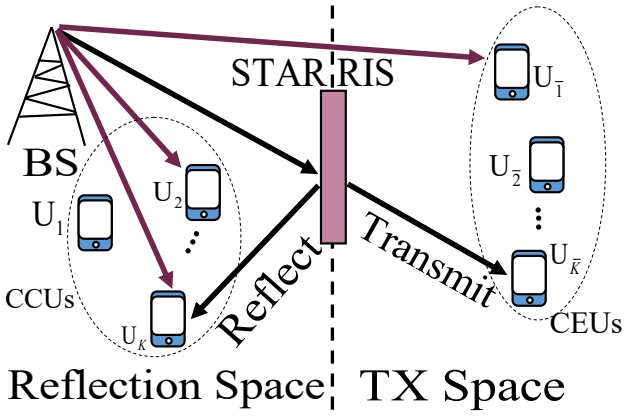
Fig. 1: A broadcast channel assisted by a STAR-RIS.

switching (MS) scheme.

Our results show that the NOMA-based IGS scheme can significantly outperform the NOMA-based proper Gaussian signaling (PGS) scheme and/or the IGS scheme with treating interference as noise (TIN). Moreover, we show that STAR-RIS can considerably increase the minimum rate of users, even with a relatively low number of STAR-RIS components per users. Furthermore, we show that the MS scheme performs close to the ES schemes with different feasibility sets for STAR-RIS components. MS schemes have lower computational and implementation complexities than the ES schemes, which may make them more practical if their performance is close to ES schemes. Additionally, we show that IQI can reduce the minimum rate of users even if it is compensated by IGS, and neglecting IQI in the design can cause a significant performance loss.

II. System Model

We consider a single-cell MIMO STAR-RIS-assisted BC, consisting of a BS with N_{BS} antennas and $2K$ users with N_u antennas each, as shown in Fig. **??**. We assume that there is a passive STAR-RIS that assists the BS to serve the users. We further assume that the transceivers suffer from IQI based on the model in [**?**]. We consider a system in which users are located into two clusters with noticeably different distances to the BS, similar to [**?**]. Indeed, we assume that there are two types of users in each cell: cell-centric users (CCUs) and cell-edge users (CEUs), where each cluster has K users. We randomly divide the users into K pairs (one CCU and one CEU), as in [**?**], each pair being served by NOMA. Note that the proposed scheme in this paper can be applied to other scenarios with a given user pairing/ordering scheme. In this work, our main focus is on optimizing the transmission parameters and STAR-RIS components and leave obtaining the optimal user pairing/ordering for future work.

A. STAR-RIS model

A STAR-RIS can provide a $360°$ coverage since it can simultaneously transmit and reflect. In a STAR-RIS-assisted

system, there can be two spaces for each STAR-RIS: transmission and reflection, as depicted in Fig. **??**. In the reflection space, users receive signals through a reflection from STAR-RIS, while in the transmission space, the STAR-RIS transmits signals to users. Thus, each user belongs to either a transmission space or a reflection space of a STAR-RIS. Hence, the channel between the BS and user k is [**?**, Eq. (2)]

$$\mathbf{H}_k\big(\{\boldsymbol{\Theta}^r, \boldsymbol{\Theta}^t\}\big) = \mathbf{G}_k \boldsymbol{\Theta}^{r/t} \mathbf{G} + \mathbf{F}_k \in \mathbb{C}^{N_u \times N_{BS}}, \quad (1)$$

where $\mathbf{F}_k$ is the direct link between the BS and user k, $\mathbf{G}$ is the channel matrix between the BS and the STAR-RIS, $\mathbf{G}_k$ is the channel vector between the STAR-RIS and user k, $\boldsymbol{\Theta}^r = \text{diag}\left(\theta_1^r, \theta_2^r, \cdots, \theta_{N_{RIS}}^r\right)$, is a diagonal matrix, containing all the reflection coefficients of the STAR-RIS and $\boldsymbol{\Theta}^t = \text{diag}\left(\theta_1^t, \theta_2^t, \cdots, \theta_{N_{RIS}}^t\right)$ is a diagonal matrix, containing all the transmission coefficients of the STAR-RIS. If user k is in the reflection (or transmission) space of the STAR-RIS, its channel can be optimized by $\boldsymbol{\Theta}^r$ (or $\boldsymbol{\Theta}^t$). Hereafter, for notational simplicity, we drop the dependency of the channels on the RIS components and represent channels as $\mathbf{H}_k$ for all k.

