

STARS-Enabled Full-Duplex Two-Way mMIMO System Under Spatially-Correlated Channels

Anastasios Papazafeiropoulos, Pandelis Kourtessis, Symeon Chatzinotas, Dimitra I. Kaklamani, Iakovos S. Venieris

Abstract—Low-cost reconfigurable intelligent surfaces (RISs) are being considered as promising physical-layer technology for next-generation wireless networks due to their ability to re-engineer the propagation environment by tuning their elements. Although RIS-aided systems have many advantages, one of their major bottlenecks is that they can provide coverage only in front of the surface. Fortunately, simultaneous transmitting and reflecting surface (STARS)-assisted systems have emerged to fill this gap by providing 360° wireless coverage. In parallel, full-duplex (FD) communication offers a higher achievable rate through efficient spectrum utilization compared to the half-duplex (HD) counterpart. Moreover, two-way/bi-directional communications in an FD system can further enhance the system's spectral efficiency. Hence, in this paper, we propose a STARS-enabled massive MIMO deployment in an FD two-way communication network for highly efficient spectrum utilization, while covering the dead zones around the STARS. This model enables simultaneous information exchange between multiple nodes, while *potentially* doubling the spectral efficiency (SE). By invoking the use-and-then-forget (UaTF) combining scheme, we derive a closed-form expression for an achievable SE at each user of the system considering both uplink and downlink communications based on statistical channel state information (CSI), while also accounting for imperfect CSI and correlated fading conditions. Moreover, we formulate an optimization problem to obtain an optimal passive beamforming matrix design at the STARS that maximizes the sum achievable SE. The considered problem is non-convex and we propose a provably-convergent low-complexity algorithm, termed as projected gradient ascent method (ProGrAM), to obtain a stationary solution. Extensive numerical results are provided to establish the performance superiority of the FD STARS-enabled system over the HD STARS-enabled and FD conventional RIS (cRIS)-enabled counterparts, and also to show the effect of different parameters of interest on the system performance.

Index Terms—Simultaneously transmitting and reflecting surface (STARS), massive MIMO, full-duplex communications, two-way communications, correlated Rayleigh fading.

I. INTRODUCTION

RECONFIGURABLE intelligent surface (RIS) has gained immense popularity in the past few years as a groundbreaking physical-layer hardware technology for beyond-fifth-generation (B5G) wireless communications [1]–[3]. RISs consist of a large number of sub-wavelength (nearly) passive

elements made from software-controlled meta-materials [4], [5], which makes them capable of re-engineering the wireless propagation environment. The performance superiority of RIS-aided wireless communication systems over their conventional (non-RIS-aided) counterparts has been well established in terms of spectral efficiency [6]–[9], energy efficiency [10], [11], and transmit power minimization [12], [13] among other performance metrics.

Unfortunately, in a conventional RIS (cRIS) enabled system, the elements are only able to reflect the incident electromagnetic waves (EM) to one side of the surface, leading to limited coverage and topological constraints that restrict practical deployments. A constraint example is that the transmitter and the receiver should be located on the same side of the cRIS. This limitation has recently been overcome by a new hardware known as simultaneously transmitting and reflecting surface (STARS) which is a metasurface being capable of simultaneously transmitting and reflecting the impinging EM waves to both sides of the surface [14]–[18]. The authors in [14] provided a general hardware model and two channel models corresponding to the near- and far-field regions of STARS with two user equipments (UEs). Therein, it was shown that the coverage and diversity gain is greater than the corresponding cRIS-assisted system. Also, in [15], three operating protocols for adjusting the transmission and reflection coefficients of the transmitted and reflected signals from the STARS were presented, namely, energy splitting (ES), mode switching (MS), and time switching (TS). It is noteworthy that STARS has attracted significant interest not only in academia but also in industry [19].

In parallel, the in-band FD communication has gained much interest as a potential technology for B5G wireless systems due to its ability to allow both uplink (UL) and downlink (DL) operations throughout the entire frequency band which leads to a higher SE compared to its half-duplex (HD) counterpart [20], [21]. Although FD is inherently impaired by loop interference (LI), this can be partially suppressed by appropriate interference mitigation techniques. It is interesting to note that although cRIS/STARS naturally supports full-duplex (FD) communications, the literature on FD cRIS-/STARS-enabled communications is rather limited. Recently, in this direction, FD has been incorporated to cRIS in [22]–[26]. More specifically, the authors in [22] obtained an optimal transmit structure and passive beamforming matrix (PBM) using a majorization-minimization (MM) algorithm to maximize the weighted sum rate for a cRIS-aided system consisting of a multi-antenna FD BS and multiple dual-antenna FD UEs. Similarly, the problem of obtaining optimal transmit structure and PBM for a cRIS-aided FD system consisting of two multi-transmit-antenna and

Anastasios Papazafeiropoulos is with the Communications and Intelligent Systems Research Group, University of Hertfordshire, Hatfield AL10 9AB, U. K., and with SnT at the University of Luxembourg, Luxembourg (e-mail: tapapazaf@gmail.com).

Pandelis Kourtessis is with the Communications and Intelligent Systems Research Group, University of Hertfordshire, Hatfield AL10 9AB, U. K. (e-mail: p.kourtessis@herts.ac.uk).

Symeon Chatzinotas is with the SnT at the University of Luxembourg, Luxembourg (e-mail: symeon.chatzinotas@uni.lu).

Dimitra I. Kaklamani is with the Microwave and Fiber Optics Laboratory, and Iakovos S. Venieris is with the Intelligent Communications and Broadband Networks Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens, Zografou, 15780 Athens, Greece.

single-receive-antenna nodes was presented in [23]. On the other hand, the problem of achievable sum rate maximization for a two-user FD cRIS-aided system where both the users were equipped with multiple transmit and receive antennas was considered in [24]. The analysis of the achievable sum rate in a hardware-impaired cRIS-assisted multipair FD system was presented in [25], where the authors also proposed an optimal PBM design to maximize the sum rate using particle swarm optimization. The authors in [26] considered an FD multi-pair cRIS system assuming an FD BS with multiple transmit and receive antennas and multiple dual antenna FD users, while also considering the impact of correlated channels and statistical CSI availability.

On the other hand, some of the recent contributions on FD STARS-enabled communications include [27]–[31]. In particular, the authors in [27] considered a STARS-enabled system consisting of a two-antenna FD transmitter and two HD UEs, where they derived a closed-form expression for sum SE, and also proposed an algorithm to obtain optimal PBM design to maximize the sum SE, for both ES and MS protocols. It is noteworthy that the derived closed-form expression for SE in [27] was based on statistical channel state information (CSI) which reduces the channel estimation overhead, and the impact of channel correlation at STARS was also taken into consideration. Similarly, in [28], closed-form expressions for UL and DL outage probability were derived for a STARS-enabled system (operating in MS protocol) consisting of one FD base station (BS) equipped with two antennas, and two HD users. In [29], the authors considered the problem of weighted sum rate maximization under quality-of-service (QoS) constraints in a STARS-enabled FD system consisting of a two-antenna FD BS and two single-antenna HD UEs, for both ES and MS protocols. The authors then reformulated the original problem to a quadratically-constrained quadratic programming (QCQP) problems using successive convex approximation (SCA) for the ES protocol and a penalty-based method for MS protocol and then obtained a solution using interior-point methods. Moreover, in [30], the authors derived outage probability and ergodic rate expressions in closed-form for an FD two-way STARS-enabled device-to-device communication system, where multiple dual-antenna UEs located in the transmission regime of the STARS communicated to its paired dual-antenna UE in the reflection regime of the STARS via MS protocol. In particular, the authors partitioned the STARS elements such that only a subset of elements serve one transmitter-receiver pair, which significantly restricts the STARS capability. The problem of transmit power minimization in a STARS-enabled FD system consisting of a dual-antenna FD BS and two single-antenna HD users was presented in [31], where the authors proposed an alternating optimization (AO) based algorithm to obtain optimal beamforming designs. In summary, none of the works reported in [27]–[31] have considered the impact of channel correlation at the STARS (except [27]), and all are limited to an FD dual-antenna BS (except [30], where the role of BS was limited to managing some control signals) and two HD users.

It is noteworthy that since instantaneous CSI acquisition is a challenging task in cRIS-/STARS-aided system [32], [33], con-

sideration of instantaneous CSI in [22]–[25] for cRIS-based systems and in [28]–[31] for STARS-based systems, where the optimization of PBM should take place at every coherence interval of the channel, is of less practical importance. On the other hand, the work presented in [27], [34]–[36] rely on statistical CSI being dependent on the large-scale statistics, which change at every several coherence intervals. The design of a STARS-aided system based on statistical CSI is crucial because it significantly reduces the signaling overhead. It is worthwhile to mention that the overhead increases further in the case of the STARS [14], [15]. Also, this effect becomes even more pronounced in high-speed scenarios, where the channels vary rapidly [37]. Nonetheless, massive multiple-input multiple-output (mMIMO) systems have been the key enabler for 5G wireless, and it is highly anticipated that mMIMO will continue to be the backbone for B5G networks [38], [39]. However, none of the works have presented the performance of the STARS-aided FD two-way system with mMIMO architecture.

