

# Joint Spatial Division and Multiplexing for FDD in Intelligent Reflecting Surface-assisted Massive MIMO Systems

Anastasios Papazafeiropoulos, Pandelis Kourtessis, Konstantinos Ntontin, Symeon Chatzinotas

**Abstract**—Intelligent reflecting surface (IRS) is a promising technology to deliver the higher spectral and energy requirements in fifth-generation (5G) and beyond wireless networks while shaping the propagation environment. Such a design can be further enhanced with massive multiple-input-multiple-output (mMIMO) characteristics towards boosting the network performance. However, channel reciprocity, assumed in 5G systems such as mMIMO, appears to be questioned in practice by recent studies on IRS. Hence, contrary to previous works, we consider frequency division duplexing (FDD) to study the performance of an IRS-assisted mMIMO system. However, FDD is not suitable for large number of antennas architectures. For this reason we employ the joint spatial division and multiplexing (JSDM) approach exploiting the structure of the correlation of the channel vectors to reduce the channel state information (CSI) uplink feedback, and thus, allowing the use even of a large number of antennas at the base station. JSDM entails dual-structured precoding and clustering the user equipments (UEs) with the same covariance matrix into groups. Specifically, we derive the sum spectral efficiency (SE) based on statistical CSI in terms of large-scale statistics by using the deterministic equivalent (DE) analysis while accounting for correlated Rayleigh fading. Subsequently, we formulate the optimization problem concerning the sum SE with respect to the reflecting beamforming matrix (RBM) and the total transmit power, which can be performed at every several coherence intervals by taking advantage of the slow-time variation of the large-scale statistics. This notable property contributes further to the decrease of the feedback overhead. Numerical results, verified by Monte-Carlo (MC) simulations, enable interesting observations by elucidating how fundamental system parameters such as the rank of the covariance matrix and the number of groups of UEs affect the performance. For example, the selection of a high rank improves the channel conditioning but increases the feedback overhead.

**Index Terms**—Intelligent reflecting surface (IRS), frequency division duplexing (FDD), achievable spectral efficiency, deterministic equivalents, beyond 5G networks.

## I. INTRODUCTION

The stringent requirements of fifth-generation (5G) and beyond wireless networks in terms of spectral efficiency (SE) and energy efficiency (EE) could be achieved by relying on disruptive technologies such as intelligent reflecting surface (IRS) and massive multiple-input multiple-output (mMIMO) systems [1]–[3]. In particular, IRS has been recently emerged as a revolutionary solution supporting high data rates and low energy consumption [4]–[12]. It consists of nearly passive

elements, which can create favourable propagation conditions, especially against obstacles, by a smart adjustment of the elements' phase shifts on the impinging waves. Besides IRS, the mMIMO technology, which has been already deployed in 5G systems [13], assumes a large number of antennas at each base station (BS) by applying low-complexity linear precoding techniques. Among its benefits, we meet high directional beamforming and multiplexing gains while almost cancelling intra-cell interference in the large antenna regime due to channel orthogonality.

With respect to IRS, previous works are classified in two categories depending on whether the phase shifts optimization relies on the instantaneous channel state information (I-CSI) [4]–[6] or statistical CSI (S-CSI) [7], [8], [14]–[22]. For example, in the first category, requiring optimization at each coherence interval, active and passive beamforming were jointly considered for single cell multi-antenna systems in [4] while the EE maximization problem was addressed in [5].

In the other category, phase shifts, optimized based on S-CSI, do not need to change at every coherence interval but every several intervals. Compared to I-CSI, this method offers lower feedback overhead due to less frequent phase optimization, especially when the numbers of BS antennas and IRS elements are large. Moreover, the feedback overhead reduction results in less power consumption by the IRS controller and release of the capacity requirement for the IRS control link. Furthermore, a significant computational complexity reduction is achieved at the BS, which needs to update the phase shift matrix at a much larger time scale compared to phase shift matrix update in I-CSI schemes that occurs at a much lower frequency. As a result, I-CSI designs could be applied to scenarios with low mobility or fixed location while, in the cases of short coherence intervals such as high mobility environments, phase shifts tuning based on S-CSI tends to be an advantageous solution. On the ground of these benefits, S-CSI design has attracted a lot of interest [7], [8], [14]–[22]. For instance, in [8], optimal IRS phase shift design was presented based on maximization of the ergodic rate. Further, in [14], the authors studied IRS-assisted multi-pair systems and applied a genetic algorithm during the phase shifts optimization. Taking into account for random matrix theory, the authors in [18] studied the minimum signal-to-interference-plus-noise ratio (SINR) maximization. In [19], a novel two-timescale beamforming optimization scheme was proposed, where first S-CSI was used to optimize the passive beamforming, and then I-CSI was used to design the active beamforming. In [20], both IRS and additive transceiver hardware impairments (HIs) were studied. Note that IRS HIs appear due to the lack of infinite precision at the IRS phase shifts [23]. In [22], two insightful scenarios,

A. Papazafeiropoulos is with the Communications and Intelligent Systems Research Group, University of Hertfordshire, Hatfield AL10 9AB, U. K., and with SnT at the University of Luxembourg, Luxembourg. P. Kourtessis is with the Communications and Intelligent Systems Research Group, University of Hertfordshire, Hatfield AL10 9AB, U. K. K. Ntontin and S. Chatzinotas are with the SnT at the University of Luxembourg, Luxembourg. E-mails: tapapazaf@gmail.com, p.kourtessis@herts.ac.uk, {kostantinos.ntontin,symeon.chatzinotas}@uni.lu.

namely, a finite number of large IRS and a large number of finite size IRS were considered towards deriving the coverage probability and it was shown that the latter implementation method is more advantageous.

Although most works have relied on a time division duplex (TDD) design, recent results regarding IRS showed that the phase shifts depend on the angle of the impinging electromagnetic waves, which renders the assumption of channel reciprocity questionable [24]–[26]. In particular, in [24], it was shown that channel reciprocity holds only when the incident angles are small while, in [26], several IRS-assisted approaches such as the time-varying control and the structural asymmetries, realizing non-reciprocal channels, were introduced. Hence, the study of the design and the performance of IRS-assisted systems, operating in frequency division duplexing (FDD), is required and is the motivation for the work presented in this paper.<sup>1</sup> Note that many current wireless networks are based on FDD systems, which contributes further to our interest in such systems. Moreover, FDD operates more effectively in the cases of symmetric traffic and delay-sensitive applications [27]. However, FDD cannot be easily deployed in 5G and beyond systems such as mMIMO because the multi-antenna channel acquisition makes their implementation prohibitive [28]. In particular, FDD includes downlink training, which together with the CSI feedback, present a significant bottleneck as the number of BS antennas increases. Fortunately, the use of the joint spatial division and multiplexing (JSDM) approach allows achieving SE gains similar to mMIMO [29] while it has not been applied yet in IRS systems.

The concept of JSDM relies on the clustering of UEs into groups having approximately the same covariance matrix and the application of dual structured precoding to reduce the feedback overhead efficiently. The structure of the precoder consists of a prebeamforming matrix minimizing the inter-group interference and a linear precoding that depend on large and small-scale fading, respectively. Specifically, the large-scale fading is slowly varying because it depends on the channel second-order statistics such as the channel covariance matrices, which can be acquired accurately with a low feedback overhead while the small-scale fading involves instantaneous CSI. It is worthwhile to mention that the application of JSDM is particularly justified in IRS-assisted systems since a large number of IRS elements results in large dimensional channel matrices that incur further high feedback overhead in FDD IRS systems.

#### A. Contributions

The main contributions are summarized as follows.

- Contrary to existing works such as [14]–[22], which relied on time division duplexing (TDD) transmission for IRS-assisted systems, we consider FDD. To the best of our knowledge, there are no prior works accounting for FDD in IRS and studying its performance except of [30]–[32]. Especially, [30] proposed a cascaded codebook and an adaptive bit partitioning strategy, [31] considered a two-way passive beamforming design, and [32] proposed a

<sup>1</sup>This observation has been also recognised in [25] as one of the most interesting topics for future research.

dimension reduced channel feedback scheme to reduce the channel feedback overhead in RIS assisted systems.

- This is the first work applying the concept of JSDM [29] on IRS-assisted systems to reduce the feedback overhead being increasingly significant with the numbers of antennas and IRS elements. In particular, not only do we study the performance but we also provide system guidelines concerning the choice of JSDM parameters. Moreover, we have generalized [29] since we do not assume equal power among the UEs but perform power optimization.
- Although many previous works were based on I-CSI knowledge and independent Rayleigh fading such as [4]–[6], we move towards consideration of realistic conditions and we hinge on S-CSI (more advantageous) and correlated Rayleigh fading, which is unavoidable as shown in [33]. Remarkably, based on the use of S-CSI and the effective channel, consisting of the cascaded and direct channels, we achieve to obtain the achievable SE of IRS-assisted systems similar to conventional systems. Notably, works considering S-CSI and correlated Rayleigh fading exist (e.g., [18]–[20], [22]) but none of them has accounted for the JSDM approach.
- By employing the per-group processing (PGP) method of the JSDM approach to reduce the feedback overhead [29], we derive the sum SE of the IRS-assisted mMIMO systems based on FDD in closed-form by leveraging results from the deterministic equivalent (DE) analysis [34]–[36]. The importance of the results in terms of DEs is noteworthy since DEs provide an analytical tool resulting in a deterministic expression of the SE based on a convergent system of fixed-point equations while avoiding lengthy Monte Carlo simulations.
- We formulate the optimization problem concerning the sum SE subject to reflecting beamforming matrix (RBM) and total transmit power constraints. Note that the PGP method includes the regularized zero-forcing (RZF) precoding and this is the unique work providing RBM optimization based on S-CSI with such a complex precoder. Other works such as [18] considered a simple linear precoding, e.g., maximum ratio transmission (MRT) precoding. Furthermore, the proposed optimization is based on deterministic expressions dependent only on large-scale statistics, and thus, can be performed at every several coherence intervals and reduce considerably the signal overhead as required especially in FDD systems.
- We verify the results by Monte-Carlo (MC) simulations and we shed light on the impact of the system parameters on the sum SE such as the signal-to-noise ratio (SNR), the effective covariance rank, the effective channel dimension, and the number of IRS elements. Especially, the effective covariance rank and effective channel dimension play a prominent role by presenting a trade-off between the feedback reduction and the performance.

#### B. Paper Outline

The remainder of this paper is organized as follows. Section II presents the system model of an IRS-assisted massive

Intelligent reflecting surface	IRS
Frequency division duplexing	FDD
Time division duplexing	TDD
Joint spatial division and multiplexing	JSDM
Statistical CSI	S-CSI
Instantaneous channel state information	I-CSI
Joint spatial division and multiplexing	JSDM
Per-group processing	PGP
Joint-group processing	JGP
Reflecting beamforming matrix	RBM
Regularized zero-forcing	RZF

Table I: Table of main abbreviations

MIMO system with correlated Rayleigh fading operating in FDD and employing the JSDM. Section III provides the asymptotic performance analysis. Section IV presents the sum SE maximization with respect to the IRS RBM and the transmit power. The numerical results are placed in Section V, and Section VI concludes the paper.

### C. Notation

Vectors and matrices are denoted by boldface lower and upper case symbols, respectively. The notations  $(\cdot)^T$ ,  $(\cdot)^H$ , and  $\text{tr}(\cdot)$  represent the transpose, Hermitian transpose, and trace operators, respectively. The expectation operator is denoted by  $\mathbb{E}[\cdot]$  while  $\text{diag}(\mathbf{a})$  represents an  $n \times n$  diagonal matrix with diagonal elements being the elements of vector  $\mathbf{a}$ . In the case of a matrix  $\mathbf{A}$ ,  $\text{diag}(\mathbf{A})$  denotes a vector with elements the diagonal elements of  $\mathbf{A}$ . Also, we denote  $[x]^+ = \max(0, x)$  and the notations  $\xrightarrow[M \rightarrow \infty]{\text{a.s.}}$  and  $a_n \asymp b_n$  with  $a_n$  and  $b_n$  being two infinite sequences denote almost sure convergence as  $M \rightarrow \infty$ . The notations  $\text{Span}(\mathbf{X})$  and  $\text{Span}^\perp(\mathbf{X})$  denote the column space of  $\mathbf{X}$  and its orthogonal complement, respectively. Also, the notation  $\mathbf{X}_{[k]}$  denotes the matrix obtained upon removing the  $k$ th column of  $\mathbf{X}$ , while  $\mathbf{b} \sim \mathcal{CN}(\mathbf{0}, \Sigma)$  represents a circularly symmetric complex Gaussian vector with zero mean and covariance matrix  $\Sigma$ . Finally, Table I provides the main abbreviations for the sake of convenience.

## II. SYSTEM MODEL

We consider the downlink of an IRS-aided multi-user (MU) multiple-input, single-output (MISO) system with one BS deployed with  $M$  antennas that serve  $K$  single-antenna UEs as shown in Fig. 1. Additionally to possibly existing direct links between the BS and the UEs, one IRS, implemented with  $N$  nearly passive reflecting elements introducing phase shifts onto the impinging signal waves, assists the communication. The size of each IRS element is  $d_H \times d_V$  with  $d_V$  and  $d_H$  expressing its vertical height and its horizontal width, respectively. The management of the phase shifts takes place by a controller that communicates with the BS through a perfect backhaul link.

As a reasonable design, we assume that the IRS has a line-of-sight (LoS) with the BS. This assumption can be justified by assuming that both the BS and IRS are deployed at high altitude and their locations are fixed. Moreover, we assume that the UEs are spatially clustered into  $G$  groups each having  $K_g$  UEs with the same spatial covariance matrix. The UEs in the same group are nearly co-located while different groups are

well-separated. Note that this is a practical assumption since UEs tend to confine in small regions such as buildings. Hence, there exist  $K = \sum_{g=1}^G K_g$  UEs in total, where the index  $g$  refers to UEs in group  $g$ .

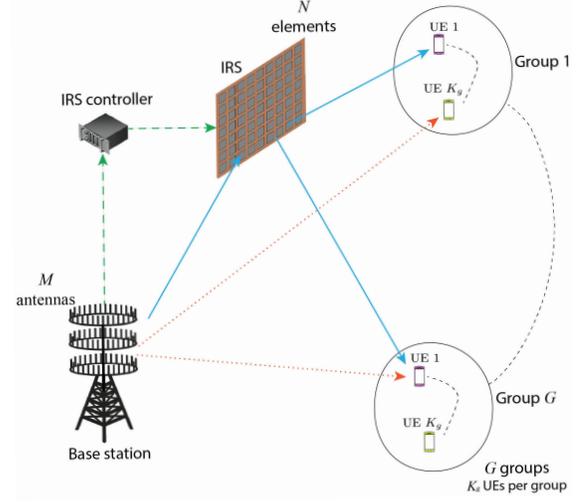


Figure 1: An IRS-assisted downlink MU-MISO communication system with  $M$  BS antennas,  $N$  IRS elements, and  $K$  UEs clustered into  $G$  groups.