There are three different assumptions for modeling the relation between the reflection and transmission coefficients of each STAR-RIS component. First, the amplitudes of the reflection and transmission coefficients are related as [**?**, Eq. (2)]

$$|\theta_i^r|^2 + |\theta_i^t|^2 \le 1, \forall i, \quad (2)$$

which is a convex constraint. We call the feasibility sets for this model $\mathcal{T}_U$, given by [**?**, Eq. (9)]. Second, the amplitudes of the reflection and transmission coefficients are related as [**?**, Eq. (1)]: $|\theta_i^r|^2 + |\theta_i^t|^2 = 1$ for all i. This non-convex constraint can be written as the two constraints: (**??**) and

$$|\theta_i^r|^2 + |\theta_i^t|^2 \ge 1, \forall i. \quad (3)$$

Indeed, in this model, there is the additional constraint (**??**), comparing to $\mathcal{T}_U$, which implies that the second model is more stringent. We call the feasibility set of the second model $\mathcal{T}_I$, given by [**?**, Eq. (10)]. Finally, in the third model, the amplitudes of the reflection and transmission coefficients are related to as in $\mathcal{T}_I$, and the phases are related to as $\angle\theta_i^r = \angle\theta_i^r \pm \pi/2$ for all i [**?**], [**?**]. The phase relation can be also modeled as the following convex constraint [**?**, Lemma 1]

$$|\theta_i^r \pm \theta_i^t|^2 \le 1, \forall i. \quad (4)$$

We represent the feasibility set for the third model by $\mathcal{T}_N$, given by [**?**, Eq. (11)]. It is evident that $\mathcal{T}_N \subset \mathcal{T}_I \subset \mathcal{T}_U$. Thus, $\mathcal{T}_U$ should perform not worse than $\mathcal{T}_I$, which performs not worse than $\mathcal{T}_N$. The set $\mathcal{T}_U$ is convex; however, $\mathcal{T}_N$ and $\mathcal{T}_I$ are non-convex because of the constraint (**??**).

B. Signal model

The BS aims at transmitting the signal

$$\mathbf{x} = \sum_{k=1}^{2K} \mathbf{x}_k \in \mathbb{C}^{N_{BS} \times 1}, \quad (5)$$

where $\mathbf{x}_k$ is the signal intended for user k. However, due to IQI, the actual transmitted signal at the BS is a WLT of $\mathbf{x}$ as $\mathbf{x}^t = \boldsymbol{\Gamma}_{1,t}\mathbf{x} + \boldsymbol{\Gamma}_{2,t}\mathbf{x}^*$, where $\boldsymbol{\Gamma}_{1,t} = \frac{\mathbf{I}+\mathbf{A}_T e^{j\phi_T}}{2}$ and $\boldsymbol{\Gamma}_{2,t} = \mathbf{I} - \boldsymbol{\Gamma}_{1,t}^*$ [?]. The matrices $\mathbf{A}_T$ and ϕ_T are diagonal matrices that correspond to, respectively, the amplitude and phase errors between the two quadrature signal paths at the transmitter side [?]. If $\mathbf{A}_T = \mathbf{I}$ and $\phi_T = \mathbf{0}$, there is no IQI. Otherwise, $\mathbf{x}^t$ can be improper even if $\mathbf{x}$ is proper.

The received signal at user k is $\mathbf{y}_k = \mathbf{H}_k\mathbf{x}^t + \mathbf{n}_k$, where $\mathbf{n}_k$ is zero-mean additive white Gaussian noise with variance $\sigma^2\mathbf{I}$. Due to IQI at the receiver of user k, the output signal is a widely linear transformation of $\mathbf{y}_k$ as $\mathbf{y}_k^r = \boldsymbol{\Gamma}_{1,r}\mathbf{y}_k + \boldsymbol{\Gamma}_{2,r}\mathbf{y}_k^*$, where $\boldsymbol{\Gamma}_{1,r} = \frac{\mathbf{I}+\mathbf{A}_R e^{j\phi_R}}{2}$ and $\boldsymbol{\Gamma}_{2,r} = \mathbf{I} - \boldsymbol{\Gamma}_{1,r}^*$ [?]. Similarly, the matrices $\mathbf{A}_R$ and ϕ_R are diagonal matrices, corresponding to the amplitude and phase errors of each branch at the receiver side, respectively [?].

