Hybrid Relay-Reflecting Intelligent Surface-Assisted Wireless Communications

Nhan Thanh Nguyen, *Member, IEEE*, Quang-Doanh Vu, *Member, IEEE*, Kyungchun Lee, *Senior Member, IEEE*, and Markku Juntti, *Fellow, IEEE*

***Abstract*—Reconfigurable intelligent surface (RIS) has emerged as a cost- and energy-efficient solution to enhance the wireless communications capacity. However, recent studies show that a very large surface is required for a RIS-assisted communications system; otherwise, they may be outperformed by the conventional relay. Furthermore, the performance gain of a RIS can be considerably degraded by hardware impairments such as limited-resolution phase shifters. To overcome those challenges, we propose a hybrid relay-reflecting intelligent surface (HR-RIS) architecture, in which a single or few elements serve as active relays, while the remaining only reflect the incident signals. We propose two HR-RIS architectures, namely, the fixed and dynamic HR-RIS. The joint transmit beamforming at the base station and hybrid relay-reflecting at the HR-RIS are designed based on alternating optimization and successive convex approximation methods. The simulation results for a 4 $\times$ 2 multiple-input multiple-output system show that the proposed HR-RIS with only a single active element can achieve about 40% and 25% improvement in spectral efficiency and energy efficiency, respectively, with respect to the conventional RIS-aided system. The results also show that the HR-RIS can outperform the conventional relaying scheme in most of the considered scenarios.**

***Index Terms*—MIMO systems, hybrid relay-reflecting intelligent surface, spectral efficiency.**

I. Introduction

THE *reconfigurable intelligent surface* (RIS), also known as intelligent reflecting surface (IRS) [1] or large intelligent surface (LIS) [2], has recently emerged as a cost- and energy-efficient technology that can customize and program the physical propagation environment by reflecting radio waves in preferred directions [3]–[6]. RIS is a planar meta-surface consisting of a large number of *passive reflecting elements*, which are connected to a controller, allowing adjusting the phases of the incident signals. As a result, its phase shifts can be optimized to make the wireless channel between the transmitter and receiver more favorable for communications [1], [4], [5], [7]. Several potential technologies for RIS implementation exist. It can be constructed as an array of discrete phase shifters, which control the impedance values of the elements. Thereby the phase shifts can steer the reflected signal to the desired direction of the receiver. Such RIS is called the *discrete RIS* with no baseband processing capability [4]–[6]. The phase shift control is typically assumed to be designed in some active network element such as a base station (BS) or access point (AP). The phase shift values are then transferred via some control link from the BS/AP to the RIS. Furthermore, a RIS with active receive elements introduced by Taha *et al.* in [8], [9] has also been considered to enable the channel estimation at the RIS.

A. Related Works

RIS has attracted great interest in the literature with a particular focus on investigating its performance and design aspects. Specifically, Hu *et al.* in [10], [11] show that the capacity per RIS area-unit converges to $\frac{\bar{P}}{2\sigma^2}$ as the wavelength goes to zero, where $\bar{P}$ is the transmit power per area-unit, and σ^2 is the additive white Gaussian noise (AWGN) power spectral density. In [12], an adaptive phase shifter design based on hierarchical codebooks and limited feedback from the mobile station (MS) is proposed for a RIS-aided millimeter wave (mmWave) multiple-input multiple-output (MIMO) system for both accurate positioning and high data rate transmission. In [13], the asymptotic achievable rate of a RIS-aided downlink system is examined under practical reflection coefficients, and a passive beamformer and a modulation scheme that can be used in a RIS without interfering with existing users are proposed to increase the achievable system sum-rate. Furthermore, in [14], the propagation channel of point-to-point MIMO systems is enriched by using the RIS so that more paths with different spatial angles are added. As a result, the rank of the channel matrix is enhanced, and the multiplexing gain is achieved even when the direct path has low rank. In [15] and [16], the sub-6 GHz and mmWave multiple-input-single-output (MISO) downlink systems aided by RISs are investigated, respectively. These works show that a RIS equipped with N reflecting elements can offer a total beamforming gain of N^2 [15] and allows the received signal power to increase quadratically with N [16]. Moreover, the works in [17]–[22]

Manuscript received October 22, 2021; revised January 30, 2022; accepted March 4, 2022. Date of publication March 11, 2022; date of current version June 24, 2022. This work was supported in part by the Academy of Finland under 6G Flagship under Grant 318927, in part by EERA Project under Grant 332362, and in part by the University of Oulu Graduate School through Infotech Program. This article was presented in part at the IEEE International Conference on Communications Workshop, Montreal, Quebec, Canada, June 2021. The review of this article was coordinated by Prof. Daniel Benevides da Costa. *(Corresponding author: Nhan Thanh Nguyen.)*

Nhan Thanh Nguyen and Markku Juntti are with the Centre for Wireless Communications, University of Oulu, FI-90014 Oulu, Finland (e-mail: nhan.nguyen@oulu.fi; markku.juntti@oulu.fi).

Quang-Doanh Vu was with the Centre for Wireless Communications, University of Oulu, FI-90014 Oulu, Finland. He is now with the Mobile Networks, Nokia, 90650 Oulu, Finland (e-mail: quang-doanh.vu@nokia.com).

Kyungchun Lee is with the Department of Electrical and Information Engineering, Seoul National University of Science and Technology, Seoul 01811, Republic of Korea (e-mail: kclee@seoultech.ac.kr).

Digital Object Identifier 10.1109/TVT.2022.3158686

characterize the performance gains of practical RIS with limited resolution phase shifts and/or hardware impairments.

Another important line of studies on RIS focuses on the capacity/data rate maximization of RIS-aided communication systems [23]. In [24], a RIS-aided single-input-single-output (SISO) orthogonal frequency division multiplexing (OFDM) system is considered. Specifically, the achievable rate maximization problem has been solved by jointly optimizing the transmit power allocation and the RIS passive coefficients. In contrast, the MISO systems have been considered in [19], [25]–[28]. In particular, [25]–[27] consider the joint optimization of transmit beamforming and passive reflecting of RISs with ideal phase shifts to attain remarkable performance improvement. In contrast, Han *et al.* in [19] show that only two-bit-resolution phase shifts are sufficient for the RIS to provide a satisfactory capacity, which is similar to the finding in [28]. The capacity/rate optimization of RIS-assisted MIMO systems has been recently considered in [7], [21], [29]. In [7], an alternating optimization (AO) method has been proposed to efficiently find the locally optimal phase shifts of the RIS for both the narrowband frequency-flat and broadband frequency-selective MIMO OFDM systems. It is shown that with the proposed AO-based RIS design, the channel total power, rank, and condition number can be significantly improved for capacity enhancement [7]. Furthermore, the projected gradient method proposed in [30] can perform comparable with the AO scheme but requires lower computational cost. In [21], a RIS-enhanced full-duplex (FD) MIMO two-way communication system is considered. The system sum-rate is maximized through jointly optimizing the transmit beamforming and the RIS coefficient matrix. The same design is conducted in [31] but for simultaneous wireless information and power transfer (SWIPT). It is shown that with the aid of RISs, both the weighted sum-rate and harvested power significantly increase with the number of RIS elements [31]. The secure transmission is considered in [32] for an artificial noise (AN)-aided MIMO system with the presence of a multi-antenna eavesdropper. In this work, the transmit precoding matrix at the BS, covariance matrix of AN, and the RIS phase shifts are jointly optimized, resulting to a remarkable improvement in the secrecy rate. To realize the practical deployment of RISs, Wang *et al.* in [33] propose a novel three-phase channel estimation framework for RIS-assisted multiuser communication systems. Whereas, Liu *et al.* [34] leverage the learning ability of convolutional neural network for this purpose.

B. Motivations and Contributions

In this paper, we propose a novel *semi-passive RIS-aided beamforming* concept in which active beamforming (relaying) can be applied together with passive beamforming (reflecting). The idea is to activate a few elements of the RIS by connecting them with radio frequency (RF) chains and power amplifiers (PAs) or by deploying the reflection amplifiers (RAs) [35], [36], allowing them to not only modify the phases but also amplify the power of the incident signals. In this respect, these few elements become active FD amplify-and-forward (AF) relaying elements, and the conventional RIS becomes a *hybrid relay-reflecting intelligent surface* (HR-RIS). We use this term throughout the paper to refer to the proposed architecture. We note that the proposed HR-RIS herein is similar to the *hybrid active/passive RIS* architectures introduced in [8] and [37] in the sense that a subset of elements in the surface is capable of performing active processing. However, they have different purposes and operations. Specifically, the introduction of active elements in [8] and [37] enables efficient channel sensing and/or passive beamforming of the RIS. In contrast, those in the HR-RIS offer the capability of signal amplification as relay elements to significantly improve the system performance. The proposal of HR-RIS is motivated by recent comparisons between the conventional relay and RIS and the practical deployment of RIS with hardware impairments. First, a main limitation of the RIS compared to the active relays is the fact that its pure reflection limits the degrees of freedom in the beamforming. Second, based on the performance comparisons of relay and RIS in [15] and [38], we found that the HR-RIS is promising to provide a remarkable performance improvement. Specifically, it is shown in both [38] and [15] that a very large-sized RIS is required to outperform decode-and-forward (DF) relaying. Unless, it can be easily surpassed by a small-sized half-duplex relay. Furthermore, for a sufficiently large RIS, a small increase in the number of elements does not result in a significant performance improvement. These observations imply that replacing a few passive elements of the RIS by active ones can provide a remarkable active relaying gain to the system. In this regard, HR-RIS with only a few active elements can leverage the advantages of both RIS and relay, i.e., the passive reflecting and active relaying capabilities, with the requirement of low power consumption. Furthermore, the relaying coefficients of the active elements can be optimized to compensate for the considerable performance loss due to the limited-resolution phase shifts of practical RISs [18]–[21].

We recognize that there are several practical open problems related to the practical and efficient implementation of HR-RIS. Furthermore, we assume perfect channel state information (CSI) [7], [15], [18], [39] for the beamforming design. However, with the proposed HR-RIS architecture, the assumption is less restrictive than with a passive RIS, because the active processing chains can be readily used for channel estimation using their built-in active elements. Specifically, the CSI acquisition for the HR-RIS-aided systems can be conducted by the approaches proposed in [8], [40]. Thereby, our paper focuses on proposing the concept and demonstrates its significant system-level potential. This gives more understanding on the fundamental tradeoffs between RIS and relaying as part of a MIMO communication link. The main contributions of the paper are as follows:

- We propose the novel HR-RIS architecture, which enables a hybrid active-passive beamforming scheme rather than fully-passive beamforming as in the conventional RIS. It requires only a single or few active elements which serve as active relays to achieve significant performance gain. Particularly, HR-RIS exploits the benefits of both the relaying and passive reflecting schemes.
- We propose two HR-RIS architectures, namely, *fixed* and *dynamic* HR-RIS. In the former, the active elements are

fixed in manufacture; in contrast, those in the latter can be adaptively configured to improve the performance and save power consumption. The coefficient matrix of the HR-RIS and that of the active transmit beamforming matrix are optimized in the formulated SE maximization problem, which is solved by the AO and successive convex approximation (SCA) schemes. Furthermore, the favorable deployment and performance gains of the HR-RIS are derived analytically and numerically verified by computer simulations.

- The total power consumption and energy efficiency (EE) of the system with the proposed HR-RIS is analyzed and compared to that with the conventional RIS. It is shown that the former requires higher power consumption than the latter because additional power is consumed for the active processing. However, with a small number of active elements in the proposed HR-RIS, the improvement in both the SE and EE is guaranteed, as numerically demonstrated by the simulation results.
- Furthermore, we numerically show that the proposed HR-RIS generally outperforms the conventional FD-AF relaying scheme. The reverse only occurs when the relay is equipped with numerous relaying antennas and has a sufficiently large power budget. The paper bridges the theoretical gap between a passive RIS and active AF relay aiding a MIMO communication link. This gives more insight into the fundamental performance of each of the approaches and attainable SE and EE.

We note that a part of this work has been presented in [41]. Specifically, in [41], we introduce the fixed HR-RIS architecture and show its potential in improving the SE performance of the MIMO system. In this work, we further introduce the dynamic HR-RIS. Moreover, we consider a more general system model where the direct BS-MS channel is taken into consideration, and both the transmit beamforming at the BS and the semi-passive beamforming at the HR-RIS are optimized. We also investigate the power consumption and EE of the proposed HR-RIS architectures.

Structure: The rest of the paper is organized as follows. In Section II, we introduce the concept of the HR-RIS and the system model of the HR-RIS-aided MIMO system, and the problem of SE maximization is formulated. Its efficient solution is found in Section III. Based on that, the beamforming matrices of the fixed and dynamic HR-RIS-aided MIMO systems are derived in Section IV. Their power consumption is investigated in Section V. Simulation results are shown in Section VI, and finally, conclusions are drawn in Section VII.

Notations: Throughout this paper, numbers, vectors, and matrices are denoted by lower-case, bold-face lower-case, and bold-face upper-case letters, respectively. $(\cdot)^*$ and $(\cdot)^H$ denote the conjugate of a complex number and the conjugate transpose of a matrix or vector, respectively, and $(\cdot)^{\frac{H}{2}}$ represents $[(\cdot)^{\frac{1}{2}}]^H$. $\boldsymbol{I}_N$ denotes the identity matrix of size $N \times N$, and $\text{diag}\{a_1, \ldots, a_N\}$ represents a diagonal matrix with diagonal entries $a_1, \ldots, a_N$. Furthermore, $|\cdot|$ denotes either the absolute value of a scalar or determinant of a matrix, and the expectation operator is denoted by $\mathbb{E}\{\cdot\}$.

II. Proposed HR-RIS Architecture, System Model, and SE Problem Formulation

A. HR-RIS Beamforming Architecture

The HR-RIS is equipped with N elements, including K active relaying and M passive reflecting elements, with $M + K = N$. We consider the case that the amplitude of a passive element is fixed to unity to maximize the received signal power and to enhance the channel capacity, as widely considered in the literature [42]. In contrast, the active elements of HR-RISs can be optimized to not only adjust the phase but also amplify the incident signal for effective relaying. For $K = 0$, the HR-RIS serves as the conventional RIS. We note that an active element would require more power consumption for processing than the passive one. Therefore, we are interested in the case $1 \leq K \ll M$ for practical deployment. We now introduce notations to mathematically model the HR-RIS. In particular, let $\mathcal{A}$ denote the index set of the K active elements, $\mathcal{A} \subset \{1, 2, \ldots, N\}$, and let α_n denote the coefficient of the nth element, which is given as

$$\alpha_n = \begin{cases} |\alpha_n| \, e^{j\theta_n}, & \text{if } n \in \mathcal{A} \\ e^{j\theta_n}, & \text{otherwise} \end{cases}, \tag{1}$$

where $\theta_n \in [0, 2\pi)$ is the phase shift and $|\alpha_n| = 1$ if $n \notin \mathcal{A}$. Let $\boldsymbol{\Upsilon} = \text{diag}\{\alpha_1, \ldots, \alpha_N\} \in \mathbb{C}^{N \times N}$ be the diagonal matrix of the coefficients. We define an additive decomposition for it as $\boldsymbol{\Upsilon} = \boldsymbol{\Phi} + \boldsymbol{\Psi}$, where $\boldsymbol{\Psi} = \mathbb{1}_N^{\mathcal{A}} \circ \boldsymbol{\Upsilon}$ and $\boldsymbol{\Phi} = (\boldsymbol{I}_N - \mathbb{1}_N^{\mathcal{A}}) \circ \boldsymbol{\Upsilon}$ contain the active relaying and passive reflecting coefficients, respectively. Here, $\mathbb{1}_N^{\mathcal{A}}$ is an $N \times N$ diagonal matrix whose non-zero diagonal elements are all unity and have positions determined by $\mathcal{A}$, and $\circ$ represents a Hadamard product.

We consider two HR-RIS architectures, including the fixed and dynamic HR-RIS. In the fixed one, the number of active elements and their positions, i.e., set $\mathcal{A}$, are predefined and fixed as illustrated in Fig. 1(a). In contrast, in the dynamic HR-RIS architecture, the number and positions of active elements can be dynamically changed according to the propagation condition. In other words, set $\mathcal{A}$ can be a design parameter in this architecture. More specifically, there are a number of RF-PA chains each can be turned on/off. An element of HR-RIS is active if it connects to an "ON" RF-PA chain; otherwise, it serves as a passive reflecting element [8], [9], [43]. Such connection can be done via a switching (SW) network as illustrated in Fig. 1(b). Furthermore, the signal processing of the active elements is controlled by an HR-RIS controller and baseband unit [8], [9]. In this paper, both the fixed and dynamic HR-RIS architectures are considered for the SE maximization problem.