Contributions: The previous observations motivate the topic of this paper. Specifically, we propose to employ STARS in an FD mMIMO two-way network to provide full coverage for UEs in blind areas found in front and behind the STARS as depicted in Fig. 1. Unlike other works in cRIS assisted two-way mMIMO communications, this is the only study integrating a STARS. Note that [30] studies a device-to-device communication system. Moreover, we account for correlated fading and imperfect CSI. The marriage of STARS with the FD architecture improves the STARS's capability to shape the propagation environment, and thus, improves the sum SE of the system under consideration. Herein, our objective is to design the PBM at the STARS such that the sum SE of the network is maximized. The coupling of the design variables results in a non-convex optimization problem, which is challenging to solve. In this regard, we propose a projected gradient ascent method (ProGrAM), where each iteration is performed using derived closed-form expressions which maximizes the achievable sum SE. More specifically,

- Compared to [26], we have considered a STARS, while compared to [22]–[24], [40], we have considered correlated fading and imperfect CSI additionally to the STARS. We show that in the case of spatially-correlated channels, a stronger correlation at the STARS reduces the sum SE of the network. Also, contrary to STARS-assisted system in [14], we have assumed multiple UEs on each side of the STARS and FD communications in a unified way.
- Despite that many previous works have assumed that the optimization of the PBM occurs at each coherence interval because their expressions are dependent on small-scale fading, in this paper, we optimize the PBM at every several coherence intervals since it depends on large-scale statistics.
- We perform channel estimation (CE) by means of linear minimum mean-square error (LMMSE) with a low overhead compared to previous works such as [41] or other works that do not provide analytical expressions

for the channel estimate. Also, previous works derive the individual channel estimates concerning each cRIS element and leave the inter-element correlation unknown, while we obtain analytical closed-form expressions that take into account the full STARS correlation.

- We derive closed-form expressions for the average achievable UL and DL SE for the STARS-enabled FD two-way system with a maximal-ratio transmitter (MRT) and a maximal-ratio combining (MRC) receiver, that depend only on the path-losses and covariance matrices (large-scale statistics). To obtain the optimal PBM design, we use ProGrAM where each iteration is performed in closed form.
- We present extensive simulation results to establish the advantage of the STARS-aided FD system over the corresponding STARS-aided HD system and cRIS-aided FD system. Note that the FD mode increases the difficulty in deriving the closed-form solutions compared to the HD since the former requires the derivations of the self-interference and loop-interference terms as well as the corresponding gradients. Moreover, we also analyze the impact of different parameters of interest on the system performance.

Notations: Bold uppercase and lowercase letters are used to denote matrices and vectors respectively. $\mathbf{X}^T$, $\mathbf{X}^H$, $\|\mathbf{X}\|$, and $\text{Tr}(\mathbf{X})$ denote the (ordinary) transpose, conjugate transpose, Frobenius norm and trace of the matrix $\mathbf{X}$. A square diagonal matrix whose main diagonal consists of the elements of vector $\mathbf{x}$ is represented by $\text{diag}(\mathbf{x})$, while the elements of the main diagonal of matrix $\mathbf{X}$ are represented by the column-vector $\text{vec}_d(\mathbf{X})$. The vectorization operator which stacks the columns of $\mathbf{X}$ to create a single long column vector is denoted by $\text{vec}(\mathbf{X})$. The complex-valued gradient of $f(\cdot)$ with respect to (w.r.t.) $\mathbf{X}^*$ is denoted by $\nabla_{\mathbf{X}} f(\cdot)$. By $\Pi_{\mathcal{X}}(\mathbf{x})$ we denote the Euclidean projection of $\mathbf{x}$ onto the set $\mathcal{X}$, i.e., $\Pi_{\mathcal{X}}(\mathbf{x}) \triangleq \arg\min_{\mathbf{y} \in \mathcal{X}} \|\mathbf{y} - \mathbf{x}\|$. The vector space of all complex-valued matrices of size $M \times N$ is denoted by $\mathbb{C}^{M \times N}$. The amplitude and phase of a complex number x are denoted by $|x|$ and $\angle x$, respectively. The statistical expectation operator is denoted using $\mathbb{E}\{\cdot\}$.

II. SYSTEM MODEL

We consider a STARS-enabled FD two-way mMIMO system as shown in Fig. 1, where an FD BS serves K FD UEs, denoted by $\text{UE}_k, k \in \mathcal{K} \triangleq \{1, 2, \dots, K\}$, with the uplink and downlink transmissions occurring simultaneously at the same frequency.¹ For this purpose, we assume that the BS is equipped with M_T transmit and M_R receive antennas, while depending on the implementation, each UE is equipped either with a single antenna or two antennas for signal transmission and reception [44]. Without the loss of generality, we assume that the BS is located in the reflection region (i.e., in the front side) of the STARS. Moreover, we consider the scenario

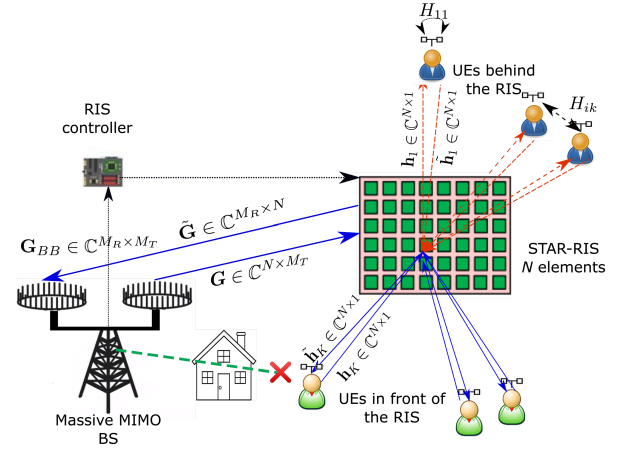


Fig. 1. A STARS-Enabled FD two-way mMIMO system with multiple UEs in transmission and reflection regions.

where out of the K UEs, K_r UEs are located in the reflection region and K_t UEs are located in the transmission region (i.e., in the back side) of the STARS, such that $K_r + K_t = K$. The set of indices of the UEs in the reflection and transmission regions are respectively denoted by $\mathcal{K}_r$ and $\mathcal{K}_t$, where $\bigcap_{m \in \{r, t\}} \mathcal{K}_m = \emptyset$ and $\bigcup_{m \in \{r, t\}} \mathcal{K}_m = \mathcal{K}$. Moreover, we define $\mathcal{W} = \{w_1, w_2, \dots, w_K\}$ being the set describing the STARS operation for UEs, where $w_k = r$ if $k \in \mathcal{K}_r$ and $w_k = t$ otherwise. Regarding the STARS, we assume that it consists of a uniform planar array (UPA) with $N \triangleq N_h \times N_v$ (nearly) passive elements, where N_h and N_v represent the number of columns and rows of the UPA, respectively.

Different from the half-space coverage of a cRIS, the advantage of the STARS is that it can exploit the whole space by configuring the transmitted and reflected signals in terms of two independent coefficients. In this work, we consider the ES protocol, which assumes that all STARS elements simultaneously serve all the UEs in both transmission and reflection regions. Let s_n being the incident signal on the n^{th} element of the STARS ($n \in \mathcal{N} \triangleq \{1, 2, \dots, N\}$), the reflected and transmitted signals from the n^{th} elements are respectively denoted by $\theta_{r,n}s_n$ and $\theta_{t,n}s_n$, respectively, where $|\theta_{r,n}|, |\theta_{t,n}| \in [0, 1]$, $\angle \theta_{r,n}, \angle \theta_{t,n} \in [0, 2\pi]$, and $|\theta_{r,n}|^2 + |\theta_{t,n}|^2 = 1, \forall n \in \mathcal{N}$. Therefore, the PBM for UE $_k$ is given by $\Theta_{w_k} \triangleq \text{diag}(\theta_{1,w_k}, \theta_{2,w_k}, \dots, \theta_{N,w_k})$.²

Next, we describe the channel model for the system under consideration.

A. Channel Model

Based on the narrowband quasi-static block fading channel models, we account for both small- and large-scale fading. The latter includes path-loss and STARS correlation that change slowly, i.e., the large-scale fading can be assumed to be constant for several coherence intervals contrary to small-scale fading, which remains static only during one coherence

¹Note that this paper considers the passive type of RIS, which is well known to suffer from the double path fading effect. To address this issue, researchers have proposed the novel concept of active RIS [42], [43], which is equipped with amplifiers. Hence, it would be an interesting direction for future work to study active STARS with FD two-way relay.

²A more accurate model for purely passive STARS was recently proposed in [45], where the transmission and reflection phase shifts are coupled with each other. However, in this paper, we consider the conventional modeling of the STARS phase shifts to keep the complexity within tractable limits.

interval and then changes independently. It is worthwhile to mention that in this paper, we aim to present a unified analysis with respect to the channel estimation and data transmission phases for UL and DL scenarios.