We employ the real-decomposition method for modeling and optimizing improper signals, similar to [?], [?]. The output signal at the receiver side, using the real-decomposition method, is

$$\underline{\mathbf{y}}_k = \underline{\mathbf{H}}_k \sum_{i=1}^{2K} \underline{\mathbf{x}}_i + \underline{\mathbf{n}}_k, \tag{6}$$

where $\underline{\mathbf{y}}_k = \begin{bmatrix} \Re\{\mathbf{y}_k\}^T & \Im\{\mathbf{y}_k\}^T \end{bmatrix}^T$, and $\underline{\mathbf{x}}_k = \begin{bmatrix} \Re\{\mathbf{x}_k\}^T & \Im\{\mathbf{x}_k\}^T \end{bmatrix}^T$ for all k. Moreover, $\underline{\mathbf{n}}_k$ is the effective noise at the output of user k, and $\underline{\mathbf{H}}_k$ is the effective channel in the real domain, given by [?, Lemma 1]. The signals $\underline{\mathbf{x}}_k$ and $\underline{\mathbf{n}}_k$ are zero-mean. We represent the covariance matrices of $\underline{\mathbf{x}}_k$ and $\underline{\mathbf{n}}_k$ by, respectively, $\mathbf{P}_k$ and $\underline{\mathbf{C}}_n$, where $\mathbf{P}_k$ is an optimization parameter, while $\underline{\mathbf{C}}_n$ is given by [?, Lemma 1]. The matrix $\underline{\mathbf{C}}_n$ is equal to $\frac{\sigma^2}{2}\mathbf{I}$ if there is no IQI at the receivers. Otherwise, $\underline{\mathbf{n}}_k$ is improper and its covariance matrix is not equal to a scaled identity matrix.

C. NOMA

We treat the inter-cluster interference as noise and employ IGS and spatial dimensions to handle it. However, we employ NOMA to manage intra-cluster interference. That is, we assume that the CCU k first decodes the message of its paired CEU $\bar{k} = K + k$ and cancels the message from the received signal. Thus, the achievable rate of user k is

$$r_k = \frac{1}{2}\log_2\left|\mathbf{I} + \mathbf{D}_k^{-1}\mathbf{S}_k\right| = \underbrace{\frac{1}{2}\log_2|\mathbf{D}_k + \mathbf{S}_k|}_{r_{k,1}} - \underbrace{\frac{1}{2}\log_2|\mathbf{D}_k|}_{\bar{r}_{k,2}}, \tag{7}$$

where $\mathbf{D}_k = \sum_{j=1,j\neq k,\bar{k}}^{2K} \underline{\mathbf{H}}_k\mathbf{P}_j\underline{\mathbf{H}}_k^T + \underline{\mathbf{C}}_n$, and $\mathbf{S}_k = \underline{\mathbf{H}}_k\mathbf{P}_k\underline{\mathbf{H}}_k^T$. On the contrary, the CEU $\bar{k}$ treats all the received signals as noise while decoding its own signal. The message for user $\bar{k}$ should be transmitted at a rate that is decodable for both users k and $\bar{k}$. Thus, the maximum transmission rate for user $\bar{k}$ is

$$r_{\bar{k}} = \min\left(\bar{r}_{\bar{k}}, \bar{r}_{k\to\bar{k}}\right), \tag{8}$$

where $\bar{r}_{\bar{k}}$ is the maximum decoding rate of $\mathbf{x}_{\bar{k}}$ at user $\bar{k}$ treating all other signals as noise, which is given by

$$\bar{r}_{\bar{k}} = \frac{1}{2}\log_2\left|\mathbf{I} + \mathbf{D}_{\bar{k}}^{-1}\mathbf{S}_{\bar{k}}\right| = \underbrace{\frac{1}{2}\log_2|\mathbf{D}_{\bar{k}} + \mathbf{S}_{\bar{k}}|}_{\bar{r}_{\bar{k},1}} - \underbrace{\frac{1}{2}\log_2|\mathbf{D}_{\bar{k}}|}_{\bar{r}_{\bar{k},2}}, \tag{9}$$

where $\mathbf{D}_{l\bar{k}} = \sum_{j=1,j\neq\bar{k}}^{2K} \underline{\mathbf{H}}_{\bar{k}}\mathbf{P}_j\underline{\mathbf{H}}_{\bar{k}}^T + \underline{\mathbf{C}}_n$, and $\mathbf{S}_{\bar{k}} = \underline{\mathbf{H}}_{\bar{k}}\mathbf{P}_{\bar{k}}\underline{\mathbf{H}}_{\bar{k}}^T$. Moreover, $\bar{r}_{k\to\bar{k}}$ is the maximum decoding rate of $\mathbf{x}_{\bar{k}}$ at user k, while treating all other signals as noise, given by

$$\bar{r}_{k\to\bar{k}} = \underbrace{\frac{1}{2}\log_2|\mathbf{D}_k + \mathbf{S}_k + \mathbf{S}_{k\to\bar{k}}|}_{\bar{r}_{k\to\bar{k},1}} - \underbrace{\frac{1}{2}\log_2|\mathbf{D}_k + \mathbf{S}_k|}_{\bar{r}_{lk\to\bar{k},2}}$$
$$= \frac{1}{2}\log_2\left|\mathbf{I} + (\mathbf{D}_k + \mathbf{S}_k)^{-1}\mathbf{S}_{k\to\bar{k}}\right|, \tag{10}$$

where $\mathbf{S}_{k\to\bar{k}} = \underline{\mathbf{H}}_k\mathbf{P}_{\bar{k}}\underline{\mathbf{H}}_k^T$.