### A. Channel and Signal Models

Assuming a flat-fading channel, the received  $K_g \times 1$  complex baseband signal by the UEs at the  $g$ th group is given by

$$\mathbf{y}_g = (\mathbf{H}_{d,g}^H + \mathbf{H}_{2,g}^H \Phi \mathbf{H}_1) \mathbf{x} + \mathbf{w}_g, \quad (1)$$

where  $\mathbf{H}_{d,g} = [\mathbf{h}_{d,g_1}, \dots, \mathbf{h}_{d,g_{K_g}}] \in \mathbb{C}^{M \times K_g}$  is the channel matrix describing the direct channels between the BS and the UEs in group  $g$  with  $\mathbf{h}_{d,g_i} \in \mathbb{C}^{M \times 1}$ ,  $i = 1, \dots, K_g$  being the individual UEs channels. Note that the index  $g_i$  denotes UE  $i$  in group  $g$ . The subscripts 1 and 2 correspond to the BS-IRS and IRS-UEs links, respectively. Hence,  $\mathbf{H}_{2,g} = [\mathbf{h}_{2,g_1}, \dots, \mathbf{h}_{2,g_{K_g}}] \in \mathbb{C}^{N \times K_g}$  describes the channel between the IRS and the  $g$ th group with  $\mathbf{h}_{2,g_i} \in \mathbb{C}^{N \times 1}$ ,  $i = 1, \dots, K_g$ . Also,  $\mathbf{H}_1 = [\mathbf{h}_{1,1}, \dots, \mathbf{h}_{1,N}] \in \mathbb{C}^{M \times N}$  describes the LoS channel between the BS and the IRS with  $\mathbf{h}_{1,i} \in \mathbb{C}^{M \times 1}$ ,  $i = 1, \dots, N$ . Moreover,  $\mathbf{x} = \sum_{g=1}^G \mathbf{V}_g \mathbf{d}_g$ , satisfying the power constraint  $\mathbb{E}[\|\mathbf{x}\|^2] = \text{tr}(\mathbf{P} \mathbf{V}^H \mathbf{V}) \leq P_{\max}$ , is the  $M \times 1$  linearly precoded transmit signal vector, which means that  $\rho = P_{\max}/\sigma^2$  expresses the transmit SNR. Also,  $\mathbf{d}_g = \mathbf{P}_g^{1/2} \mathbf{s}_g \in \mathbb{C}^{K_g \times 1}$ , where  $\mathbf{s}_g \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{K_g}) \in \mathbb{C}^{K_g \times 1}$  expresses the vector of data symbols in group  $g$ . Note that  $\mathbf{V} = [\mathbf{V}_1, \dots, \mathbf{V}_G] \in \mathbb{C}^{M \times K}$  and  $\mathbf{P} = \text{diag}(\mathbf{P}_1, \dots, \mathbf{P}_G)$  with  $\mathbf{P}_g = \text{diag}(p_{g_1}, \dots, p_{g_{K_g}})$  for  $g = 1, \dots, G$ , where  $\mathbf{V}_g \in \mathbb{C}^{M \times K_g}$  is the linear precoding matrix for group  $g$  and  $P_g = \sum_{i=1}^{K_g} p_{g_i}$  is the total available transmit power for group  $g$  with  $p_{g_k} \geq 0$  expressing the signal power of UE  $k$ . The vector  $\mathbf{w}_g \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_{K_g})$  expresses the additive white Gaussian noise (AWGN) at the BS. Furthermore,  $\Phi = \text{diag}(\alpha_1 e^{j\phi_1}, \dots, \alpha_N e^{j\phi_N}) \in \mathbb{C}^{N \times N}$  is the diagonal RBM, which expresses the response of the  $N$  elements with

$\phi_n \in [0, 2\pi]$  and  $\alpha_n \in [0, 1]$  describing the phase and amplitude coefficient for element  $n$ , respectively. Henceforth, we make the common assumption of maximum reflection ( $\alpha_n = 1 \forall n$ ) based on recent advances in lossless metasurfaces [37].

Despite the majority of existing works, e.g., [4], [5], which assumed independent Rayleigh model, we consider the practical effect of spatial correlation, which is unavoidable in realistic IRS-assisted systems [33].<sup>2</sup> Also, according to the JSDM method, despite that the individual effects are different across different UEs, the covariances matrices, describing the aggregate result of path-loss and correlation, have been assumed equal across different UEs of the same group since they are clustered accordingly. Thus,  $\mathbf{h}_{d,g_k}$  and  $\mathbf{h}_{2,g_k}$ , concerning UE  $k$  in group  $g$ , are expressed as

$$\mathbf{h}_{d,g_k} = \mathbf{R}_{BS,g}^{1/2} \mathbf{z}_{d,g_k}, \quad \mathbf{h}_{2,g_k} = \mathbf{R}_{IRS,g}^{1/2} \mathbf{z}_{2,g_k}, \quad (2)$$

where  $\mathbf{R}_{BS,g} = \beta_{d,g_k} \mathbf{R}_{BS,g_k} \in \mathbb{C}^{M \times M}$  and  $\mathbf{R}_{IRS,g} = \beta_{2,g_k} \mathbf{R}_{IRS,g_k} \in \mathbb{C}^{N \times N}$  represent the deterministic Hermitian-symmetric positive semi-definite aggregate covariance matrices at the BS and the IRS respectively corresponding to group  $g$  with  $\text{tr}(\mathbf{R}_{BS,g}) = M$  and  $\text{tr}(\mathbf{R}_{IRS,g}) = N$ . Note that  $\beta_{d,g_k}$ ,  $\mathbf{R}_{BS,g_k}$  and  $\beta_{2,g_k}$ ,  $\mathbf{R}_{IRS,g_k}$  are the path-losses, correlations of the BS-UE  $k$  and IRS-UE  $k$  links in group  $g$ , respectively. This channel modeling is very versatile since it can account for both correlation ( $\mathbf{R}_{BS,g_k}$ ,  $\mathbf{R}_{IRS,g_k}$ ) and path loss ( $\beta_{d,g_k}$ ,  $\beta_{2,g_k}$ ) simultaneously or either of the two. Certain estimation methods (see e.g., [38], [39]) allow to obtain the correlation matrices and the path-losses, which can, thus, be assumed known by the network. Moreover,  $\mathbf{z}_{d,g_k} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$  and  $\mathbf{z}_{2,g_k} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$  express the corresponding fast-fading vectors.

The channel matrix  $\mathbf{H}_1$  can be expressed as

$$[\mathbf{H}_1]_{m,n} = \sqrt{\beta_1} \exp\left(j \frac{2\pi}{\lambda} (m-1) d_{BS} \sin \theta_{1,n} \sin \phi_{1,n} + (n-1) d_{IRS} \sin \theta_{2,m} \sin \phi_{2,m}\right). \quad (3)$$

Here,  $\lambda$  is the carrier wavelength,  $\beta_1$  is the path-loss between the BS and the IRS. Also,  $d_{BS}$  and  $d_{IRS}$  are the inter-antenna separation at the BS and inter-element separation at the IRS, respectively [40]. The elevation and azimuth LoS angles of departure (AoD) at the BS with respect to IRS element  $n$  are described by  $\theta_{1,n}$  and  $\varphi_{1,n}$ . The elevation and azimuth LoS angles of arrival (AoA) at the IRS are described by  $\theta_{2,n}$  and  $\varphi_{2,n}$ . The phases corresponding to the response of the surface elements adjusted in a way so that  $\text{rank}(\mathbf{H}_{d,g}^H + \mathbf{H}_{2,g}^H \Phi \mathbf{H}_1) \geq K$ , which is the necessary condition for supporting multi-stream transmission to the  $K$  users of a group. This approach of rank improvement through an intelligent surface has been proposed and validated in [41] for a scenarios with a 2-antenna BS and a 2-antenna UE. In our case, we generalize this approach to a system with an arbitrary number of UEs.

Given the RBM, the overall channel vector between the BS and UE  $k$  in group  $g$  is written according to (1) as  $\mathbf{h}_{g_k} = \mathbf{h}_{d,g_k} + \mathbf{H}_1 \Phi \mathbf{h}_{2,g_k}$ , which is distributed as  $\mathbf{h}_{g_k} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_g)$  with  $\mathbf{R}_g = \mathbf{R}_{BS,g} + \mathbf{H}_1 \Phi \mathbf{R}_{IRS,g} \Phi^H \mathbf{H}_1^H$ .

<sup>2</sup>Contrary to most works on IRS that model correlation by using models for conventional antenna arrays, we adopt the recently presented correlation model by the authors in [33], being more suitable for IRS.

It is worthwhile to mention that according to the FDD design with training for channel estimation in both the uplink and the downlink, the following procedure takes place, which consists of two frames. In the uplink frame, each UE first transmits orthogonal pilot symbols to the BS, where these are used to estimate the underlying channels. Following this, all the UEs simultaneously transmit their data to the BS. The data is detected at the BS using the CSI acquired during the pilot transmission phase. During the downlink training frame, the BS transmits orthogonal pilot symbols from its antennas which are used by the UEs to estimate the corresponding downlink channel. These channel estimates are then fed back to the BS by the UEs, and are used by the latter to appropriately precode the data symbols transmitted simultaneously to all the UEs during the data transmission phase.

In terms of implementation, two approaches could be used that present a trade-off between performance and cost. The first approach assumes that makes no compromise on the performance but has a higher cost. It requires  $2N$  IRS elements, where  $N$  elements are used for the uplink transmission and  $N$  elements are used for the downlink transmission. The downlink/uplink group of elements operates at the corresponding uplink/downlink frequency. Note that this approach requires all the corresponding hardware to be doubled.<sup>3,4</sup> Without significant performance loss, the whole procedure can be facilitated by the use of duplexers that allow the simultaneous processing of two different frequencies in common FDD systems and also enable the separate phase shifts optimizations that happen during the uplink and downlink transmissions. In particular, the use of radio-frequency (RF)-micro-electromechanical system (MEMS) single-pole double-throw (SPDT) switches to realize phased array transceivers has been well-developed in antenna engineering with applications from microwave frequency to mm-wave frequency ranges. As shown Fig. 2 in [45], the uplink/downlink links, can be realized

<sup>3</sup>It is true that the gap between the downlink and uplink in 4G/5G systems is not big. It is even lower than 0.3 GHz. For instance, in LTE Band 1 the uplink is defined in the range 1.92 – 1.98 GHz, whereas the downlink is defined in the range 2.11 – 2.17 GHz [42]. Hence the gap between the uplink and downlink is only 0.13 GHz. However, please have in mind that normally the unit cells of the metasurface that a RIS consists of elements having a narrowband frequency response, unless special designs are manufactured that create a wideband response. For instance, in Section II of [43] it is stated that: “For instance, reflective metasurfaces composed of microstrip patch resonators typically exhibit a fractional bandwidth of less than 5%.” Even if we assume the worst-case scenario of 5% fractional bandwidth, this would mean that the frequency response in a downlink FR2 LTE case with carrier frequency 2.14 GHz is limited in the range  $0.05 * 2.14 \text{ GHz} = 0.107 \text{ GHz}$ , which is smaller than the aforementioned 0.13 GHz gap between the uplink and downlink in LTE. Hence, with the proper design of the intelligent surface unit cells, the frequency response of the unit cells dedicated for the two bands, uplink and downlink, can be such that there is no interference from one band to the other.

<sup>4</sup>The proposed design concerns 5G and beyond networks, where possibly the frequency interval between the uplink and downlink will be greater than 0.3 GHz, which is in 4G/5G systems. Hence, there will be no coupling or interference between the uplink and downlink. Despite this, even for smaller intervals, there is no issue because of the following reasons. Doubling the hardware means that there will be separated beamforming networks/control circuits for the uplink RIS and downlink RIS; thus, the uplink RIS and downlink RIS can operate independently. We would like to mention that similar techniques have been well-developed for the phased array design. For example, in [44], an architecture of using a separate transmitter and receiver array operating at 94 GHz is presented. Second, RF filters with high selectivity are used to suppress any unwanted signals from other channels belonging to other frequencies.

on a single monolithic microwave integrated circuit (MMIC) with a shared vector modulator (where the phase shifter is located). This well-developed technology is compatible with IRS design and can be applied to IRS to support simultaneous FDD channels. Notably, the switching time of modern SPDT is in the scale of picoseconds. Thus, the uplink and downlink signals can be recovered as far as the speed of the switch can provide enough samples to recover the signals, which is possible according to [45]. Moreover, another technique is through the use of a microwave circulator [46], where the transmit signal will be delivered to the antenna while the received signal from the same antenna will only be delivered to the receiver port.

### B. Advantages/disadvantages of FDD in IRS-assisted systems

FDD has several advantages over TDD that become more pronounced in IRS-assisted systems. For example, FDD design is accompanied with lower latency since communication takes place simultaneously while in TDD there is a switching between transmission and reception. Also, in IRS-assisted systems, where the distance between the transmitter and the receiver is increased, the guard period in TDD systems, being proportional to the distance, increases and affects the performance, while FDD does not appear a problem with large distances. Furthermore, the medium access control (MAC) layer of TDD systems is more complex because it requires accurate time synchronization between uplink and downlink, while FDD systems do not require any uplink/downlink switching mechanism at timescale. However, FDD systems come with higher implementation cost demands such as the duplexers that are required to make the operation in different frequencies possible. Moreover, given that in practical systems, most network volume is consumed in the downlink, in TDD, it is possible to balance the traffic between uplink and downlink by utilising more time slots for downlink. In addition, FDD, using two frequencies instead of one used in TDD, allocates more spectrum.

### C. PGP approach

The JSMD approach was presented in terms of two methods, the joint-group processing (JGP) and PGP [29]. JGP can be applied in the case that the estimation and feedback of the whole matrix from all UEs in all groups are affordable. Normally, the JGP method is accompanied by prohibitive overhead for channel estimation and feedback while PGP is more advantageous due to lower overhead. For this reason, the PGP method, summarized below, is suggested as a more practical approach and is employed in this work.

1) *Signal model*: PGP suggests a two-stage precoding matrix for group  $g$ , resulting in feedback overhead and the computational complexity reductions, expressed as

$$\mathbf{V}_g = \mathbf{B}_g \mathbf{F}_g, \quad (4)$$

where  $\mathbf{B}_g \in \mathbb{C}^{M \times b_g}$  is the preprocessing matrix based on the long-term channel statistics with  $b_g \geq K_g$  being an integer parameter to be optimized. Also,  $\mathbf{F}_g \in \mathbb{C}^{b_g \times K_g}$ ,  $g = 1, \dots, G$  is the precoding matrix based on the instantaneous channel (short-term CSI) of group  $g$ .

Hence, the received signal in (1) with dual precoding described by (4) can be written as

$$\mathbf{y}_g = \mathbf{H}_g^H \mathbf{B}_g \mathbf{F}_g \mathbf{d}_g + \sum_{i \neq g} \mathbf{H}_g^H \mathbf{B}_i \mathbf{F}_i \mathbf{d}_i + \mathbf{w}_g, \quad (5)$$

where  $\mathbf{H}_g = [\mathbf{h}_{g_1}, \dots, \mathbf{h}_{g_{K_g}}] \in \mathbb{C}^{M \times K_g}$ .

Based on the Karhunen-Loeve representation, the channel vector  $\mathbf{h}_{g_k}$  can be written as

$$\mathbf{h}_{g_k} = \mathbf{U}_g \boldsymbol{\Lambda}_g^{\frac{1}{2}} \mathbf{z}_{g_k}, \quad (6)$$

where  $\mathbf{U}_g \in \mathbb{C}^{M \times r_g}$  is a tall unitary matrix including the eigenvectors of  $\mathbf{R}_g$  that correspond to the non-zero eigenvalues,  $\boldsymbol{\Lambda}_g$  is an  $r_g \times r_g$  diagonal matrix with elements the non-zero eigenvalues of  $\mathbf{R}_g$ , and  $\mathbf{z}_{g_k} \in \mathbb{C}^{r_g \times 1} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{r_g})$ . Generally,  $M \gg r_g$ . Thus, we have  $\mathbf{R}_g = \mathbf{U}_g \boldsymbol{\Lambda}_g \mathbf{U}_g^H$ .