The HR-RIS may potentially be realized by recently introduced low-power complementary metal-oxide semiconductor (CMOS)-based technologies, such as the reflection amplifier [35], [36], which is capable of not only turning the phases but also amplifying the amplitudes of the incident signals. Thereby a possible implementation of the proposed architecture could produce a meta-surface with reflection amplifier elements. However, this method can be cost-inefficient because the number of elements in the HR-RIS is large, while only a few active

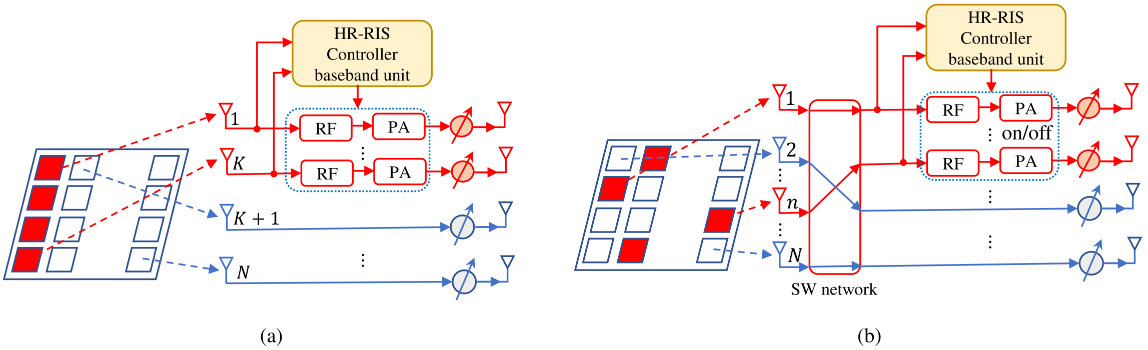


Fig. 1. Illustration of the fixed and dynamic HR-RIS architectures. (a) *Fixed HR-RIS architecture.* All *K* RF-PA chains are ON. (b) *Dynamic HR-RIS architecture.* $\mathbb{A}^{\star}$ out of *K* RF-PA chains are ON.

elements are required, as will be proven in the sequel. Therefore, a more practical solution appears to be upgrading the conventional RIS by replacing some elements with the reflection amplifiers. Compared to the RIS, HR-RIS requires additional complexity for hardware implementation and signal processing of the active elements. However, those are only required for a single or few active elements. Considering that the total number of elements in the conventional RIS is very large, the hardware and computational costs increase only moderately. However, the detailed HR-RIS implementation is out of the scope of this paper. We focus on proposing it and analyzing its potential for system performance improvement.

B. *System Model*

We consider a downlink transmission between a BS and an MS which is aided by the HR-RIS. Let us denote by N_t and N_r the numbers of antennas equipped at the BS and MS, respectively, where $N_t \geq 1$ and $N_r \geq 1$. Let $\boldsymbol{H}_t \in \mathbb{C}^{N \times N_t}$, $\boldsymbol{H}_r \in \mathbb{C}^{N_r \times N}$, and $\boldsymbol{H}_d \in \mathbb{C}^{N_r \times N_t}$ denote the channel between the BS and the HR-RIS, between the HR-RIS and the MS, and the direct BS-MS channel, respectively. The transmitted signal vector at the BS can be written as $\boldsymbol{s} = \boldsymbol{F}\boldsymbol{x}$ ($\boldsymbol{s} \in \mathbb{C}^{N_t \times 1}$), where $\boldsymbol{x} \in \mathbb{C}^{N_s \times 1}$ is the vector of N_s symbols such that $\mathbb{E}\{\boldsymbol{x}\boldsymbol{x}^H\} = \boldsymbol{I}_{N_s}$, and $\boldsymbol{F} \in \mathbb{C}^{N_t \times N_s}$ be the transmit beamforming matrix. Without loss of generality, we assume that $N_s = \min\{N_r, N_t\}$, i.e., the number of data streams is equal to the rank of the channel matrix. The average transmit power at the BS is constrained by $\mathbb{E}\{\|\boldsymbol{s}\|^2\} \leq P_{\mathrm{BS}}$, which is equivalent to $\|\boldsymbol{F}\|^2 \leq P_{\mathrm{BS}}$, where P_{BS} is the maximum transmit power at the BS. As a result, the received signals at the MS can be written as

$$\boldsymbol{y} = (\boldsymbol{H}_d + \boldsymbol{H}_r\boldsymbol{\Upsilon}\boldsymbol{H}_t)\boldsymbol{F}\boldsymbol{x} + \boldsymbol{n}, \tag{2}$$

where

$$\boldsymbol{n} \triangleq \boldsymbol{H}_r\boldsymbol{\Psi}\left(\boldsymbol{n}_{\mathrm{H}} + \boldsymbol{n}_{\mathrm{SI}}\right) + \boldsymbol{n}_{\mathrm{MS}} \tag{3}$$

is the total effective noise vector at the MS, $\boldsymbol{n} \in \mathbb{C}^{N_r \times 1}$. Here, $\boldsymbol{n}_{\mathrm{H}} = [n_{\mathrm{H},1}, n_{\mathrm{H},2}, \ldots, n_{\mathrm{H},N}]^T \in \mathbb{C}^{N \times 1}$ and $\boldsymbol{n}_{\mathrm{SI}} = [n_{\mathrm{SI},1}, n_{\mathrm{SI},2}, \ldots, n_{\mathrm{SI},N}]^T \in \mathbb{C}^{N \times 1}$ represent the AWGN vector and possible residual self-interference caused by the HR-RIS, respectively. Here, the self-interference is generated due to the full-duplex amplify-and-forward operation of the active elements at the HR-RIS. We note that at the HR-RIS, only active elements cause noise and self-interference. Therefore, we have $n_{\mathrm{H},i} \sim \mathcal{CN}(0, \sigma_{\mathrm{H}}^2)$ if $i \in \mathcal{A}$; otherwise, $n_{\mathrm{H},i} = 0$, $i \notin \mathcal{A}$. Furthermore, we adopt the residual self-interference model in [44], where the residual self-interference components are modeled as zero-mean additive Gaussian noise samples [1], i.e., $n_{\mathrm{SI},i} \sim \mathcal{CN}(0, \sigma_{\mathrm{SI}}^2)$ if $i \in \mathcal{A}$; otherwise, $n_{\mathrm{SI},i} = 0$, $i \notin \mathcal{A}$. Consequently, $\boldsymbol{H}_r\boldsymbol{\Psi}(\boldsymbol{n}_{\mathrm{H}} + \boldsymbol{n}_{\mathrm{SI}}) \in \mathbb{C}^{N_r \times 1}$ represents the effective noise vector at the MS received from the HR-RIS. Furthermore, in (3), $\boldsymbol{n}_{\mathrm{MS}} \sim \mathcal{CN}(\boldsymbol{0}, \sigma_{\mathrm{MS}}^2\boldsymbol{I}_{N_r})$ is the AWGN vector at the MS. For notational simplicity, we assume that $\sigma_{\mathrm{H}}^2 = \sigma_{\mathrm{MS}}^2 = \sigma^2$, and $\sigma_{\mathrm{SI}}^2 = \eta\sigma^2$. Thus, we have $\boldsymbol{n} \sim \mathcal{CN}(\boldsymbol{0}, \tilde{\boldsymbol{R}})$, where

$$\tilde{\boldsymbol{R}} \triangleq \sigma_{\mathrm{MS}}^2\boldsymbol{I}_{N_r} + \boldsymbol{H}_r\boldsymbol{\Psi}\left(\sigma_{\mathrm{H}}^2\boldsymbol{I}_N + \sigma_{\mathrm{SI}}^2\boldsymbol{I}_N\right)\boldsymbol{\Psi}^H\boldsymbol{H}_r^H = \sigma^2\boldsymbol{R} \tag{4}$$

is the aggregate noise covariance matrix, with $\boldsymbol{R} \triangleq \boldsymbol{I}_{N_r} + (\eta + 1)\boldsymbol{H}_r\boldsymbol{\Psi}\boldsymbol{\Psi}^H\boldsymbol{H}_r^H$.

We note that in an active element, the incident signal is only amplified without decoding. Therefore, the delay when the signals go through active elements is much smaller than the coherence interval and has a negligible impact on the channel estimation and signal combining at the receiver as in [46]–[50]. Similarly, the reflection amplifier-based AF relay does not cause significant residual self-interference levels [47], but we still include it in our analysis for completeness.

C. *Problem Formulation*

Based on (2) and (4), the SE of an HR-RIS-aided MIMO system can be expressed as

$$f_0(\boldsymbol{F}, \{\alpha_n\}) = \log_2\left|\boldsymbol{I}_{N_r} + \rho(\boldsymbol{H}_d + \boldsymbol{H}_r\boldsymbol{\Upsilon}\boldsymbol{H}_t)\boldsymbol{F}\boldsymbol{F}^H \times (\boldsymbol{H}_d + \boldsymbol{H}_r\boldsymbol{\Upsilon}\boldsymbol{H}_t)^H\boldsymbol{R}^{-1}\right|, \tag{5}$$

where $\{\alpha_n\} \triangleq \{\alpha_1, \alpha_2, \ldots, \alpha_N\}$ represents the set of all the HR-RIS relay/reflecting coefficients, and $\rho = \frac{1}{\sigma^2}$. Let $\tilde{\sigma}^2 \triangleq (\eta +$

[1] This residual self-interference model is based on the fact that there are numerous sources of imperfections in the RF chain [44], and the experiment in [45] shows that the residual self-interference can be eliminated to be as low as 1 dB independent of the transmit power and the number of transmit antennas.

$1)\sigma^2$. Then, the transmit power of the HR-RIS active elements is given as

$$P_{\mathrm{a}}(\boldsymbol{F},\{\alpha_n\}) \triangleq \mathrm{trace}\boldsymbol{\Psi}\left(\boldsymbol{H}_{\mathrm{t}}\mathbb{E}\left\{\boldsymbol{s}\boldsymbol{s}^{\mathrm{H}}\right\}\boldsymbol{H}_{\mathrm{t}}^{\mathrm{H}} + \tilde{\sigma}^2\boldsymbol{I}_{\mathrm{N}}\right)\boldsymbol{\Psi}^{\mathrm{H}}$$
$$= \mathrm{trace}\boldsymbol{\Psi}\left(\boldsymbol{H}_{\mathrm{t}}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}_{\mathrm{t}}^{\mathrm{H}} + \tilde{\sigma}^2\boldsymbol{I}_{\mathrm{N}}\right)\boldsymbol{\Psi}^{\mathrm{H}}. \quad (6)$$

We aim at investigating the potential performance of the HR-RIS in terms of SE (compared with the conventional passive RIS). Therefore, we focus on the the problem of jointly designing the transmit beamformer at the BS and the relay/reflection coefficients of the HR-RIS to maximize the SE, i.e.,

$$(\mathrm{P0}) \quad \underset{\boldsymbol{F},\{\alpha_n\}}{\mathrm{maximize}} \quad f_0(\boldsymbol{F},\{\alpha_n\}) \quad (7\mathrm{a})$$
$$\text{subject to} \quad \|\boldsymbol{F}\|^2 \leq P_{\mathrm{BS}} \quad (7\mathrm{b})$$
$$|\alpha_n| = 1 \text{ for } n \notin \mathcal{A} \quad (7\mathrm{c})$$
$$P_{\mathrm{a}}(\boldsymbol{F},\{\alpha_n\}) \leq P_{\mathrm{a}}^{\max}, \quad (7\mathrm{d})$$

where $P_{\mathrm{a}}^{\max}$ is the power budget of the HR-RIS active elements. Function $f_0(\boldsymbol{F},\{\alpha_n\})$ is nonconvex with respect to $(\boldsymbol{F},\{\alpha_n\})$. In addition, the feasible set of (P0) is nonconvex due to the unit-modulus constraint (7c). Thus, problem (P0) is intractable, and it is challenging to find an optimal solution.

III. Efficient Solution to (P0)

To overcome the challenges in the problem of joint active and semi-passive beamforming design, i.e., jointly optimizing $\boldsymbol{F}$ and $\{\alpha_n\}$ in (P0), we adopt an AO framework and decouple the optimization of these two variables. Specifically, we will alternately solve for $\boldsymbol{F}$ and $\{\alpha_n\}$ while fixing the other, as presented in the subsequent subsections.

A. Active Transmit Beamforming Design With Given HR-RIS Coefficients

In this section, we first aim at solving the active transmit beamforming problem associated with the variable $\boldsymbol{F}$ with the given HR-RIS relay/reflecting coefficients $\{\alpha_n\}$. For ease of exposition, we denote by $\boldsymbol{G}$ the effective channel between the BS and MS, i.e., $\boldsymbol{G} \triangleq \boldsymbol{H}_d + \boldsymbol{H}_r\boldsymbol{\Upsilon}\boldsymbol{H}_t$. Then, given $\{\alpha_n\}$, the problem (P_0) can be rewritten as

$$(\mathrm{P}_{\mathrm{tx}}) \quad \underset{\boldsymbol{F}}{\mathrm{maximize}} \quad f_0(\boldsymbol{F}) = \log_2\left|\boldsymbol{I}_{N_r} + \rho\boldsymbol{G}\boldsymbol{F}\boldsymbol{F}^H\boldsymbol{G}^H\boldsymbol{R}^{-1}\right|$$
$$\text{subject to} \quad (7\mathrm{b}), (7\mathrm{d}).$$

Because $\boldsymbol{R}$ is semi-definite, we can express $\boldsymbol{R} = \boldsymbol{R}^{\frac{H}{2}}\boldsymbol{R}^{\frac{1}{2}}$, and thus, the objective function of $(\mathrm{P}_{\mathrm{tx}})$ can be rewritten as

$$f_0(\boldsymbol{F}) = \log_2\left|\boldsymbol{I}_{N_r} + \rho\boldsymbol{G}\boldsymbol{F}\boldsymbol{F}^H\boldsymbol{G}^H\boldsymbol{R}^{-\frac{H}{2}}\boldsymbol{R}^{-\frac{1}{2}}\right|$$
$$= \log_2\left|\boldsymbol{I}_{N_r} + \rho\tilde{\boldsymbol{G}}\boldsymbol{F}\boldsymbol{F}^H\tilde{\boldsymbol{G}}^H\right|,$$

where $\tilde{\boldsymbol{G}} \triangleq \boldsymbol{R}^{-\frac{1}{2}}\boldsymbol{G}$. Given $\{\alpha_n\}$, both $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Psi}$ are determined, and so is $\tilde{\boldsymbol{G}}$. Problem $(\mathrm{P}_{\mathrm{tx}})$ is convex with respect to $\boldsymbol{F}$, and its optimal solution can be obtained based on the eigenmode transmission. Specifically, let $\boldsymbol{V} \in \mathbb{C}^{N_t \times N_s}$ consist of N_s columns as N_s right-singular vectors associated with the N_s largest singular values of $\tilde{\boldsymbol{G}}$. Then, the optimal solution to $\boldsymbol{F}$ is given by

$$\boldsymbol{F}^{\star} = \boldsymbol{V}\boldsymbol{\Sigma}^{1/2}, \quad (8)$$

where $\boldsymbol{\Sigma}^{1/2} = \sqrt{p_1}, \sqrt{p_2}, \ldots, \sqrt{p_{N_s}}$; here, p_i represents the power amount allocated to the ith data stream satisfying (7b) and (7d). With the solution in (8), we have $f_0(\boldsymbol{F}^{\star}) = \sum_{i=1}^{N_s}\log_2(1 + \rho\varkappa_i^2 p_i)$, where $\varkappa_i$ is the ith largest singular value of $\tilde{\boldsymbol{G}}$ [7]. Furthermore, because $\boldsymbol{\Sigma}$ is diagonal, it follows that $\|\boldsymbol{F}^{\star}\|^2 = \sum_{i=1}^{N_s} p_i$ and

$$P_{\mathrm{a}}(\boldsymbol{F}^{\star},\{\alpha_n\}) = \mathrm{trace}\boldsymbol{\Sigma}\tilde{\boldsymbol{V}} + \mathrm{trace}\tilde{\sigma}^2\boldsymbol{\Psi}\boldsymbol{\Psi}^{\mathrm{H}}$$
$$= \sum_{i=1}^{N_s} p_i\tilde{v}_i + \tilde{\sigma}^2\sum_{n\in\mathcal{A}}|\alpha_n|^2,$$

where $\tilde{\boldsymbol{V}} = \boldsymbol{V}^H\boldsymbol{H}_t^H\boldsymbol{\Psi}^H\boldsymbol{\Psi}\boldsymbol{H}_t\boldsymbol{V}$, and $\tilde{v}_i$ is the ith diagonal element of $\tilde{\boldsymbol{V}}$. As a result, $\{p_i\}$ can be obtained by solving problem

$$(\mathrm{P}_p) \quad \underset{\{p_i\}}{\mathrm{maximize}} \quad \sum_{i=1}^{N_s}\log_2(1 + \rho\varkappa_i^2 p_i) \quad (9\mathrm{a})$$
$$\text{subject to} \quad \sum_{i=1}^{N_s} p_i \leq P_{\mathrm{BS}} \quad (9\mathrm{b})$$
$$\sum_{i=1}^{N_s} p_i\tilde{v}_i \leq P_{\mathrm{a}}^{\max} - \tilde{\sigma}^2\sum_{n\in\mathcal{A}}|\alpha_n|^2, \quad (9\mathrm{c})$$

which is convex and can be solved with existing optimization tools such as CVX or YALMIP-MOSEK.[2]

B. Semi-Passive Beamforming Design With Given Active Transmit Beamformer

Here, we find efficient solution to the reflecting/relay coefficients of the HR-RIS, i.e., $\{\alpha_n\}$, with given transmit beamformer $\boldsymbol{F}^{\star}$. Specifically, we consider the problem

$$(\mathrm{P}_{\mathrm{H}}) \quad \underset{\{\alpha_n\}}{\mathrm{maximize}} \quad f_0(\{\alpha_n\})$$
$$\text{subject to} \quad (7\mathrm{c}), (7\mathrm{d}).$$

This problem is intractable because $f_0(\{\alpha_n\})$ is nonconvex with respect to $\{\alpha_n\}$, and the constraint (7c) leads to a nonconvex feasible set. As the first step of developing an efficient solution to $(\mathrm{P}_{\mathrm{H}})$, we approximate the problem into a more tractable one. Specifically, $f_0(\{\alpha_n\})$ is upper bounded by $f(\{\alpha_n\})$, as follows:

$$f_0(\{\alpha_n\}) = \log_2\left|\left(\boldsymbol{R} + \rho\boldsymbol{G}\boldsymbol{F}^{\star}\boldsymbol{F}^{\star H}\boldsymbol{G}^H\right)\boldsymbol{R}^{-1}\right|$$
$$= \log_2\left|\boldsymbol{R} + \rho\boldsymbol{G}\boldsymbol{F}^{\star}\boldsymbol{F}^{\star H}\boldsymbol{G}^H\right| - \log_2|\boldsymbol{R}| \quad (10)$$
$$\leq \log_2\left|\boldsymbol{I}_{N_r} + (\eta+1)\boldsymbol{H}_r\boldsymbol{\Psi}\boldsymbol{\Psi}^H\boldsymbol{H}_r^H + \rho\boldsymbol{G}\boldsymbol{F}^{\star}\boldsymbol{F}^{\star H}\boldsymbol{G}^H\right| \quad (11)$$
$$\triangleq f(\{\alpha_n\}),$$