D. Problem statement

In this paper, we aim at maximizing the minimum weighted rate, which can be written as

$$\max_{r, \boldsymbol{\Theta}\in\mathcal{T}, \{\mathbf{P}\}\in\mathcal{P}} r, \quad \text{s.t.} \quad \lambda_k r_k \geq r, \quad k = 1, \cdots, 2K, \tag{11}$$

where λ_k is the corresponding weight for user k, $\{\mathbf{P}\}$ is the set of covariance matrices, and $\mathcal{P}$ is the feasibility set of the covariance matrices, which is given by [?, Eq. (23)] for IGS and by [?, Eq. (24)] for PGS schemes.

III. PROPOSED SCHEME

To solve (??), we employ alternating optimization (AO) and an MM-based technique. That is, we first fix $\{\boldsymbol{\Theta}^{(t-1)}\}$, and optimize over covariance matrices to obtain $\{\mathbf{P}^{(t)}\}$. We then alternate and optimize over RIS components for fixed covariance matrices $\{\mathbf{P}^{(t)}\}$ to find $\{\boldsymbol{\Theta}^{(t)}\}$. We iterate this procedure until a convergence criterion is met.

A. Optimizing transmit covariance matrices

Solving (??) for fixed $\{\boldsymbol{\Theta}^{(t-1)}\}$ is very similar to the scheme in [?, Sec. III.A]. For the sake of completeness, we briefly describe the scheme and refer the reader to [?] for more details. The optimization problem (??) for fixed $\{\boldsymbol{\Theta}^{(t-1)}\}$ is non-convex. Thus, we employ an MM-based algorithm to find a solution for this problem. That is, we first find suitable surrogate functions for the rates. To this end, we employ [?, Lemma 3] to approximate the convex part of the rates by an affine function, which results in the following concave lower bounds

$$r_k \geq \tilde{r}_k = r_{k,1} - r_{k,2}^{(t-1)}$$
$$- \sum_{j=1,j\neq k,\bar{k}}^{2K} \text{Tr}\left(\frac{\underline{\mathbf{H}}_k^T(\mathbf{D}_k^{(t-1)})^{-1}\underline{\mathbf{H}}_k}{2\ln 2}\left(\mathbf{P}_j - \mathbf{P}_j^{(t-1)}\right)\right), \tag{12}$$
$$\tilde{r}_{\bar{k}}^l = \bar{r}_{\bar{k},1} - \bar{r}_{\bar{k},2}^{(t-1)}$$

$$-\sum_{j=1,j\neq k}^{2K}\mathrm{Tr}\left(\frac{\mathbf{H}_{\bar{k}}^T(\mathbf{D}_{\bar{k}}^{(t-1)})^{-1}\mathbf{H}_{\bar{k}}}{2\ln 2}\left(\mathbf{P}_j-\mathbf{P}_j^{(t-1)}\right)\right),\quad(13)$$

$$\tilde{r}_{k\to\bar{k}}^l=\bar{r}_{k\to\bar{k},1}-\bar{r}_{k\to\bar{k},2}^{(t-1)}-\sum_{j=1,j\neq\bar{k}}^{2K}\mathrm{Tr}\frac{\mathbf{H}_k^T}{2\ln 2}$$

$$\times\left(\mathbf{D}_k^{(t-1)}+\mathbf{S}_k^{(t-1)}\right)^{-1}\underline{\mathbf{H}}_k\left(\mathbf{P}_{ij}-\mathbf{P}_{ij}^{(t-1)}\right)\Bigg),\quad(14)$$

where $r_{k,2}^{(t-1)}$, $r_{\bar{k},2}^{(t-1)}$, $r_{k\to\bar{k},2}^{(t-1)}$, $\mathbf{S}_k^{(t-1)}$, $\mathbf{D}_k^{(t-1)}$, and $\mathbf{D}_{\bar{k}}^{(t-1)}$ are the corresponding initial values at this optimization step. Note that the concave lower bound for the rate of the CEU $\bar{k}$ is $\tilde{r}_{\bar{k}}=\min\left(\tilde{r}_{\bar{k}}^l,\tilde{r}_{k\to\bar{k}}^l\right)$. Substituting the concave lower bounds for the rates in the optimization problem (??) for fixed $\{\boldsymbol{\Theta}^{(t-1)}\}$ results in a convex optimization problem, which can be efficiently solved by existing numerical tools.