2) *Imperfect CSI*: Certain reasons such as quantization can result in imperfect feedback of the CSI to the BS (see Rem. 1 below), which is accompanied with significant overhead. Especially, the imperfect CSI at the BS from group  $g$  can be expressed as

$$\hat{\mathbf{Z}}_g = \sqrt{1 - \tau_g^2} \mathbf{Z}_g + \tau_g \mathbf{E}_g, \quad (7)$$

where  $\mathbf{Z}_g = [\mathbf{z}_{g_1}, \dots, \mathbf{z}_{g_{K_g}}] \in \mathbb{C}^{M \times K_g}$  is the perfect CSI and  $\mathbf{E}_g = [\mathbf{e}_{g_1}, \dots, \mathbf{e}_{g_{K_g}}] \in \mathbb{C}^{M \times K_g}$  with  $\mathbf{e}_{g_k}$  denoting the corresponding error that consists of elements with zero mean and unit variance. The parameter  $\tau_g \in [0, 1]$  describes the accuracy of the available CSI in group  $g$ . Hence, if  $\tau_g = 0$ , we obtain perfect CSI while  $\tau_g = 1$  implies no CSI knowledge. Without loss of generality, we assume that the levels of CSI accuracy for all groups are the same, i.e.,  $\tau_1 = \dots = \tau_G = \tau$ . From (7),  $\hat{\mathbf{H}}_g$  can be defined as the imperfect CSI knowledge of  $\mathbf{H}_g$  by using  $\hat{\mathbf{Z}}_g$ . Note that the imperfect CSI  $\hat{\mathbf{H}}_g$ , given in terms of  $\mathbf{H}_g$  is based on a given RBM. The assumption of a fixed RBM during one coherence interval is justified based on the fact that RBM is expected to change according to large-scale statistics expressed by  $\mathbf{R}_g$ , which changes at every several coherence intervals.

*Remark 1*: Having assumed that both the BS and UEs are aware of large-scale statistics, the model in (7) can describe the scenario, where the  $k$ th UE in the group  $g$  quantizes  $\mathbf{h}_{g_k}$  with the help of a random codebook and sends back the codeword index to the BS [28]. We provide an example of the advantage of using this method in Sec. V.

3) *Design of pre-beamforming*: Using block diagonalization (BD) preprocessing to cancel out the inter-group interference, we design the pre-beamforming matrix as

$$\mathbf{H}_g^H \mathbf{B}_i \approx 0, \quad \text{for } i \neq g, \quad (8)$$

where the index  $i$  corresponds to group  $i$ . In the case of exact BD, (5) becomes

$$\mathbf{y}_g = \mathbf{H}_g^H \mathbf{B}_g \mathbf{F}_g \mathbf{d}_g + \mathbf{w}_g. \quad (9)$$

Exact BD can be achieved, if  $\text{Span}(\mathbf{U}_g) \not\subseteq \text{Span}(\{\mathbf{U}_i : i \neq g\})$  while the necessary multiplexing gain can be achieved if  $\dim(\text{Span}(\mathbf{U}_g) \cap \text{Span}^\perp(\{\mathbf{U}_i : i \neq g\})) \geq K_g$  for all groups  $g = 1, \dots, G$ . In other words, for given sets of  $\{b_g\}$

and  $\{K_g\}$  satisfying  $K_g \leq b_g \leq r_g$ ,  $\mathbf{B}_g$  can be designed if  $\mathbf{U}_i^H \mathbf{B}_g = 0$  for all  $i \neq g$  and  $\text{rank}(\mathbf{U}_g^H \mathbf{B}_g) \geq K_g$ . If the dimension of  $\text{Span} \perp(\{\mathbf{U}_i : i \neq g\})$  is less than  $K_g$ , the rank condition is not fulfilled. In such case, there are two options: i) reduction of  $K_g$ , or ii) resort to approximate BD.

Herein, we choose approximate BD to design  $\{\mathbf{B}_g\}$ . For this reason, we choose  $r_g^*$  dominant eigenvalues of  $\mathbf{R}_g$ . On this ground, we write  $\mathbf{U}_g = [\mathbf{U}_g^*, \mathbf{U}_g']$ , where  $\mathbf{U}_g^*$  is the  $M \times r_g^*$  matrix consisted of the dominant eigenvectors and  $\mathbf{U}_g'$  is the  $M \times (r_g - r_g^*)$  matrix including the eigenvectors corresponding to the weakest eigenvalues. To achieve approximate BD, we have to choose the dominant eigenmodes  $r_g^*$  for each group  $g$  subject to the condition  $\text{Span}(\mathbf{U}_g^*) \not\subseteq \text{Span}(\{\mathbf{U}_i^* : i \neq g\})$  for all groups  $g = 1, \dots, G$ . Notably, the design of  $\{\mathbf{B}_g\}$  has to take into account a feasible choice of the parameters  $\{r_g^*\}$ ,  $\{b_g\}$ , and  $\{K_g\}$  that should be optimized to maximize the SE under a given system setup. The matrix  $\mathbf{B}_g$  is constructed as in [29]. Herein, we briefly summarize the main steps. Hence, first, we define the matrix

$$\mathbf{U}_{-g} = [\mathbf{U}_1^*, \dots, \mathbf{U}_{g-1}^*, \mathbf{U}_{g+1}^*, \dots, \mathbf{U}_G^*], \quad (10)$$

whose size is  $M \times \sum_{i \neq g} r_i^*$  and its rank is  $\sum_{i \neq g} r_i^*$ . Based on SVD, let a set of left eigenvectors of  $\mathbf{U}_{-g}$  denoted by  $[\mathbf{E}_g^{(1)}, \mathbf{E}_g^{(0)}]$ , where  $\mathbf{E}_g^{(1)}$  expresses the left singular vectors corresponding to the  $\sum_{i \neq g} r_i^*$  dominant singular values  $\Lambda_g^{(1)}$  while  $\mathbf{E}_g^{(0)}$  corresponds to the left singular vectors with  $M - \sum_{i \neq g} r_i^*$  non-dominant singular values  $\Lambda_g^{(0)}$ . Now, we define the matrix  $\tilde{\mathbf{H}}_g = (\mathbf{E}_g^{(0)})^H \mathbf{H}_g$ , which has the property to be orthogonal to the dominant eigenspace spanned by the channel of other groups. The covariance matrix of  $\tilde{\mathbf{H}}_g$  is given by

$$\tilde{\mathbf{R}}_g = (\mathbf{E}_g^{(0)})^H \mathbf{R}_g \mathbf{E}_g^{(0)} \quad (11)$$

$$= (\mathbf{E}_g^{(0)})^H \mathbf{U}_g \Lambda_g \mathbf{U}_g^H \mathbf{E}_g^{(0)} \quad (12)$$

$$= \mathbf{G}_g \mathbf{Q}_g \mathbf{G}_g^H, \quad (13)$$

where (12) is based on (6) while (13) is the SVD of  $\tilde{\mathbf{R}}_g$ , i.e.,  $\mathbf{G}_g$  describes the eigenvectors of  $\tilde{\mathbf{R}}_g$ . The pre-beamforming matrix  $\mathbf{B}_g$  is obtained by setting  $\mathbf{G}_g = [\mathbf{G}_g^{(1)}, \mathbf{G}_g^{(0)}]$ , where  $\mathbf{G}_g^{(1)}$  corresponds to the dominant  $b_g$  eigenmodes of  $\tilde{\mathbf{R}}_g$ . In particular, we have that

$$\mathbf{B}_g = \mathbf{E}_g^{(0)} \mathbf{G}_g^{(1)}, \quad (14)$$

which means that  $\mathbf{B}_g$  coincides with the  $b_g$  dominant eigenmodes of  $\tilde{\mathbf{R}}_g$  while it is orthogonal to the dominant  $r_i^*$  eigenmodes of groups  $i \neq g$ . Gathering all constraints, the condition  $K_g \leq b_g \leq \text{rank}(\tilde{\mathbf{R}}_g)$ , where  $\text{rank}(\tilde{\mathbf{R}}_g) = \min\{M - \sum_{i \neq g} r_i^*, r_g\}$  should be satisfied.

### III. ASYMPTOTIC PERFORMANCE ANALYSIS WITH DUAL PRECODING

In this section, we first provide the analysis towards the derivation of the downlink achievable SE of IRS-assisted systems based on FDD while employing the JSDM method in the case of the PGP approach.

The derivation of the DE SINR requires the following assumptions concerning the correlation matrices and the power allocation matrix as in [34].

*Assumption 1:* All correlation matrices, i.e.,  $\mathbf{R}_{\text{IRS},g}$  and  $\mathbf{R}_{\text{BS},g}$ ,  $g = 1, \dots, G$  have uniformly bounded spectral norm on  $N$  and  $M$ , respectively, i.e.,

$$\limsup_{N,K \rightarrow \infty} \sup_{1 \leq k \leq K} \|\mathbf{R}_{\text{IRS},g}\| < \infty, \quad \limsup_{M,K \rightarrow \infty} \sup_{1 \leq k \leq K} \|\mathbf{R}_{\text{BS},g}\| < \infty \quad (15)$$

*Assumption 2:* The maximum power of group  $g$ , i.e.,  $\max(p_{g1}, \dots, p_{gK_g})$  is of the order  $\mathcal{O}(1/K_g)$ , being equivalent to

$$\|\mathbf{P}_g\| = \mathcal{O}(1/K_g). \quad (16)$$

According to (5), the effective channel of group  $g$  is  $\bar{\mathbf{H}}_g = \mathbf{B}_g^H \mathbf{H}_g$  with covariance matrix for UE  $g_k$  being  $\bar{\mathbf{R}}_g = \mathbf{B}_g^H \mathbf{R}_g \mathbf{B}_g$ . For the sake of exposition, we assume the same number of dominant eigenvalues per group, the same number of UEs per group, and the same dimension of the pre-beamforming matrix per group, i.e.,  $r_1^* = \dots = r_G^* = \bar{r}$ ,  $K_1 = \dots = K_g = \bar{K}$ , and  $b_1 = \dots = b_g = \bar{b}$ . The extension to the general case is immediate and straightforward. Hence, the dimensions of  $\bar{\mathbf{H}}_g$  are  $\bar{b} \times \bar{K}$ . The precoding matrix for group  $g$  is designed in terms of the instantaneous imperfect CSI to cancel out the intra-group interference on that group as

$$\mathbf{F}_g = \sqrt{\lambda_g} \Sigma_g \hat{\mathbf{H}}_g, \quad (17)$$

where  $\Sigma_g = (\hat{\mathbf{H}}_g \hat{\mathbf{H}}_g^H + \bar{b} \alpha \mathbf{I}_{\bar{b}})^{-1}$  with  $\hat{\mathbf{H}}_g = \mathbf{B}_g^H \hat{\mathbf{H}}_g$  being the effective channel estimate at the BS while  $\hat{\mathbf{H}}_g$  is the imperfect CSI knowledge of  $\mathbf{H}_g$ . Also  $\alpha$ , selected as  $\alpha = \frac{M}{b P_{\text{max}}}$  to be equivalent with the RZF linear filter [47], is a regularization factor that makes expressions converge to a constant, and  $\lambda_g$  is the normalization parameter that fulfils the power constraint for group  $g$ . In particular, the normalization parameter becomes

$$\lambda_g = \frac{P_{\text{max}}}{\text{tr}(\mathbf{P} \hat{\mathbf{H}}_g^H \Sigma_g \mathbf{B}_g^H \mathbf{B}_g \Sigma_g \hat{\mathbf{H}}_g)} \quad (18)$$

$$= \frac{P_{\text{max}}}{\text{tr}(\mathbf{P}_g \hat{\mathbf{H}}_g^H \Sigma_g^2 \hat{\mathbf{H}}_g)} \quad (19)$$

where (18) is obtained because  $\mathbf{B}_g^H \mathbf{B}_g = \mathbf{I}_{\bar{b}}$  as the product of two tall matrices. Hence, based on (5) and (19), the SINR of the  $k$ th UE in group  $g$  with perfect receiver CSI is given by

$$\gamma_{gk, \text{PGP}} = \frac{\text{DS}_{gk}}{\text{SGI}_{gk} + \text{IGI}_{gk} + \sigma^2}, \quad (20)$$

where

$$\text{DS}_{gk} = p_{gk} \lambda_g |\mathbf{h}_{gk}^H \mathbf{B}_g \Sigma_g \hat{\mathbf{h}}_{gk}|^2, \quad (21)$$

$$\text{SGI}_{gk} = \lambda_g \sum_{j \neq k} p_{gj} |\mathbf{h}_{gk}^H \mathbf{B}_g \Sigma_g \hat{\mathbf{h}}_{gj}|^2, \quad (22)$$

$$\text{IGI}_{gk} = \sum_{l \neq g} \sum_j \lambda_l p_{lj} |\mathbf{h}_{gk}^H \mathbf{B}_l \Sigma_l \hat{\mathbf{h}}_{lj}|^2 \quad (23)$$

with the numerator expressing the desired signal power while the first and second terms in the denominator express the self-group and inter-group interferences, respectively.

Under the Assumptions 1-2 and by assuming that  $M \rightarrow \infty$  while  $\bar{K}$ ,  $\bar{r}$ , and  $\bar{b}$  also go to infinity by keeping their ratio with  $M$  fixed, the DE SINR  $\bar{\gamma}_{gk,\text{PGP}}$  fulfils

$$\gamma_{gk,\text{PGP}} - \bar{\gamma}_{gk,\text{PGP}} \xrightarrow[M \rightarrow \infty]{\text{a.s.}} 0. \quad (24)$$

*Theorem 1:* The DE of the downlink SINR of UE  $g_k$  with PGP in IRS-assisted mMIMO systems with FDD, accounting for imperfect CSI, is given by

$$\bar{\gamma}_{gk,\text{PGP}} = \frac{\bar{S}_{gk}}{\bar{I}_{gk}} \quad (25)$$

where  $\bar{S}_{gk} = p_{gk} (1 - \tau^2) \bar{\delta}_g^2$  and  $\bar{I}_{gk} = \bar{\mathbf{Y}}_{gg} \left( 1 - \tau^2 \left( 1 - (1 + \bar{\delta}_g^2)^2 \right) \right) + \left( 1 + \sum_{l \neq g} \bar{\lambda}_l \bar{\mathbf{Y}}_{gl} \right) \frac{(1 + \bar{\delta}_g)^2}{\bar{\lambda}_g}$  with  $\bar{\lambda}_g = \frac{p_g}{\bar{\Psi}_g}$ . The expressions of  $\bar{\delta}_g$ ,  $\bar{\Psi}_g$ ,  $\bar{\mathbf{Y}}_{gg}$ ,  $\bar{\mathbf{Y}}_{gl}$  are the unique solutions of

$$\bar{\delta}_g = \frac{1}{\bar{b}} \text{tr}(\bar{\mathbf{R}}_g \mathbf{T}_g), \quad \mathbf{T}_g = \left( \frac{\bar{K}}{\bar{b}} \frac{\bar{\mathbf{R}}_g}{1 + \bar{\delta}_g} + \alpha \mathbf{I}_b \right)^{-1}, \quad (26)$$

$$\bar{\Psi}_g = \frac{P_g}{\bar{b}} \frac{\bar{m}_g}{(1 + \bar{\delta}_g)^2}, \quad \bar{\mathbf{Y}}_{gg} = \frac{P_g}{\bar{b}} \left( 1 - \frac{p_{gk}}{P_g} \right) \frac{\bar{m}_{gg}}{(1 + \bar{\delta}_g)^2}, \quad (27)$$

$$\bar{\mathbf{Y}}_{gl} = \frac{1}{\bar{b}} \sum_{k=1}^{\bar{K}} p_{lk} \frac{m_{gl}}{(1 + \bar{\delta}_l)^2}, \quad \bar{m}_g = \frac{\frac{1}{\bar{b}} \text{tr}(\bar{\mathbf{R}}_g \mathbf{T}_g \mathbf{B}_g^H \mathbf{B}_g \mathbf{T}_g)}{1 - \frac{\bar{K}}{\bar{b}} \frac{\text{tr}(\bar{\mathbf{R}}_g \mathbf{T}_g \bar{\mathbf{R}}_g \mathbf{T}_g)}{\bar{b}(1 + \bar{\delta}_g)^2}}, \quad (28)$$

$$\bar{m}_{gg} = \frac{\frac{1}{\bar{b}} \text{tr}(\bar{\mathbf{R}}_g \mathbf{T}_g \bar{\mathbf{R}}_g \mathbf{T}_g)}{1 - \frac{\bar{K}}{\bar{b}} \frac{\text{tr}(\bar{\mathbf{R}}_g \mathbf{T}_g \bar{\mathbf{R}}_g \mathbf{T}_g)}{\bar{b}(1 + \bar{\delta}_g)^2}}, \quad \bar{m}_{gl} = \frac{\frac{1}{\bar{b}} \text{tr}(\bar{\mathbf{R}}_l \mathbf{T}_l \mathbf{B}_l^H \mathbf{R}_g \mathbf{B}_l \mathbf{T}_l)}{1 - \frac{\bar{K}}{\bar{b}} \frac{\text{tr}(\bar{\mathbf{R}}_l \mathbf{T}_l \bar{\mathbf{R}}_l \mathbf{T}_l)}{\bar{b}(1 + \bar{\delta}_l)^2}}. \quad (29)$$

*Proof:* The proof is provided in Appendix A. ■

Notably, Assumption 2 allows the omission of term  $\frac{p_{gk}}{P}$  in  $\bar{\mathbf{Y}}_{gg}$  since the convergence in (24) is still true.