[2] We note that for the conventional passive RIS, we have $\mathcal{A} = \emptyset$ and $\boldsymbol{\Psi} = \boldsymbol{0}$, leading to $\tilde{v}_i = 0, \forall i$. Thus, in this case, constraint (9c) always satisfies and the optimal power allocation at the BS is given as $p_i^{\star} = \max\{\frac{1}{\bar{p}} - \frac{\sigma^2}{\varkappa_i^2}, 0\}$, where $\bar{p}$ satisfies $\sum_{i=1}^{N_s} p_i = P_{\mathrm{BS}}$ [7].

where the inequality in (11) follows by substituting the expression of $\boldsymbol{R}$ to (10). Clearly, the equality occurs when $\mathcal{A} = \emptyset$. In addition, the smaller the term $\log_2 |\boldsymbol{R}|$ is, the tighter the bound is. We can also observe that $\log_2 |\boldsymbol{R}|$ is small when the path loss is large and the number of active elements is small (since $\boldsymbol{\Psi}$ becomes very sparse when only a single or few active elements are employed at the HR-RIS). These observations motivate us to arrive at the following approximation of $(\mathrm{P_H})$:

$$(\mathrm{P'_H}) \quad \underset{\{\alpha_n\}}{\text{maximize}} \quad f(\{\alpha_n\})$$
$$\text{subject to} \quad (7\text{c}), (7\text{d}).$$

Although $f(\{\alpha_n\})$ is still nonconvex with respect to $\{\alpha_n\}$, its structure allows the development of an efficient solution via the AO method. It results in closed-form solutions not only to the HR-RIS phase shifts but also to the power amplification coefficients of the active relay elements, providing important insights into the design and deployment of the HR-RIS. Therefore, we utilize the AO method in [7] to solve the problem $(\mathrm{P'_H})$. Here, it should be mentioned that the derivations in [7] cannot be directly applied, because the objective function and constraints in $(\mathrm{P'_H})$ are different from those in [7]. Specifically, the term $\boldsymbol{H}_r \boldsymbol{\Psi} \boldsymbol{\Psi}^H \boldsymbol{H}_r^H$ in $f(\{\alpha_n\})$, variables α_n with $|\alpha_n| \neq 1$ for $n \in \mathcal{A}$, as well as power constraint (7d) do not appear in [7]. Therefore, we find the solutions $\{\alpha_n^\star\}$ below.

The proposed solution is a sequential procedure: In each iteration, a specific coefficient of HR-RIS is updated when the others are fixed. We first further expand $f(\{\alpha_n\})$ in (11) to extract the role of variable α_n, for any $n \in \{1, \ldots, N\}$. Specifically, by recalling that $\boldsymbol{G} = \boldsymbol{H}_d + \boldsymbol{H}_r \boldsymbol{\Upsilon} \boldsymbol{H}_t$ and by denoting $\tilde{\boldsymbol{H}}_d \triangleq \boldsymbol{H}_d \boldsymbol{F}^\star$ and $\tilde{\boldsymbol{H}}_t \triangleq \boldsymbol{H}_t \boldsymbol{F}^\star$, we can write

$$\begin{aligned} f(\{\alpha_n\}) = \log_2 \big| &\boldsymbol{I}_{N_r} + (\eta+1) \boldsymbol{H}_r \boldsymbol{\Psi} \boldsymbol{\Psi}^H \boldsymbol{H}_r^H \\ &+ \rho (\tilde{\boldsymbol{H}}_d + \boldsymbol{H}_r \boldsymbol{\Upsilon} \tilde{\boldsymbol{H}}_t)(\tilde{\boldsymbol{H}}_d + \boldsymbol{H}_r \boldsymbol{\Upsilon} \tilde{\boldsymbol{H}}_t)^H \Big| . \end{aligned} \tag{12}$$

Furthermore, let $\boldsymbol{t}_n^H \in \mathbb{C}^{1 \times N_t}$ be the nth row of $\tilde{\boldsymbol{H}}_t$ and $\boldsymbol{r}_n \in \mathbb{C}^{N_r \times 1}$ be the nth column of $\boldsymbol{H}_r$, i.e., $\tilde{\boldsymbol{H}}_t = [\boldsymbol{t}_1, \ldots, \boldsymbol{t}_N]^H$ and $\boldsymbol{H}_r = [\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N]$. Since $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Psi}$ are diagonal matrices, we have $\boldsymbol{H}_r \boldsymbol{\Upsilon} \tilde{\boldsymbol{H}}_t = \sum_{n=1}^N \alpha_n \boldsymbol{r}_n \boldsymbol{t}_n^H$ and $\boldsymbol{H}_r \boldsymbol{\Psi} = [\psi_1 \boldsymbol{r}_1, \ldots, \psi_N \boldsymbol{r}_N]$, where ψ_n is the nth diagonal element of $\boldsymbol{\Psi}$, and we have

$$\psi_n = \begin{cases} \alpha_n, & \text{if } n \in \mathcal{A} \\ 0, & \text{otherwise} \end{cases},$$

based on the definition of $\boldsymbol{\Psi}$. Consequently, we can express $f(\{\alpha_n\})$ as

$$f(\{\alpha_n\}) = \log_2 \left| \boldsymbol{A}_n + |\alpha_n|^2 \boldsymbol{B}_n + \alpha_n \boldsymbol{C}_n + \alpha_n^* \boldsymbol{C}_n^H \right|, \tag{13}$$

where matrices $\boldsymbol{A}_n$, $\boldsymbol{B}_n$, and $\boldsymbol{C}_n$ are given by

$$\boldsymbol{A}_n \triangleq \begin{cases} \boldsymbol{I}_{N_r} + \rho \Big(\tilde{\boldsymbol{H}}_d + \sum\limits_{i \neq n}^{N} \alpha_i \boldsymbol{r}_i \boldsymbol{t}_i^H \Big) \Big(\tilde{\boldsymbol{H}}_d^H + \sum\limits_{i \neq n}^{N} \alpha_i^* \boldsymbol{t}_i \boldsymbol{r}_i^H \Big) \\ \qquad + (\eta+1) \Big(\sum\limits_{i \in \mathcal{A} \backslash n} |\alpha_i|^2 \boldsymbol{r}_i \boldsymbol{r}_i^H \Big), \ n \in \mathcal{A} \\ \boldsymbol{I}_{N_r} + \rho \Big(\tilde{\boldsymbol{H}}_d + \sum\limits_{i \neq n}^{N} \alpha_i \boldsymbol{r}_i \boldsymbol{t}_i^H \Big) \Big(\tilde{\boldsymbol{H}}_d^H + \sum\limits_{i \neq n}^{N} \alpha_i^* \boldsymbol{t}_i \boldsymbol{r}_i^H \Big) \\ \qquad + (\eta+1) \Big(\sum\limits_{i \in \mathcal{A}} |\alpha_i|^2 \boldsymbol{r}_i \boldsymbol{r}_i^H \Big), \ n \notin \mathcal{A} \end{cases} \tag{14}$$

with $\mathcal{A} \backslash n = \{i : i \in \mathcal{A}, i \neq n\}$, and

$$\boldsymbol{B}_n \triangleq \begin{cases} (\eta+1) \boldsymbol{r}_n \boldsymbol{r}_n^H + \rho \boldsymbol{r}_n \boldsymbol{t}_n^H \boldsymbol{t}_n \boldsymbol{r}_n^H, & n \in \mathcal{A} \\ \rho \boldsymbol{r}_n \boldsymbol{t}_n^H \boldsymbol{t}_n \boldsymbol{r}_n^H, & n \notin \mathcal{A} \end{cases} \tag{15}$$

$$\boldsymbol{C}_n \triangleq \rho \boldsymbol{r}_n \boldsymbol{t}_n^H \Big(\tilde{\boldsymbol{H}}_d^H + \sum_{i \neq n}^{N} \alpha_i^* \boldsymbol{t}_i \boldsymbol{r}_i^H \Big), \qquad \forall n. \tag{16}$$

We note that matrices $\boldsymbol{A}_n$, $\boldsymbol{B}_n$, and $\boldsymbol{C}_n$ do not contain variable α_n. This means that if all variables $\{\alpha_i\}_{i=1, i \neq n}^N$ are fixed, the three matrices are determined. Similarly, we also extract the role of variable α_n in $P_{\mathrm{a}}(\boldsymbol{F}^\star, \{\alpha_n\})$. Specifically, recalling that $\tilde{\boldsymbol{H}}_t = \boldsymbol{H}_t \boldsymbol{F}^\star$, we can write

$$\begin{aligned} P_{\mathrm{a}}(\{\alpha_n\}) &= \mathrm{trace} \boldsymbol{\Psi} \left(\boldsymbol{H}_{\mathrm{t}} \boldsymbol{F}^\star \boldsymbol{F}^{\star \mathrm{H}} \boldsymbol{H}_{\mathrm{t}}^{\mathrm{H}} + \tilde{\sigma}^2 \boldsymbol{I}_{\mathrm{N}} \right) \boldsymbol{\Psi}^{\mathrm{H}} \\ &= \tilde{\sigma}^2 \mathrm{trace} \boldsymbol{\Psi} \boldsymbol{\Psi}^{\mathrm{H}} + \mathrm{trace} \boldsymbol{\Psi} \tilde{\boldsymbol{H}}_{\mathrm{t}} \tilde{\boldsymbol{H}}_{\mathrm{t}}^{\mathrm{H}} \boldsymbol{\Psi}^{\mathrm{H}} \\ &= \tilde{\sigma}^2 \sum_{n \in \mathcal{A}} |\alpha_n|^2 + \sum_{n \in \mathcal{A}} |\alpha_n|^2 \|\boldsymbol{t}_n\|^2 = \sum_{n \in \mathcal{A}} |\alpha_n|^2 \vartheta_n, \end{aligned} \tag{17}$$

where

$$\vartheta_n \triangleq \tilde{\sigma}^2 + \|\boldsymbol{t}_n\|^2 . \tag{18}$$

Let $\tilde{P}_{\mathrm{a},n} \triangleq \sum_{i \in \mathcal{A}, i \neq n} |\alpha_i|^2 \vartheta_i$, $n \in \mathcal{A}$. We can see that $\tilde{P}_{\mathrm{a},n}$ is a constant if all variables $\{\alpha_i\}_{i=1, i \neq n}^N$ are fixed. Therefore, the role of α_n in $P_{\mathrm{a}}(\{\alpha_n\})$ can be seen in the following expression of $P_{\mathrm{a}}(\{\alpha_n\})$:

$$P_{\mathrm{a}}(\alpha_n) = |\alpha_n|^2 \vartheta_n + \tilde{P}_{\mathrm{a},n}. \tag{19}$$

1) Problem of Updating α_n: In the AO approach, in an iteration of updating α_n, $\{\alpha_i\}_{i=1, i \neq n}^N$ are fixed. Therefore, the objective function $f(\{\alpha_n\})$ in (13) can be rewritten in the explicit form

$$f_n(\alpha_n) = \log_2 \left| \boldsymbol{A}_n + |\alpha_n|^2 \boldsymbol{B}_n + \alpha_n \boldsymbol{C}_n + \alpha_n^* \boldsymbol{C}_n^H \right|, \tag{20}$$

which has only one variable, i.e., α_n. Noting that $\mathrm{rank}(\boldsymbol{A}_n) = N_r$, i.e., $\boldsymbol{A}_n$ is of full-rank and invertible, $f_n(\alpha_n)$ can be rewrite as

$$f_n(\alpha_n) = \log_2 |\boldsymbol{A}_n| + g_n(\alpha_n), \tag{21}$$

where

$$\begin{aligned} g_n(\alpha_n) \triangleq \log_2 \Big| &\boldsymbol{I}_{N_r} + |\alpha_n|^2 \boldsymbol{A}_n^{-1} \boldsymbol{B}_n \\ &+ \alpha_n \boldsymbol{A}_n^{-1} \boldsymbol{C}_n + \alpha_n^* \boldsymbol{A}_n^{-1} \boldsymbol{C}_n^H \big| . \end{aligned} \tag{22}$$

In (21), $\log_2|\boldsymbol{A}_n|$ is a constant with α_n. Therefore, the problem of updating α_n, denoted by (P_α), is given by

$$(\mathrm{P}_\alpha) \quad \begin{cases} \max\limits_{\alpha_n} g_n(\alpha_n) \text{ s.t. } |\alpha_n| = 1, & \text{for } n \notin \mathcal{A} \\ \max\limits_{\alpha_n} g_n(\alpha_n) \text{ s.t. } |\alpha_n|^2 \leq \frac{P_{\mathrm{a}}^{\max} - \tilde{P}_{\mathrm{a},n}}{\vartheta_n}, & \text{for } n \in \mathcal{A} \end{cases} \tag{23}$$

2) Solution to (P_α)*:* (P_α) in (23) admits a closed-form solution, and, thus, it is efficient for practical implementation. To derive the solution, we rewrite $g_n(\alpha_n)$ as

$$\begin{aligned} g_n(\alpha_n) &= \log_2\left|\boldsymbol{D}_n + \alpha_n \boldsymbol{A}_n^{-1}\boldsymbol{C}_n + \alpha_n^* \boldsymbol{A}_n^{-1}\boldsymbol{C}_n^H\right| \\ &= \log_2|\boldsymbol{D}_n| + \underbrace{\log_2\left|\boldsymbol{I}_{N_r} + \alpha_n \boldsymbol{E}_n^{-1}\boldsymbol{C}_n + \alpha_n^* \boldsymbol{E}_n^{-1}\boldsymbol{C}_n^H\right|}_{\triangleq g_n'(\alpha_n)}, \end{aligned} \tag{24}$$

where $\boldsymbol{D}_n = \boldsymbol{I}_{N_r} + |\alpha_n|^2\boldsymbol{A}_n^{-1}\boldsymbol{B}_n$, and $\boldsymbol{E}_n = \boldsymbol{A}_n\boldsymbol{D}_n$. We start investigating the objective function $g_n(\alpha_n)$ by considering the first term in (24), i.e., $\log_2|\boldsymbol{D}_n|$. Specifically, we note that in $\boldsymbol{D}_n$, rank$(\boldsymbol{A}_n^{-1}\boldsymbol{B}_n) \leq$ rank$(\boldsymbol{B}_n) = 1$. In addition, the probability that rank$(\boldsymbol{A}_n^{-1}\boldsymbol{B}_n) = 0$ is almost zero (it only happens when $\boldsymbol{A}_n^{-1}\boldsymbol{B}_n = \boldsymbol{0}$). Consequently, we generally have rank$(\boldsymbol{A}_n^{-1}\boldsymbol{B}_n) = 1$. Also, we observe that $\boldsymbol{A}_n^{-1}\boldsymbol{B}_n$ is non-diagonalizable iff trace$(\boldsymbol{A}_n^{-1}\boldsymbol{B}_n) = 0$ [7], which rarely happens in general. Thus, we almost surely have trace$(\boldsymbol{A}_n^{-1}\boldsymbol{B}_n) \neq 0$ and $\boldsymbol{A}_n^{-1}\boldsymbol{B}_n$ is diagonalizable. As a result, it can be factorized as $\boldsymbol{A}_n^{-1}\boldsymbol{B}_n = \boldsymbol{T}_n\boldsymbol{\Gamma}_n\boldsymbol{T}_n^{-1}$ based on the eigenvalue decomposition (EVD) [51], where $\boldsymbol{T}_n \in \mathbb{C}^{N_r \times N_r}$ and $\boldsymbol{\Gamma}_n = \gamma_n, 0, \ldots, 0$ with γ_n being the sole non-zero eigenvalue of $\boldsymbol{A}_n^{-1}\boldsymbol{B}_n$. Finally, because both $\boldsymbol{A}_n$ and $\boldsymbol{B}_n$ are positive semi-definite, γ_n has non-negative and real value. Thus, we can write

$$\begin{aligned} \log_2|\boldsymbol{D}_n| &= \log_2\left|\boldsymbol{I}_{N_r} + |\alpha_n|^2\boldsymbol{T}_n\boldsymbol{\Gamma}_{\boldsymbol{n}}\boldsymbol{T}_n^{-1}\right| \\ &= \log_2\left|\boldsymbol{T}_n\left(\boldsymbol{I}_{N_r} + |\alpha_n|^2\boldsymbol{\Gamma}_{\boldsymbol{n}}\right)\boldsymbol{T}_n^{-1}\right| \\ &= \log_2\left|\boldsymbol{I}_{N_r} + |\alpha_n|^2\boldsymbol{\Gamma}_{\boldsymbol{n}}\right| = \log_2\left(1 + |\alpha_n|^2\gamma_n\right). \end{aligned} \tag{25}$$