B. Optimizing RIS components

In this subsection, we update the STAR-RIS coefficients by solving (??) for fixed $\{\mathbf{P}^{(t)}\}$. Unfortunately, (??) for fixed $\{\mathbf{P}^{(t)}\}$ is not convex since the rates are not concave in $\{\boldsymbol{\Theta}\}$. Moreover, the feasibility set for STAR-RIS components is not convex for the sets $\mathcal{T}_I$ and $\mathcal{T}_N$. To update $\{\boldsymbol{\Theta}\}$, we propose a suboptimal scheme based on MM. In the following, we first consider the ES scheme. Then we mention how the scheme can be modified to be applied to MS schemes.

To solve (??) for fixed $\{\mathbf{P}^{(t)}\}$, we should first obtain suitable concave lower bounds for the rates. To this end, we employ the bounds in [?, Lemma 4], which results in the following lower bounds

$$\hat{r}_k=r_k^{(t-1)}-\frac{1}{2\ln 2}\mathrm{Tr}\left(\bar{\mathbf{S}}_k\bar{\mathbf{D}}_k^{-1}\right)+\frac{1}{\ln 2}\mathrm{Tr}\left(\mathbf{A}_k\underline{\mathbf{H}}_k\right)$$
$$-\frac{1}{2\ln 2}\mathrm{Tr}\big((\bar{\mathbf{D}}_k^{-1}-(\bar{\mathbf{S}}_k+\bar{\mathbf{D}}_k)^{-1})(\mathbf{D}_k+\mathbf{S}_k)\big),\quad(15)$$

$$\hat{r}_{\bar{k}}^l=r_{\bar{k}}^{(t-1)}-\frac{1}{2\ln 2}\mathrm{Tr}\left(\bar{\mathbf{S}}_{\bar{k}}\bar{\mathbf{D}}_{\bar{k}}^{-1}\right)+\frac{1}{\ln 2}\mathrm{Tr}\left(\mathbf{A}_{\bar{k}}\underline{\mathbf{H}}_{\bar{k}}\right)$$
$$-\frac{1}{2\ln 2}\mathrm{Tr}\big((\bar{\mathbf{D}}_{\bar{k}}^{-1}-(\bar{\mathbf{S}}_{\bar{k}}+\bar{\mathbf{D}}_{\bar{k}})^{-1})(\mathbf{D}_{\bar{k}}+\mathbf{S}_{\bar{k}})\big),\quad(16)$$

$$\hat{r}_{k\to\bar{k}}^l=r_{k\to\bar{k}}^{(t-1)}-\frac{1}{2\ln 2}\big[\,\mathrm{Tr}\left(\bar{\mathbf{S}}_{k\to\bar{k}}(\bar{\mathbf{D}}_k+\bar{\mathbf{S}}_k)^{-1}\right)$$
$$+2\,\mathrm{Tr}\left(\mathbf{A}_{k\to\bar{k}}\underline{\mathbf{H}}_{\bar{k}}\right)-\mathrm{Tr}\left(((\bar{\mathbf{D}}_k+\bar{\mathbf{S}}_k)^{-1}\right.$$
$$-(\mathbf{S}_{k\to\bar{k}}+\bar{\mathbf{S}}_k+\bar{\mathbf{D}}_k)^{-1})(\mathbf{D}_k+\mathbf{S}_k+\mathbf{S}_{k\to\bar{k}}))\big],\quad(17)$$

where $\bar{\mathbf{S}}_k$, $\bar{\mathbf{S}}_{\bar{k}}$, $\bar{\mathbf{S}}_{k\to\bar{k}}$, $\bar{\mathbf{D}}_k$, and $\bar{\mathbf{D}}_{\bar{k}}$ are the corresponding initial values at this optimization step. Moreover, $\mathbf{A}_k=\mathbf{P}_k\mathbf{H}_k^T(\{\boldsymbol{\Theta}^{(t-1)}\})\bar{\mathbf{D}}_k^{-1}$, $\mathbf{A}_{\bar{k}}=\mathbf{P}_{\bar{k}}\mathbf{H}_{\bar{k}}^T(\{\boldsymbol{\Theta}^{(t-1)}\})\bar{\mathbf{D}}_{\bar{k}}^{-1}$, and $\mathbf{A}_{k\to\bar{k}}=\mathbf{P}_{\bar{k}}\underline{\mathbf{H}}_k^T(\{\boldsymbol{\Theta}^{(t-1)}\})(\bar{\mathbf{D}}_k+\bar{\mathbf{S}}_k)^{-1}$. Finally, the concave lower bound for the CEU $\bar{k}$ is $\hat{r}_{\bar{k}}=\min\left(\hat{r}_{\bar{k}}^l,\hat{r}_{k\to\bar{k}}^l\right)$. Substituting the concave lower bounds for the rates in the problem (??), we have