We start by providing the channel expressions in mathematical terms. In particular, the channel matrices for the BS-STARS and STARS-BS links assuming correlation at both sides are respectively given by

$$\begin{aligned}\mathbf{G} &= \sqrt{\delta_g} \mathbf{R}_s^{1/2} \mathbf{D} \mathbf{R}_b^{1/2} \in \mathbb{C}^{N \times M_T}, \\ \tilde{\mathbf{G}} &= \sqrt{\tilde{\delta}_g} \tilde{\mathbf{R}}_b^{1/2} \tilde{\mathbf{D}} \mathbf{R}_s^{1/2} \in \mathbb{C}^{M_R \times N},\end{aligned}\quad (1)$$

where $\mathbf{R}_b \in \mathbb{C}^{M_T \times M_T}$, $\tilde{\mathbf{R}}_b \in \mathbb{C}^{M_R \times M_R}$, and $\mathbf{R}_s \in \mathbb{C}^{N \times N}$ express the *deterministic* positive semi-definite correlation matrices at the BS transmit array, BS receive array, and the STARS, respectively, which are assumed to be known by the network since they can be obtained by existing estimation methods [46], [47].³ Note that many previous works have relied only on line-of-sight (LoS) components between the BS and the STARS, while our modeling is more general because we assume that all links are correlated Rayleigh fading distributed. Also, δ_g and $\tilde{\delta}_g$ represent the path-loss, and $\text{vec}(\mathbf{D}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{NM_T})$ and $\text{vec}(\tilde{\mathbf{D}}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{M_R N})$ express the fast-fading vectors. Similarly the STARS-UE_k and UE_k-STARS channel vectors are respectively given by

$$\begin{aligned}\mathbf{h}_k &= \sqrt{\delta_{h_k}} \mathbf{c}_k \mathbf{R}_s^{1/2} \in \mathbb{C}^{1 \times N}, \\ \tilde{\mathbf{h}}_k &= \sqrt{\tilde{\delta}_{h_k}} \tilde{\mathbf{R}}_s^{1/2} \tilde{\mathbf{c}}_k \in \mathbb{C}^{N \times 1},\end{aligned}\quad (2)$$

where δ_{h_k} , and $\tilde{\delta}_{h_k}$ represent the path-loss, and $\mathbf{c}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ and $\tilde{\mathbf{c}}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ express the fast-fading vectors.⁴ Furthermore, we assume that no direct links exist between the BS and UEs due to blockages or severe path-loss. The direct link between UE_k and UE_j is given by $H_{kj} \sim \mathcal{CN}(0, \sigma_{kj}^2) \forall k, j \in \mathcal{K}$, and the LI channel between the transmit and receive antenna arrays at the BS is modeled as $\mathbf{G}_b = \mathbf{R}_b^{1/2} \tilde{\mathbf{G}}_b \mathbf{R}_b^{1/2}$ with the entries in $\tilde{\mathbf{G}}_b$ being independent and identically distributed (i.i.d.) following $\mathcal{CN}(0, \sigma_L^2)$.

As the BS is assumed to be in the reflection region, the signal traveling from the STARS to the BS is controlled by Θ_r . As a result, the UL cascaded channel vector for the UE_k-STARS-BS link $\tilde{\mathbf{u}}_k \triangleq \tilde{\mathbf{G}} \Theta_r \tilde{\mathbf{h}}_k \in \mathbb{C}^{M_R \times 1}$ has a covariance matrix $\tilde{\mathbf{R}}_k = \mathbb{E}\{\tilde{\mathbf{u}}_k \tilde{\mathbf{u}}_k^H\}$ given by

$$\tilde{\mathbf{R}}_k = \tilde{\delta}_{g_k} \text{Tr}(\mathbf{A}_r \Theta_r^H) \tilde{\mathbf{R}}_b, \quad (3)$$

where $\tilde{\delta}_{g_k} \triangleq \tilde{\delta}_g \tilde{\delta}_{h_k}$, $\mathbf{A}_r \triangleq \mathbf{R}_s \Theta_r \mathbf{R}_s$, and we have used the independence between $\tilde{\mathbf{G}}$ and $\tilde{\mathbf{h}}_k$. Moreover, $\mathbb{E}\{\tilde{\mathbf{h}}_k \tilde{\mathbf{h}}_k^H\} = \tilde{\delta}_{h_k} \mathbf{R}_s$ and we have used the property $\mathbb{E}\{\mathbf{V} \mathbf{B} \mathbf{V}^H\} = \text{Tr}(\mathbf{B}) \mathbf{I}_M$ with $\mathbf{B}$ being a deterministic square matrix, and $\mathbf{V}$ being any matrix with i.i.d. entries of zero mean and unit variance.

³ Another way of practical calculation of the covariance matrices follows. In particular, the covariance matrices depend on the distances between the RIS elements and between the BS antennas, respectively. Moreover, they depend on the angles. Generally, the distances are known from the construction of the RIS and the BS, while the angles can be calculated when the locations are given. Thus, the covariance matrices can be considered to be known.

⁴ The analysis with correlated Rician fading that includes a LoS component could be the topic of future work.

Similarly, the DL cascaded channel vector for the BS-STARS-UE_k link $\mathbf{u}_k \triangleq \mathbf{h}_k \Theta_{w_k} \mathbf{G} \in \mathbb{C}^{1 \times M_T}$ has a covariance matrix $\mathbf{R}_k = \mathbb{E}\{\mathbf{u}_k^H \mathbf{u}_k\}$ given by

$$\mathbf{R}_k = \delta_{g_k} \text{Tr}(\mathbf{A}_{w_k} \Theta_{w_k}^H) \mathbf{R}_b, \quad (4)$$

where $\delta_{g_k} \triangleq \delta_g \delta_{h_k}$ and $\mathbf{A}_{w_k} \triangleq \mathbf{R}_s \Theta_{w_k} \mathbf{R}_s$. It is noteworthy that in the case of no correlation at the STARS, we have $\mathbf{R}_s = \mathbf{I}_N$, and thus $\tilde{\mathbf{R}}_k$ and $\mathbf{R}_k$ are independent of the phase shifts but depend only on the amplitudes, which was also observed in [48].

In the next section, we discuss the signal transmission and reception for the uplink and downlink scenarios.

III. SIGNAL TRANSMISSION AND RECEPTION

During the data transmission phase, the STARS transmits and reflects the signals from the simultaneously transmitting antennas of the BS and UEs. The signal transmitted by the FD BS is given by $\mathbf{x}_b = \sqrt{\beta} \sum_{k \in \mathcal{K}} \sqrt{p_{bk}} \mathbf{f}_k s_{bk}$, where p_{bk} is the power allocated for UE_k, s_{bk} is the symbol intended for UE_k, and $\mathbf{f}_k$ is the corresponding precoding vector. Define $\mathbf{F} \triangleq [\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_K] \in \mathbb{C}^{M_T \times K}$ as the precoding matrix, and the scaling factor β ensures $\mathbb{E}\{\mathbf{x}_b^H \mathbf{x}_b\} = p_b$, where p_b is the total transmit power budget at the BS. For the sake of tractability and following the mMIMO literature, we consider the equal-power allocation among the UEs, i.e., $p_{bk} = p_b/K$, $\forall k \in \mathcal{K}$, resulting in $\beta = K/\mathbb{E}\{\text{Tr}(\mathbf{F} \mathbf{F}^H)\}$. On the other hand, the signal transmitted from UE_k is given by $x_{uk} = \sqrt{p_{uk}} s_{uk}$, where p_{uk} is the transmit power and s_{uk} is the transmitted symbol from UE_k. With this background, the signal received at the BS is given by

$$\mathbf{y}_b = \tilde{\mathbf{G}} \Theta_r \mathbf{y}_s + \mathbf{G}_b \mathbf{x}_b + \mathbf{z}_b, \quad (5)$$

where $\mathbf{y}_s$ is the signal received at the STARS due to the signals transmitted from the BS and UEs, the second term accounts for the BS LI, and $\mathbf{z}_b \sim \mathcal{CN}(\mathbf{0}, \sigma_b^2 \mathbf{I}_{M_R})$ is the additive White Gaussian noise (AWGN) at the BS. The signal received at the STARS is given by

$$\mathbf{y}_s = \mathbf{G} \mathbf{x}_b + \sum_{k \in \mathcal{K}} \tilde{\mathbf{h}}_k x_{uk}. \quad (6)$$

The first and second terms in the preceding equation correspond to the signal received from the BS and K UEs, respectively.