We now focus on the second term in (24). Following the similar arguments for the first term, we also almost surely have $\boldsymbol{E}_n^{-1}\boldsymbol{C}_n$ diagonalizable. Thus, we have $\boldsymbol{E}_n^{-1}\boldsymbol{C}_n = \boldsymbol{U}_n\boldsymbol{\Lambda}_n\boldsymbol{U}_n^{-1}$ based on the EVD [51], where $\boldsymbol{U}_n \in \mathbb{C}^{N_r \times N_r}$ and $\boldsymbol{\Lambda}_n = \lambda_n, 0, \ldots, 0$ with λ_n being the sole non-zero eigenvalue of $\boldsymbol{E}_n^{-1}\boldsymbol{C}_n$. With the note that $\boldsymbol{E}_n = \boldsymbol{A}_n\boldsymbol{D}_n = \boldsymbol{A}_n + |\alpha_n|^2\boldsymbol{B}_n$ is symmetric, we can express $g_n'(\alpha_n)$ in (24) as

$$\begin{aligned} g_n'(\alpha_n) = &\ \log_2\left|\boldsymbol{I}_{N_r} + \alpha_n\boldsymbol{E}_n^{-1}\boldsymbol{C}_n + \alpha_n^*\boldsymbol{E}_n^{-1}(\boldsymbol{E}_n^{-1}\boldsymbol{C}_n)^H\boldsymbol{E}_n\right| \\ = &\ \log_2\left|\boldsymbol{I}_{N_r} + \alpha_n\boldsymbol{U}_n\boldsymbol{\Lambda}_n\boldsymbol{U}_n^{-1} + \alpha_n^*\boldsymbol{E}_n^{-1}\boldsymbol{U}_n^{-H}\boldsymbol{\Lambda}_n^H\boldsymbol{U}_n^H\boldsymbol{E}_n\right| \\ = &\ \log_2\Big|\boldsymbol{U}_n\boldsymbol{U}_n^{-1} + \alpha_n\boldsymbol{U}_n\boldsymbol{\Lambda}_n\boldsymbol{U}_n^{-1} \\ &\ + \alpha_n^*\boldsymbol{U}_n\boldsymbol{U}_n^{-1}\boldsymbol{E}_n^{-1}\boldsymbol{U}_n^{-H}\boldsymbol{\Lambda}_n^H\boldsymbol{U}_n^H\boldsymbol{E}_n\boldsymbol{U}_n\boldsymbol{U}_n^{-1}\Big| \\ = &\ \log_2\Big(|\boldsymbol{U}_n|\Big|\boldsymbol{I}_{N_r} + \alpha_n\boldsymbol{\Lambda}_n \end{aligned}$$

$$\begin{aligned} &+ \alpha_n^*\boldsymbol{U}_n^{-1}\boldsymbol{E}_n^{-1}\boldsymbol{U}_n^{-H}\boldsymbol{\Lambda}_n^H\boldsymbol{U}_n^H\boldsymbol{E}_n\boldsymbol{U}_n\Big|\left|\boldsymbol{U}_n^{-1}\right|\Big) \\ \overset{(a)}{=} &\ \log_2\left|\boldsymbol{I}_{N_r} + \alpha_n\boldsymbol{\Lambda}_n + \alpha_n^*\boldsymbol{U}_n^{-1}\boldsymbol{E}_n^{-1}\boldsymbol{U}_n^{-H}\boldsymbol{\Lambda}_n^H\boldsymbol{U}_n^H\boldsymbol{E}_n\boldsymbol{U}_n\right| \\ = &\ \log_2\left|\boldsymbol{I}_{N_r} + \alpha_n\boldsymbol{\Lambda}_n + \alpha_n^*\boldsymbol{Z}_n^{-1}\boldsymbol{\Lambda}_n^H\boldsymbol{Z}_n\right|, \end{aligned}$$

where equality $\overset{(a)}{=}$ follows $|\boldsymbol{A}||\boldsymbol{B}||\boldsymbol{A}^{-1}| = |\boldsymbol{A}||\boldsymbol{A}|^{-1}|\boldsymbol{B}| = |\boldsymbol{B}|$, and $\boldsymbol{Z}_n \triangleq \boldsymbol{U}_n^H\boldsymbol{E}_n\boldsymbol{U}_n$. Let $\bar{\boldsymbol{z}}_n^T$ be the first row of $\boldsymbol{Z}_n$, and $\bar{\boldsymbol{z}}_n'$ be the first column of $\boldsymbol{Z}_n^{-1}$. Recalling that $\boldsymbol{\Lambda}_n = \lambda_n, 0, \ldots, 0$, we can further expand $g_n'(\alpha_n)$ as

$$\begin{aligned} g_n'(\alpha_n) &= \log_2\left|\boldsymbol{I}_{N_r} + \alpha_n\boldsymbol{\Lambda}_n + \alpha_n^*\bar{\boldsymbol{z}}_n'\lambda_n^*\bar{\boldsymbol{z}}_n^T\right| \\ &\overset{(b)}{=} \log_2\left(\left(1 + \alpha_n^*\lambda_n^*\bar{\boldsymbol{z}}_n^T(\boldsymbol{I}_{N_r} + \alpha_n\boldsymbol{\Lambda}_n)^{-1}\bar{\boldsymbol{z}}_n'\right)|\boldsymbol{I}_{N_r} + \alpha_n\boldsymbol{\Lambda}_n|\right) \\ &= \log_2\Big[(1 + \alpha_n\lambda_n)\Big(1 + \alpha_n^*\lambda_n^*\bar{\boldsymbol{z}}_n^T \\ &\qquad \times \left(\boldsymbol{I}_{N_r} - \frac{\alpha_n\lambda_n}{1 + \alpha_n\lambda_n}, 0, \ldots, 0\right)\bar{\boldsymbol{z}}_n'\Big)\Big] \\ &\overset{(c)}{=} \log_2\left[(1 + \alpha_n\lambda_n)\left(1 + \alpha_n^*\lambda_n^* - \frac{\alpha_n^*\lambda_n^* z_n\alpha_n\lambda_n z_n'}{1 + \alpha_n\lambda_n}\right)\right] \\ &= \log_2\left[(1 + \alpha_n\lambda_n)(1 + \alpha_n^*\lambda_n^*) - \alpha_n^*\lambda_n^* z_n\alpha_n\lambda_n z_n'\right] \\ &= \log_2\left(1 + |\alpha_n|^2|\lambda_n|^2 + 2\Re(\alpha_n\lambda_n) - |\alpha_n|^2|\lambda_n|^2 z_n z_n'\right) \\ &= \log_2\left(1 + |\alpha_n|^2|\lambda_n|^2(1 - z_n z_n') + 2\Re(\alpha_n\lambda_n)\right), \end{aligned} \tag{26}$$

where equality $\overset{(b)}{=}$ follows the fact that $|\boldsymbol{I} + \boldsymbol{a}\boldsymbol{b}^T| = 1 + \boldsymbol{b}^T\boldsymbol{a}$; in equality $\overset{(c)}{=}$, z_n and z_n' are the first elements of $\bar{\boldsymbol{z}}_n$ and $\bar{\boldsymbol{z}}_n'$, respectively; $\Re(\cdot)$ represents the real part of a complex number.

In summary, based on (25) and (26), we obtain

$$\begin{aligned} g_n(\alpha_n) = &\log_2\left(1 + |\alpha_n|^2\gamma_n\right) \\ &+ \log_2\left(1 + |\alpha_n|^2|\lambda_n|^2(1 - z_n z_n') + 2\Re(\alpha_n\lambda_n)\right). \end{aligned} \tag{27}$$

As a result, an optimal solution to (P_α), denoted by $\alpha_n^\star$, admits the form

$$\alpha_n^\star = \begin{cases} |\alpha_n^\star|\, e^{-j\arg\{\lambda_n\}}, & n \in \mathcal{A} \\ e^{-j\arg\{\lambda_n\}}, & \text{otherwise} \end{cases}. \tag{28}$$

It is observed that the amplitude $|\alpha_n^\star|$ is not available at this stage because it requires a determined $\mathcal{A}$. To this end, it is natural to consider two scenarios: $\mathcal{A}$ is predetermined and fixed in the manufacture, and $\mathcal{A}$ is dynamic and optimized based on the propagation condition, associated with the fixed and dynamic HR-RIS architectures illustrated in Fig. 1(a) and (b), respectively. The solution $\{\alpha_n^\star\}$ dedicated to these architectures will be derived in the next section.

IV. Joint Beamforming in Fixed/Dynamic HR-RIS-Assisted MIMO System

In this section, we propose two algorithms to obtain the active transmit beamforming matrix at the BS and the hybrid relay/reflecting coefficients at the fixed and dynamic HR-RISs aiding the MIMO system.

A. Fixed HR-RIS-Aided MIMO System

In the fixed HR-RIS, $\mathcal{A}$ is available to determine $\{|\alpha_n^\star|\}_{n\in\mathcal{A}}$. Specifically, $|\alpha_n^\star|$, $n \in \mathcal{A}$, can be determined based on the fact that, given the optimal form in (28), $g_n(\alpha_n^\star)$ monotonically increases with $|\alpha_n^\star|$. Therefore, from (23), we obtain

$$|\alpha_n^\star| = \sqrt{\frac{P_{\mathrm{a}}^{\max} - \tilde{P}_{\mathrm{a},n}}{\vartheta_n}}, n \in \mathcal{A}. \tag{29}$$

As a result, the optimal solution to (P_α) is given as

$$\alpha_n^\star = \begin{cases} \sqrt{\frac{P_{\mathrm{a}}^{\max} - \tilde{P}_{\mathrm{a},n}}{\vartheta_n}} e^{-j\arg\{\lambda_n\}}, & n \in \mathcal{A} \\ e^{-j\arg\{\lambda_n\}}, & \text{otherwise} \end{cases}. \tag{30}$$

The solutions in (8), (30), and the given $\mathcal{A}$ are readily used to obtain $\boldsymbol{F}^\star$ and $\{\alpha_n^\star\}$, as outlined in Algorithm 1. Specifically, at the initial stage, coefficients $\{\alpha_n\}$ are randomly generated to have arbitrary phases and amplitudes satisfying $|\alpha_n| = 1, n \notin \mathcal{A}$ and $\sum_{n\in\mathcal{A}} |\alpha_n|^2\vartheta_n = P_{\mathrm{a}}^{\max}$, based on (7d), (17), and (29). During steps 2–15, the transmit beamforming matrix $\boldsymbol{F}^\star$ and the coefficients $\alpha_n^\star$ are alternatively updated based on (8) and (30), respectively. It terminates when a convergence criterion on the objective $f(\boldsymbol{F}, \{\alpha_n\})$ is reached. The iteration procedure in Algorithm 1 guarantees the convergence. To see this, we note that objective function $f_0(\boldsymbol{F}, \{\alpha_n\})$ is upper bounded. In addition, $(\mathrm{P}_{\mathrm{tx}})$ and (P_α) are solved optimally by (8) and (30), respectively. Thus, the resultant sequences of $f_0(\boldsymbol{F})$ and $f(\{\alpha_n\})$ are non-decreasing over iterations [7]. Although $f(\{\alpha_n\})$ converges, the convergence of $f_0(\{\alpha_n\})$ is not guaranteed. However, we note that $f_0(\{\alpha_n\})$ is a upper bound of $f(\{\alpha_n\})$, and the bound is tight when the number of active elements is small and when the HR-RIS has low power budget. Therefore, it is expected that $f_0(\{\alpha_n\})$ also converges well considering that K and $P_{\mathrm{a}}^{\max}$ are small in the proposed HR-RIS. We will further verify this by numerically results in Section VI. In the following, we make two remarks on the design of the fixed HR-RIS.

Remark 1: Inserting $\tilde{P}_{\mathrm{a},n} \triangleq \sum_{i=1,i\neq n}^{K} |\alpha_i|^2\vartheta_i,\ n \in \mathcal{A}$ to (29), it is observed that a larger K results in a smaller $|\alpha_n^\star|$. Therefore, increasing K does not always guarantee the SE improvement of the fixed HR-RIS compared to the conventional RIS. In particular, with a limited power budget $P_{\mathrm{a}}^{\max}$, the HR-RIS can have $|\alpha_n^\star| < 1$, causing signal attenuation, and hence degrades the SE. In this case, the fixed HR-RIS with a few active elements, i.e., small K, is easier to attain SE gains than that with large K.

Remark 2: In (29), both $\tilde{P}_{\mathrm{a},n}$ and ϑ_n increase with P_{BS}. Therefore, for a fixed $P_{\mathrm{a}}^{\max}$, a lower transmit power P_{BS} and/or a smaller channel power gain, i.e., $\|\boldsymbol{t}_n\|^2$, from the transmitter results in a larger $|\alpha_n^\star|$, and equivalently, a more significant performance improvement can be achieved.

Algorithm 1: Find $\boldsymbol{F}^\star$ and $\{\alpha_n^\star\}$ for the Fixed HR-RIS-Aided MIMO System.

```
    Input: $\boldsymbol{H}_d, \boldsymbol{H}_t, \boldsymbol{H}_r, \mathcal{A}$.
    Output: $\boldsymbol{F}^\star$ and $\{\alpha_1^\star, \dots, \alpha_N^\star\}$.
 1: Randomly generate $\{\alpha_n\}$ with $|\alpha_n| = 1, n \notin \mathcal{A}$, and
    $\sum_{n\in\mathcal{A}} |\alpha_n|^2\vartheta_n = P_{\mathrm{a}}^{\max}$. Set $\{\alpha_n^\star\}$ to $\{\alpha_n\}$.
 2: while objective value does not converge do
 3:   Obtain $\boldsymbol{F}^\star$ by solving (9) with $\alpha_n = \alpha_n^\star, \forall n$.
 4:   for $n = 1 \to N$ do
 5:     Obtain $\boldsymbol{A}_n, \boldsymbol{B}_n, \boldsymbol{C}_n$ based on (14)-(16) with
        $\boldsymbol{F} = \boldsymbol{F}^\star$.
 6:     if $n \in \mathcal{A}$ then
 7:       $\boldsymbol{D}_n = \boldsymbol{I}_{N_r} + |\alpha_n|^2 \boldsymbol{A}_n^{-1}\boldsymbol{B}_n$
 8:     else
 9:       $\boldsymbol{D}_n = \boldsymbol{I}_{N_r} + \boldsymbol{A}_n^{-1}\boldsymbol{B}_n$
10:     end if
11:     $\boldsymbol{E}_n = \boldsymbol{A}_n\boldsymbol{D}_n$
12:     Obtain $\lambda_n$ as the sole non-zero eigenvalue of
        $\boldsymbol{E}_n^{-1}\boldsymbol{C}_n$.
13:     Update $\alpha_n^\star$ as (30).
14:   end for
15: end while
```

B. Dynamic HR-RIS-Aided MIMO System

Unlike the fixed HR-RIS architecture, in the dynamic HR-RIS illustrated in Fig. 1(b), the number and positions of the active elements can be cast as design parameters. In particular, given $\boldsymbol{F}^\star$, the problem of SE maximization for the dynamic HR-RIS scheme can be formulated as

$$(\mathrm{P1}) \quad \underset{\{\alpha_n\},\mathcal{A}}{\text{maximize}} \quad f_0(\{\alpha_n\}) \tag{31b}$$

$$\text{subject to} \quad (7\mathrm{c}), (7\mathrm{d}), \tag{31c}$$

$$|\mathcal{A}| \leq K, \mathcal{A} \subset \{1, 2, \dots, N\}, \tag{31d}$$

where the constraint (31c) means that the number of active elements in the HR-RIS does not exceed the number of RF-PA chains. Problem (P1) does not only inherit the numerical challenge of $(\mathrm{P}_{\mathrm{H}})$ but also includes the difficulty from cardinality constraint (31c). In the following, we develop an efficient solution to (P1) for practical employment.

Similar to problem $(\mathrm{P}_{\mathrm{H}})$, we first employ the upper bound of $f_0(\{\alpha_n\})$ (see (11)) to obtain an approximate but more tractable problem of (P1), which is given as

$$(\mathrm{P}_{\mathrm{dyn}}) \quad \underset{\{\alpha_n\},\mathcal{A}}{\text{maximize}} \quad f(\{\alpha_n\}) \tag{32b}$$

$$\text{subject to} \quad (7c), (7d), \tag{32c}$$

$$|\mathcal{A}| \leq K, \mathcal{A} \subset \{1, 2, \dots, N\}. \tag{32d}$$

At this point, $(\mathrm{P}_{\mathrm{dyn}})$ can be solved based on an exhaustive search, i.e., using the result presented in Section III to determine the coefficients $\{\alpha_n\}$ corresponding to each valid set $\mathcal{A}$, then choosing the one providing the best performance as the final solution. Such an approach requires high complexity due to an excessively large number of combinations of $\mathcal{A}$.

We now propose a computationally reasonable scheme for $(\mathrm{P}_{\mathrm{dyn}})$. Recall that we can write $\alpha_n = |\alpha_n|e^{j\theta_n}$. Based on this, we introduce two matrices as $\boldsymbol{\Theta} = e^{j\{\theta_1\}}, \dots, e^{j\{\theta_N\}}$ and $\hat{\boldsymbol{\Upsilon}} = |\alpha_1|, \dots, |\alpha_N|$ with $|\alpha_n| = 1, n \notin \mathcal{A}$. Then we can write $\boldsymbol{\Upsilon} = \boldsymbol{\Theta}\hat{\boldsymbol{\Upsilon}}$. This motivates us develop an alternating procedure applied on three sets of variables $\{\mathbf{\Theta}, \hat{\mathbf{\Upsilon}}, \mathcal{A}\}$. Specifically, each of $\{\mathbf{\Theta}, \hat{\mathbf{\Upsilon}}, \mathcal{A}\}$ is determined when the others are fixed, which are the key steps in our proposed solution.