$$\max_{r,\boldsymbol{\Theta}\in\mathcal{T}}\ r,\quad\text{s.t.}\quad\lambda_k\hat{r}_k\geq r,\quad k=1,\cdots,2K,\quad(18)$$

which is convex only for $\mathcal{T}_U$. However, since the sets $\mathcal{T}_I$ and $\mathcal{T}_N$ are not convex sets (because of (??)), the problem (??) is not convex for these sets. To convexify (??) for $\mathcal{T}_I$ and $\mathcal{T}_N$, we can apply CCP to the constraint in (??) since $|\theta_i^r|^2+|\theta_i^t|^2$

for all i is a convex function and should be converted to a concave function. Additionally, we can relax (??) for a faster convergence. Thus, (??) can be approximated as

$$|\theta_i^{r^{(t-1)}}|^2+2\mathfrak{R}\left(\theta_i^{r^{(t-1)}}(\theta_i^r-\theta_i^{r^{(t-1)}})^*\right)+|\theta_i^{t^{(t-1)}}|^2$$
$$+2\mathfrak{R}\left(\theta_i^{t^{(t-1)}}(\theta_i^t-\theta_i^{t^{(t-1)}})^*\right)\geq 1-\epsilon,\forall i,\quad(19)$$

where $\epsilon>0$. The constraint (??) is convex since it is linear in θ_i^r and θ_i^t. Therefore, (??) for $\mathcal{T}_I$ can be approximated as

$$\max_{r,\boldsymbol{\Theta}}\ r,\quad\text{s.t.}\quad\lambda_k\hat{r}_k\geq r,\quad k=1,\cdots,2K,\quad(20a)$$

$$(??),(??).\quad(20b)$$

Moreover, (??) for $\mathcal{T}_N$ can be approximated as

$$\max_{r,\boldsymbol{\Theta}}\ r,\quad\text{s.t.}\quad\lambda_k\hat{r}_k\geq r,\quad k=1,\cdots,2K,\quad(21a)$$

$$(??),(??),(??).\quad(21b)$$

The problems (??) and (??) are convex and can be efficiently solved. However, their solutions may not satisfy the constraint $|\theta_i^r|^2+|\theta_i^t|^2=1$ due to the relaxation in (??). Thus, to ensure a valid solution, we should normalize the solutions of (??) (or (??)) as

$$\hat{\theta}_i^t=\frac{\theta_i^{t^{(\star)}}}{\sqrt{|\theta_i^{t^{(\star)}}|^2+|\theta_i^{r^{(\star)}}|^2}},\quad\hat{\theta}_i^r=\frac{\theta_i^{r^{(\star)}}}{\sqrt{|\theta_i^{t^{(\star)}}|^2+|\theta_i^{r^{(\star)}}|^2}},\quad(22)$$

for all i. Although $\hat{\theta}_i^t$ and $\hat{\theta}_i^r$ are feasible, they may not result in a non-decreasing sequence of the minimum rate of users. To address this issue, we update $\boldsymbol{\Theta}^r$ and $\boldsymbol{\Theta}^t$ such that

$$\{\boldsymbol{\Theta}^{(t)}\}=\begin{cases}\{\hat{\boldsymbol{\Theta}}\}&\text{if}\quad\min_k\left\{\lambda_k r_k\left(\{\hat{\boldsymbol{\Theta}}\}\right)\right\}\geq\\&\quad\min_k\left\{\lambda_k r_k(\{\boldsymbol{\Theta}^{(t-1)}\})\right\}\\\{\boldsymbol{\Theta}^{(t-1)}\}&\text{Otherwise},\end{cases}\quad(23)$$

where $\{\hat{\boldsymbol{\Theta}}\}=\{\hat{\boldsymbol{\Theta}}^r,\hat{\boldsymbol{\Theta}}^t\}$ is given by (??). This updating step ensures the convergence of the proposed scheme since the framework generates a non-decreasing sequence of the minimum rates of the users.

Now we consider the MS scheme in which each STAR-RIS component either reflects or transmits at a time. In this case, we set $\theta_i^t=0$ (or $\theta_i^r=0$) for STAR-RIS components that operate in the reflection (or transmission) mode. Then the other steps remain the same. Specifically, to obtain new $\{\boldsymbol{\Theta}\}$, we should solve (??) for $\mathcal{T}_U$, (??) for $\mathcal{T}_I$, and (??) for $\mathcal{T}_N$. Then we should normalize the solutions for the feasibility sets $\mathcal{T}_I$ and $\mathcal{T}_N$ according to (??) and update $\{\boldsymbol{\Theta}\}$ based on (??). Note that the feasibility sets $\mathcal{T}_I$ and $\mathcal{T}_N$ are equivalent for the MS scheme since each STAR-RIS component either transmits or reflects in which the phase relation does not have any impact.