In this paper, we adopt linear precoding for DL and linear combining for UL communications. Specifically, for the detection of signal from UE_k to BS, we use the combining vector $\mathbf{v}_k$ at the BS, which yields

$$\begin{aligned}y_{uk} &= \mathbf{v}_k^H \mathbf{y}_b \\ &= \mathbf{v}_k^H \left[\tilde{\mathbf{G}} \Theta_r \mathbf{G} \sqrt{\frac{\beta p_b}{K}} \sum_{j \in \mathcal{K}} \mathbf{f}_j s_{bj} + \tilde{\mathbf{G}} \Theta_r \right. \\ &\quad \times \sum_{j \in \mathcal{K}} \tilde{\mathbf{h}}_j \sqrt{p_{uj}} s_{uj} + \mathbf{G}_b \sqrt{\frac{\beta p_b}{K}} \sum_{j \in \mathcal{K}} \mathbf{f}_j s_{bj} + \mathbf{z}_b \left. \right] \\ &= \sqrt{p_{uk}} \mathbf{v}_k^H \tilde{\mathbf{u}}_k s_{uk} + \sum_{i \in \mathcal{K} \setminus \{k\}} \sqrt{p_{ui}} \mathbf{v}_k^H \tilde{\mathbf{u}}_i s_{ui} + \sqrt{\frac{\beta p_b}{K}} \mathbf{v}_k^H \tilde{\mathbf{G}} \\ &\quad \times \Theta_r \mathbf{G} \sum_{j \in \mathcal{K}} \mathbf{f}_j s_{bj} + \sqrt{\frac{\beta p_b}{K}} \mathbf{v}_k^H \mathbf{G}_b \sum_{j \in \mathcal{K}} \mathbf{f}_j s_{bj} + \mathbf{v}_k^H \mathbf{z}_b\end{aligned}$$

$$\begin{aligned}
&= \underbrace{\sqrt{p_{uk}} \mathbb{E}\{\mathbf{v}_k^H \tilde{\mathbf{u}}_k\} s_{uk}}_{\text{desired signal over known channel}} + \underbrace{\sqrt{p_{uk}} (\mathbf{v}_k^H \tilde{\mathbf{u}}_k - \mathbb{E}\{\mathbf{v}_k^H \tilde{\mathbf{u}}_k\}) s_{uk}}_{\text{desired signal over unknown channel}} \\
&+ \underbrace{\sum_{i \in \mathcal{K} \setminus \{k\}} \sqrt{p_{ui}} \mathbf{v}_k^H \tilde{\mathbf{u}}_i s_{ui}}_{\text{UL MUI}} + \underbrace{\sqrt{\frac{\beta p_b}{K}} \mathbf{v}_k^H \tilde{\mathbf{G}} \Theta_r \mathbf{G} \sum_{j \in \mathcal{K}} \mathbf{f}_j s_{bj}}_{\text{BS SI}} \\
&+ \underbrace{\sqrt{\frac{\beta p_b}{K}} \mathbf{v}_k^H \mathbf{G}_b \sum_{j \in \mathcal{K}} \mathbf{f}_j s_{bj}}_{\text{BS LI}} + \underbrace{\mathbf{v}_k^H \mathbf{z}_b}_{\text{post-combining AWGN}}, \quad (7)
\end{aligned}$$

where the third term is the UL multiuser interference (MUI) from other UEs, the fourth term is the BS self-interference (SI) from its own transmit signal reflected by the STARS, and the fifth term is the BS LI.

On the other hand, the DL signal received by UE_k is given by

$$\begin{aligned}
y_{dk} &= \mathbf{h}_k \Theta_{w_k} \mathbf{y}_s + \sum_{j \in \mathcal{K}} H_{kj} \sqrt{p_{uj}} s_{uj} + z_{dk} \\
&= \sqrt{\frac{\beta p_b}{K}} \mathbf{u}_k \mathbf{f}_k s_{bk} + \sqrt{\frac{\beta p_b}{K}} \sum_{i \in \mathcal{K} \setminus \{k\}} \mathbf{u}_k \mathbf{f}_i s_{bi} + \mathbf{h}_k \Theta_{w_k} \\
&\times \sum_{j \in \mathcal{K}} \sqrt{p_{uj}} \tilde{\mathbf{h}}_j s_{uj} + \sum_{i \in \mathcal{K} \setminus \{k\}} H_{ki} \sqrt{p_{ui}} s_{ui} + H_{kk} \sqrt{p_{uk}} s_{uk} + z_{dk} \\
&= \underbrace{\sqrt{\frac{\beta p_b}{K}} \mathbb{E}\{\mathbf{u}_k \mathbf{f}_k\} s_{bk}}_{\text{desired signal over known channel}} + \underbrace{\sqrt{\frac{\beta p_b}{K}} (\mathbf{u}_k \mathbf{f}_k - \mathbb{E}\{\mathbf{u}_k \mathbf{f}_k\}) s_{bk}}_{\text{desired signal over unknown channel}} \\
&+ \underbrace{\sum_{i \in \mathcal{K} \setminus \{k\}} H_{ki} \sqrt{p_{ui}} s_{ui}}_{\text{IUI}} + \underbrace{H_{kk} \sqrt{p_{uk}} s_{uk}}_{\text{UE-LI}} + z_{dk}, \quad (8)
\end{aligned}$$

where the third, fourth, fifth, and sixth terms refer to DL MUI, UE co-channel interference, interuser interference, and UE_k's LI. Moreover, $z_{dk} \sim \mathcal{CN}(0, \sigma_{dk}^2)$ is the AWGN at UE_k. For ease of exposition, in the rest of this paper, we assume $\sigma_{dk}^2 = \sigma_d^2 \forall k \in \mathcal{K}$.

Before proceeding further, we present the channel estimation for the system under consideration first, in the next section.

IV. CHANNEL ESTIMATION

In practice, perfect CSI is not available at the BS. Additionally, a STARS is implemented by means of nearly passive elements that cannot send pilots to the BS and allow channel estimation or receive pilots and estimate the corresponding channel.

In this direction, given that the FD BS is deployed with separate arrays for transmission and reception, it will estimate the cascaded UL and DL channels separately, in order to design the decoding and precoding vectors $\mathbf{v}_k$ and $\mathbf{f}_k$, respectively. Note that the physical separation of transmit and receive antennas is a common passive cancellation technique to combat LI, and the corresponding UL and DL channels, i.e., $\tilde{\mathbf{G}}$ and $\mathbf{G}$ are not identical [49], [50]. In parallel, each FD UE is equipped with separate transmit and receive antennas. Hence, the DL channel between the STARS and UE_k receive antenna, and the UL channel between UE_k and the STARS are given by $\mathbf{h}_k$ and $\tilde{\mathbf{h}}_k$, respectively. This means that the

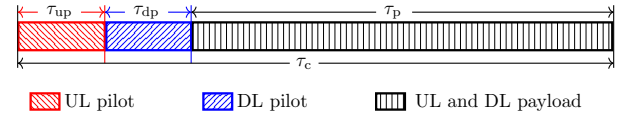


Fig. 2. A coherence interval.

effective DL and UL channels from the BS to the UE receive antenna and from the transmit UE antenna to the BS are not the same. Different from [26] and [41], which assumed N phases for channel estimation, we propose a scheme where channel estimation is to be performed in a single phase that achieves a higher SE due to a higher pre-log factor and can be applied not only for low-speed UEs but also for fast UEs since the training does not take a long time.

We assume that each block has a duration of $\tau_c = B_c T_c$ channel uses, where B_c is the coherence bandwidth, and T_c is the coherence time. From this duration, $\tau_{up} \in \mathbb{Z}$ and $\tau_{dp} \in \mathbb{Z}$ channel uses are considered for UL and DL pilots, respectively, and $\tau_c - (\tau_{up} + \tau_{dp})$ channel uses are considered for simultaneous transmission for UL and DL data, as shown in Fig. 2.

During the UL training phase of duration τ_{up} symbols, UEs transmit orthogonal pilot sequences to the BS via STARS. Especially, $\tilde{\mathbf{x}}_{pk} = [\tilde{x}_{pk,1}, \tilde{x}_{pk,2}, \dots, \tilde{x}_{pk,\tau_{up}}]^T \in \mathbb{C}^{\tau_{up} \times 1}$ denotes the pilot sequence of UE_k with $\tilde{\mathbf{x}}_{pk}^H \tilde{\mathbf{x}}_{pl} = 0 \forall k, l \in \mathcal{K}, k \neq l$ and $\tilde{\mathbf{x}}_{pk}^H \tilde{\mathbf{x}}_{pk} = \tau_{up} p_{\text{Train}}$ joules with $p_{\text{Train}} = |\tilde{x}_{pk,i}|^2, \forall k \in \mathcal{K}, i \in \{1, 2, \dots, \tau_{up}\}$ being the common average transmit power per UE during the UL training phase.

The received signal at the BS during the UL training period is given by

$$\mathbf{Y}_u = \sum_{i \in \mathcal{K}} \tilde{\mathbf{u}}_i \tilde{\mathbf{x}}_{pi}^H + \tilde{\mathbf{W}}, \quad (9)$$

where $\tilde{\mathbf{W}} \in \mathbb{C}^{M_R \times \tau_{up}}$ is the AWGN matrix at the BS with independent columns, each one distributed as $\mathcal{CN}(\mathbf{0}, \sigma_u^2 \mathbf{I}_{M_R})$. The received training signal at the BS is multiplied with the transmitted training sequence from UE_k, given by (9), to eliminate the interference caused by other UEs which yields

$$\tilde{\mathbf{r}}_k = \tilde{\mathbf{u}}_k + \frac{\tilde{\mathbf{w}}}{\tau_{up} p_{\text{Train}}}, \quad (10)$$

where $\tilde{\mathbf{w}} = \tilde{\mathbf{W}} \tilde{\mathbf{x}}_{pk} \sim \mathcal{CN}(\mathbf{0}, \tau_{up} p_{\text{Train}} \sigma_u^2 \mathbf{I}_{M_R})$.

Proposition 1: The MMSE estimate of the UL cascaded channel $\tilde{\mathbf{u}}_k$ is given by

$$\hat{\tilde{\mathbf{u}}}_k = \tilde{\mathbf{R}}_k \tilde{\mathbf{Q}}_k \tilde{\mathbf{r}}_k, \quad (11)$$

where

$$\tilde{\mathbf{Q}}_k = \left(\tilde{\mathbf{R}}_k + \frac{\sigma_u^2}{\tau_{up} p_{\text{Train}}} \mathbf{I}_{M_R} \right)^{-1}, \quad (12)$$

and $\tilde{\mathbf{r}}_k$, given by (10), is the noisy observation of $\tilde{\mathbf{u}}_k$.