1) Determine $\{\theta_n\}$ (Or Equivalently $\mathbf{\Theta}$): For this step, we suppose that $|\alpha_n| = 1, \forall n$, which corresponds to the conventional RIS where all elements are passive. Then, we determine $\{\theta_n\}$ via solving the following problem extracted from $(\mathrm{P}_{\text{dyn}})$

$$(\mathrm{P}'_{\text{dyn}}) \quad \{\alpha_n\}\text{maximize} \quad f(\{\alpha_n\})$$
$$\text{subject to} \quad |\alpha_n| = 1, \forall n.$$

A solution to $(\mathrm{P}'_{\text{dyn}})$ can be obtained by using the same approach presented in Section IV-A (i.e., see (28)) for the case $n \notin \mathcal{A}$.

2) Determine $\mathcal{A}$ and $\{|\alpha_n|\}$ (Or Equivalently $\hat{\mathbf{\Upsilon}}$): We recall that the passive reflecting elements have unit modulus, i.e., $|\alpha_n| = 1, \forall n \notin \mathcal{A}$. Thus, we only need to determine $|\alpha_n|$, for $n \in \mathcal{A}$. Furthermore, an active element only results in performance gain with respect to a passive element if $|\alpha_n| > 1$; otherwise, it causes performance loss due to signal attenuation. Therefore, to minimize the loss, in the dynamic HR-RIS architecture, the RF-PA chain connected to the nth element should be turned off if $|\alpha_n| < 1$. As a result, in the dynamic HR-RIS scheme, we have $|\alpha_n| > 1, \forall n \in \mathcal{A}$, and the solution in (28) becomes

$$\alpha_n^\star = \begin{cases} |\alpha_n^\star|\, e^{-j\arg\{\lambda_n\}}, & |\alpha_n^\star| > 1, n \in \mathcal{A} \\ e^{-j\arg\{\lambda_n\}}, & \text{otherwise} \end{cases}, \tag{33}$$

with $\{\arg\{\lambda_n\}\}$ being determined in the previous step. With this solution, it is obvious that $\Re(\alpha_n^\star \lambda_n) = |\alpha_n^\star||\lambda_n|$. For ease of exposition, let $\bar{a}_n = |\alpha_n^\star| > 1$, and (27) can be rewritten as

$$g_n(\bar{a}_n) = \log_2(1 + \gamma_n \bar{a}_n^2) + \log_2(1 + |\lambda_n|^2 (1 - z_n z'_n)\bar{a}_n^2 + 2|\lambda_n|\bar{a}_n). \tag{34}$$

The problem of choosing optimal active elements and optimizing their coefficients can be formulated as

$$\underset{\mathcal{A}, \{\bar{a}_n\}}{\text{maximize}} \quad \sum_{n \in \mathcal{A}} g_n(\bar{a}_n) \tag{35a}$$
$$\text{subject to} \quad \sum_{n \in \mathcal{A}} \bar{a}_n^2 \vartheta_n \le P_{\text{a}}^{\max} \text{ and } (32\text{c}). \tag{35b}$$

We note that in the second term of (34), the factor $|\lambda_n|^2(1 - z_n z'_n)$ can be either positive or negative depending on $1 - z_n z'_n$. Therefore, based on the first term of (34) and the objective function in (35a), the optimal set of active elements, i.e., $\mathcal{A}^\star$, can be determined as the set of indices of the K largest elements in $\{\gamma_1, \ldots, \gamma_N\}$.

With the determined $\mathcal{A}^\star$, we aim at solving $\{\bar{a}_n\}$ below. To tackle the nonconvexity of $g_n(\bar{a}_n)$, we introduce slack variables $\{x_n\}$ and $\{y_n\}$ (as the lower bounds of $\gamma_n \bar{a}_n^2$ and $|\lambda_n|^2(1 - z_n z'_n)\bar{a}_n^2 + 2|\lambda_n|\bar{a}_n$ in (34), respectively). Then, problem (35) is equivalent to

$$\underset{\{\bar{a}_n, x_n, y_n\}}{\text{maximize}} \quad \sum_{n \in \mathcal{A}^\star} \log_2(1 + x_n) + \log_2(1 + y_n) \tag{36a}$$
$$\text{subject to} \quad \gamma_n \bar{a}_n^2 \ge x_n, \ \forall n \in \mathcal{A}^\star \tag{36b}$$
$$|\lambda_n|^2 (1 - z_n z'_n)\bar{a}_n^2 + 2|\lambda_n|\bar{a}_n \ge y_n, \ \forall n \in \mathcal{A}^\star \tag{36c}$$
$$\sum_{n \in \mathcal{A}^\star} \bar{a}_n^2 \vartheta_n \le P_{\text{a}}^{\max}, \tag{36d}$$

where constraint (36b) is concave, and (36c) is convex only for $z_n z'_n > 1$. To tackle the nonconvexity of (36), we apply the SCA approach. Specifically, in each iteration of this iterative procedure, problem (36) is successively approximated to a convex one and solved via convex optimization solver. For the ease of exposition, let us denote functions $h_1(\bar{a}_n) \triangleq \gamma_n \bar{a}_n^2$ and $h_2(\bar{a}_n) \triangleq |\lambda_n|^2(1 - z_n z'_n)\bar{a}_n^2$. Convex lower bounds of $h_1(\bar{a}_n)$ and $h_2(\bar{a}_n)$ can be found based on the first-order Taylor approximations around $\bar{a}_n^{(i)}$ as follows:

$$h_1(\bar{a}_n) \ge h_1^{\text{lb}}(\bar{a}_n; \bar{a}_n^{(i)}) \triangleq \gamma_n (\bar{a}_n^{(i)})^2 + 2\gamma_n \bar{a}_n^{(i)}(\bar{a}_n - \bar{a}_n^{(i)}), \tag{37}$$

$$h_2(\bar{a}_n) \ge h_2^{\text{lb}}(\bar{a}_n; \bar{a}_n^{(i)}) \triangleq |\lambda_n|^2 (1 - z_n z'_n)(\bar{a}_n^{(i)})^2 + 2|\lambda_n|^2 (1 - z_n z'_n)\bar{a}_n^{(i)}(\bar{a}_n - \bar{a}_n^{(i)}), \text{for} z_n z'_n > 1, \tag{38}$$

respectively. Denote

$$\tilde{h}_2^{\text{lb}}(\bar{a}_n; \bar{a}_n^{(i)}) = \begin{cases} h_2^{\text{lb}}(\bar{a}_n; \bar{a}_n^{(i)}), & \text{if\textasciitilde} z_n z'_n > 1 \\ h_2(\bar{a}_n), & \text{otherwise.} \end{cases}$$

As a result, problem (35) can be approximated to the following program at iteration i:

$$\underset{\{\bar{a}_n, x_n, y_n\}}{\text{maximize}} \quad \sum_{n \in \mathcal{A}^\star} \log_2(1 + x_n) + \log_2(1 + y_n) \tag{39a}$$
$$\text{subject to} \quad h_1^{\text{lb}}(\bar{a}_n; \bar{a}_n^{(i)}) \ge x_n, \ \forall n \in \mathcal{A}^\star \tag{39b}$$
$$\tilde{h}_2^{\text{lb}}(\bar{a}_n; \bar{a}_n^{(i)}) + 2|\lambda_n|\bar{a}_n \ge y_n, \ \forall n \in \mathcal{A}^\star \tag{39c}$$
$$\sum_{n \in \mathcal{A}^\star} \bar{a}_n^2 \vartheta_n \le P_{\text{a}}^{\max}, \tag{39d}$$

which is convex and can be solved with existing optimization tools such as CVX or YALMIP-MOSEK.

The proposed procedure for finding an efficient solution to the problem of SE maximization in the dynamic HR-RIS-aided MIMO system is outlined in Algorithm 2. It starts with randomly generating $\{\alpha_n\}$ with unit modules. Then steps 2–11 are executed to determine $\boldsymbol{F}^\star$ and $\{\theta_n^\star\}$. After that, $\mathcal{A}^\star$ is determined in steps 12, and amplitudes $\{\alpha_n^\star\}$ are iteratively solved in steps 13–17. Specifically, we first initialize $\bar{a}_n^{(0)} = |\alpha_n^\star| = 1, \forall n$ because $\alpha_n^\star$ obtained in step 9 has unit modulus. Then, in each iteration, $\{\bar{a}_n^\star\}$ is solved and $\bar{a}_n^{(0)}$ are updated as in step 16. This iterative process terminates once the objective function of problem (39) converges. At this point, the solution to $\{|\alpha_n^\star|\}$ is obtained in step 18. Here, a scalar $\varsigma > 0$ is employed to avoid the activation of elements with amplitudes being close to unity, whose active relaying gains may be insignificant. Furthermore, it is clear that $|\alpha_n^\star| > 1, \forall n \in \mathcal{A}^\star$, and the number of active elements in the dynamic HR-RIS is not larger than that in the fixed one. Finally, the coefficients of the dynamic HR-RIS are derived in step 19 based on (28).

Algorithm 2 Find $\boldsymbol{F}^\star$ and $\{\alpha_n^\star\}$ for the Dynamic HR-RIS-Aided MIMO System

```
    Require: $\boldsymbol{H}_d, \boldsymbol{H}_t, \boldsymbol{H}_r, \varsigma$.
    Ensure: $\boldsymbol{F}^\star$, $\mathcal{A}^\star$, $\{\alpha_n^\star\}$.
 1: Randomly generate $\boldsymbol{\Theta} = e^{j\theta_1}, \dots, e^{j\theta_N}$ and assign
    $\boldsymbol{\Upsilon} = \boldsymbol{\Theta}$. Set $\{\alpha_n^\star\}$ to $\{e^{j\theta_N}\}$.
 2: while objective value does not converge do
 3:   Obtain $\boldsymbol{F}^\star$ by solving (9) with $\alpha_n = \alpha_n^\star, \forall n$.
 4:   for $n = 1 \to N$ do
 5:     Obtain $\boldsymbol{A}_n, \boldsymbol{B}_n$, and $\boldsymbol{C}_n$ based on (14)-(16) for
        the case $n \notin \mathcal{A}$, with $\boldsymbol{F} = \boldsymbol{F}^\star$.
 6:     $\boldsymbol{D}_n = \boldsymbol{I}_{N_r} + \boldsymbol{A}_n^{-1}\boldsymbol{B}_n, \boldsymbol{E}_n = \boldsymbol{A}_n\boldsymbol{D}_n$.
 7:     Set $\boldsymbol{Z}_n = \boldsymbol{U}_n^H\boldsymbol{E}_n\boldsymbol{U}_n$, where $\boldsymbol{U}_n$ satisfies
        $\boldsymbol{E}_n^{-1}\boldsymbol{C}_n = \boldsymbol{U}_n\boldsymbol{\Lambda}_n\boldsymbol{U}_n^{-1}$. Obtain $z_n$ and $z_n'$ as the
        first entries of $\boldsymbol{Z}_n$ and $\boldsymbol{Z}_n^{-1}$, respectively
 8:     Obtain $\lambda_n$ and $\gamma_n$ as the sole non-zero eigenvalue
        of $\boldsymbol{E}_n^{-1}\boldsymbol{C}_n$ and $\boldsymbol{A}_n^{-1}\boldsymbol{B}_n$, respectively.
 9:     $\theta_n^\star = -\arg\{\lambda_n\}, \alpha_n^\star = e^{-j\theta_n^\star}$.
10:   end for
11: end while
12: Set $\mathcal{A}^\star$ to the indices of the $K$ largest positive values
    in $\{\gamma_1, \dots, \gamma_N\}$.
13: Set $\bar{a}_n^{(0)} = 1, \forall n \in \mathcal{A}^\star$, and set $i = 0$.
14: while objective value of problem (39) does not
    converge do
15:   $i = i + 1$.
16:   Solve problem (39) to obtain $\{\bar{a}_n^\star\}$. Set
      $\bar{a}_n^{(i)} = \bar{a}_n^\star, \forall n \in \mathcal{A}^\star$.
17: end while
18: Set $|\alpha_n^\star| = \begin{cases} \bar{a}_n^\star, \text{if } \bar{a}_n^\star > 1 + \varsigma \\ 1, \text{otherwise} \end{cases}, \forall n \in \mathcal{A}^\star$.
19: Obtain $\{\alpha_n^\star\}$ based on (28).
```

C. Complexity Analysis

It is observed that most of the complex operations in Algorithms 1 and 2 are to obtain $\boldsymbol{F}^\star$ and $\{\alpha_n^\star\}$, which are required in both algorithms. Therefore, they approximately have the same complexity. First, to obtain $\boldsymbol{F}^\star$ in (8), $\tilde{\boldsymbol{G}}$ is computed and its singular value decomposition (SVD) is performed, requiring the complexity of $\mathcal{O}(N_r^3 + N_r^2 N_t + N_r N_t \min(N_r, N_t))$. To update each coefficient $\alpha_n^\star$ through steps 5–13 in Algorithm 1 and to update each phase shift $\theta_n^\star$ through steps 3–9 in Algorithm 2, matrix inversions/multiplications, eigenvalues, and $\{\boldsymbol{A}_n, \boldsymbol{B}_n, \boldsymbol{C}_n\}$ are computed. Here, we note that $\boldsymbol{r}_i\boldsymbol{r}_i^H$, $\boldsymbol{r}_i\boldsymbol{t}_i^H$ $(i = 1, \dots, N)$, and $\boldsymbol{B}_n$ in (14)–(16) are only computed once because they do not change over iterations. Therefore, the complexities required for updating $\{\alpha_n^\star\}$ and $\{\theta_n^\star\}, n = 1, \dots, N$ in Algorithms 1 and 2, respectively, are the same in terms of $\mathcal{O}$-complexity, given as $\mathcal{O}(N(5N_r^3 + 2N_r^2 N_t))$. The complexity of the iterative procedure in steps 14–17 of Algorithm 2 is $\mathcal{C}_{\bar{a}} = \mathcal{O}(\mathcal{I}_{\bar{a}}(2K+1)^{0.5}(3\tilde{}K)^3)$, where $\mathcal{I}_{\bar{a}}$ is the number of iterations to solve $\{\bar{a}_n^\star\}$. In conclusion, the complexities of Algorithms 1 and 2 are $\mathcal{C}_{\text{fix. HR-RIS}} = \mathcal{O}(\mathcal{I}\mathcal{C})$ and $\mathcal{C}_{\text{dyn. HR-RIS}} = \mathcal{O}(\mathcal{I}\mathcal{C} + \mathcal{C}_{\bar{a}})$, respectively, where $\mathcal{C} = N_r^3 + N_r^2 N_t + N_r N_t \min(N_r, N_t) + N(5N_r^3 + 2N_r^2 N_t)$, and $\mathcal{I}$ is the number of outer iterations required in Algorithm 1 and steps 2–11 of Algorithm 2. Here, we note that $\mathcal{C}_{\bar{a}} \ll \mathcal{I}\mathcal{C}$ because $K \ll N$. In other words, Algorithms 1 and 2 approximately have the same complexity of $\mathcal{C}_{\text{HR-RIS}} = \mathcal{O}(\mathcal{I}\mathcal{C})$.

To show the complexity difference between the HR-RIS and the passive RIS schemes, we note that the latter also requires obtaining $\boldsymbol{F}^\star$ and $\{\theta_n^\star\}$. With the same AO approach, the complexity of updating $\{\theta_n^\star\}$ is still $\mathcal{O}(5N_r^3 + 2N_r^2 N_t)$. However, with the passive RIS, $\boldsymbol{R} = \boldsymbol{I}_{N_r}$, yielding $\tilde{\boldsymbol{G}} = \boldsymbol{G}$. In this case, only the SVD is performed to obtain $\boldsymbol{F}^\star$ with the complexity of $\mathcal{O}(N_r N_t \min(N_r, N_t))$. As a result, the complexity of the conventional passive RIS-aided MIMO system is $\mathcal{C}_{\text{RIS}} = \mathcal{O}(\mathcal{I}(N_r N_t \min(N_r, N_t) + N(5N_r^3 + 2N_r^2 N_t)))$, which is lower than $\mathcal{C}_{\text{HR-RIS}}$ an amount of $\Delta\mathcal{C} = \mathcal{O}(\mathcal{I}(N_r^3 + N_r^2 N_t))$. Noting that $N_r, N_t \ll N$, we have $\Delta\mathcal{C} \ll \mathcal{C}_{\text{HR}-\text{RIS}}, \mathcal{C}_{\text{RIS}}$, i.e., the increase in the complexity of the HR-RIS schemes with respect to the passive RIS scheme is just marginal.

V. Power Consumption Analysis

A. Fixed HR-RIS Scheme

In the fixed HR-RIS, the number of active and passive elements are fixed to K and M, respectively. Therefore, the total power consumption of the whole MIMO system aided by the fixed HR-RIS can be modeled as [38], [52], [53]

$$P_{\text{H}}^{\text{fix.}} = \frac{P_{\text{BS}}}{\tau_{\text{BS}}} + \frac{P_{\text{a}}}{\tau_{\text{a}}} + P_{\text{c,H}}^{\text{fix.}} + MP_{\text{p}}, \tag{40}$$

where $\tau_{\text{BS}}, \tau_{\text{a}} \in (0, 1]$ are the power amplifier efficiencies of the BS and active elements at the HR-RIS, respectively, P_{a} is given in (19), and P_{p} is the power required for a passive reflecting element [38]. Furthermore, $P_{\text{c,H}}^{\text{fix.}}$ is the total circuit power consumption of the fixed HR-RIS-aided MIMO system and can be computed as $P_{\text{c,H}}^{\text{fix.}} = N_t P_{\text{BS,dynamic}} + KP_{\text{a,dynamic}} + P_{\text{BS,static}} + P_{\text{a,static}}$, with $P_{\text{BS,dynamic}}$ denoting the dynamic power consumption of each RF chain of the BS, and $P_{\text{BS,static}}$ denoting the static power overhead of the BS, including baseband processing, power suply, and cooling power consumption [52]. Similarly, $P_{\text{a,dynamic}}$ and $P_{\text{a,static}}$ are the dynamic and static power consumption of the active relay elements in the HR-RIS.