IV. NUMERICAL RESULTS

In this section, we present some numerical results. To this end, we consider a scenario in which the CCUs are in the reflection space of the RIS, while the CEUs are in the transmission space of the RIS. Thus, a regular RIS can assist
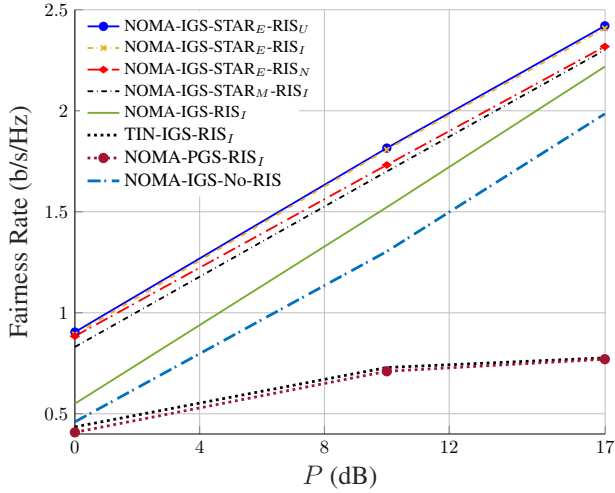
Fig. 2: The average fairness rate versus P for $N_{BS} = N_u = 2$, $N_{RIS} = 60$, and $K = 3$.

Fig. 3: The average performance improvement by employing STAR-RIS versus K for $N_{BS} = N_u = 2$, and $N_{RIS} = 60$.

Fig. 4: IGS with respect to PGS.

Fig. 5: The average performance improvement by employing IGS versus K for $N_{BS} = N_u = 2$, and $N_{RIS} = 60$.

only the CCUs and cannot cover the CEUs. However, a STAR-RIS can cover all the users due to its $360°$ coverage. In order to have a fair comparison, we assume that the number of RIS components for both the regular and STAR-RISs are the same. To better investigate the role of IGS, we divide the numerical results into two subsection. In the first subsection, we investigate the role of IGS as an interference-management technique. To this end, we consider perfect devices since IQI can increase the benefits of IGS, as shown in [**?**], [**?**]. Then, we consider the impact of IQI and the importance of HWI-aware techniques in Section IV.B.

A. Impact of interference on spectral efficiency

Fig. **??** shows the average fairness rate versus P for $N_{BS} = N_u = 2$, $N_{RIS} = 60$, and $K = 3$. As can be observed, the proposed NOMA-based IGS scheme with ES and the feasibility set $\mathcal{T}_U$ outperforms the other schemes, while the NOMA-based IGS scheme with ES and $\mathcal{T}_I$ performs very close to the ES scheme with $\mathcal{T}_U$. Moreover, we observe that the MS scheme performs very close to the ES schemes, and the performance gap between the MS scheme and the ES scheme with the feasibility set $\mathcal{T}_N$ is minor, especially at higher SNR regimes. Note that the computational and/or implementation complexities of MS schemes are less than ES schemes, which can compensate the small performance gap, comparing to the ES schemes.

In Fig. **??**, we can also observe that IGS and/or NOMA can highly improve the system performance. Interestingly, the NOMA-based IGS scheme without RIS can significantly outperform the NOMA-based PGS scheme as well as the IGS scheme with TIN for RIS-assisted systems. This result, indeed, shows the importance of interference-management techniques in highly overloaded STAR-RIS-assisted systems. Moreover, it shows that RIS alone cannot handle interference in over-loaded systems, and we have to employ powerful interference-management techniques to fully reap (STAR-)RIS benefits.
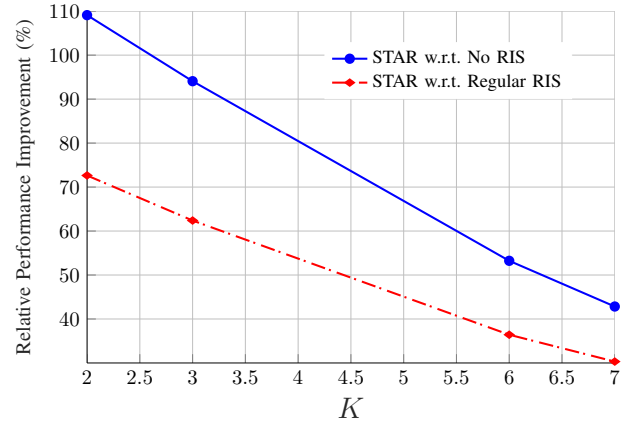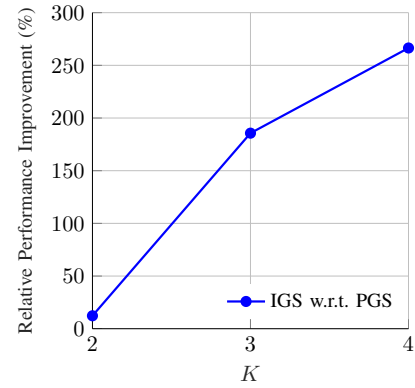
Fig. **??** shows the benefits of employing STAR-RIS versus K for $N_{BS} = N_u = 2$, and $N_{RIS} = 60$. As can be observed, STAR-RIS can significantly improve the system performance even over a regular RIS. For a fixed number of STAR-RIS components, the benefits of STAR-RIS decrease with the number of users. However, the benefits are still significant when there are less than five components per users, i.e., when there are 14 users in the network (for $K = 7$). Note that the total number of users is $2K$.