Proof: The proof follows the common steps of MMSE channel estimation [51]. ■

Based on the property of orthogonality of MMSE estimation, the perfect cascaded UL channel is given by

$$\tilde{\mathbf{u}}_k = \hat{\tilde{\mathbf{u}}}_k + \tilde{\mathbf{e}}_k, \quad (13)$$

where $\hat{\mathbf{u}}_k$ and $\tilde{\mathbf{e}}_k$, being the estimated channel and the channel error vector, having zero mean and variances $\tilde{\Psi}_k = \tilde{\mathbf{R}}_k \tilde{\mathbf{Q}}_k \tilde{\mathbf{R}}_k$ and $\tilde{\mathbf{E}}_k = \tilde{\mathbf{R}}_k - \tilde{\Psi}_k$, respectively.

In a similar way, the following proposition provides the MMSE estimate of the DL cascaded channel $\mathbf{u}_k$, where $w_k \sim \mathcal{CN}(0, \sigma_{dk}^2)$ is the AWGN at UE $_k$.

Proposition 2: The MMSE estimate of the DL cascaded channel $\mathbf{u}_k$ is given by

$$\mathbf{u}_k = \hat{\mathbf{u}}_k + \mathbf{e}_k, \quad (14)$$

where $\hat{\mathbf{u}}_k$ and $\mathbf{e}_k$ have zero mean and variances $\Psi_k = \mathbf{R}_k \mathbf{Q}_k \mathbf{R}_k$ and $\mathbf{E}_k = \mathbf{R}_k - \Psi_k$ with $\mathbf{Q}_k = \left(\mathbf{R}_k + \frac{\sigma_{dk}^2}{\tau_{dp} p_{\text{Train}}} \mathbf{I}_{M_T} \right)^{-1}$, respectively.

Proof: The proof follows the common steps of MMSE channel estimation [51]. ■

Remark 1: The proposed method has a number of advantages compared to other methods. First, it comes with lower complexity and requires less hardware complexity compared to estimating the individual channels, which is a quite demanding procedure as the number of surface elements increases. Note that if we would like to estimate the individual channels, one solution would be to assume receive radio frequency (RF) chains integrated into the STARS as in [52]. Second, we have achieved to derive the channel estimates in closed forms in terms of large-scale statistics. This means that channel estimation can be performed at every several coherence intervals. Other existing methods do not capture the correlation of the whole surface but they obtain the estimated channel per element [41], or they do not provide analytical expressions, e.g., [53].

V. PROBLEM FORMULATION

In this section, we first derive closed-form expressions for UL and DL SE, and then formulate an optimization problem to find the optimal PBM that maximizes the achievable sum SE. The self-interference, loop interference, and the co-channel interference for a STAR-RIS system induces certain difficulties as can be seen by the mathematical derivations below. Notably, similar difficulties are met during the optimization of the sum SE in Sec. VI, where we obtain specific derivatives.

A. Average Uplink SE

Considering the worst-case additive noise for all the terms in (7) except the desired signal, the expression for average achievable UL sum SE using the use-and-then-forget (UaTF) technique [51] is given by

$$\text{SE}_u = \zeta \sum_{k \in \mathcal{K}} \log_2 (1 + \gamma_{uk}), \quad (15)$$

where $\zeta = \frac{\tau_c - \tau_p}{\tau_c}$, $\gamma_{uk} = S_{uk}/I_{uk}$ with

$$S_{uk} = p_{uk} |\mathbb{E}\{\mathbf{v}_k^H \tilde{\mathbf{u}}_k\}|^2, \quad (16)$$

and

$$\begin{aligned} I_{uk} = & \text{var}\{\sqrt{p_{uk}} \mathbf{v}_k^H \tilde{\mathbf{u}}_k\} + \sum_{i \in \mathcal{K} \setminus \{k\}} p_{ui} \mathbb{E}\{|\mathbf{v}_k^H \tilde{\mathbf{u}}_i|^2\} \\ & + \frac{\beta p_b}{K} \sum_{j \in \mathcal{K}} \mathbb{E}\{|\mathbf{v}_k^H (\tilde{\mathbf{G}} \Theta_r \mathbf{G} + \mathbf{G}_b) \mathbf{f}_j|^2\} + \sigma_b^2 \mathbb{E}\{\|\mathbf{v}_k\|^2\}. \end{aligned} \quad (17)$$

Theorem 1: Considering MRC decoding at the BS with $\mathbf{v}_k = \hat{\mathbf{u}}_k$, closed-form expressions for S_{uk} and I_{uk} are respectively given by

$$S_{uk} = p_{uk} \text{Tr}^2(\tilde{\Psi}_k), \quad (18)$$

and

$$\begin{aligned} I_{uk} = & \text{Tr}(\tilde{\Psi}_k \tilde{\mathbf{R}}_{\text{sum}}) + \frac{p_b}{\text{Tr}(\tilde{\Psi}_{\text{sum}})} \text{Tr}(\tilde{\Psi}_k \tilde{\mathbf{R}}_b) \text{Tr}(\mathbf{R}_b \tilde{\Psi}_{\text{sum}}) \\ & \times \left\{ \delta_g \tilde{\delta}_g \text{Tr}(\mathbf{A}_r \Theta_r^H) + \sigma_L^2 \right\} - p_{uk} \text{Tr}(\tilde{\Psi}_k^2) + \sigma_b^2 \text{Tr}(\tilde{\Psi}_k), \end{aligned} \quad (19)$$

where $\tilde{\mathbf{R}}_{\text{sum}} \triangleq \sum_{j \in \mathcal{K}} p_{uj} \tilde{\mathbf{R}}_j$, $\tilde{\Psi}_{\text{sum}} \triangleq \sum_{j \in \mathcal{K}} \tilde{\Psi}_j$.

Proof: See Appendix A. ■

B. Average Downlink SE

Analogously, using the UaTF scheme, the expression for average achievable DL sum SE following (8) is given by

$$\text{SE}_d = \zeta \sum_{k \in \mathcal{K}} \log_2 (1 + \gamma_{dk}), \quad (20)$$

where $\gamma_{dk} = S_{dk}/I_{dk}$ with

$$S_{dk} = \frac{\beta p_b}{K} |\mathbb{E}\{\mathbf{u}_k \mathbf{f}_k\}|^2, \quad (21)$$

and

$$\begin{aligned} I_{dk} = & \frac{\beta p_b}{K} \text{var}\{\mathbf{u}_k \mathbf{f}_k\} + \frac{\beta p_b}{K} \sum_{i \in \mathcal{K} \setminus \{k\}} \mathbb{E}\{|\mathbf{u}_k \mathbf{f}_i|^2\} \\ & + \sum_{j \in \mathcal{K}} p_{uj} (\sigma_{kj}^2 + \mathbb{E}\{|\mathbf{h}_k \Theta_{w_k} \tilde{\mathbf{h}}_j|^2\}) + \sigma_d^2. \end{aligned} \quad (22)$$

Theorem 2: Assuming that the BS uses MRT/conjugate beamforming based on the estimated CSI, i.e., $\mathbf{f}_k = \hat{\mathbf{u}}_k^H$, closed-form expressions for S_{dk} and I_{dk} are respectively given by

$$S_{dk} = \frac{p_b}{\text{Tr}(\tilde{\Psi}_{\text{sum}})} \text{Tr}^2(\tilde{\Psi}_k), \quad (23)$$

and

$$\begin{aligned} I_{dk} = & \sum_{j \in \mathcal{K}} \left[\frac{p_b}{\text{Tr}(\tilde{\Psi}_{\text{sum}})} \text{Tr}(\mathbf{R}_k \tilde{\Psi}_j) + p_{uj} \{\sigma_{kj}^2 + \delta_{h_k} \tilde{\delta}_{h_j}\} \right. \\ & \left. \times \text{Tr}(\mathbf{A}_{w_k} \Theta_{w_k}^H) \right] - \frac{p_b}{\text{Tr}(\tilde{\Psi}_{\text{sum}})} \text{Tr}(\tilde{\Psi}_k^2) + \sigma_d^2. \end{aligned} \quad (24)$$

Proof: See Appendix B. ■

Note that in what follows, with a slight abuse of notation, we write

$$S_{dk} = \text{Tr}^2(\tilde{\Psi}_k), \quad (25)$$

$$\begin{aligned} I_{dk} = & \text{Tr}(\mathbf{R}_k \tilde{\Psi}_{\text{sum}}) + \frac{\text{Tr}(\tilde{\Psi}_{\text{sum}})}{p_b} \sum_{j \in \mathcal{K}} \left[p_{uj} \{\sigma_{kj}^2 + \delta_{h_k} \tilde{\delta}_{h_j}\} \right. \\ & \left. \times \text{Tr}(\mathbf{A}_{w_k} \Theta_{w_k}^H) \right] - \text{Tr}(\tilde{\Psi}_k^2) + \frac{\sigma_d^2}{p_b} \text{Tr}(\tilde{\Psi}_{\text{sum}}). \end{aligned} \quad (26)$$

It can easily be seen that (25) and (26) lead to the same γ_{dk} as achieved using (23) and (24). It is evident from (18), (19), (25), and (26) that both UL and DL SE depend only on large-scale statistics in terms of correlation matrices and path-losses, which follow wide-sense stationarity and are constant for many coherence intervals [54]. In the following, we formulate an optimization problem, which aims to maximize the average sum SE.