B. Dynamic HR-RIS Scheme

In the dynamic HR-RIS architecture, the number of active and passive elements vary depending on $P_{\text{a}}^{\max}$ and ϑ_n. Let $|\mathcal{A}^\star|$ denote the number of actual active elements obtained in Algorithm 2. As a result, the number of passive elements is given as $N - |\mathcal{A}^\star|$, and the total power consumption of a dynamic HR-RIS system is given by

$$P_{\text{H}}^{\text{dyn.}} = \frac{P_{\text{BS}}}{\tau_{\text{BS}}} + \frac{P_{\text{a}}}{\tau_{\text{a}}} + P_{\text{c,H}}^{\text{dyn.}} + (N - |\mathcal{A}^\star|)P_{\text{p}} + NP_{\text{SW}}, \tag{41}$$

where the last term accounts for the total power consumed by the N switches in Fig. 1(b), each requires a power of P_{SW}, and

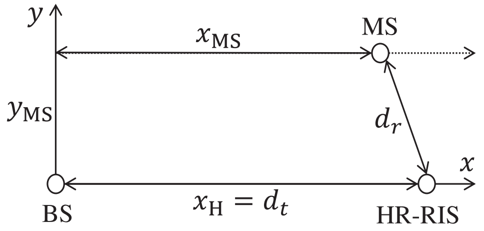


Fig. 2. Horizontal locations of the HR-RIS, BS, and MS.

$P_{\mathrm{c,H}}^{\mathrm{dyn.}}$ is the total circuit power consumption of the dynamic HR-RIS-aided MIMO system, given as $P_{\mathrm{c,H}}^{\mathrm{dyn.}} = N_t P_{\mathrm{BS,dynamic}} + |\mathcal{A}^\star| P_{\mathrm{a,dynamic}} + P_{\mathrm{BS,static}} + P_{\mathrm{a,static}}$ [52].

C. Conventional RIS Scheme

We compare the total power consumption of the proposed HR-RIS-assisted system to that of the conventional RIS-assisted system. We note that in the latter, there is no active relay element, but $N = M + K$ passive reflecting elements are used at the RIS. Therefore, the total power consumption of the RIS-aided system can be expressed as

$$P_{\mathrm{RIS}} = \frac{P_{\mathrm{BS}}}{\tau_{\mathrm{BS}}} + P_{\mathrm{c,RIS}} + N P_{\mathrm{p}}, \tag{42}$$

where $P_{\mathrm{c,RIS}} = N_t P_{\mathrm{BS,dynamic}} + P_{\mathrm{BS,static}}$ is the total circuit power of the dynamic HR-RIS-aided MIMO system

By comparing (40) and (41) to (42), the increased total power consumption of the fixed and dynamic HR-RIS-aided systems with respect to the RIS-assisted system can be respectively given as $\Delta P_{\mathrm{H}}^{\mathrm{fix.}} = \frac{P_{\mathrm{a}}}{\tau_{\mathrm{a}}} + K(P_{\mathrm{a,dynamic}} - P_{\mathrm{p}}) + P_{\mathrm{a,static}}$ and $\Delta P_{\mathrm{H}}^{\mathrm{dyn.}} = \frac{P_{\mathrm{a}}}{\tau_{\mathrm{a}}} + |\mathcal{A}^\star|(P_{\mathrm{a,dynamic}} - P_{\mathrm{p}}) + P_{\mathrm{a,static}} + N P_{\mathrm{p}} + N P_{\mathrm{SW}}$, respectively, which linearly increases with the number of active elements, i.e., K in the fixed HR-RIS and $|\mathcal{A}^\star|$ in the dynamic HR-RIS. The numerical comparison of these architectures will be presented in the next section.

VI. Simulation Results

In this section, numerical results are provided to validate the proposed HR-RIS schemes. We assume that uniform linear arrays (ULAs) are deployed at the BS and MS, respectively. In contrast, an uniform planar array (UPA) of N elements is employed for the HR-RIS. Furthermore, half-wavelength distancing between array elements are assumed for the BS, MS, and HR-RIS. As shown in Fig. 2, in a two-dimensional coordinate, we assume that the BS is deployed at a fixed location (0,0). In contrast, the HR-RIS and MS are located at $(x_{\mathrm{H}}, 0)$ and $(x_{\mathrm{MS}}, y_{\mathrm{MS}})$, respectively, where x_{H} and x_{MS} can vary. Thus, the BS-MS distance is $d_0 = \sqrt{x_{\mathrm{MS}}^2 + y_{\mathrm{MS}}^2}$, while those between the BS and the HR-RIS and between the HR-RIS and the MS are $d_t = x_{\mathrm{H}}$ and $d_r = \sqrt{(x_{\mathrm{H}} - x_{\mathrm{MS}})^2 + y_{\mathrm{MS}}^2}$, respectively. The path loss of a link distance d is given by [7], [15] $\beta(d) = \beta_0 (\frac{d}{1\mathrm{m}})^{-\epsilon}$, where β_0 is the path loss at the reference distance of 1 m (m), and ϵ is the path loss exponent.

For small-scale fading, we assume the Rician fading channel model [7], [15]. As a result, the small-scale fading channel from the BS to the HR-RIS can be modeled as [7], [15]

$$\bar{\boldsymbol{H}}_t = \left(\sqrt{\frac{\kappa_t}{1+\kappa_t}} \boldsymbol{H}_t^{\mathrm{LoS}} + \sqrt{\frac{1}{1+\kappa_t}} \boldsymbol{H}_t^{\mathrm{NLoS}} \right), \tag{43}$$

where $\boldsymbol{H}_t^{\mathrm{LoS}}$ and $\boldsymbol{H}_t^{\mathrm{NLoS}}$ represent the deterministic LoS and non-LoS (NLoS) components, respectively. The NLoS channel is modeled by the Rayleigh fading, with the entry on the ith row and jth column of $\boldsymbol{H}_t^{\mathrm{NLoS}}$ being given as $h_{t,ij}^{\mathrm{NLoS}} \sim \mathcal{CN}(0,1)$. The LoS component for each channel is modeled as a deterministic component, i.e., $\boldsymbol{H}_t^{\mathrm{LoS}} = \boldsymbol{a}_{\mathrm{H}}(\theta_{\mathrm{H}}, \phi_{\mathrm{H}}) \boldsymbol{a}_{\mathrm{BS}}^H(\theta_{\mathrm{BS}})$, where $\boldsymbol{a}_{\mathrm{BS}}(\theta_{\mathrm{BS}})$ and $\boldsymbol{a}_{\mathrm{H}}(\theta_{\mathrm{H}}, \phi_{\mathrm{H}})$ are the array response vectors at the BS and HR-RIS, respectively. Here, the nth element of $\boldsymbol{a}_{\mathrm{BS}}(\theta_{\mathrm{BS}})$ and $\boldsymbol{a}_{\mathrm{H}}(\theta_{\mathrm{H}}, \phi_{\mathrm{H}})$ are given as $a_{\mathrm{BS},n}(\theta_{\mathrm{BS}}) = e^{j\pi(n-1)\sin\theta_{\mathrm{BS}}}, n = 1, \ldots, N_t$ and $a_{\mathrm{H},n}(\theta_{\mathrm{H}}, \phi_{\mathrm{H}}) = e^{j\pi(\lfloor \frac{n}{N_x} \rfloor \sin\theta_{\mathrm{H}} \sin\phi_{\mathrm{H}} + (n - \lfloor \frac{n}{N_x} \rfloor N_x) \sin\theta_{\mathrm{H}} \cos\phi_{\mathrm{H}})}$, $n = 1, \ldots, N$ [7]. θ_{BS}, $\theta_{\mathrm{H}} \in [0, 2\pi)$ denote the angle-of-departure (AoD) at the BS and the azimuth angle-of-arrival (AoA) at the HR-RIS, respectively, and $\phi_{\mathrm{H}} \in [-\pi/2, \pi/2)$ denotes the elevation AoA at the HR-RIS. Furthermore, in (43), κ_t is the Rician factor. With $\kappa_t \to 0$, $\bar{\boldsymbol{H}}_t$ approaches the Rayleigh fading channel, and with $\kappa_t \to \infty$, $\bar{\boldsymbol{H}}_t$ becomes the LoS channel. The channel matrix between the BS and the HR-RIS is obtained by $\boldsymbol{H}_t = \sqrt{\beta(d_t)} \bar{\boldsymbol{H}}_t$. Those between the HR-RIS and MS and between the BS and MS are modeled similarly.

In this work, we set $\beta_0 = -30$ dB, and the noise power is computed as $\sigma^2 = -169$ dBm/Hz $+ 10\log_{10}$ BW + NF, where BW = 20 MHz and NF = 10 dB are set for the system bandwidth and noise figure, respectively, based on [7]. In the simulations, the normalized residual self-interference power is set to 1 dB, i.e., $\eta = 1$ dB [44], [45]. Unless otherwise stated, we assume that the MS and HR-RIS are deployed at $(x_{\mathrm{MS}}, y_{\mathrm{MS}}) = (45, 2)$ m and $(x_{\mathrm{H}}, y_{\mathrm{H}}) = (50, 0)$, respectively. Note that the BS is placed at (0,0) m. Therefore, we set the Rician factors of the channels between the BS and MS, between the BS and HR-RIS, and between the HR-RIS and MS to $\{\kappa_d, \kappa_t, \kappa_r\} = \{0, 1, \infty\}$, respectively [7], [15]. This implies that the channel between the HR-RIS and the MS is dominated by the LoS link, while the other channels are dominated by the NLoS components. Furthermore, the path loss of the channels between the BS and MS, between the BS and HR-RIS, and between the HR-RIS and MS are set to $\{\epsilon_d, \epsilon_t, \epsilon_r\} = \{3.5, 2.2, 2.0\}$, respectively [7]. The component power consumption is assumed as follows: $P_{\mathrm{BS,dynamic}} = 40$ dBm, $P_{\mathrm{a,dynamic}} = 35$ dBm, $P_{\mathrm{BS,static}} = 35$ dBm, $P_{\mathrm{a,static}} = 30$ dBm, $P_{\mathrm{p}} = P_{\mathrm{SW}} = 5$ mW, and $\tau_{\mathrm{a}} = \tau_{\mathrm{BS}} = 0.5$ [38], [53], [54]. Then, the EE of a scheme is given as EE $= \frac{\mathrm{BW} \times \mathrm{SE}}{P}$ [bps/W], where P is the total power consumption. In all the simulation results for the fixed HR-RIS scheme, $\mathcal{A}$ is fixed to $\{1, \ldots, K\}$, while to obtain the results of the dynamic HR-RIS, we set $\varsigma = 0.1$ in Algorithm 2 and use modeling tool YALMIP with MOSEK solver. We note that the HR-RIS requires an additional power of $P_{\mathrm{a}}(\boldsymbol{F}, \{\alpha_n\})$. Therefore, for fair comparisons, the transmit power at the BS of the system without the HR-RIS

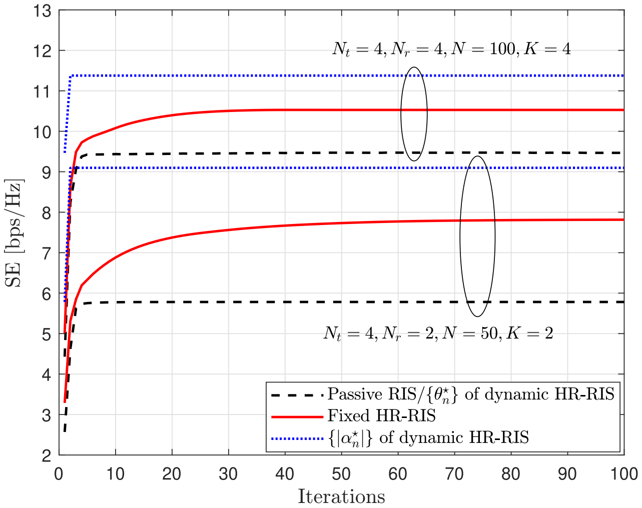


Fig. 3. Convergence of Algorithms 1 and 2 in a 4×2 and 4×4 MIMO systems with $N = \{50, 100\}$, $K = \{2, 4\}$, $P_{\text{BS}} = 20$ dBm, and $P_{\text{a}}^{\max} = 5$ dBm.

is allocated an additional amount of $P_{\text{a}}(\boldsymbol{F}, \{\alpha_n\})$ so that all the compared schemes have the same total transmit power.

We first show in Fig. 3 the convergence of Algorithms 1 and 2 with $N_t = 4, N_r = \{2, 4\}, N = \{50, 100\}, K = \{2, 8\}$, $P_{\text{BS}} = 20$ dBm, and $P_{\text{a}}^{\max} = 5$ dBm. The SE is computed as $f_0(\boldsymbol{F}, \{\alpha_n\})$ in (5). Note that in Algorithm 1, both the phases $\{\theta_n\}$ and amplitudes $\{|\alpha_n|\}$ of $\{\alpha_n\}$ are iteratively updated simultaneously. Unlike that, in Algorithm 2, $\{\theta_n\}$ are obtained first via the AO approach (steps 2–11) with given $|\alpha_n| = 1, \forall n$, which is the same as the passive RIS scheme. Then, with given $\{\theta_n^\star\}$, $\{|\alpha_n|\}, n \in \mathcal{A}^\star$ are solved based on the SCA approach (steps 13–18). Thus, for the dynamic HR-RIS scheme in Algorithm 2, we show the convergence of the solutions to both $\{\theta_n\}$ and $\{|\alpha_n|\}$. It is observed that the passive RIS and dynamic HR-RIS schemes converge fast to reach the convergence after only several iterations. In contrast, the fixed HR-RIS scheme in Algorithm 1 converges slower due to the joint optimization of the transmit beamforming matrix $\boldsymbol{F}$, phases $\{\theta_n\}, \forall n$, and amplitude $\{|\alpha_n|\}, \forall n \in \mathcal{A}$.

A. *Performance Improvement of the Proposed HR-RIS*

In Fig. 4, we show the SE and EE performance improvement of the proposed HR-RIS architectures for a 4×2 MIMO system. Both the HR-RIS and RIS are equipped with $N = 50$ elements, but in the former, only a single active element is deployed, i.e., $K = 1$, $M = N - K = 49$. The simulation results are shown for $P_{\text{a}}^{\max} = \{-5, 5\}$ dBm, and $P_{\text{BS}} = [10, 40]$ dBm. For comparisons, we consider the conventional passive RIS with phases being randomly generated or optimized via the AO approach [7], which are referred to as the *"RIS, random phase"* and *"RIS, AO"*, respectively, in the figures. It is shown that the HR-RIS schemes achieve significant improvement in both the SE and EE compared to the conventional RIS, especially at low and moderate P_{BS}. This numerically justifies the observation in Remark 2, which states that a higher performance gain can be obtained by the HR-RIS for limited P_{BS}. Furthermore, the HR-RIS only requires a small power budget to achieve remarkable improvement in SE and EE for almost all the considered range of P_{BS}. For example, with $P_{\text{BS}} = 20$ dBm, the dynamic HR-RIS only requires $P_{\text{a}}^{\max} = 5$ dBm to attain 40% and 25% improvement in SE and EE, respectively, compared with the passive RIS. However, with a small power budget at the HR-RIS but a large transmit power at the BS, e.g., with $P_{\text{a}}^{\max} = -5$ dBm and $P_{\text{BS}} > 35$ dBm, the performance gain of the HR-RIS becomes less significant, and it performs comparably with the conventional RIS in both SE and EE. This agrees with the discussion in Remark 2.

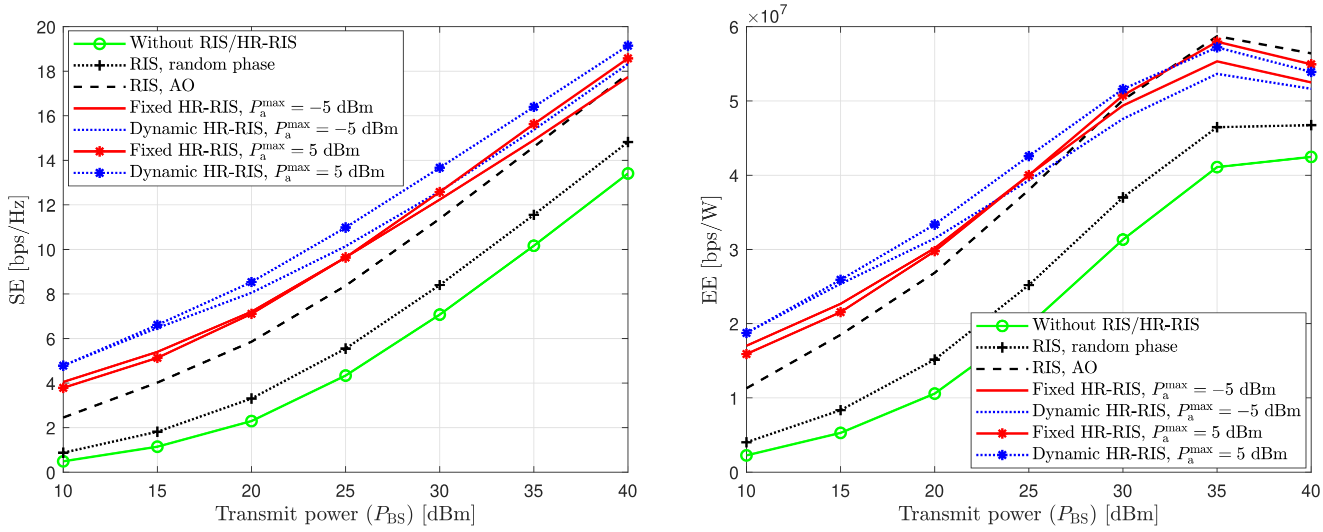
In Fig. 5, we investigate the SEs of the HR-RIS for different locations of the RIS/HR-RIS by varying x_{H}. We note that for a fair comparison, the RIS and HR-RIS are set to be placed at the same position, i.e., $(0, x_{\text{H}})$, with $x_{\text{H}} \in [10, 100]$ m. In this simulation, we set $N_t = 4$, $N_r = 2$, $N = 50$, $K = 2$, $P_{\text{BS}} = 30$ dBm, and $P_{\text{a}}^{\max} = 5$ dBm. The BS and MS are placed at the coordinate (0,0) and (45,2) m, respectively. It is observed that the highest SE is attained when the HR-RIS is closest to the MS, i.e., at $x_{\text{H}} = x_{\text{MS}} = 45$ m, which is similar to the conventional RIS [38]. In particular, when the RIS moves far away from both the BS and MS, e.g., for $x_{\text{H}} \geq 80$ m, the RIS loses the reflecting gains and performs close to the system without RIS. In contrast, the proposed HR-RIS still exhibits considerable performance gains in this harsh scenario.