Based on the dominated convergence [48] and the continuous mapping theorem [49], the DE of the sum-rate  $\sum_{g=1}^G \sum_{k=1}^{\bar{K}} \log_2(1 + \gamma_{gk,\text{PGP}})$  is given by

$$\bar{\text{SR}}_{\text{PGP}} = \sum_{g=1}^G \sum_{k=1}^{\bar{K}} \log_2(1 + \bar{\gamma}_{gk,\text{PGP}}), \quad (30)$$

where  $\sum_{g=1}^G \sum_{k=1}^{\bar{K}} \log_2(1 + \gamma_{gk,\text{PGP}}) - \bar{\text{SR}}_{\text{PGP}} \xrightarrow[M \rightarrow \infty]{\text{a.s.}} 0$ .

The nature of the multicast mechanism indicates that the DE sum-rate of group  $g$ , allocated with total power  $P_g$  such that  $\sum_{g=1}^G P_g = P_{\max}$ , is determined by  $\bar{\text{SR}}_{\text{PGP},g} = \sum_{k=1}^{\bar{K}} \bar{\text{SR}}_{\text{PGP},gk}$ , where  $\bar{\text{SR}}_{\text{PGP},gk} = \log_2(1 + \bar{\gamma}_{gk,\text{PGP}})$  is the DE rate of UE  $k$  in group  $g$ .

*Proposition 1:* In IRS-assisted mMIMO systems with PGP, all UEs in group  $g$  are allocated with equal power  $P_g/\bar{K}$  and the DE sum-rate of group  $g$  is  $\bar{\text{SR}}_{\text{PGP},g} = \bar{K} R_g^*$  since all these UEs have equal achievable rate  $R_g^*$ .

*Proof:* The proof is provided in Appendix B. ■

*Remark 2:* The results, presented by Prop. 1, rely on the fact that all UEs in the group  $g$  have the same covariance matrix.

*Corollary 1:* Based on Proposition 1, the DE rate of UE  $k$  in group  $g$  can be written as

$$\bar{\gamma}_{g,\text{PGP}} = \frac{\bar{S}_g}{\bar{I}_g}, \quad (31)$$

where  $\bar{S}_g = \frac{P_g}{\bar{K}} (1 - \tau^2) \bar{\delta}_g^2$  and  $\bar{I}_g = \bar{\mathbf{Y}}_{gg} (1 - \tau^2 (1 - (1 + \bar{\delta}_g^2)^2)) + (1 + \sum_{l \neq g} \bar{\lambda}_l \bar{\mathbf{Y}}_{gl}) \frac{(1 + \bar{\delta}_g)^2}{\bar{\lambda}_g}$  with  $\bar{\mathbf{Y}}_{gg} = \frac{P_g}{\bar{b}} (1 - \frac{1}{\bar{K}}) \frac{\bar{m}_{gg}}{(1 + \bar{\delta}_g)^2}$  and  $\bar{\mathbf{Y}}_{gl} = \frac{1}{\bar{b}} \frac{P_l}{\bar{K}} \frac{m_{gl}}{(1 + \bar{\delta}_l)^2}$  while all other variables, given by Theorem 1, remain the same.

#### IV. SUM SE MAXIMIZATION DESIGN

In this section, we perform optimization of sum SE of the overall IRS-assisted mMIMO system. In particular, the sum-rate maximization problem is expressed as follows

$$(\mathcal{P}2) \quad \max_{\Phi, \mathbf{p}} \bar{\text{SR}}_{\text{PGP}} = \max_{\Phi, \mathbf{p}} \bar{K} \sum_{g=1}^G \log_2(1 + \bar{\gamma}_{g,\text{PGP}}) \quad (32a)$$

$$\text{s.t.} \quad \sum_{g=1}^G P_g \leq P_{\max}, \quad (32b)$$

$$|\tilde{\phi}_n| = 1, \quad n = 1, \dots, N, \quad (32c)$$

where constraint (32b) guarantees that the BS transmit power is kept below the maximum power  $P_{\max}$  and constraint (32c) expresses that each IRS element results in only a phase shift without any amplification of the incoming signal. Note that for the sake of exposition, we have defined the vector  $\mathbf{p} = [P_1, \dots, P_G]^T$  and  $\tilde{\phi}_n = \exp(j\phi_n)$  for all  $n$  corresponding to the elements of  $\Phi$ .

*Corollary 2:* Let  $\mathbf{R}_{\text{IRS},g} = \mathbf{I}_N$ , then  $\mathbf{R}_g$  does not depend on  $\Phi$ . Equivalently,  $\bar{\mathbf{R}}_g$  is independent on the phase shifts, and thus,  $\bar{\gamma}_{g,\text{PGP}}$  cannot be optimized in this case.

The optimization problem ( $\mathcal{P}2$ ) is non-convex and subject to a unit-modulus constraint regarding  $\tilde{\phi}_n$ , which make its solution challenging. To address this, we rely on the common approach in IRS literature, being the application of the alternating optimization (AO) technique, where  $\Phi$  and  $\mathbf{p}$  are going to be solved separately and iteratively. Hence, first, we solve for  $\Phi$  given a fixed  $\mathbf{p}$ . Next, we focus on finding the optimum  $\mathbf{p}$  with  $\Phi$  fixed. The iteration of this process achieves the increase of  $\bar{\text{SR}}_{\text{PGP}}$  at each iteration step until convergence of the objective to its optimum value since it is upper-bounded due to the power constraint (32b).

##### A. IRS Design

So far, we have considered the RBM fixed. However, to exploit the IRS towards the maximization of the sum SE, the RBM has to be optimized. Notably, we observe its presence inside the covariance matrices that appear in the DE expression of the SE with PGP. By assuming infinite resolution phase shifters and under imperfect CSI conditions, the RBM optimization problem for single-cell FDD systems employing the JSMD method is formulated as

$$(\mathcal{P}3) \quad \max_{\Phi} \bar{\text{SR}}_{\text{PGP}} \quad (33)$$

$$\text{s.t.} \quad |\tilde{\phi}_n| = 1, \quad n = 1, \dots, N,$$

where  $\overline{\text{SR}}_{\text{PGP}}$  is given by (32a). The maximization problem ( $\mathcal{P}3$ ) is non-convex with respect to  $\Phi$  while it is subject to a unit-modulus constraint regarding  $\tilde{\phi}_n$ . A local optimal solution to this problem can be obtained by applying the projected gradient ascent algorithm until converging to a stationary point, which means that, at every step, we project the solution onto the closest feasible point satisfying the unit-modulus constraint concerning  $\tilde{\phi}_n$  [18]. Specifically, at step  $l$ , the phases are included in the vectors  $\mathbf{s}_l = [\tilde{\phi}_{l,1}, \dots, \tilde{\phi}_{l,N}]^\top$ . The next step of the iteration towards convergence increases  $\overline{\text{SR}}_{gk,1}$ , and is described by

$$\tilde{\mathbf{s}}_{l+1} = \mathbf{s}_l + \mu \mathbf{q}_l, \quad \mathbf{s}_{l+1} = \exp(j \arg(\tilde{\mathbf{s}}_{l+1})), \quad (34)$$

where the parameter  $\mu$  describes the step size and  $\mathbf{q}_l$  expresses the ascent direction at step  $l$ . For the computation of the suitable step size at each iteration, we apply the backtracking line search [50]. Note that  $\mathbf{q}_l = \frac{\partial \overline{\text{SR}}_{\text{PGP}}}{\partial \mathbf{s}_l^*} = \frac{\bar{K}}{\ln 2} \sum_{g=1}^G \frac{\partial \bar{\gamma}_{g,\text{PGP}}}{\partial \mathbf{s}_l^*}$  is provided by Proposition 2 below. The projection problem  $\min_{|\phi_n|=1, n=1, \dots, N} \|\mathbf{s} - \tilde{\mathbf{s}}\|^2$  under the unit-modulus constraint provides the solution of the problem described by (34). In Algorithm 1, we present the outline of this procedure.

---

**Algorithm 1** Projected Gradient Ascent Algorithm for the IRS Design

---

1. **Initialisation:**  $\mathbf{s}_0 = \exp(j\pi/2) \mathbf{1}_N$ ,  $\Phi_0 = \text{diag}(\mathbf{s}_0)$ ,  $\overline{\text{SR}}_{\text{PGP}}^0 = f(\Phi_0)$  given by (32a);  $\epsilon > 0$
  2. **Iteration  $l$ :** for  $l = 0, 1, \dots$ , do
  3.  $\mathbf{q}_l = \frac{\bar{K}}{\ln 2} \sum_{g=1}^G \frac{\partial \bar{\gamma}_{g,\text{PGP}}}{\partial \mathbf{s}_l^*}$ , where  $\frac{\partial \bar{\gamma}_{g,\text{PGP}}}{\partial \mathbf{s}_l^*}$  is given by Proposition 2;
  4. **Find**  $\mu$  by backtracking line search( $f(\Phi_0)$ ,  $\mathbf{q}_l, \mathbf{s}_l$ ) [50];
  5.  $\tilde{\mathbf{s}}_{l+1} = \mathbf{s}_l + \mu \mathbf{q}_l$ ;
  6.  $\mathbf{s}_{l+1} = \exp(j \arg(\tilde{\mathbf{s}}_{l+1}))$ ;  $\Phi_{l+1} = \text{diag}(\mathbf{s}_{l+1})$ ;
  7.  $\overline{\text{SR}}_{\text{PGP}}^{l+1} = f(\Phi_{l+1})$ ;
  8. **Until**  $\|\overline{\text{SR}}_{\text{PGP}}^{l+1} - \overline{\text{SR}}_{\text{PGP}}^l\| < \epsilon$ ; **Obtain**  $\Phi^* = \Phi_{l+1}$ ;
  9. **end for**
- 

*Proposition 2:* The derivative of  $\bar{\gamma}_{gk,\text{PGP}}$  with respect to  $\mathbf{s}_l^*$  is provided by

$$\frac{\partial \bar{\gamma}_{g,\text{PGP}}}{\partial \mathbf{s}_l^*} = \frac{\frac{\partial S_g}{\partial \mathbf{s}_l^*} I_g - S_g \frac{\partial I_g}{\partial \mathbf{s}_l^*}}{I_g^2}, \quad (35)$$

where

$$\frac{\partial S_g}{\partial \mathbf{s}_l^*} = 2 \frac{P_g}{\bar{K}} (1 - \tau_g^2) \bar{\delta}_g \bar{\delta}'_g, \quad (36)$$

$$\begin{aligned} \frac{\partial I_k}{\partial \mathbf{s}_l^*} &= \bar{\mathbf{Y}}_g' (1 - \tau_g^2 (1 - (1 + \bar{\delta}_g)^2)) + 2 \bar{\mathbf{Y}}_{gg} \tau_g^2 (1 + \bar{\delta}_g) \bar{\delta}'_g \\ &+ \sum_{l \neq g} (\bar{\lambda}'_l \bar{\mathbf{Y}}_{gl} + \bar{\lambda}_l \bar{\mathbf{Y}}'_{gl}) \frac{(1 + \bar{\delta}_g)^2}{\bar{\lambda}_g} \\ &+ \left( 1 + \sum_{l \neq g} \bar{\lambda}_l^2 \bar{\mathbf{Y}}_{gl} \right) \frac{(1 + \bar{\delta}_g) (2 \bar{\delta}'_g \bar{\lambda}_g - (1 + \bar{\delta}_g) \bar{\lambda}'_g)}{\bar{\lambda}_g^2} \end{aligned} \quad (37)$$

with

$$\bar{\delta}'_g = \frac{1}{\bar{b}} \text{tr}(\bar{\mathbf{R}}_g' \mathbf{T}_g + \bar{\mathbf{R}}_g \mathbf{T}_g'), \quad \mathbf{T}_g' = -\mathbf{T}_g (\mathbf{T}_g^{-1})' \mathbf{T}_g, \quad (38)$$

$$(\mathbf{T}_g^{-1})' = \frac{K \bar{\mathbf{R}}_g' (1 + \bar{\delta}_g) - \bar{\mathbf{R}}_g \bar{\delta}'_g}{(1 + \bar{\delta}_g)^2}, \quad (39)$$

$$\bar{\mathbf{Y}}_{gg}' = \frac{P_g}{\bar{b}} \left( 1 - \frac{1}{\bar{K}} \right) \frac{\bar{m}'_{gg} (1 + \bar{\delta}_g) - 2 \bar{m}_{gg} \bar{\delta}'_g}{(1 + \bar{\delta}_g)^3}, \quad (40)$$

$$\bar{\mathbf{Y}}_{gl}' = \frac{1}{\bar{b} \bar{K}} \frac{P_l \bar{m}'_{gl} (1 + \bar{\delta}_l) - 2 \bar{m}_{gl} \bar{\delta}'_l}{(1 + \bar{\delta}_l)^3}, \quad (41)$$

$$\bar{\lambda}'_g = \frac{\rho_g P_g}{\bar{b}} \frac{2 m_g \bar{\delta}'_g - m'_g (1 + \bar{\delta}_g)}{(1 + \bar{\delta}_g)^3 \bar{\Psi}_g^2}, \quad (42)$$

$$\bar{m}'_g = f'_g(\mathbf{B}_g^H \mathbf{B}_g), \quad \bar{m}'_{gg} = f'_g(\bar{\mathbf{R}}_g), \quad \bar{m}'_{gl} = f'_l(\mathbf{B}_l^H \bar{\mathbf{R}}_g \mathbf{B}_l), \quad (43)$$

with

$$\begin{aligned} f'_g(\mathbf{A}) &= \frac{\frac{\bar{K}}{b} (1 + \bar{\delta}_g) (q'(\mathbf{A}) (1 + \bar{\delta}_g) + 2q(\mathbf{A}) \bar{\delta}'_g)}{\left( b (1 + \bar{\delta}_g)^2 - \frac{\bar{K}}{b} q(\bar{\mathbf{R}}_g) \right)} \\ &+ \frac{\frac{\bar{K}}{b} (1 + \bar{\delta}_g)^2 q(\mathbf{A}) \left( 2b (1 + \bar{\delta}_g) \bar{\delta}'_g - \frac{\bar{K}}{b} q'(\bar{\mathbf{R}}_g) \right)}{\left( b (1 + \bar{\delta}_g)^2 - \frac{\bar{K}}{b} q(\bar{\mathbf{R}}_g) \right)^2}, \end{aligned} \quad (44)$$

and  $q'(\mathbf{C})$  for  $\mathbf{C} = \mathbf{A}, \bar{\mathbf{R}}_g$  written as  $q'(\mathbf{C}) = \text{tr}(\bar{\mathbf{R}}_g' \mathbf{T}_g \mathbf{C} \mathbf{T}_g + \bar{\mathbf{R}}_g \mathbf{T}_g' \mathbf{C} \mathbf{T}_g + \bar{\mathbf{R}}_g \mathbf{T}_g \mathbf{C}' \mathbf{T}_g + \bar{\mathbf{R}}_g \mathbf{T}_g \mathbf{C} \mathbf{T}_g')$ , where each term is computed by Lemma 2. The expressions of  $\bar{\delta}_g$ ,  $\Psi_g$ ,  $\bar{\mathbf{Y}}_{gg}$ , and  $\bar{\mathbf{Y}}_{gl}$ , given in Theorem 1 and Corollary 1 together with  $\bar{\delta}'_g$ ,  $\bar{\lambda}'_g$ ,  $\bar{\mathbf{Y}}_{gg}'$ ,  $\bar{\mathbf{Y}}_{gl}'$ ,  $\bar{m}'_g$ ,  $\bar{m}'_{gg}$ , and  $\bar{m}'_{gl}$  are obtained by solving the system of fixed-point equations (26)-(29), (39)-(43).