C. Problem Formulation to Maximize the Sum SE

With the help of (15) and (20), the average sum SE (considering both UL and DL communications), is given by⁵

$$SE = SE_u + SE_d = \zeta \sum_{k \in \mathcal{K}} \{ \log_2(1 + \gamma_{uk}) + \log_2(1 + \gamma_{dk}) \}. \quad (27)$$

Note that the STARS-aided system requires the optimal design of both the receive and transmit PBMs (i.e., Θ_r, Θ_t) to provide the maximum performance. Notably, we account for correlated fading and imperfect CSI in an FD architecture while we assume infinite-resolution phase shifters. Therefore, the problem of finding optimal PBM designs that maximize SE is formulated as

$$(\mathcal{P}) \quad \underset{\Theta_r, \Theta_t}{\text{maximize}} \quad f(\Theta_r, \Theta_t) \triangleq SE, \quad (28a)$$

$$\text{subject to } |\theta_{r,n}|^2 + |\theta_{t,n}|^2 = 1 \quad \forall n \in \mathcal{N}. \quad (28b)$$

In the next section, we propose a simple, yet efficient algorithm to obtain a stationary solution to $(\mathcal{P})$.

VI. PROPOSED SOLUTION

It can easily be noted that $(\mathcal{P})$ is non-convex due to the coupling between the design variables θ_r and θ_t in both (28a) and (28b), and the non-convex constraints in (28b); therefore is challenging to solve. Before proceeding further, we define the set

$$\Theta \triangleq \{ \theta = [\theta_r^T, \theta_t^T]^T = [\theta_{r,1}, \dots, \theta_{r,N}, \theta_{t,1}, \dots, \theta_{t,N}]^T \in \mathbb{C}^{2N \times 1} \mid |\theta_{r,n}|^2 + |\theta_{t,n}|^2 = 1 \quad \forall n \in \mathcal{N} \}. \quad (29)$$

Exploiting the fact that for a given $\hat{\theta}$, its Euclidean projection onto Θ can be obtained via simple scaling, we now discuss an efficient solution using the projected gradient ascent method [55, Ch. 2]. Using (27) it is straightforward to note that for $m \in \{r, t\}$, we have

$$\begin{aligned} \nabla_{\theta_m} f(\theta_r, \theta_t) &= \zeta \sum_{k \in \mathcal{K}} \nabla_{\theta_m} [\log_2(1 + \gamma_{uk}) + \log_2(1 + \gamma_{dk})] \\ &= \zeta \log_2(e) \sum_{k \in \mathcal{K}} \left[\frac{\nabla_{\theta_m} \gamma_{uk}}{1 + \gamma_{uk}} + \frac{\nabla_{\theta_m} \gamma_{dk}}{1 + \gamma_{dk}} \right] \\ &= \zeta \log_2(e) \sum_{k \in \mathcal{K}} \left[\frac{1}{(1 + \gamma_{uk}) I_{uk}^2} (I_{uk} \nabla_{\theta_m} S_{uk} - S_{uk} \nabla_{\theta_m} I_{uk}) \right. \\ &\quad \left. + \frac{1}{(1 + \gamma_{dk}) I_{dk}^2} (I_{dk} \nabla_{\theta_m} S_{dk} - S_{dk} \nabla_{\theta_m} I_{dk}) \right]. \quad (30) \end{aligned}$$

In the theorem shown below, we obtain the expression for different gradient terms involved in (30), which are obtained in closed forms in terms of products of covariance matrices.

Theorem 3: Considering $m \in \{r, t\}$, closed-form expressions for $\nabla_{\theta_m} S_{uk}$, $\nabla_{\theta_m} S_{dk}$, $\nabla_{\theta_m} I_{uk}$, and $\nabla_{\theta_m} I_{dk}$ are given by (31) – (37):

$$\nabla_{\theta_m} S_{uk} = \begin{cases} \tilde{\nu}_k \text{diag}(\mathbf{A}_r), & \text{if } m = r \\ \mathbf{0}, & \text{otherwise} \end{cases}, \quad (31)$$

⁵In the case of the HD mode, the self-interference and loop-interference terms are missing, and the pre-log factors of the SEs are different.

Algorithm 1: The Proposed ProGrAM to solve $(\mathcal{P})$.

Input: $\theta_r^{(0)}, \theta_t^{(0)}, \mu_1 > 0, \epsilon$
Output: $\theta_r^{\text{opt}}, \theta_t^{\text{opt}}$

```

1  $\ell \leftarrow 1$ ;
2 repeat
    /* Find the gradients */
3   Obtain  $\nabla_{\theta_r} f(\theta_r^{(\ell-1)}, \theta_t^{(\ell-1)})$  and
      $\nabla_{\theta_t} f(\theta_r^{(\ell-1)}, \theta_t^{(\ell-1)})$  using (30) – (37);
    /* Move  $\theta_r$  along  $\nabla_{\theta_r} f(\cdot)$  direction */
4    $\hat{\theta}_r^{(\ell)} \leftarrow \theta_r^{(\ell-1)} + \mu_\ell \nabla_{\theta_r} f(\theta_r^{(\ell-1)}, \theta_t^{(\ell-1)})$ ;
    /* Move  $\theta_t$  along  $\nabla_{\theta_t} f(\cdot)$  direction */
5    $\hat{\theta}_t^{(\ell)} \leftarrow \theta_t^{(\ell-1)} + \mu_\ell \nabla_{\theta_t} f(\theta_r^{(\ell-1)}, \theta_t^{(\ell-1)})$ ;
    /* concatenate  $\hat{\theta}_r^{(\ell)}$  and  $\hat{\theta}_t^{(\ell)}$  */
6    $\hat{\theta}^{(\ell)} \triangleq [(\hat{\theta}_r^{(\ell)})^T, (\hat{\theta}_t^{(\ell)})^T]^T$ ;
    /* Project onto feasible  $\Theta$  */
7    $\theta^{(\ell)} \leftarrow \Pi_\Theta \{ \hat{\theta}^{(\ell)} \}$ ;
8    $\mu_{\ell+1} \leftarrow \mu_\ell$ ;
9    $\ell \leftarrow \ell + 1$ ; // update iteration index
10 until  $\{ f(\theta_r^{(\ell)}, \theta_t^{(\ell)}) - f(\theta_r^{(\ell-1)}, \theta_t^{(\ell-1)}) \} / f(\theta_r^{(\ell-1)}, \theta_t^{(\ell-1)}) \geq \epsilon$ ;
11  $\theta_r^{\text{opt}} \leftarrow \theta_r^{(\ell)}, \theta_t^{\text{opt}} \leftarrow \theta_t^{(\ell)}$ ;

```

$$\nabla_{\theta_r} S_{dk} = \begin{cases} \nu_k \text{diag}(\mathbf{A}_r), & \text{if } w_k = r \\ \mathbf{0}, & \text{otherwise} \end{cases}, \quad (32)$$

$$\nabla_{\theta_t} S_{dk} = \begin{cases} \nu_k \text{diag}(\mathbf{A}_t), & \text{if } w_k = t \\ \mathbf{0}, & \text{otherwise} \end{cases}, \quad (33)$$

$$\nabla_{\theta_t} I_{uk} = \left[\delta_g \sum_{i \in \mathcal{K}_t} \delta_{h_i} \text{Tr} \{ \mathbf{R}_b \tilde{\mathbf{L}}_{ki} \} \right] \text{diag}(\mathbf{A}_t), \quad (35)$$