Fig. **??** shows the benefits of employing IGS versus K for $N_{BS} = N_u = 2$, and $N_{RIS} = 60$. The IGS benefits are computed by comparing the performance of the NOMA-based IGS scheme with the NOMA-based PGS scheme for RIS-assisted systems. As can be observed, IGS can significantly improve the system performance. Furthermore, the benefits of IGS and/or NOMA increase with the number of users, which is in line with our previous studies in [**?**], [**?**]. The reason is that, the higher the number users is, the higher interference exists, which makes the use of advanced interference management techniques more necessary and their benefit more significant.
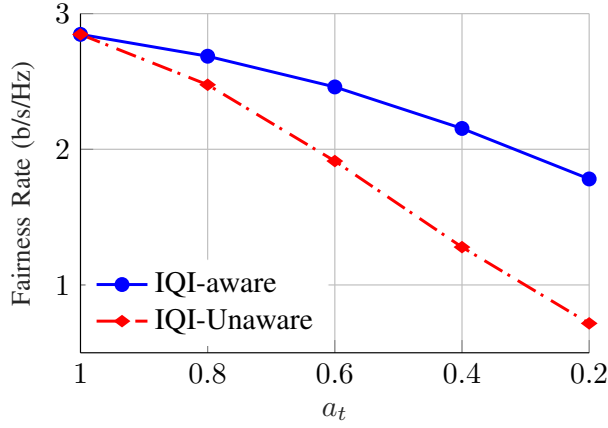
Fig. 6: The impact of IQI on the minimum rate of users for $N_{BS} = N_u = 2$, $K = 2$ and $N_{RIS} = 30$.

B. Impact of IQI on spectral efficiency

In this subsection, we consider the impact of IQI on the system performance. We assume that the IQI parameters at all users are similar. In other words, we set $\mathbf{A}_T = \mathbf{A}_R$ and $\boldsymbol{\phi}_T = \boldsymbol{\phi}_R$. Similar to [?], we assume that $\mathbf{A}_T = a_t\mathbf{I}$ and $\boldsymbol{\phi}_T = \phi\mathbf{I}$, where $\phi = 5°$. In the previous subsection, we study the impact of IGS and NOMA as interference-management techniques. Moreover, we investigate the performance of STAR-RIS. thus, in this subsection, we consider only the NOMA-based IGS scheme for STAR-RIS-assisted systems with MS to study the impact of IQI.

Fig. **??** shows the impact of IQI on the minimum rate of users for $N_{BS} = N_u = 2$, $K = 2$ and $N_{RIS} = 30$. As can be observed, the minimum rate of users decrease with the IQI level, where there is more than around 35% performance loss for $a_t = 0.2$ even when the IQI is considered and compensated in the design. Such performance loss can even increase up to more than 70% if IQI is overlooked in the system design.

V. CONCLUSION

This paper proposed a scheme to improve the minimum rate of a MIMO STAR-RIS-assisted BC with HWI by employing IGS and NOMA. We considered three feasibility sets for STAR-RIS components and proposed different schemes to optimize STAR-RIS components. We showed that STAR-RIS can significantly improve the minimum rate of the users. Furthermore, we showed that the MS scheme can perform very close to the ES scheme, especially when the phases of the reflection and transmission coefficients are dependent. Additionally, we showed that the NOMA-based IGS scheme can highly outperform the NOMA-based PGS and/or the IGS scheme with TIN, and the benefits of IGS increases with the number of users, which is in line with our previous study in [?]. Finally, we showed that IQI can highly degrade the system performance even if it is compensated in the system design.