*Proof:* The proof is given in Appendix C.  $\blacksquare$

The RBM beamforming design is based on the gradient ascent and offers a significant advantage because the gradient ascent is derived in a closed-form. Note that it has low computational complexity because it consists of simple matrix operations. Specifically, the complexity of (35) is  $\mathcal{O}(G(MN^2 + N + M))$ . Obviously, the derivative is a function of the fundamental system parameters  $G$ ,  $M$ , and  $N$  with the number of IRS elements having the higher (square) impact.

### B. Power Allocation Optimization

Now, for a fixed RBM  $\Phi$ , the objective is the optimization over  $\mathbf{p}$ . In particular, we have

$$\begin{aligned} (\mathcal{P}4) \quad & \max_{\mathbf{p} \geq 0} \overline{\text{SR}}_{\text{PGP}} \\ & \text{s.t.} \quad \sum_{g=1}^G P_g \leq P_{\max}, \end{aligned} \quad (45)$$

where  $\overline{\text{SR}}_{\text{PGP}}$  is given by (32a). This problem is not convex but a local optimal solution can be obtained by using a weighted minimum mean square error (WMMSE) reformulation of the sum SE maximization. By denoting  $\mathbf{c}_g = [c_{g1}, \dots, c_{gG}]^\top$ , the SINR  $\bar{\gamma}_{g,\text{PGP}}$  can be expressed as a function of the downlink power coefficients given by the vector  $\mathbf{p}$  as

$$\bar{\gamma}_{g,\text{PGP}} = \frac{P_g q_g}{\mathbf{c}_g^\top \mathbf{p} + t_g^2}, \quad (46)$$

where

$$q_g = \frac{1}{\bar{K}} (1 - \tau^2) \bar{\delta}_g^2, \quad t_g^2 = \frac{(1 + \bar{\delta}_g)^2}{\bar{\lambda}_g}, \quad (47)$$

$$c_{gg} = \frac{1}{\bar{b}} \left( 1 - \frac{1}{\bar{K}} \right) \frac{\bar{m}_{gg}}{(1 + \bar{\delta}_g)^2} \left( 1 - \tau^2 \left( 1 - (1 + \bar{\delta}_g^2) \right) \right), \quad \forall g \quad (48)$$

$$c_{gi} = \frac{1}{\bar{b}} \frac{1}{\bar{K}} \frac{m_{gl}}{(1 + \bar{\delta}_l)^2} \frac{\bar{\lambda}_l (1 + \bar{\delta}_g)^2}{\bar{\lambda}_g}, \quad \forall g, \quad \forall i \neq g. \quad (49)$$

Now, the optimization problem becomes

$$(\mathcal{P}5) \quad \max_{\mathbf{p} \geq 0} \quad \bar{K} \sum_{g=1}^G \log_2 \left( 1 + \frac{P_g q_g}{\mathbf{c}_g^T \mathbf{p} + t_g^2} \right) \quad (50)$$

$$\text{s.t.} \quad \sum_{g=1}^G P_g \leq P_{\max}.$$

For the MMSE reformulation, we assume the single-input and single-output (SISO) channel that corresponds to this SINR given by

$$\tilde{y}_g = \sqrt{P_g q_g} s_g + \sum_{i=1}^G \sqrt{P_g c_{gi}} s_i + n_g, \quad (51)$$

where  $\tilde{y}_g$  is the received signal,  $s_g \in \mathbb{C}$  denotes the normalized and independent random data signal with  $\mathbb{E}[|s_g|^2] = 1$ , and  $n_g \sim \mathcal{CN}(0, t_g^2)$ . In such case, the receiver can compute an estimate  $\hat{s}_g = v_g^* \tilde{y}_g$  of the desired signal  $s_g$ , where  $v_g$  as a scalar combining coefficient. The resulting MSE  $e_g(\mathbf{p}, v_g) = \mathbb{E}[|\hat{s}_g - s_g|^2]$  is written as

$$e_g(\mathbf{p}, v_g) = v_g^2 (P_g q_g + \mathbf{c}_g^T \mathbf{p} + t_g^2) - 2v_g \sqrt{q_g P_g} + 1. \quad (52)$$

The coefficient  $v_g$ , minimizing the MSE  $e_g(\mathbf{p}, v_g)$  for a given  $\mathbf{p}$ , becomes

$$v_g = \frac{\sqrt{P_g q_g}}{P_g q_g + \sum_{i=1}^G P_g c_{gi} + t_g^2}. \quad (53)$$

Plugging  $v_g$  into (52),  $e_g$  becomes  $1/(1 + \bar{\gamma}_{g,\text{PGP}})$ . Based on the weighted MMSE method, we introduce the auxiliary weight  $d_g \geq 0$  for the MSE  $e_g$  and solve the following problem

$$(\mathcal{P}6) \quad \min_{\substack{\mathbf{p} \geq 0, \\ \{v_g, d_g \geq 0: g=1, \dots, G\}}} \quad \bar{K} \sum_{g=1}^G d_g e_g(\mathbf{p}, \mathbf{v}_g) - \ln(d_g) \quad (54)$$

$$\text{s.t.} \quad \sum_{g=1}^G P_g \leq P_{\max}.$$

Note that both  $(\mathcal{P}5)$  and  $(\mathcal{P}6)$  are equivalent, meaning that they have the same optimal solution. The equivalence stems from the fact that the optimal  $d_g$  in (54) is  $1/e_g = (1 + \bar{\gamma}_{g,\text{PGP}})$ . The advantage of the reformulation results in the following lemma adapted from [51, Th. 3].

*Lemma 1:* The block descent coordinate algorithm, provided in Algorithm 2, converges to a local optimum of  $(\mathcal{P}6)$  in terms of AO among three blocks of variables  $\{v_g : g = 1, \dots, G\}$ ,  $\{d_g : g = 1, \dots, G\}$ , and  $\mathbf{p}$ .

---

**Algorithm 2** Block coordinate descent algorithm for solving  $(\mathcal{P}6)$

---

1. **Initialisation:** Set  $\mathbf{p} = \frac{P_{\max}}{G} \mathbf{1}_G$  and the solution accuracy  $\epsilon > 0$ ,
2. **while** the objective function in (54) is not improved more than  $\epsilon$  **do**
3.  $v_g = \frac{\sqrt{P_g q_g}}{P_g q_g + \sum_{i=1}^G P_g c_{gi} + t_g^2}$ ,  $g = 1, \dots, G$
4.  $d_g = 1/e_g(\mathbf{p}, v_g)$ ,  $g = 1, \dots, G$
5. Solve the following problem for the current values of  $v_g$  and  $d_g$ :

$$(\mathcal{P}7) \quad \min_{\mathbf{p} \geq 0} \quad \bar{K} \sum_{g=1}^G d_g e_g(\mathbf{p}, \mathbf{v}_g) \quad (55)$$

$$\text{s.t.} \quad \sum_{g=1}^G P_g \leq P_{\max},$$

6. Update  $\mathbf{p}$  by the obtained solution to (55)
  7. **end while**
  8. **Output:**  $\mathbf{p}^*$
- 

Algorithm 2 describes the whole procedure for the power optimization and Step 5 includes a subproblem that needs to be solved in every iteration. It can be solved in closed-form as

$$P_g = \min \left( P_{\max}, \frac{q_g d_g^2 v_g^2}{\left( q_g d_g v_g^2 + \sum_{i=1}^G d_i v_i^2 c_{ig} \right)^2} \right), \quad (56)$$

where  $\sqrt{P_g}$ ,  $g = 1, \dots, G$  are treated as optimization variables. Basically, the problem is decomposed into  $G$  independent subproblems, where each one of them concerns a quadratic minimization under a bound constraint.

The power allocation presents a similar complexity to the RBM design since Algorithm 2 consists of similar matrix operations, i.e., its complexity is  $\mathcal{O}(G(MN^2 + N + M))$ .

*Remark 3:* Both Algorithms 1 and 2 converge quickly and have low computation complexity. Also, since they achieve to obtain just a local optimum, different initializations are expected to lead to different solutions of the overall algorithm.

## V. NUMERICAL RESULTS

In this section, we elaborate on the numerical results of the sum SE in IRS-assisted systems under the FDD protocol by applying PGP of the JSDM method. MC simulations in terms of  $10^3$  independent channel realizations verify the deterministic equivalent analysis.

The simulation setup includes a uniform linear array (ULA) of  $M = 100$  antennas at the BS, which serves  $K = 30$  UEs while aided by an IRS with a uniform planar array (UPA) of  $N = 100$  elements. UEs are equally shared among  $G = 6$  groups, which means that each group consists of 5 UEs. The correlation matrices  $\mathbf{R}_{\text{BS},g_k}$  and  $\mathbf{R}_{\text{IRS},g_k}$  are obtained similar to [35] and [33], respectively. Note that the size of each IRS element is given by  $d_H = d_V = \lambda/4$ . Furthermore, the path-losses for the BS-to-IRS and IRS-to-UE  $k$  links are given by [4], [18]

$$\beta_1 = \frac{C_1}{d_1^{\alpha_1}}, \quad \beta_{2,k} = \frac{C_2}{d_{2,k}^{\alpha_2}}, \quad (57)$$

where  $\alpha_1$  and  $d_1$  are the path-loss exponent and distance concerning the former link while  $\alpha_2$  and  $d_{2,k}$  are the path-loss exponent and distance concerning the latter link. In the case of  $\beta_{d,k}$ , we assume the same parameters as for  $\beta_{2,k}$ , and additionally, we consider a penetration loss of 15 dB. The choice of these values relies on the 3GPP Urban Micro (UMi) scenario from TR36.814 for a carrier frequency of 2.5 GHz and a noise level of  $-80$  dBm. Especially, the path-losses for  $\mathbf{H}_1$  and  $\mathbf{h}_{2,k}$  are generated according to the LOS and NLOS versions [52]. Also,  $C_1 = 26$  dB and  $C_2 = 28$  dB, which are the path-losses at a reference distance of 1m while each UE includes a single 0dBi antenna [53]. Also, we have chosen  $\bar{b} = 12$  and  $\bar{r} = 12$ . The latter value is justified because under these setting the rank of the channel covariance matrix  $\mathbf{R}_g$  is  $\bar{r} = 27$  while only 12 of them are important. Unless otherwise stated, this set of parameters values is used during the simulations. Note that  $\sigma^2 = -174 + 10 \log_{10} B_c$ , where  $B_c = 200$  kHz.

Figs. 2.(a) and 2.(b) illustrate the impact of the effective rank of the channel covariance matrix of each group on the performance of JSDM in IRS-assisted systems in the cases of perfect and imperfect CSI. Under perfect CSI conditions ("solid" lines) with  $\tau = 0$ , we notice that in the case of choosing  $\bar{r} = 5$  (Fig. 2.(a)), PGP saturates at high SNR since this effective rank value is too small. In particular, such a choice does not allow for consideration of the substantial eigenmodes by the pre-beamforming matrix and results in large inter-group interference. On the contrary, a better (higher) choice for  $\bar{r}$ , i.e.,  $\bar{r} = 15$  (see Fig. 2.(b)) allows the inclusion of the significant eigenmodes and no interference-limited behaviour is noticed until 30 dB. Of course, the rate saturates at some larger SNR, belonging outside the scope of practical applications. Obviously, the choice of the rank value depends on the channel covariance matrix. In the case of imperfect CSI ( $\tau = 0.1$ ), met in practice, saturation is observed in both figures ("dashed" lines). However, when we have  $\bar{r} = 5$ , the saturation starts much earlier (25 dB) while a choice of  $\bar{r} = 15$  (Fig. 2.(b)) is more robust regarding the performance since the saturation is encountered after 21.7 dB. Similar observations hold when  $\tau = 0.3$ . The inference from these two figures is that a good selection of  $\bar{r}$ , where the major eigenmodes are taken into account, defines the performance. A small value results in severe performance degradation while a value close to the full rank  $\bar{r}$  will result in a dimensionality bottleneck while not achieving any profit concerning the interference. In these figures, we also show the degradation of the sum SE when no correlation ("dotted" lines) is taken into account because in such a case the IRS cannot be optimized.

To elaborate further on the saving advantage of the JSDM method, we would like to highlight that after the downlink training of all groups ( $G = 6$ ) with effective channel dimension  $\bar{b} = 12$ , each of the  $\bar{K}$  UEs in each group has to feedback its  $12 \times 1$  channel vector, which equals to  $5 \times 6 \times 12 = 360$  quantized complex channel coefficients.<sup>5</sup> However, in this case, where the BS is deployed with  $M = 100$  antenna aimed

<sup>5</sup>This conclusion assumes that downlink training has already taken place at some previous stage. Note that this work does not perform downlink training but this task and its effects are left for future study.

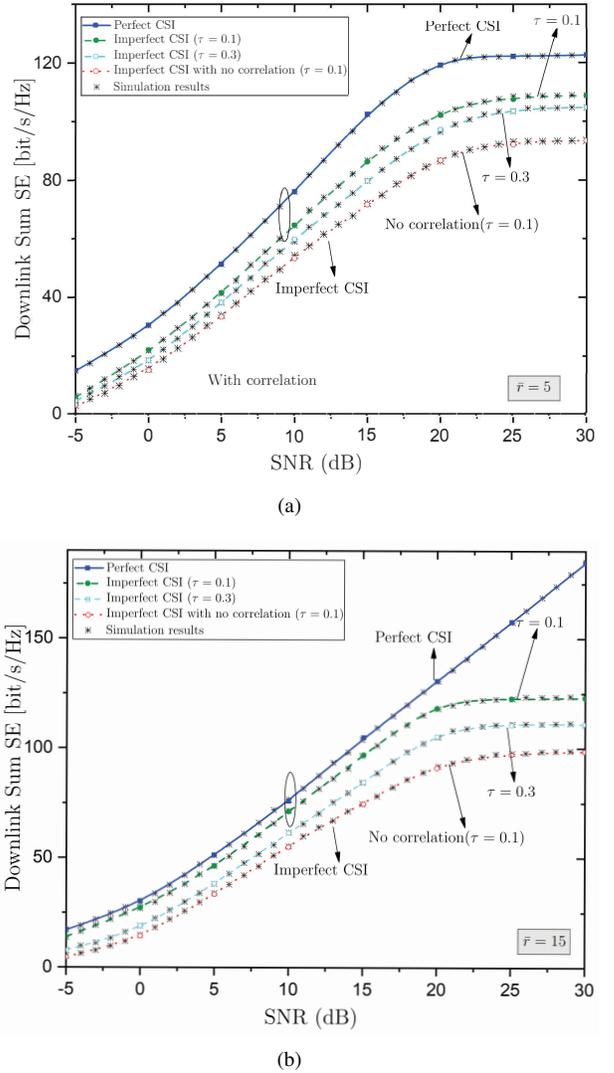


Figure 2: Sum SE of an IRS-assisted MIMO system with FDD in the cases of perfect/imperfect CSI and correlated/independent Rayleigh fading versus the SNR: (a)  $\bar{r} = 5$  ; (b)  $\bar{r} = 12$ .

at serving  $K = 30$  UEs in total, 100 orthogonal downlink training symbols are required while a total of  $30 \times 100$  channel coefficients will be fed back, which means the achievement of saving by a factor of 10.