where $\tilde{\nu}_k \triangleq 2\delta_g \tilde{\delta}_{h_k} p_{uk} \text{Tr}(\tilde{\Psi}_k) \text{Tr}(\tilde{\mathbf{R}}_b \tilde{\mathbf{C}}_k)$, $\tilde{\mathbf{C}}_k \triangleq \tilde{\mathbf{Q}}_k \tilde{\mathbf{R}}_k - \tilde{\mathbf{Q}}_k \tilde{\mathbf{R}}_k^2 \tilde{\mathbf{Q}}_k + \mathbf{R}_k \mathbf{Q}_k$, $\nu_k \triangleq 2\delta_g \delta_{h_k} \text{Tr}(\Psi_k) \text{Tr}(\mathbf{R}_b \mathbf{C}_k)$, $\mathbf{C}_k \triangleq \mathbf{Q}_k \mathbf{R}_k - \mathbf{Q}_k \mathbf{R}_k^2 \mathbf{Q}_k + \mathbf{R}_k \mathbf{Q}_k$, $\tilde{\mathbf{B}}_{k1} \triangleq \tilde{\mathbf{Q}}_k \tilde{\mathbf{R}}_k \tilde{\mathbf{R}}_{\text{sum}} - \tilde{\mathbf{Q}}_k \tilde{\mathbf{R}}_k \tilde{\mathbf{R}}_{\text{sum}} \tilde{\mathbf{R}}_k \tilde{\mathbf{Q}}_k + \tilde{\mathbf{R}}_{\text{sum}} \tilde{\mathbf{R}}_k \tilde{\mathbf{Q}}_k$, $\varpi \triangleq \frac{p_b}{\text{Tr}(\Psi_{\text{sum}})} \text{Tr}(\mathbf{R}_b \Psi_{\text{sum}}) \{ \delta_g \delta_g \text{Tr}(\mathbf{A}_r \Theta_r^H) + \sigma_L^2 \}$, $\tilde{\Xi}_k \triangleq \frac{p_b}{\text{Tr}(\Psi_{\text{sum}})} \text{Tr}(\tilde{\Psi}_k \tilde{\mathbf{R}}_b) \{ \delta_g \delta_g \text{Tr}(\mathbf{A}_r \Theta_r^H) + \sigma_L^2 \} [\mathbf{R}_b - \frac{\text{Tr}(\mathbf{R}_b \Psi_{\text{sum}})}{\text{Tr}(\Psi_{\text{sum}})}]$, $\tilde{\chi}_k \triangleq \frac{p_b \delta_g \tilde{\delta}_g}{\text{Tr}(\Psi_{\text{sum}})} \text{Tr}(\tilde{\Psi}_k \tilde{\mathbf{R}}_b) \text{Tr}(\mathbf{R}_b \Psi_{\text{sum}})$, $\tilde{\mathbf{B}}_{k2} \triangleq \varpi (\tilde{\mathbf{Q}}_k \tilde{\mathbf{R}}_k \tilde{\mathbf{R}}_b - \tilde{\mathbf{Q}}_k \tilde{\mathbf{R}}_k \tilde{\mathbf{R}}_b \tilde{\mathbf{Q}}_k + \tilde{\mathbf{R}}_b \tilde{\mathbf{R}}_k \tilde{\mathbf{Q}}_k)$, $\tilde{\mathbf{L}}_{kj} \triangleq \mathbf{Q}_j \mathbf{R}_j \tilde{\Xi}_k - \mathbf{Q}_j \mathbf{R}_j \tilde{\Xi}_k \mathbf{R}_j \mathbf{Q}_j + \tilde{\Xi}_k \mathbf{R}_j \mathbf{Q}_j$, $\tilde{\mathbf{B}}_{k3} = 2p_{uk} (\mathbf{Q}_k \mathbf{R}_k \tilde{\Psi}_k - \mathbf{Q}_k \tilde{\mathbf{R}}_k \tilde{\Psi}_k \tilde{\mathbf{R}}_k \mathbf{Q}_k + \tilde{\Psi}_k \tilde{\mathbf{R}}_k \mathbf{Q}_k)$, $\mathbf{L}_{kj} = \mathbf{Q}_j \mathbf{R}_j \mathbf{R}_k - \mathbf{Q}_j \mathbf{R}_j \mathbf{R}_k \mathbf{R}_j \mathbf{Q}_j + \mathbf{R}_k \mathbf{R}_j \mathbf{Q}_j$, $\mathbf{B}_k \triangleq 2(\mathbf{Q}_k \mathbf{R}_k \Psi_k - \mathbf{Q}_k \mathbf{R}_k \Psi_k \mathbf{R}_k \mathbf{Q}_k + \Psi_k \mathbf{R}_k \mathbf{Q}_k)$, $\chi_{k1} \triangleq \frac{1}{p_b} \sum_{j \in \mathcal{K}} p_{uj} \{ \sigma_{kj}^2 + \delta_{h_k} \tilde{\delta}_{h_j} \text{Tr}(\mathbf{A}_{w_k} \Theta_{w_k}^H) \} + \frac{\sigma_a^2}{p_b}$, and $\chi_{k2} \triangleq \frac{\delta_{h_k}}{p_b} \text{Tr}(\Psi_{\text{sum}}) \sum_{j \in \mathcal{K}} p_{uj} \tilde{\delta}_{h_j}$.

Proof: See Appendix C. ■

With the background above, we are now in a position to discuss ProGrAM as shown in **Algorithm 1**. In this algorithm, in the ℓ^{th} iteration with given the initial points $\theta_r^{(\ell)}, \theta_t^{(\ell)}$, we first obtain $\nabla_{\theta_r} f(\theta_r^{(\ell-1)}, \theta_t^{(\ell-1)})$ and $\nabla_{\theta_t} f(\theta_r^{(\ell-1)}, \theta_t^{(\ell-1)})$ (see line 3). After obtaining the gradients, we ascent in the gradient direction to obtain $\hat{\theta}_r^{(\ell)}$ and $\hat{\theta}_t^{(\ell)}$ (see lines 4

$$\begin{aligned} \nabla_{\theta_r} I_{uk} = & \left[\tilde{\delta}_g \tilde{\delta}_{h_k} \text{Tr} \{ \tilde{\mathbf{R}}_b (\tilde{\mathbf{B}}_{k1} + \tilde{\mathbf{B}}_{k2} - \tilde{\mathbf{B}}_{k3} + \sigma_u^2 \tilde{\mathbf{C}}_k) \} \right. \\ & \left. + \tilde{\delta}_g \left(\sum_{j \in \mathcal{K}} \tilde{\delta}_{h_j} p_{uj} \right) \text{Tr} \{ \tilde{\mathbf{R}}_b \tilde{\Psi}_k \} + \delta_g \sum_{i \in \mathcal{K}_r} \delta_{h_i} \text{Tr} \{ \mathbf{R}_b \tilde{\mathbf{L}}_{ki} \} + \tilde{\chi}_k \right] \text{diag}(\mathbf{A}_r), \end{aligned} \quad (34)$$

$$\nabla_{\theta_r} I_{dk} = \begin{cases} \left[\delta_g \delta_{h_k} \text{Tr} \{ \mathbf{R}_b (\Psi_{\text{sum}} - \mathbf{B}_k) \} + \delta_g \sum_{i \in \mathcal{K}_r} \delta_{h_i} \text{Tr} \{ \mathbf{R}_b (\mathbf{L}_{ki} + \chi_{k1} \mathbf{C}_i) \} + \chi_{k2} \right] \text{diag}(\mathbf{A}_r), & \text{if } w_k = r \\ \left[\delta_g \sum_{i \in \mathcal{K}_r} \delta_{h_i} \text{Tr} \{ \mathbf{R}_b (\mathbf{L}_{ki} + \chi_{k1} \mathbf{C}_i) \} \right] \text{diag}(\mathbf{A}_r), & \text{otherwise} \end{cases}, \quad (36)$$

$$\nabla_{\theta_t} I_{dk} = \begin{cases} \left[\delta_g \sum_{i \in \mathcal{K}_t} \delta_{h_i} \text{Tr} \{ \mathbf{R}_b (\mathbf{L}_{ki} + \chi_{k1} \mathbf{C}_i) \} \right] \text{diag}(\mathbf{A}_t), & \text{if } w_k = r \\ \left[\delta_g \delta_{h_k} \text{Tr} \{ \mathbf{R}_b (\Psi_{\text{sum}} - \mathbf{B}_k) \} + \delta_g \sum_{i \in \mathcal{K}_t} \delta_{h_i} \text{Tr} \{ \mathbf{R}_b (\mathbf{L}_{ki} + \chi_{k1} \mathbf{C}_i) \} + \chi_{k2} \right] \text{diag}(\mathbf{A}_t), & \text{otherwise} \end{cases}, \quad (37)$$

$$\{\theta_{r,n}^{(\ell)}, \theta_{t,n}^{(\ell)}\} = \begin{cases} \left\{ \frac{\hat{\theta}_{r,n}^{(\ell)}}{\sqrt{(\hat{\theta}_{r,n}^{(\ell)})^2 + (\hat{\theta}_{t,n}^{(\ell)})^2}}, \frac{\hat{\theta}_{t,n}^{(\ell)}}{\sqrt{(\hat{\theta}_{r,n}^{(\ell)})^2 + (\hat{\theta}_{t,n}^{(\ell)})^2}} \right\}, & \text{if } \sqrt{(\hat{\theta}_{r,n}^{(\ell)})^2 + (\hat{\theta}_{t,n}^{(\ell)})^2} \neq 0 \\ \{\sqrt{0.5} \exp(j\phi), \sqrt{0.5} \exp(j\phi)\}, \phi \in [0, 2\pi], & \text{otherwise} \end{cases}, \forall n \in \mathcal{N}. \quad (38)$$

and 5). Next, we project the new vector $\hat{\theta}^{(\ell)}$ (see line 6) onto the set Θ (as shown in line 7). Note that for a given $\hat{\theta}^{(\ell)} = [\hat{\theta}_{r,1}^{(\ell)}, \hat{\theta}_{r,2}^{(\ell)}, \dots, \hat{\theta}_{r,N}^{(\ell)}, \hat{\theta}_{t,1}^{(\ell)}, \hat{\theta}_{t,2}^{(\ell)}, \dots, \hat{\theta}_{t,N}^{(\ell)}]^T$, its projection onto Θ , i.e., $\Pi_{\Theta}(\hat{\theta}^{(\ell)})$, is given by $\theta^{(\ell)} = [\theta_{r,1}^{(\ell)}, \theta_{r,2}^{(\ell)}, \dots, \theta_{r,N}^{(\ell)}, \theta_{t,1}^{(\ell)}, \theta_{t,2}^{(\ell)}, \dots, \theta_{t,N}^{(\ell)}]^T$, where $\{\theta_{r,n}^{(\ell)}, \theta_{t,n}^{(\ell)}\}$ is given by (38). It is noteworthy that choosing an appropriate value of step-size μ_{ℓ} is critical to control the speed of convergence for the algorithm, which can efficiently be obtained by using the Barzilai-Borwein (BB) rule (see [56], and the discussion therein).