In Fig. 6, the SEs and EEs are shown for different physical sizes of the RIS/HR-RIS. Specifically, we set $K = 2$, $P_{\text{BS}} = 30$ dBm, $P_{\text{a}}^{\max} = 5$ dBm, and $N = [20, 200]$ for a 4×2 MIMO system. It is seen that the HR-RIS schemes attain a significant improvement in SE for all the considered values of N compared to the conventional RIS. Although this improvement becomes less significant as N increases, the gains of the proposed HR-RIS are still considerable in terms of SE and comparable in terms of EE compared to the RIS. The degradation in the performance gains of the HR-RIS is because at large N, its relaying gain offered by only $K = 2$ active elements is dominated by the reflecting gain of M ($\gg K$) passive elements.

B. *SE-EE Tradeoff of the Proposed HR-RIS*

In this section, we focus on exploring the SE-EE tradeoff of the proposed HR-RIS by varying the number of active elements, i.e., K. Furthermore, we also consider the RIS with $N - K$ elements and the FD-AF relaying schemes with K relay elements for comparison. The relay precoder is obtained based on the well-known singular value decomposition [55]. We consider a 4×2 MIMO system with $N = 50$ and $K = [0, 20]$. It is obvious that for $K = 0$, the HR-RIS performs the same as the conventional passive RIS. Based on Remarks 1 and 2, we consider $(P_{\text{BS}}, P_{\text{a}}^{\max}) = \{(45, -10), (30, 5), (20, 5)\}$ dBm to show the SE-EE tradeoff of the proposed HR-RIS and to compare the HR-RIS with the relay.

We first investigate the SE performance of the HR-RIS as K increases, as shown in Fig. 7. The SEs of the systems without RIS or with $N-$element RIS are constant with K. Furthermore, the following observations are noted:



(a)

(b)

Fig. 4. SEs and EEs of the HR-RIS schemes for a 4 × 2 MIMO system with $N = 50$, $K = 1$, $M = N - K = 49$, and $P_a^{max} = \{-5, 5\}$ dBm. (a) SE performance. (b) EE performance.

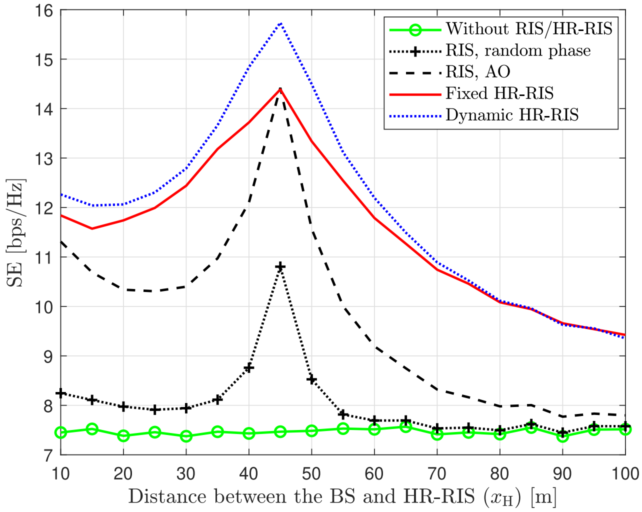


Fig. 5. SE improvement for different positions of the HR-RIS aided a 4 × 2 MIMO system with $N = 50$, $K = 2$, $M = 48$, $P_{BS} = 30$ dBm, and $P_a^{max} = 5$ dBm.

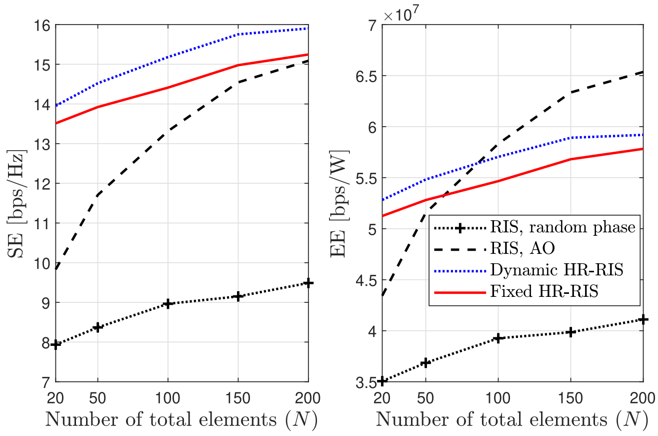
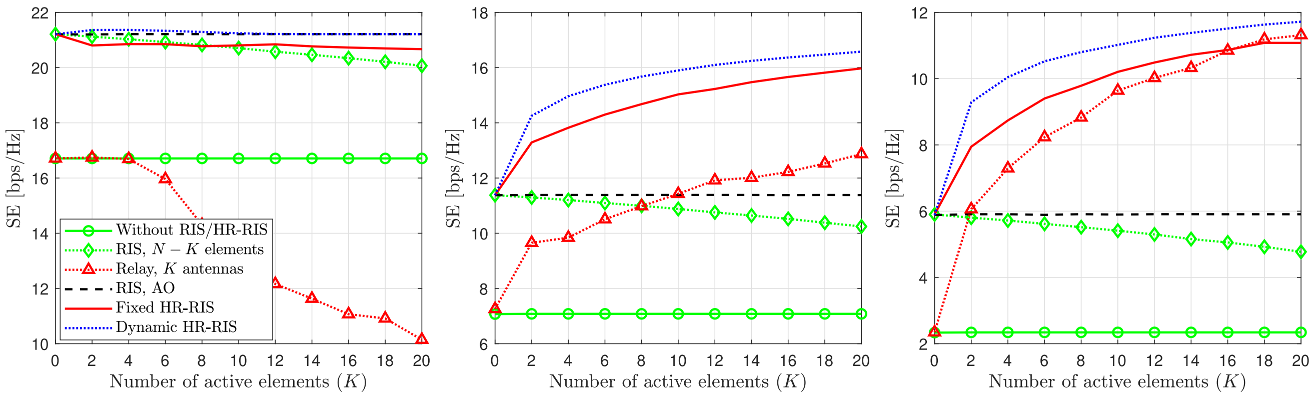


Fig. 6. SEs and EEs of the HR-RIS schemes for a 4 × 2 MIMO system with $K = 2$, $P_{BS} = 30$ dBm, $P_a^{max} = 5$ dBm, and $N = [20, 200]$.

- First, we compare the performances of RIS with $N - K$ passive elements and the HR-RIS with K active elements. Obviously, as K increases, both these two offer less passive reflecting gains, but at the same time, the latter provides more active relaying gains. It can be observed for small K in Fig. 7(b)–(c) that the SE gains of the latter are much more significant than the SE loss of the former. This verifies the motivation of the HR-RIS discussed earlier, i.e., when a few elements are activated, the active relaying gains attained by the HR-RIS are remarkable while the losses in passive reflecting gains are just marginal.
- In Fig. 7(b) and (c), the HR-RIS achieves a great performance improvement compared to the conventional RIS. However, with low P_a^{max} as in Fig. 7(a), the fixed HR-RIS has performance loss, especially at large K. This agrees with the finding in Remark 1, i.e., increasing K does not always guarantee the SE improvement, especially for low P_a^{max}. Therefore, in this case, HR-RISs with a few active elements should be used, which is also sufficient to achieve significant improvement with respect to the conventional RIS for the other cases in Fig. 7(b) and (c). The advantage of the dynamic HR-RIS in terms of SE compared to the fixed HR-RIS can be clearly seen in Fig. 7(a)–(c). In particular, a large number of active elements causes no performance loss for the dynamic HR-RIS because the active elements can be deactivated to serve as passive reflecting ones, as seen in Fig. 7(a).
- For all the considered scenarios, the HR-RIS generally outperforms the relay. The reason is that the HR-RIS can offer not only the relaying gains but also the reflecting gains as a passive RIS. The relay only outperforms the

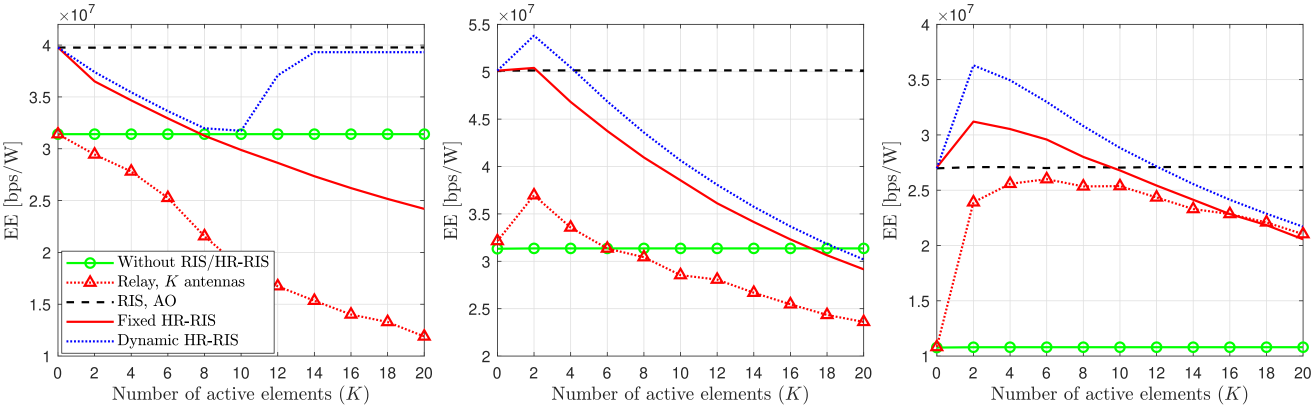


(a)

(b)

(c)

Fig. 7. SEs of the HR-RIS, passive RIS, and relay schemes with K active elements, $N_t = 4$, $N_r = 2$, $N = 50$ and $K = [0, 20]$. (a) $(P_{\text{BS}}, P_a^{\max}) = (45, -10)$ dBm. (b) $(P_{\text{BS}}, P_a^{\max}) = (30, 5)$ dBm. (c) $(P_{\text{BS}}, P_a^{\max}) = (20, 5)$ dBm.



(a)

(b)

(c)

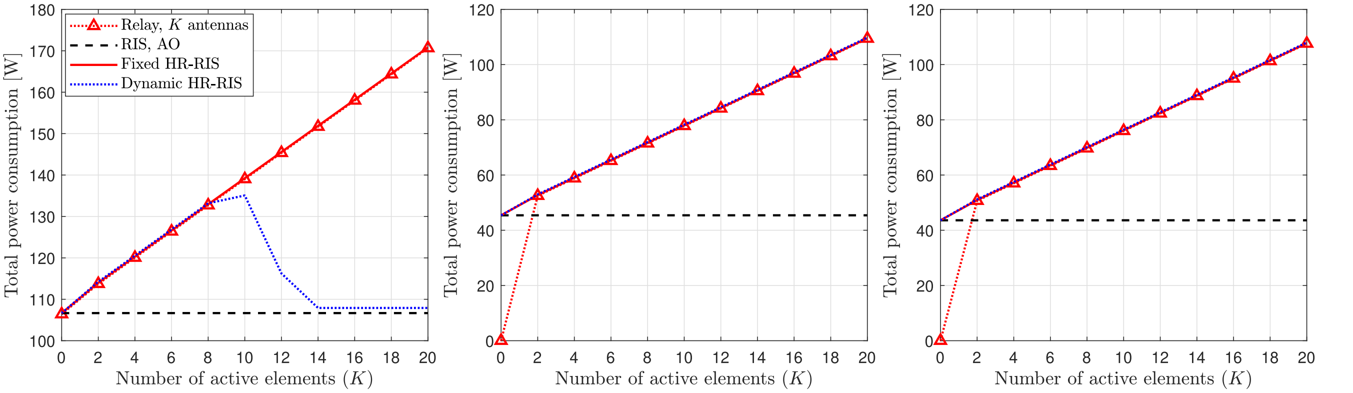
Fig. 8. EEs of the HR-RIS, passive RIS, and relay schemes with K active elements, $N_t = 4$, $N_r = 2$, $N = 50$ and $K = [0, 20]$. (a) $(P_{\text{BS}}, P_a^{\max}) = (45, -10)$ dBm. (b) $(P_{\text{BS}}, P_a^{\max}) = (30, 5)$ dBm. (c) $(P_{\text{BS}}, P_a^{\max}) = (20, 5)$ dBm.

fixed HR-RIS for $K \geq 16$ in Fig. 7(c). However, such a large number of active elements is not suggested for the HR-RIS, because it can attain a satisfactory performance improvement with a small K. For example, in the case of moderate P_{BS} and $P_{\text{a}}^{\max}$ as in Fig. 7(b), the HR-RIS requires only $K = 2$ active elements to outperform a relay with $K = 20$ antennas.

We further investigate the EEs of the HR-RIS versus the numbers of active elements in Fig. 8. We assume the same simulation results as those in Fig. 7. It is observed for $K > 0$ that, the EE of the HR-RIS rapidly decreases as K increases due to the increased power consumption. For small K, the proposed HR-RIS schemes achieves much higher or comparable EEs than the conventional AO-based RIS and the relay. At large K, both the HR-RIS and relay schemes can be outperformed by the conventional passive RIS in terms of EE. This is reasonable because while the passive RIS has relatively low power consumption, that of the HR-RIS and relay almost linearly increases with K. We note an observation from Fig. 8(a), that is, at moderate and large K, the EE of the dynamic HR-RIS is nondecreasing with K. This is because with a small power budget, the dynamic HR-RIS only selected the best elements to serve as active ones. Thus, significant power can be saved.

C. Power Consumption of the Proposed HR-RISs

We compare the HR-RIS, RIS, and relay schemes in terms of total power consumption in Fig. 9. We assume the same simulation parameters as those in Figs. 7 and 8. The total power consumption of the MIMO system aided by the HR-RIS schemes and RIS, i.e., $P_{\text{H}}^{\text{fix.}}$, $P_{\text{H}}^{\text{dyn.}}$, and P_{RIS}, are computed based on (40), (41) and (42), respectively. As analyzed in Section V, it is clear that for $K > 0$, the proposed HR-RIS architectures and the relay require higher power consumption than the RIS. Furthermore, in the HR-RIS, the overall power consumption is dominated by that of the active elements. Therefore, with the same number of active elements, the HR-RIS architectures and the relay approximately have the same total power consumption, which almost linearly increases with K. In Fig. 9(a), $P_{\text{H}}^{\text{dyn.}}$ increases at first, then starts to decrease. The reason is that, with numerous active elements,



(a) (b) (c)

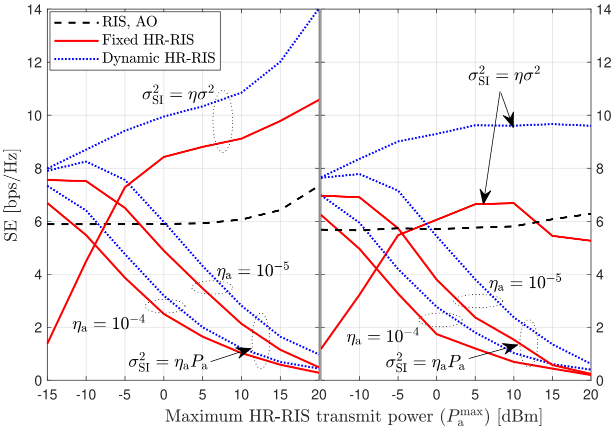
Fig. 9. Total power consumption of the HR-RIS, passive RIS, and relay schemes with K active elements, $N_t = 4$, $N_r = 2$, $N = 50$ and $K = [0, 20]$. (a) $(P_{\text{BS}}, P_a^{\max}) = (45, -10)$ dBm. (b) $(P_{\text{BS}}, P_a^{\max}) = (30, 5)$ dBm. (c) $(P_{\text{BS}}, P_a^{\max}) = (20, 5)$ dBm.

$P_{\text{a}}^{\max} = -5$ dBm is not sufficient to share among all the active elements such that $|\alpha_n| > 1, \forall n \in \mathcal{A}^\star$. Therefore, the dynamic HR-RIS deactivates a subset of the elements in $\mathcal{A}^\star$ to save power and improve the SE-EE tradeoff, as justified in Figs. 7(a) and 8(a). In contrast, when the power budget $P_{\text{a}}^{\max}$ is sufficiently large, as in Fig. 9(b) and (c), all the active elements are employed in the dynamic HR-RIS. In this case, the dynamic HR-RIS has a slightly higher power consumption than its fixed counterpart due to the employment of the switches, as observed in (41).