In Fig. 3, we depict the impact of the parameter  $\bar{b}$  for a specific  $\bar{K}$  and when  $\tau = 0.1$  (imperfect CSI) for varying SNR values, being SNR = 5, 15, 25 dB. Specifically, after selecting  $\bar{r} = 15$  according to the discussion in the previous paragraph and subject to the constraint  $\bar{K} \leq \bar{b} \leq M - \bar{r}(G - 1)$ , we observe that the sum SE does not increase monotonically but there is an optimal value  $\bar{b}$  because its increase results in a trade-off between a larger dimensionality overhead and a better channel conditioning. Moreover, we observe that the higher the SNR, the smaller the optimal  $\bar{b}$  becomes. Also, we notice that at higher SNR, the range of sum SE values takes higher values but the impact of  $\bar{b}$  variation is smaller with increasing SNR as witnessed by the slopes of the corresponding lines.

Fig. 4 shows the achievable sum SE versus the number of

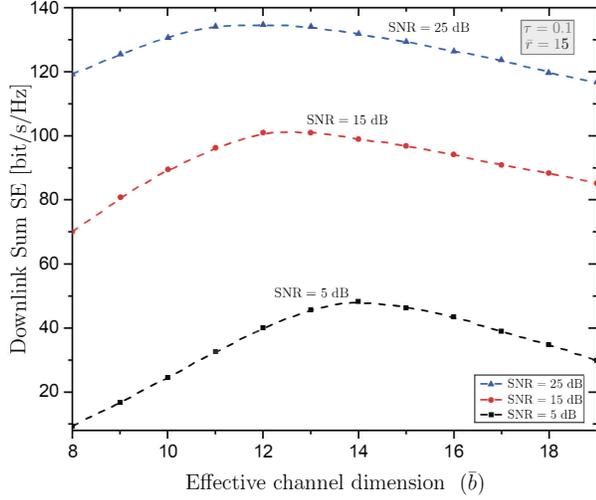


Figure 3: Sum SE of an IRS-assisted MIMO system with FDD in the cases of imperfect CSI ( $\tau = 0.1$ ) versus the effective channel dimension  $\bar{b}$  for varying SNR values.

BS antennas  $M$  for both perfect and imperfect CSI, when  $N = 100$  elements. We observe that  $\overline{\text{SR}}_{\text{PGP}}$  exhibits a similar dependence on  $M$  as on  $N$  in the previous figure. Thus, when  $M$  grows large, the sum SE increases without limit in both cases of perfect and imperfect CSI. In general, these two figures indicate that an IRS-assisted system performs better with larger values of IRS elements and BS antennas. In particular, the latter is further interesting because it agrees with the massive MIMO design trend that has already started to be implemented. In addition, we verify that the performance increases in the presence of the IRS.

Fig. 5 illustrates the achievable sum SE versus the number of IRS elements  $N$  for both cases of perfect and imperfect CSI, when  $M = 100$  antennas. We notice that the sum SE increases with the increasing number of IRS elements in both cases, but the increase is slower in the case of imperfect CSI. Also, disregarding the IRS size, defined by its elements, it always provides better performance compared to its absence. In other words, its application is beneficial because it results in a further enhancement of the channel additionally to the direct signal. Moreover, we consider the scenario of "random" phase shifts and we show that the RBM optimization enhances significantly the performance.

Fig. 6 depicts the sum SE versus the number of groups in the case of imperfect CSI ( $\tau = 0.1$ ) while varying the effective rank  $\bar{r}$ , the effective channel dimension  $\bar{b}$ , and UEs per group  $\bar{K}$ . We notice that the sum SE increases with the number of groups but this increase becomes slower when the number of UEs per group grows due to increased multiuser interference while the rest of the parameters are the same. In addition, for a specific  $\bar{K}$ , an increase of the channel rank results in an increase of the sum SE but caution should be taken since the overhead (dimensionality) increases too. Moreover, by increasing  $\bar{b}$ , the sum SE increases until a specific  $\bar{b}$  equal to 10 while keeping increasing  $\bar{b}$  results in lower sum SE due to larger dimensionality cost again.

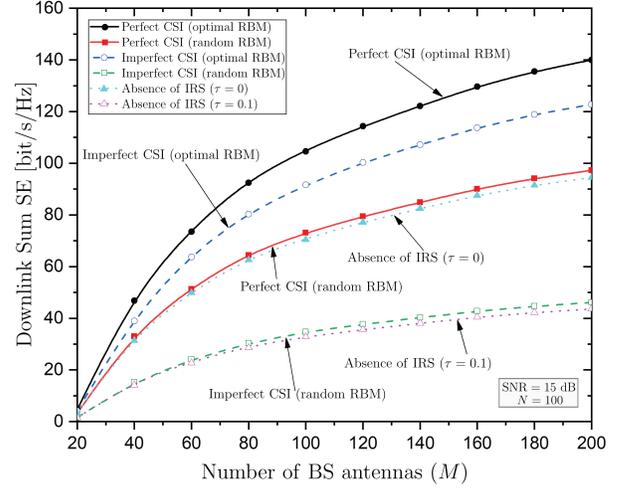


Figure 4: Sum SE of an IRS-assisted MIMO system with FDD versus the number of BS antennas  $M$  in the cases of i) perfect/imperfect CSI ( $\tau = 0.1$ ) with optimal and random RBM, ii) without IRS and perfect/imperfect CSI.

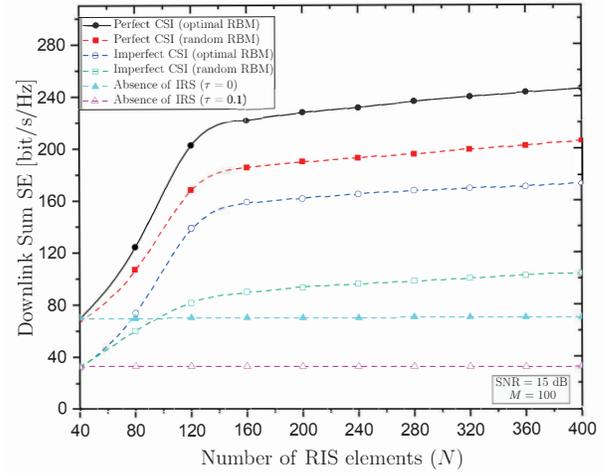


Figure 5: Sum SE of an IRS-assisted MIMO system with FDD versus the number of IRS elements  $N$  in the cases of i) perfect/imperfect CSI ( $\tau = 0.1$ ) with optimal and random RBM, ii) without IRS and perfect/imperfect CSI.

Fig. 7.(a) illustrates the convergence of the proposed overall algorithm based on AO, which consists of the subproblems of RBM and transmit power optimizations. Specifically, we have depicted the sum SE versus the number of iterations for varying numbers of IRS elements and UE groups. As can be seen, the convergence of the optimization is fast in all cases, where the algorithm has converged at most in 8 iterations (see the "dashed-square" line). Also, as expected, as the number of IRS elements and groups increases, more iterations are needed until convergence since the number of optimization variables has increased.

The non-convexity of the overall optimization problem suggests that its solution depends on the initial point, i.e., different initial points result in different locally optimal solutions. Fig.

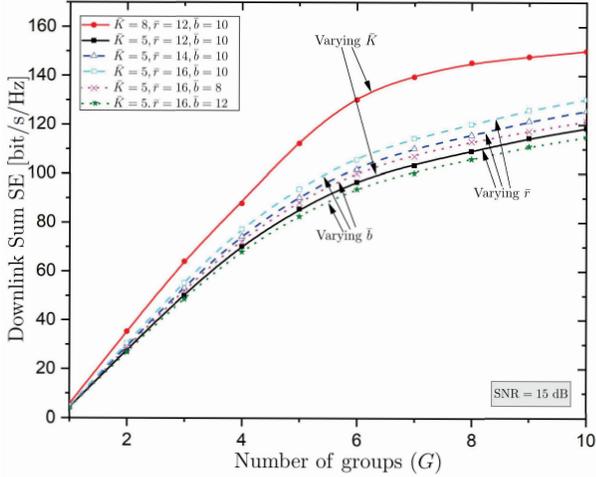


Figure 6: Sum SE of an IRS-assisted MIMO system with FDD versus the number of groups  $G$  in the case of imperfect CSI ( $\tau = 0.1$ ) for optimal and random RBM for varying effective rank  $\bar{r}$ , effective channel dimension  $\bar{b}$ , and UEs per group  $\bar{K}$ .

7.(b) investigates this dependence on the initializations by accounting for 30 channel realizations. The initialization of the overall algorithm including Algs. 1 and 2 assumes that  $\mathbf{s}_0 = \exp(j\pi/2) \mathbf{1}_N$  and  $\mathbf{p} = \frac{P_{\max}}{G} \mathbf{1}_G$  as mentioned previously. "Alg. 1-Test" in the figure assumes the best initial point out of 100 random initial points for each channel instance. We observe indeed that different initializations result in different solutions and that the sum SE in both cases is almost the same, which means that this phase shifts selection for initialization is a good choice.

## VI. CONCLUSION

In this paper, we tackled the channel non-reciprocity witnessed in IRS-assisted communication systems. For this reason, we assumed an mMIMO system assisted with an IRS operating in FDD. To this end, we applied the JSDM method by grouping the UEs with the same covariance matrix to improve the performance, which, otherwise, would be afflicted due to prohibitive feedback overhead as the number of BS antennas and IRS elements increases. Based on DE tools and S-CSI knowledge, we derived the sum SE accounting for correlated Rayleigh fading. Next, we formulated the optimization problem maximizing the sum SE under RBM and power budget constraints. To solve the nonconvex optimization, we proposed an efficient AO algorithm that can be performed at every several coherence intervals, which results in further reduction of the feedback overhead and computational complexity. Among the observations, we would like to highlight that the rank of the covariance matrix is a crucial parameter having a direct impact on the performance. Its increase improves the performance due to better channel conditioning while it burdens the feedback overhead. Other interesting directions could be the extension of this work to account for 3-dimensional beamforming with multiple beams in the elevation angle and the consideration of multiple distributed IRS.

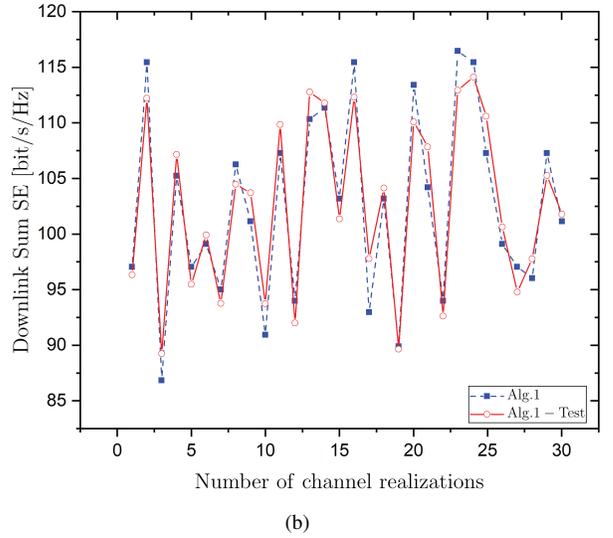
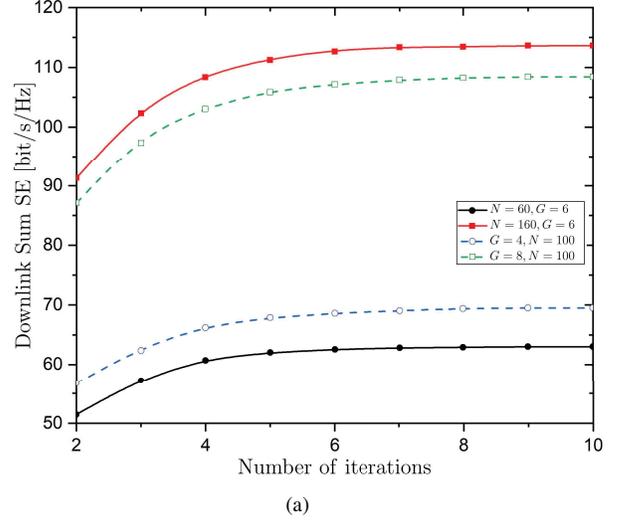


Figure 7: Sum SE of an IRS-assisted MIMO system with FDD versus: (a) the number of iterations for varying IRS elements  $N$  and groups  $G$ ; (b) 30 channel realizations.

## APPENDIX A PROOF OF THEOREM 1

The derivation of the DE SINR  $\bar{\gamma}_{g_k, \text{PGP}}$  is obtained by following the approach in [34] but with extension to imperfect CSI that brings considerable differences. Hence, taking into account for imperfect CSI, the proof consists of the derivation of the DEs of four parts: i) The power normalization term  $\Psi_g$ , ii) the desired signal power part  $\text{DS}_{g_k}$ , iii) the self-group interference power part  $\text{SGI}_{g_k}$ , and iv) the inter-group interference power part  $\text{IGI}_{g_k}$ .

Regarding  $\Psi_g$ , we have

$$\Psi_g = \sum_{k=1}^{\bar{K}} p_{g_k} \hat{\mathbf{h}}_{g_k}^H \Sigma_g^{-2} \hat{\mathbf{h}}_{g_k} \quad (58)$$

$$= \frac{1}{\bar{b}} \sum_{k=1}^{\bar{K}} p_{g_k} \frac{\frac{1}{\bar{b}} \hat{\mathbf{h}}_{g_k}^H \mathbf{B}_g \Sigma_{g[k]}^{-2} \mathbf{B}_g^H \hat{\mathbf{h}}_{g_k}}{\left(1 + \frac{1}{\bar{b}} \hat{\mathbf{h}}_{g_k}^H \mathbf{B}_g \Sigma_{g[k]}^{-1} \mathbf{B}_g^H \hat{\mathbf{h}}_{g_k}\right)^2} \quad (59)$$

$$\asymp \frac{1}{\bar{b}} \sum_{k=1}^{\bar{K}} p_{g_k} \frac{\frac{1}{\bar{b}} \text{tr}(\bar{\mathbf{R}}_g \Sigma_{g[k]}^{-2})}{(1 + \frac{1}{\bar{b}} \text{tr}(\bar{\mathbf{R}}_g \Sigma_{g[k]}^{-1}))^2} \quad (60)$$

$$\asymp \frac{1}{\bar{b}} P_g \frac{\bar{m}_g}{(1 + \bar{\delta}_g)^2}, \quad (61)$$

where  $\Sigma_{g[k]} = \frac{1}{\bar{b}} \hat{\mathbf{H}}_{g[k]} \hat{\mathbf{H}}_{g[k]}^H + \alpha \mathbf{I}_{\bar{b}}$ . In (59), we applied the matrix inversion lemma twice [35, Lem. 1] and, in (60), we used [35, Lem. 4]. The last step makes use of [35, Th. 1] with  $\bar{m}_g$  and  $\bar{\delta}_g$  defined in Theorem 1. Note that  $\bar{m}_g$  is basically the derivative of  $\bar{\delta}_g$  with respect to  $\alpha$ , i.e.,  $\bar{m}_g = \bar{\delta}_g' = \frac{1}{\bar{b}} \text{tr}(\bar{\mathbf{R}}_g \mathbf{T}_g')$ , where  $\mathbf{T}_g' = \mathbf{T}_g \left( \frac{\bar{K}}{\bar{b}} \frac{\bar{\mathbf{R}}_g \bar{\delta}_g'}{1 + \bar{\delta}_g} + \mathbf{I}_{\bar{b}} \right) \mathbf{T}_g'$ . Its expression is obtained after some algebraic manipulations.

In the case of DS $_{g_k}$ , we have

$$\mathbf{h}_{g_k}^H \mathbf{B}_g \Sigma_g \mathbf{B}_g^H \hat{\mathbf{h}}_{g_k} = \frac{\frac{1}{\bar{b}} \mathbf{h}_{g_k}^H \mathbf{B}_g \Sigma_{g[k]} \mathbf{B}_g^H \hat{\mathbf{h}}_{g_k}}{1 + \frac{1}{\bar{b}} \hat{\mathbf{h}}_{g_k}^H \mathbf{B}_g \Sigma_{g[k]} \mathbf{B}_g^H \hat{\mathbf{h}}_{g_k}} \quad (62)$$

$$= \frac{\sqrt{1 - \tau_g^2} \frac{1}{\bar{b}} \mathbf{h}_{g_k}^H \mathbf{B}_g \Sigma_{g[k]} \mathbf{B}_g^H \mathbf{h}_{g_k}}{1 + \frac{1}{\bar{b}} \hat{\mathbf{h}}_{g_k}^H \mathbf{B}_g \Sigma_{g[k]} \mathbf{B}_g^H \hat{\mathbf{h}}_{g_k}} + \frac{\tau_g \frac{1}{\bar{b}} \mathbf{h}_{g_k}^H \mathbf{B}_g \Sigma_{g[k]} \mathbf{B}_g^H \mathbf{z}_{g_k}}{1 + \frac{1}{\bar{b}} \hat{\mathbf{h}}_{g_k}^H \mathbf{B}_g \Sigma_{g[k]} \mathbf{B}_g^H \hat{\mathbf{h}}_{g_k}} \quad (63)$$

$$= \frac{\sqrt{1 - \tau_g^2} \bar{\delta}_g}{1 + \bar{\delta}_g}, \quad (64)$$

where in (62), we applied the matrix inversion lemma [35, Lem. 1], and, in (63), we used (7). In (64), we used [35, Lem. 4] by considering the independence between  $\mathbf{h}_{g_k}$  and  $\mathbf{z}_{g_k}$ . Next, we applied [35, Lem. 4], [35, Lem. 3], and [35, Th. 1].

For SGI $_{g_k}$ , we have

$$\begin{aligned} & \mathbf{h}_{g_k}^H \mathbf{B}_g \Sigma_g \mathbf{B}_g^H \hat{\mathbf{H}}_{g[k]} \mathbf{P}_{g[k]} \hat{\mathbf{H}}_{g[k]}^H \mathbf{B}_g \Sigma_g \mathbf{B}_g^H \mathbf{h}_{g_k} \\ &= \mathbf{h}_{g_k}^H \mathbf{B}_g \bar{\Sigma}_{g[k]} \mathbf{B}_g^H \hat{\mathbf{H}}_{g[k]} \mathbf{P}_{g[k]} \hat{\mathbf{H}}_{g[k]}^H \mathbf{B}_g \Sigma_g \mathbf{B}_g^H \mathbf{h}_{g_k} \\ &+ \mathbf{h}_{g_k}^H \mathbf{B}_g (\Sigma_g - \bar{\Sigma}_{g[k]}) \mathbf{B}_g^H \hat{\mathbf{H}}_{g[k]} \mathbf{P}_{g[k]} \hat{\mathbf{H}}_{g[k]}^H \mathbf{B}_g \Sigma_g \mathbf{B}_g^H \mathbf{h}_{g_k} \quad (65) \\ &= \frac{1}{\bar{b}^2} \mathbf{h}_{g_k}^H \mathbf{B}_g \mathbf{D}_g \mathbf{B}_g^H \mathbf{h}_{g_k} - \frac{c_0}{\bar{b}^2} \mathbf{h}_{g_k}^H \mathbf{B}_g \Sigma_g \mathbf{B}_g^H \mathbf{h}_{g_k} \mathbf{h}_{g_k}^H \mathbf{B}_g \mathbf{D}_g \mathbf{B}_g^H \mathbf{h}_{g_k} \\ &- \frac{c_1}{\bar{b}^2} \mathbf{h}_{g_k}^H \mathbf{B}_g \Sigma_g \mathbf{B}_g^H \mathbf{z}_{g_k} \mathbf{z}_{g_k}^H \mathbf{B}_g \mathbf{D}_g \mathbf{B}_g^H \mathbf{h}_{g_k} \\ &- \frac{c_2}{\bar{b}^2} \mathbf{h}_{g_k}^H \mathbf{B}_g \Sigma_g \mathbf{B}_g^H \mathbf{h}_{g_k} \mathbf{z}_{g_k}^H \mathbf{B}_g \mathbf{D}_g \mathbf{B}_g^H \mathbf{h}_{g_k} \\ &- \frac{c_2}{\bar{b}^2} \mathbf{h}_{g_k}^H \mathbf{B}_g \Sigma_g \mathbf{B}_g^H \mathbf{z}_{g_k} \mathbf{h}_{g_k}^H \mathbf{B}_g \mathbf{D}_g \mathbf{B}_g^H \mathbf{h}_{g_k}. \quad (66) \end{aligned}$$

In (65), we used that  $\Sigma_g - \bar{\Sigma}_{g[k]} = -\Sigma_g (\Sigma_g^{-1} - \bar{\Sigma}_{g[k]}^{-1}) \bar{\Sigma}_{g[k]}$ , where  $\Sigma_g^{-1} - \bar{\Sigma}_{g[k]}^{-1} = \mathbf{B}_g^H (c_0 \mathbf{h}_{g_k} \mathbf{h}_{g_k}^H + c_1 \mathbf{z}_{g_k} \mathbf{z}_{g_k}^H + c_2 (\mathbf{h}_{g_k} \mathbf{z}_{g_k}^H + \mathbf{z}_{g_k} \mathbf{h}_{g_k}^H)) \mathbf{B}_g$  with  $c_0 = 1 - \tau_g^2$ ,  $c_1 = \tau_g$ , and  $c_2 = \tau_g \sqrt{1 - \tau_g^2}$ . In (66), we denoted  $\mathbf{D}_g = \Sigma_g \mathbf{B}_g^H \hat{\mathbf{H}}_{g[k]} \mathbf{P}_{g[k]} \hat{\mathbf{H}}_{g[k]}^H \mathbf{B}_g \Sigma_g$ . Similar to previous derivations, we obtain

$$\frac{1}{\bar{b}^2} \mathbf{h}_{g_k}^H \mathbf{B}_g \Sigma_g \mathbf{B}_g^H \mathbf{h}_{g_k} \asymp \frac{u(1 + c_1 u)}{1 + u}, \quad (67)$$

$$\frac{1}{\bar{b}^2} \mathbf{h}_{g_k}^H \mathbf{B}_g \Sigma_g \mathbf{B}_g^H \mathbf{z}_{g_k} \asymp \frac{-c_2 u^2}{1 + u}, \quad (68)$$

$$\frac{1}{\bar{b}^2} \mathbf{h}_{g_k}^H \mathbf{B}_g \mathbf{D}_g \mathbf{B}_g^H \mathbf{h}_{g_k} \asymp \frac{u'(1 + c_1 u)}{1 + u}, \quad (69)$$

$$\frac{1}{\bar{b}^2} \mathbf{h}_{g_k}^H \mathbf{B}_g \Sigma_g \mathbf{B}_g^H \mathbf{z}_{g_k} \asymp \frac{-c_2 u u'}{1 + u}, \quad (70)$$

where  $u' = \frac{1}{\bar{b}} \text{tr}(\mathbf{P}_{g[k]} \hat{\mathbf{H}}_{g[k]}^H \mathbf{B}_g \Sigma_{g[k]} \bar{\mathbf{R}}_g \Sigma_{g[k]} \mathbf{B}_g^H \hat{\mathbf{H}}_{g[k]})$ . Hence, we have almost surely that

$$\begin{aligned} & \mathbf{h}_{g_k}^H \mathbf{B}_g \Sigma_g \mathbf{B}_g^H \hat{\mathbf{H}}_{g[k]} \mathbf{P}_{g[k]} \hat{\mathbf{H}}_{g[k]}^H \mathbf{B}_g \Sigma_g \mathbf{B}_g^H \mathbf{h}_{g_k} \\ & \asymp \frac{1}{\bar{b}} \frac{u'(1 + c_1 u)}{1 + u} - \frac{1}{\bar{b}} \frac{c_0(1 + c_1 u)^2 - c_1 c_2^2 u^2 - 2c_2^2 u}{(1 + u)^2} u u' \quad (71) \end{aligned}$$

$$= \frac{1}{\bar{b}} \frac{1 - \tau_g^2 (1 - (1 + u)^2)}{(1 + u)^2} u', \quad (72)$$

where (71) reduces to (72) after proper substitutions. Based on the rank perturbation lemma [35, Lem. 3], we have

$$u \asymp \frac{1}{\bar{b}} \text{tr}(\bar{\mathbf{R}}_g \Sigma_g). \quad (73)$$

Also, we denote  $\frac{1}{\bar{b}} u' \asymp \bar{\mathbf{Y}}_{gg}$ , where  $\bar{\mathbf{Y}}_{gg} = \frac{1}{\bar{b}^2} \text{tr}(\mathbf{P}_{g[k]} \hat{\mathbf{H}}_{g[k]}^H \mathbf{B}_g \Sigma_{g[k]} \bar{\mathbf{R}}_g \Sigma_{g[k]} \mathbf{B}_g^H \hat{\mathbf{H}}_{g[k]})$ , which can be written as

$$\bar{\mathbf{Y}}_{gg} \asymp \frac{1}{\bar{b}} \sum_{j \neq k}^{\bar{K}} p_{g_j} \frac{\frac{1}{\bar{b}} \text{tr}(\bar{\mathbf{R}}_g \Sigma_g \bar{\mathbf{R}}_g \Sigma_g)}{(1 + \frac{1}{\bar{b}} \text{tr}(\bar{\mathbf{R}}_g \Sigma_g))^2} \quad (74)$$

after applying [35, Lems. 1, 4, 3]. Note that

$$\begin{aligned} & \frac{1}{\bar{b}} \text{tr}(\bar{\mathbf{R}}_g \Sigma_g \bar{\mathbf{R}}_g \Sigma_g) = \frac{1}{\bar{b}} \text{tr}(\bar{\mathbf{R}}_g^{-1/2} \bar{\mathbf{R}}_g \bar{\mathbf{R}}_g^{-1/2} \Sigma_g^2) \\ &= \frac{d}{dz} \frac{1}{\bar{b}} \text{tr} \left( \bar{\mathbf{R}}_g \left( \frac{1}{\bar{b}} \mathbf{B}_g^H \hat{\mathbf{H}}_g \hat{\mathbf{H}}_g \mathbf{B}_g + \alpha \mathbf{I}_{\bar{b}} - z \bar{\mathbf{R}}_g \right)^{-1} \right)_{z=0} \\ &\asymp \frac{1}{\bar{b}} \text{tr}(\bar{\mathbf{R}}_g \bar{\mathbf{T}}_g'(0)), \quad (75) \end{aligned}$$

where

$$\bar{\mathbf{T}}_g(z) = \left( \frac{\bar{K}}{\bar{b}} \frac{\bar{\mathbf{R}}_g}{1 + \bar{\delta}_g} + \alpha \mathbf{I}_{\bar{b}} - z \bar{\mathbf{R}}_g \right)^{-1} \quad (76)$$

$$\bar{\mathbf{T}}_g'(0) = \bar{\mathbf{T}}_g(z) \left( \frac{\bar{K}}{\bar{b}} \frac{\bar{\mathbf{R}}_g \bar{\delta}_g'}{(1 + \bar{\delta}_g)^2} + \bar{\mathbf{R}}_g \right) \bar{\mathbf{T}}_g(z) \quad (77)$$

with  $\bar{\mathbf{T}}_g'(0)$  being the derivative  $\frac{d\bar{\mathbf{T}}_g(z)}{dz}$  at  $z = 0$ . Use of (77) into (75) and substitution into (74) gives  $\bar{\mathbf{Y}}_{gg}$ .

The derivation of IGI $_{g_k}$  follows similar lines with SGI $_{g_k}$ . In this case,  $\bar{\mathbf{Y}}_{gl}$  is obtained as given in Theorem 1. Having obtained all terms in the SINR, the proof is concluded.

## APPENDIX B

### PROOF OF PROPOSITION 1

The power optimization problem for group  $g$  is formulated as

$$\begin{aligned} (\mathcal{P}1) \quad & \max_{p_{g_k} \geq 0} \sum_{k=1}^{\bar{K}} \log_2(1 + p_{g_k} \nu_{g_k, \text{PGP}}) \\ & \text{s.t.} \quad \sum_{k=1}^{\bar{K}} p_{g_k} \leq P_g, \end{aligned} \quad (78)$$

where  $\nu_{g_k, \text{PGP}} = \bar{\gamma}_{g_k, \text{PGP}} / p_{g_k}$  does not depend on UE  $k$ . The solution to (P1) is obtained by the water-filling algorithm as

$$p_{g_k} = \left[ \mu - \frac{1}{\nu_{g_k, \text{PGP}}} \right]^+, \quad (79)$$

where  $\mu$  is a parameter defining the water level to satisfy the constraint  $\sum_{k=1}^K p_{gk} = P_g$  [54]. Since  $\nu_{gk, \text{PGP}}$  is identical for all UEs in group  $g$ , the optimal powers, maximizing (P1), are all equal and given by  $p_{gk}^* = P_g/\bar{K}$ . In such case, all user rates of group  $g$  become equal to  $R_g^*$  and the sum-rate of this group becomes  $\overline{\text{SR}}_{\text{PGP},g} = \bar{K}R_g^*$ .

## APPENDIX C PROOF OF PROPOSITION 2

First, we present the following lemma, which is required in the following derivations.