D. Effects of Self-Interference and CSI Errors

In this subsection, we investigate the SE of the proposed HR-RIS in the presence of CSI errors and when the residual self-interference power is proportional to the HR-RIS transmit power, i.e., $\sigma_{\text{SI}}^2 = \eta_{\text{a}} P_{\text{a}}$ [44]. We consider $\eta_{\text{a}} = \{10^{-4}, 10^{-5}\}$, corresponding to 40 dB and 50 dB self-interference cancellation [44], [56]. The imperfect channel estimate between the BS and the HR-RIS is modeled as [18] $\boldsymbol{H}_t = \sqrt{\beta(d_t)}(\bar{\boldsymbol{H}}_t - \tilde{\boldsymbol{H}}_t)$, where $\bar{\boldsymbol{H}}_t$ is given in (43), and $\tilde{\boldsymbol{H}}_t$ represents the channel estimation errors, whose entries have distribution $\mathcal{CN}(0, \varepsilon^2)$. The CSI errors for $\boldsymbol{H}_d$ and $\boldsymbol{H}_r$ are modeled similarly.

In Fig. 10, we show the SE versus the HR-RIS maximum transmit power, i.e., $P_{\text{a}}^{\max}$ with perfect CSI (in Fig. 10(a)) and imperfect CSI ($\varepsilon^2 = 0.2$ is assumed in Fig. 10(b)). In both scenarios, it is seen for the case $\sigma_{\text{SI}}^2 = \eta\sigma^2$ that the SE significantly increases with $P_{\text{a}}^{\max}$. In contrast, with $\sigma_{\text{SI}}^2 = \eta_{\text{a}} P_{\text{a}}$, the SE gain with respect to the conventional RIS is only achieved for small and moderate $P_{\text{a}}^{\max}$. At high $P_{\text{a}}^{\max}$, the interference power becomes significantly large and causes performance degradation. However, considerable SE improvement is achieved for small or moderate $P_{\text{a}}^{\max}$, especially with a smaller η_{a} (i.e., better self-interference cancellation). Here, we note an observation on the conventional RIS that its SE increases with $P_{\text{a}}^{\max}$. This is because an additional amount of $P_{\text{a}}^{\max}$ is allocated to its BS for fair comparison, as mentioned earlier. By comparing Fig. 10(a) and (b), it is seen that the performance is degraded in the presence of imperfect CSI. However, it is interesting to see that the performance gain of the dynamic HR-RIS compared to the fixed one is more significant when there are CSI errors.



(a) (b)

Fig. 10. SE improvement for different residual self-interference models of the HR-RIS aided a 4×2 MIMO system with $N = 50$, $K = 4$, $M = 46$, $P_{\text{BS}} = 20$ dBm, and $P_{\text{a}}^{\max} = [-20, 20]$ dBm. (a) Perfect CSI. (b) Imperfect CSI.

VII. Conclusion

We proposed a novel HR-RIS architecture to aided MIMO communication systems with substantially improved SE and EE compared to that aided by the conventional passive RIS. The HR-RIS is equipped with a few active elements, which are capable of adjusting the power of the incident signals, and numerous passive reflecting elements. The SE maximization problem in the HR-RIS-aided MIMO system is solved via the AO and power allocation strategies, resulting in two different architectures, namely, the fixed and dynamic HR-RIS. The analytical results show that the HR-RIS should be equipped with a small number of active elements because employing numerous active elements does not always guarantee the improvement in SE or EE, while consuming high power. In addition, the HR-RIS can achieve considerable SE/EE improvement in harsh scenarios such as when the transmit power is low and/or when the HR-RIS is located far away from the transmitter. The performance gain of the HR-RIS has been numerically justified via intensive simulations. The results show that the HR-RIS potentially performs better than both the passive RIS and relaying schemes in certain scenarios. For future works, the CSI acquisition and HR-RIS performance under imperfect CSI and hardware imperfections need further investigation.

References

[1] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1838–1851, Mar. 2019.

[2] S. Hu, F. Rusek, and O. Edfors, "Beyond massive MIMO: The potential of positioning with large intelligent surfaces," *IEEE Trans. Signal Process.*, vol. 66, no. 7, pp. 1761–1774, Apr. 2018.

[3] C. Liaskos, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. F. Akyildiz, "Using any surface to realize a new paradigm for wireless communications," *Commun. ACM*, vol. 61, pp. 30–33, 2018.

[4] E. Basar *et al.*, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116753–116773, 2019.

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

[6] J. He, H. Wymeersch, L. Kong, O. Silvén, and M. Juntti, "Large intelligent surface for positioning in millimeter wave MIMO systems," in *Proc. IEEE Conf. Veh. Technol.*, 2020, pp. 1–5.

[7] S. Zhang and R. Zhang, "Capacity characterization for intelligent reflecting surface aided mimo communication," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1823–1838, Aug. 2020.

[8] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling large intelligent surfaces with compressive sensing and deep learning," *IEEE Access*, vol. 9, pp. 44304–44321, 2021.

[9] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Deep learning for large intelligent surfaces in millimeter wave and massive MIMO systems," in *Proc. IEEE Conf. Glob. Commun.*, 2019, pp. 1–6.

[10] S. Hu, F. Rusek, and O. Edfors, "The potential of using large antenna arrays on intelligent surfaces," in *Proc. IEEE Conf. Veh. Technol.*, 2017, pp. 1–6.

[11] S. Hu, F. Rusek, and O. Edfors, "Beyond massive MIMO: The potential of data transmission with large intelligent surfaces," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2746–2758, May 2018.

[12] J. He, H. Wymeersch, T. Sanguanpuak, O. Silvén, and M. Juntti, "Adaptive beamforming design for mmwave RIS-aided joint localization and communication," in *Proc. IEEE Conf. Wireless Commun. Netw. Workshops*, 2020, pp. 1–6.

[13] M. Jung, W. Saad, M. Debbah, and C. S. Hong, "Asymptotic optimality of reconfigurable intelligent surfaces: Passive beamforming and achievable rate," in *Proc. IEEE Int. Conf. Commun.*, 2020, pp. 1–6.

[14] Ö. Özdogan, E. Björnson, and E. G. Larsson, "Using intelligent reflecting surfaces for rank improvement in MIMO communications," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2020, pp. 9160–9164.

[15] 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, Nov. 2019.

[16] P. Wang, J. Fang, X. Yuan, Z. Chen, and H. Li, "Intelligent reflecting surface-assisted millimeter wave communications: Joint active and passive precoding design," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 14960–14973, Dec. 2020.

[17] H. Zhang, B. Di, L. Song, and Z. Han, "Reconfigurable intelligent surfaces assisted communications with limited phase shifts: How many phase shifts are enough?," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4498–4502, Apr. 2020.

[18] H. Guo, Y.-C. Liang, J. Chen, and E. G. Larsson, "Weighted sum-rate maximization for reconfigurable intelligent surface aided wireless networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3064–3076, May 2020.

[19] Y. Han, W. Tang, S. Jin, C.-K. Wen, and X. Ma, "Large intelligent surface-assisted wireless communication exploiting statistical CSI," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 8238–8242, Aug. 2019.

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

[21] Y. Zhang, C. Zhong, Z. Zhang, and W. Lu, "Sum rate optimization for two way communications with intelligent reflecting surface," *IEEE Commun. Lett.*, vol. 24, no. 5, pp. 1090–1094, May 2020.

[22] J. V. Alegría and F. Rusek, "Achievable rate with correlated hardware impairments in large intelligent surfaces," in *Proc. IEEE Int. Workshop Comput. Adv. Multi-Sensor Adaptive Process.*, 2019, pp. 559–563.

[23] S. Gong *et al.*, "Towards smart wireless communications via intelligent reflecting surfaces: A contemporary survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 4, pp. 2283–2314, Oct.–Dec. 2020.

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

[25] X. Yu, D. Xu, and R. Schober, "MISO wireless communication systems via intelligent reflecting surfaces," in *Proc. IEEE Int. Conf. Commun.*, 2019, pp. 735–740.

[26] Y. Yang, S. Zhang, and R. Zhang, "IRS-enhanced OFDM: Power allocation and passive array optimization," in *Proc. IEEE Conf. Glob. Commun.*, 2019, pp. 1–6.

[27] J. Yuan, Y.-C. Liang, J. Joung, G. Feng, and E. G. Larsson, "Intelligent reflecting surface-assisted cognitive radio system," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 675–687, Jan. 2021.

[28] B. Di, H. Zhang, L. Li, L. Song, Y. Li, and Z. Han, "Practical hybrid beamforming with finite-resolution phase shifters for reconfigurable intelligent surface based multi-user communications," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4565–4570, Apr. 2020.

[29] K. Ying *et al.*, "GMD-based hybrid beamforming for large reconfigurable intelligent surface assisted millimeter-wave massive MIMO," *IEEE Access*, vol. 8, pp. 19530–19539, 2020.

[30] N. S. Perović, L.-N. Tran, M. Di Renzo, and M. F. Flanagan, "Achievable rate optimization for MIMO systems with reconfigurable intelligent surfaces," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3865–3882, Jun. 2021.

[31] C. Pan *et al.*, "Intelligent reflecting surface aided MIMO broadcasting for simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1719–1734, Aug. 2020.

[32] S. Hong, C. Pan, H. Ren, K. Wang, and A. Nallanathan, "Artificial-noise-aided secure mimo wireless communications via intelligent reflecting surface," *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7851–7866, 2020.

[33] Z. Wang, L. Liu, and S. Cui, "Channel estimation for intelligent reflecting surface assisted multiuser communications: Framework, algorithms, and analysis," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6607–6620, Oct. 2020.

[34] C. Liu, X. Liu, D. W. K. Ng, and J. Yuan, "Deep residual learning for channel estimation in intelligent reflecting surface-assisted multi-user communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 898–912, Feb. 2022.

[35] N. Landsberg and E. Socher, "Design and measurements of 100 GHz reflectarray and transmitarray active antenna cells," *IEEE Trans. Antennas Propag.*, vol. 65, no. 12, pp. 6986–6997, Dec. 2017.

[36] N. Landsberg and E. Socher, "A low-power 28-nm CMOS FD-SOI reflection amplifier for an active f-band reflectarray," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 10, pp. 3910–3921, Oct. 2017.

[37] Y. Lin, S. Jin, M. Matthaiou, and X. You, "Tensor-based algebraic channel estimation for hybrid IRS-Assisted MIMO-OFDM," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3770–3784, Jun. 2021.

[38] E. Björnson, O. özdogan, and E. G. Larsson, "Intelligent reflecting surface vs. decode-and-forward: How large surfaces are needed to beat relaying?," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 244–248, Feb. 2020.

[39] C. Pan *et al.*, "Multicell MIMO communications relying on intelligent reflecting surfaces," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5218–5233, Aug. 2020.

[40] R. Schroeder, J. He, and M. Juntti, "Passive RIS vs. hybrid RIS: A comparative study on channel estimation," in *Proc. IEEE Conf. Veh. Technol.*, 2021, pp. 1–7.

[41] N. T. Nguyen, Q.-D. Vu, K. Lee, and M. Juntti, "Spectral efficiency optimization for hybrid relay-reflecting intelligent surface," in *Proc. IEEE Int. Conf. Commun. Workshop*, 2021, pp. 1–6.

[42] 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, Jan. 2020.

[43] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface aided wireless communications: A tutorial," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3313–3351, May 2021.

[44] R. Malik and M. Vu, "Optimal transmission using a self-sustained relay in a full-duplex MIMO system," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 2, pp. 374–390, Feb. 2019.

[45] D. Bharadia and S. Katti, "Full duplex MIMO radios," in *Proc. 11th USENIX Symp. Netw. Syst. Des. Implement.*, 2014, pp. 359–372.

[46] G. C. Alexandropoulos and E. Vlachos, "A hardware architecture for reconfigurable intelligent surfaces with minimal active elements for explicit channel estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2020, pp. 9175–9179.

[47] R. Long, Y.-C. Liang, Y. Pei, and E. G. Larsson, “Active reconfigurable intelligent surface aided wireless communications,” *IEEE Trans. Wireless Commun.*, vol. 25, no. 11, pp. 3699–3703, Nov. 2021.

[48] M. H. Khoshafa, T. M. Ngatched, M. H. Ahmed, and A. R. Ndjiongue, “Active reconfigurable intelligent surfaces-aided wireless communication system,” *IEEE Commun. Lett.*, vol. 25, no. 11, pp. 3699–3703, Nov. 2021.

[49] C. You and R. Zhang, “Wireless communication aided by intelligent reflecting surface: Active or passive?,” *IEEE Wireless Commun. Lett.*, vol. 10, no. 12, pp. 2659–2663, 2021.

[50] K. Liu, Z. Zhang, L. Dai, S. Xu, and F. Yang, “Active reconfigurable intelligent surface: Fully-connected or sub-connected?,” *IEEE Commun. Lett.*, vol. 26, no. 1, pp. 167–171, Jan. 2022.

[51] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD, USA: Johns Hopkins Univ. Press, 1996.

[52] A. Zappone, P. Cao, and E. A. Jorswieck, “Energy efficiency optimization in relay-assisted MIMO systems with perfect and statistical CSI,” *IEEE Trans. Signal Process.*, vol. 62, no. 2, pp. 443–457, Jan. 2014.

[53] S. Gong, S. Wang, S. Chen, C. Xing, and L. Hanzo, “Robust energy efficiency optimization for amplify-and-forward MIMO relaying systems,” *IEEE Trans. Wireless Commun.*, vol. 18, no. 9, pp. 4326–4343, Sep. 2019.

[54] N. T. Nguyen and K. Lee, “Unequally sub-connected architecture for hybrid beamforming in massive MIMO systems,” *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 1127–1140, Feb. 2020.

[55] W. Guan and H. Luo, “Joint MMSE transceiver design in non-regenerative MIMO relay systems,” *IEEE Commun. Lett.*, vol. 12, no. 7, pp. 517–519, Jul. 2008.

[56] M. Heino *et al.*, “Recent advances in antenna design and interference cancellation algorithms for in-band full duplex relays,” *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 91–101, May 2015.

Nhan Thanh Nguyen (Member, IEEE) received the B.S. degree in electronics and communications engineering from the Hanoi University of Science and Technology, Hanoi, Vietnam, in 2014, and the M.S. and Ph.D. degrees in electrical and information engineering from the Seoul National University of Science and Technology, Seoul, South Korea, in 2017 and 2020, respectively. From October 2019 to March 2020, he was a Visiting Researcher with North Carolina State University, Raleigh, NC, USA. Since September 2020, he has been a Postdoctoral Researcher with the Centre for Wireless Communication, University of Oulu, Oulu, Finland. His research interests include signal processing, optimization, and applied machine learning for wireless communication with a focus on massive MIMO systems. He was the recipient of the Outstanding M.S. and Ph.D. Thesis Awards of the Seoul National University of Science and Technology, in 2017 and 2020, respectively, and Best Paper Award at the International Conference on Advanced Technologies for Communications, in 2021.

Quang-Doanh Vu (Member, IEEE) received the B.S. degree in electrical engineering from the Ho Chi Minh National University of Technology, Ho Chi Minh City, Vietnam, in 2010, and the M.S. and Ph.D. degrees in radio engineering from Kyung Hee University, Seoul, South Korea, in 2012 and 2015, respectively. From October 2015 to August 2020, he was with the Centre for Wireless Communications, University of Oulu, Oulu, Finland. Since September 2020, he has been with Nokia, Oulu, Finland. His research interests include resource allocation, energy-efficient communications, multiuser MIMO systems, and wireless power transfer.

Kyungchun Lee (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2000, 2002, and 2007, respectively. From April 2007 to June 2008, he was a Postdoctoral Researcher with the University of Southampton, Southampton, U.K. From July 2008 to August 2010, he was with Samsung Electronics, Suwon, South Korea. Since September 2010, he has been with the Seoul National University of Science and Technology, Seoul, South Korea. In 2017, he was a Visiting Assistant Professor with North Carolina State University, Raleigh, NC, USA. His research interests include wireless communications and applied machine learning. He was the recipient of the Best Paper Awards at the IEEE International Conference on Communications and IEEE Wireless Communications and Networking Conference in 2009 and 2020, respectively.

Markku Juntti (Fellow, IEEE) received the M.Sc. and Dr.Sc. degrees in electrical engineering from the University of Oulu, Oulu, Finland, in 1993 and 1997, respectively. From 1992 to 1998, he was with the University of Oulu. In academic year 1994 to 1995, he was a Visiting Scholar with Rice University, Houston, TX, USA. From 1999 to 2000, he was the Senior Specialist with Nokia Networks, Oulu, Finland. Since 2000, he has been a Professor of communications engineering with the Centre for Wireless Communications (CWC), University of Oulu, where he leads the Communications Signal Processing Research Group. He is also the Head of CWC – Radio Technologies Research Unit and an Adjunct Professor with the Department of Electrical and Computer Engineering, Rice University. He is the author or coauthor in about 500 papers published in international journals and conference records and in books *Wideband CDMA for UMTS* in 2000–2010, *Handbook of Signal Processing Systems* in 2013 and 2018, and *5G Wireless Technologies* in 2017. His research interests include signal processing for wireless networks and communication and information theory.

He is the Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He was the Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He was the Secretary of IEEE Communication Society Finland Chapter during 1996–1997 and the Chairman for years 2000–2001. He has been the Secretary of the Technical Program Committee of the 2001 IEEE International Conference on Communications, and the Chair or Co-Chair of the Technical Program Committee of several conferences, including 2006 and 2021 IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, Signal Processing for Communications Symposium of IEEE Globecom 2014, Symposium on Transceivers and Signal Processing for 5G Wireless and mm-Wave Systems of IEEE GlobalSIP 2016, ACM NanoCom 2018, and 2019 International Symposium on Wireless Communication Systems. He was also the General Chair of 2011 IEEE Communication Theory Workshop in 2011, and 2022 IEEE Workshop on Signal Processing Advances in Wireless Communications.