*Lemma 2:* The derivative of the trace  $\text{tr}(\mathbf{A}\bar{\mathbf{R}}_g)$  including the covariance matrix  $\bar{\mathbf{R}}_g$  and the matrix  $\mathbf{A}$  with respect to  $\mathbf{s}_l^*$ , where  $\mathbf{A}$  is independent of  $\mathbf{s}_l^*$ , is given by

$$\frac{\partial \text{tr}(\mathbf{A}\bar{\mathbf{R}}_g)}{\partial \mathbf{s}_l^*} = \text{diag}(\beta_{2,g} \mathbf{H}_1^H \mathbf{A} \mathbf{H}_1 \Phi \mathbf{R}_{\text{IRS},k}). \quad (80)$$

*Proof:* We have

$$\frac{\partial \text{tr}(\mathbf{A}\bar{\mathbf{R}}_g)}{\partial \mathbf{s}_l^*} = \frac{\partial \text{tr}(\beta_{2,g} \mathbf{A} \mathbf{H}_1 \Phi \mathbf{R}_{\text{IRS},k} \Phi^H \mathbf{H}_1^H)}{\partial \mathbf{s}_l^*} \quad (81)$$

$$= \frac{\partial \text{diag}(\beta_{2,g} \mathbf{H}_1^H \mathbf{A} \mathbf{H}_1 \Phi \mathbf{R}_{\text{IRS},k}) \mathbf{s}_l^*}{\partial \mathbf{s}_l^*} \quad (82)$$

$$= \text{diag}(\beta_{2,g} \mathbf{H}_1^H \mathbf{A} \mathbf{H}_1 \Phi \mathbf{R}_{\text{IRS},k}), \quad (83)$$

where (82) is obtained by using the property  $\text{tr}(\mathbf{A} \text{diag}(\mathbf{s}_l)) = (\text{diag}(\mathbf{A}))^\top \mathbf{s}_l$ . ■

Hereafter, we denote the partial derivative with respect to  $\mathbf{s}_l^*$  by  $(\cdot)'$ . Starting from the gradient of  $\bar{\gamma}_{gk, \text{PGP}}$ , with respect to  $\mathbf{s}_l^*$ , which relies on the quotient rule derivative, we obtain

$$\frac{\partial \bar{\gamma}_{g, \text{PGP}}}{\partial \mathbf{s}_l^*} = \frac{\frac{\partial S_g}{\partial \mathbf{s}_l^*} I_g - S_g \frac{\partial I_g}{\partial \mathbf{s}_l^*}}{I_g^2}, \quad (84)$$

where the calculation of the partial derivatives follows. In particular,  $S_g'$  is written as

$$S_g' = 2p_{gk} (1 - \tau^2) \bar{\delta}_g \bar{\delta}_g', \quad (85)$$

where

$$\bar{\delta}_g' = \frac{1}{b} \text{tr}(\bar{\mathbf{R}}_g' \mathbf{T}_g + \bar{\mathbf{R}}_g \mathbf{T}_g') \quad (86)$$

with  $\mathbf{T}_g'$ , being the derivative of an inverse matrix obtained by [55, Eq. 40] as  $\mathbf{T}_g' = -\mathbf{T}_g(\mathbf{T}_g^{-1})' \mathbf{T}_g$ . The derivative of  $\mathbf{T}_g^{-1}$  is given by

$$(\mathbf{T}_g^{-1})' = \frac{\bar{K} \bar{\mathbf{R}}_g' (1 + \bar{\delta}_g) - \bar{\mathbf{R}}_g \bar{\delta}_g'}{(1 + \bar{\delta}_g)^2}, \quad (87)$$

where  $\bar{\mathbf{R}}_g'$  is the derivative of  $\bar{\mathbf{R}}_g$  with respect to  $\mathbf{s}_l^*$  whose trace expression is given by Lemma 2. As a result, after substituting (86)-(87) into (85), we obtain  $S_g'$ .

The derivative of  $I_g$  is given by

$$\begin{aligned} I_g' &= \bar{\mathbf{Y}}_{gg}' \left(1 - \tau_g^2 \left(1 - (1 + \bar{\delta}_g)^2\right)\right) + 2\bar{\mathbf{Y}}_{gg} \tau_g^2 (1 + \bar{\delta}_g) \bar{\delta}_g' \\ &+ \sum_{l \neq g} (\bar{\lambda}_l \bar{\mathbf{Y}}_{gl} + \bar{\lambda}_l \bar{\mathbf{Y}}_{gl}') \frac{(1 + \bar{\delta}_g)^2}{\bar{\lambda}_g} \\ &+ \left(1 + \sum_{l \neq g} \bar{\lambda}_l^2 \bar{\mathbf{Y}}_{gl}\right) \frac{(1 + \bar{\delta}_g) (2\bar{\delta}_g' \bar{\lambda}_g - (1 + \bar{\delta}_g) \bar{\lambda}_g')}{\bar{\lambda}_g^2}, \end{aligned} \quad (88)$$

where

$$\bar{\mathbf{Y}}_{gg}' = \frac{P_g}{b} \left(1 - \frac{1}{\bar{K}}\right) \frac{\bar{m}'_{gg} (1 + \bar{\delta}_g) - 2\bar{m}_{gg} \bar{\delta}_g'}{(1 + \bar{\delta}_g)^3},$$

$$\bar{\mathbf{Y}}_{gl}' = \frac{1}{b} \frac{P_l \bar{m}'_{gl} (1 + \bar{\delta}_l) - 2\bar{m}_{gl} \bar{\delta}_l'}{(1 + \bar{\delta}_l)^3}, \quad (89)$$

$$\bar{\lambda}_g' = \frac{\rho_g}{b} P_g \frac{2m_g \bar{\delta}_g' - m_g' (1 + \bar{\delta}_g)}{(1 + \bar{\delta}_g)^3 \bar{\Psi}_g^2} \quad (90)$$

are derived easily by applying basic derivative rules. For the derivatives of  $\bar{m}_g$ ,  $\bar{m}_{gg}$ , and  $\bar{m}_{gl}$ , we notice that they have a similar expression. Thus, we resort to the definition of a new function  $f_g(\mathbf{A})$ , which includes them under specific values of its parameters  $g$  and  $\mathbf{A}$ , and we compute its derivative. Thus, by defining

$$f_g(\mathbf{A}) = \frac{\frac{1}{b} \text{tr}(\bar{\mathbf{R}}_g \mathbf{T}_g \mathbf{A} \mathbf{T}_g)}{1 - \frac{\bar{K}}{b} \text{tr}(\bar{\mathbf{R}}_g \mathbf{T}_g \bar{\mathbf{R}}_g \mathbf{T}_g)}, \quad (91)$$

we have  $\bar{m}_g = f_g(\mathbf{B}_g^H \mathbf{B}_g)$ ,  $\bar{m}_{gg} = f_g(\bar{\mathbf{R}}_g)$ , and  $\bar{m}_{gl} = f_l(\mathbf{B}_l^H \bar{\mathbf{R}}_g \mathbf{B}_l)$ . Note that the dependence on the RBM is found on  $\bar{\mathbf{R}}_g$  while  $\mathbf{B}_g$  appears no such dependence. Also, we define  $q(\mathbf{C}) = \text{tr}(\bar{\mathbf{R}}_g \mathbf{T}_g \mathbf{C} \mathbf{T}_g)$ . After some lengthy algebraic manipulations, we obtain  $f_g'(\mathbf{A})$  in (44), where  $q'(\mathbf{C})$  for  $\mathbf{C} = \mathbf{A}, \bar{\mathbf{R}}_g$  written as  $q'(\mathbf{C}) = \text{tr}(\bar{\mathbf{R}}_g' \mathbf{T}_g \mathbf{C} \mathbf{T}_g + \bar{\mathbf{R}}_g \mathbf{T}_g' \mathbf{C} \mathbf{T}_g + \bar{\mathbf{R}}_g \mathbf{T}_g \mathbf{C}' \mathbf{T}_g + \bar{\mathbf{R}}_g \mathbf{T}_g \mathbf{C} \mathbf{T}_g')$ , with  $\mathbf{T}_g'$  given above while the derivatives of the traces are obtained by using Lemma 2. Hence, after suitable substitutions in (44), we obtain  $\bar{m}'_g$ ,  $\bar{m}'_{gg}$ , and  $\bar{m}'_{gl}$ .

## REFERENCES

- [1] J. Zhang *et al.*, "Prospective multiple antenna technologies for beyond 5G," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1637–1660, 2020.
- [2] T. L. Marzetta *et al.*, *Fundamentals of Massive MIMO*. Cambridge University Press, 2016.
- [3] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106–112.
- [4] 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.
- [5] C. Huang *et al.*, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, 2019.
- [6] C. Pan *et al.*, "Multicell MIMO communications relying on intelligent reflecting surfaces," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5218–5233, 2020.
- [7] H. Guo *et al.*, "Weighted sum-rate maximization for reconfigurable intelligent surface aided wireless networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3064–3076, 2020.
- [8] Y. Han *et al.*, "Large intelligent surface-assisted wireless communication exploiting statistical CSI," *IEEE Trans. Veh. Tech.*, vol. 68, no. 8, pp. 8238–8242, 2019.
- [9] A. M. Elbir *et al.*, "Deep channel learning for large intelligent surfaces aided mm-Wave massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 9, pp. 1447–1451, 2020.
- [10] Y. Yang *et al.*, "Intelligent reflecting surface meets OFDM: Protocol design and rate maximization," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4522–4535, 2020.
- [11] Q. Wu and R. Zhang, "Joint active and passive beamforming optimization for intelligent reflecting surface assisted SWIPT under QoS constraints," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1735–1748, 2020.
- [12] W. Tang *et al.*, "Wireless communications with reconfigurable intelligent surface: Path loss modeling and experimental measurement," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 421–439, 2021.

- [13] P. von Butovitsch *et al.*, "Advanced antenna systems for 5G networks," *Ericsson White Paper*.
- [14] Z. Peng *et al.*, "Analysis and optimization for IRS-aided multi-pair communications relying on statistical CSI," *arXiv preprint arXiv:2007.11704*, 2020.
- [15] G. Yu *et al.*, "Design, analysis, and optimization of a large intelligent reflecting surface-aided B5G cellular internet of things," *IEEE Int. Things J.*, vol. 7, no. 9, pp. 8902–8916, 2020.
- [16] X. Hu *et al.*, "Location information aided multiple intelligent reflecting surface systems," *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7948–7962, 2020.
- [17] Q.-U.-A. Nadeem, A. Chaaban, and M. Debbah, "Opportunistic beamforming using an intelligent reflecting surface without instantaneous CSI," *IEEE Wireless Commun. Lett.*, vol. 10, no. 1, pp. 146–150, 2021.
- [18] A. Kammoun *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.
- [19] M.-M. Zhao *et al.*, "Intelligent reflecting surface enhanced wireless network: Two-timescale beamforming optimization," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5218–5233, 2020.
- [20] A. Papazafeiropoulos *et al.*, "Intelligent reflecting surface-assisted MU-MISO systems with imperfect hardware: Channel estimation, beamforming design," *IEEE Trans. Wireless Commun.*, 2021.
- [21] T. Van Chien *et al.*, "Outage probability analysis of IRS-assisted systems under spatially correlated channels," *IEEE Wireless Commun. Lett.*, pp. 1–1, 2021.
- [22] A. Papazafeiropoulos *et al.*, "Coverage probability of distributed IRS systems under spatially correlated channels," *IEEE Wireless Commun. Lett.*, pp. 1–1.
- [23] M.-A. Badiu and J. P. Coon, "Communication through a large reflecting surface with phase errors," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 184–188, 2019.
- [24] W. Chen *et al.*, "Angle-dependent phase shifter model for reconfigurable intelligent surfaces: Does the angle-reciprocity hold?" *IEEE Commun. Lett.*, vol. 24, no. 9, pp. 2060–2064, 2020.
- [25] C. Pan *et al.*, "Reconfigurable intelligent surfaces for 6G systems: Principles, applications, and research directions," *IEEE Commun. Mag.*, 2021.
- [26] W. Tang *et al.*, "On channel reciprocity in reconfigurable intelligent surface assisted wireless network," *arXiv preprint arXiv:2103.03753*.
- [27] Z. Jiang *et al.*, "Achievable rates of FDD massive MIMO systems with spatial channel correlation," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2868–2882, 2015.
- [28] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, 2006.
- [29] A. Adhikary *et al.*, "Joint spatial division and multiplexing-The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, 2013.
- [30] W. Chen *et al.*, "Adaptive bit partitioning for reconfigurable intelligent surface assisted FDD systems with limited feedback," *arXiv preprint arXiv:2011.14748*, 2020.
- [31] B. Guo, C. Sun, and M. Tao, "Two-way passive beamforming design for RIS-aided FDD communication systems," in *2021 IEEE Wireless Commun. Net. Conf. (WCNC)*. IEEE, 2021, pp. 1–6.
- [32] D. Shen and L. Dai, "Dimension reduced channel feedback for reconfigurable intelligent surface aided wireless communications," *IEEE Trans. Commun.*, vol. 69, no. 11, pp. 7748–7760, 2021.
- [33] E. Björnson and L. Sanguinetti, "Rayleigh fading modeling and channel hardening for reconfigurable intelligent surfaces," *IEEE Wireless Commun. Lett.*, vol. 10, no. 4, pp. 830–834, 2021.
- [34] S. Wagner *et al.*, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4537, July 2012.
- [35] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Select. Areas Commun.*, vol. 31, no. 2, pp. 160–171, February 2013.
- [36] A. K. Papazafeiropoulos and T. Ratnarajah, "Deterministic equivalent performance analysis of time-varying massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5795–5809, 2015.
- [37] T. Badloe, J. Mun, and J. Rho, "Metasurfaces-based absorption and reflection control: Perfect absorbers and reflectors," *J. of Nanomaterials*, vol. 2017, 2017.
- [38] D. Neumann, M. Joham, and W. Utschick, "Covariance matrix estimation in massive MIMO," *IEEE Signal Process. Lett.*, vol. 25, no. 6, pp. 863–867, 2018.
- [39] K. Upadhyaya and S. A. Vorobyov, "Covariance matrix estimation for massive MIMO," *IEEE Signal Process. Lett.*, vol. 25, no. 4, pp. 546–550, 2018.
- [40] Q. Nadeem *et al.*, "Intelligent reflecting surface-assisted multi-user MISO Communication: Channel estimation and beamforming design," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 661–680, 2020.
- [41] Ö. Özdoğan, E. Björnson, and E. G. Larsson, "Using intelligent reflecting surfaces for rank improvement in MIMO communications," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 9160–9164.
- [42] U. E. U. conformance specification Radio, "3rd generation partnership project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) conformance specification radio transmission and reception."
- [43] X. You, C. Fumeaux, and W. Withayachumnankul, "Tutorial on broadband transmissive metasurfaces for wavefront and polarization control of terahertz waves," *J. Applied Physics*, vol. 131, no. 6, p. 061101, 2022-02. [Online]. Available: <https://doi.org/10.1063/5.0077652>
- [44] W. Lee *et al.*, "Fully integrated 94-GHz dual-polarized TX and RX phased array chipset in SiGe BiCMOS operating up to 105° C," *IEEE J. Solid-State Circuits*, vol. 53, no. 9, pp. 2512–2531, 2018.
- [45] T. Chaloun *et al.*, "Wide-angle scanning active transmit/receive reflectarray," *IET Microwaves, Antennas & Propagation*, vol. 8, no. 11, pp. 811–818, 2014.
- [46] T. Dinc, A. Nagulu, and H. Krishnaswamy, "A millimeter-wave non-magnetic passive SOI CMOS circulator based on spatio-temporal conductivity modulation," *IEEE J. Solid-State Circuits*, vol. 52, no. 12, pp. 3276–3292, 2017.
- [47] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.
- [48] P. Billingsley, *Probability and measure*, 3rd ed. John Wiley & Sons, Inc., 2008.
- [49] A. W. van der Vaart, *Asymptotic statistics (Cambridge series in statistical and probabilistic mathematics)*. New York: Cambridge University Press, 2000.
- [50] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [51] Q. Shi *et al.*, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," vol. 59, no. 9, pp. 4331–4340, 2011.
- [52] E. Access, "Further advancements for E-UTRA physical layer aspects," *3GPP Technical Specification TR*, vol. 36, p. V2, 2010.
- [53] E. Björnson, Ö. Özdoğan, and E. G. Larsson, "Intelligent reflecting surface versus decode-and-forward: How large surfaces are needed to beat relaying?" *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 244–248, 2019.
- [54] D. P. Palomar and J. R. Fonollosa, "Practical algorithms for a family of waterfilling solutions," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 686–695.
- [55] K. B. Petersen and M. S. Pedersen, "The matrix cookbook, nov 2012," *URL <http://www2.imm.dtu.dk/pubdb/p.php>*, vol. 3274, p. 14, 2012.