Regarding the complexity (number of complex multiplications) per iteration of the proposed algorithm, we use the big- $\mathcal{O}$ notation, which is quite meaningful for large M_T and M_R used by the proposed architecture. In particular, we find that the complexity of the objective function is $\mathcal{O}(K(N(M_T^3 + M_R^3) + N^2(M_T + M_R)N^3 + N^3))$. Next, by focusing on the complexity of the computation of the gradients, we observe that each of the gradients requires a similar complexity to the objective function.

VII. NUMERICAL RESULTS AND DISCUSSION

In this section, we present extensive numerical analytical results and Monte-Carlo (MC) simulations with 10^3 independent channel realizations of (27) to obtain deeper insights into the system performance and its dependence on various system parameters of interest. For this purpose, we assume that the locations of the BS and STARS in a 2D Euclidean coordinate system are respectively given by $(x_B, y_B) = (0, 0)$ and $(x_S, y_S) = (50, 10)$, respectively, all in meter units. The UEs in the reflection regime are assumed to be distributed on two setups. The first setup assumes a straight line between $(x_S - 0.5d_0, y_S - 0.5d_0)$ and $(x_S + 0.5d_0, y_S - 0.5d_0)$, with the distance between two adjacent UEs being constant and $d_0 = 20$ m. Similarly, the UEs in the transmission region are

assumed to be located at an equal distance from the adjacent UE on the straight line between $(x_S - 0.5d_0, y_S + 0.5d_0)$ and $(x_S + 0.5d_0, y_S + 0.5d_0)$. This setup is used in figures below unless otherwise stated. ii) The second setup assumes randomly located users on circular regions of radius 10 m. The second setup is more general and is illustrated in Fig. 3. The correlation matrices $\mathbf{R}_b$ and $\mathbf{R}_s$ are obtained by following [57] and [58], respectively. We assume that the size of each of the STARS elements is $d_H = d_V$. The distance-based path loss between the transmit array of the BS and the STARS is given by $\delta_g = d_H d_V \varphi^{-\alpha}$, where φ is the distance between the BS and the STARS. A similar modeling is applied for $\tilde{\delta}_g$, δ_{h_k} and $\tilde{\delta}_{h_k}$. Unless stated otherwise, in this section we consider $M_T = M_R = 128$, $N = 144$, $\lambda = 0.1$ m, $K_r = K_t = 2$, $\tau_c = 200$, $\tau_{up} = \tau_{dp} = K$, $p_b = 30$ dBm, $p_{uk} = 15$ dBm ($\forall k \in \mathcal{K}$), $p_{\text{train}} = 15$ dBm, $\sigma^2 = \sigma_b^2 = \sigma_d^2 = -94$ dBm, $\sigma_L^2 = 0$ dB w.r.t σ^2 , $\alpha = 2.6$, $\mu_1 = 500$, and $\epsilon = 10^{-5}$. Note that $\lambda = 0.1$ m corresponds to a frequency of 3 GHz, i.e., we focus sub-6 GHz region. Moreover, the value of σ_{kj}^2 w.r.t. σ^2 is

$$\sigma_{kj}^2 = \begin{cases} 0 \text{ dB}, & \text{if } w_j = w_k \\ 0, & \text{otherwise} \end{cases}.$$

The impact of various initial points on the convergence of the ProGrAM in obtaining a stationary solution to problem ($\mathcal{P}$) is depicted in Fig. 4. Specifically, we have plotted the average sum SE versus the iteration number returned by **Algorithm 1** from 5 different randomly generated initial $(\theta_r^{(0)}, \theta_t^{(0)})$. Note that we have checked the convergence for many different initial points but we have selected 5 points for the clarity of the figure. It is observed that the proposed optimized solution is insensitive to the initialization and the algorithm converges rapidly within a small number of iterations.

In Fig. 5, we show the dependence of the average sum SE versus the number of elements in the STARS. As the

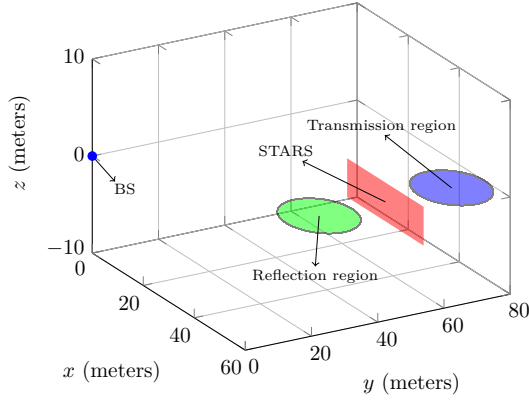


Fig. 3. Simulation setup

number of elements increases, the STARS can form narrower beams towards the UEs and/or BS by reducing interference, thus resulting in an increase in the sum SE. It is noteworthy that as the size of the STARS elements decreases (from $\lambda/4$ to $\lambda/6$), the distance between the STARS elements also decreases resulting in a stronger correlation at the STARS. This increased correlation at the STARS has a detrimental effect on system performance, as evident from the figure. Furthermore, we compare the performance of the STARS-enabled FD system with that of a corresponding STARS-enabled HD system and a cRIS-enabled FD system. In the case of the FD protocol, the BS transmits p_b units of power and all of the users transmit p_u units of power. Hence the total power used in one channel use is $(p_b + Kp_u)$. On the other hand, in the case of HD protocol, the BS transmits using p_b units of power for half channel use, and in the next half of channel use, all the users simultaneously transmit the signal using p_u units of power. Therefore, the total power used in this case is $(p_b + Kp_u)$ units of power; this indeed is the same as that used in the FD protocol. The difference between the two protocols lies in computing the spectral efficiency. In the case of the HD protocol, a factor of 0.5 appears due to half channel use. Moreover, the cRIS system is assumed to be composed of a transmit-only RIS and a reflect-only RIS with $N/2$ elements each. Regarding the optimization of the cRIS, it changes in terms of the objective function and the constraints. In particular, we have two optimization problems, one for the transmit-only RIS and one for the reflect-only RIS. The advantages of the STARS-enabled FD system over both the STARS-enabled HD system and cRIS-enabled FD systems are clearly seen in the figure. Additionally, the significantly inferior performance of the STARS-enabled FD system with random PBM, as shown in the figure, reveals the benefit of obtaining the optimal PBM. Compared to Fig. 5 corresponding to the first simulation setup with users in straight lines, in Fig. 6, we depict the impact of random realizations of the users' positions based on the second simulation setup. We observe that the rate is lower since the inter-user interference is larger due to decreased inter-user distance compared to the first geometry.

The variation of the SE of the STARS- and cRIS-enabled FD system w.r.t. p_b is shown in Fig. 8, while other quantities

in the system are fixed. Note that an increase in p_b increases the SINR at the DL UEs (see (23), which is equivalent to reduced interference as evident from (26)) while increasing the interference at the BS for UL signals (see (19)). This results in an increase in the DL SE and a decrease in the UL SE for both STARS- and cRIS-enabled systems, which results in an increase of the total SE at the beginning, while a decrease is observed at high power. Moreover, the STARS-enabled system outperforms its cRIS-enabled counterpart in terms of sum SE since the cRIS appears a decrease at high power. It is interesting to note that increasing p_b does not necessarily increase the sum SE for both the systems due to equal power allocation for all the UEs at the BS. This calls for the use of optimal power allocation at the BS, which is beyond the scope of this paper and we consider the optimal transmit beamforming design for the system under consideration as a future work. A similar analysis is shown in Fig. 7 where we vary p_u with all other system parameters kept fixed. Analogous to the previous settings, increasing p_u increases the UL SINR γ_{uk} (see (18) and (19)) while also increasing the DL interference (refer to (26)). This in turn increases the UL SE and decreases the DL SE for both STARS- and cRIS-aided systems, however, the STARS-assisted system outperforms its cRIS-assisted counterpart in terms of sum SE. Nonetheless, as shown in Fig. 7, increasing p_u does not result in a monotonic increment in sum SE for the STARS-enabled system, which calls for the need for UL power control, and is beyond the scope of this paper.

In Fig. 9, we depict the system performance for a varying number of BS receive antennas M_R while keeping all other system parameters fixed. It is clear from the figure that with a larger M_R , the MRC at the BS results in a higher UL SE, whereas the DL SE remains unchanged. It is interesting to note that the cRIS-enabled system achieves higher DL SE compared to that achieved by the STARS-enabled system because the inter-user interference in STARS is higher since it includes all the users behind and in front of the surface. On the contrary, in the case of the cRIS, the users of the transmit part of the cRIS do not interfere with the users in reflective part of the cRIS, which means that the inter-user interference is lower and the rate is higher. Analogously, in Fig. 10, we show the dependence of the SE on varying the number of transmit antennas at the BS (i.e., M_T). As expected, an increase in M_T results in a higher DL SE with UL SE being unchanged for both STARS- and cRIS-enabled systems. However, the STARS-enabled system outperforms the corresponding cRIS-assisted system in both Figs. 9 and 10.

The impact of varying transmit power during the channel estimation phase, i.e., p_{Train} on the SE of both STARS- and cRIS-aided systems is illustrated in Fig. 11. Note that an increase of p_{Train} reduces the channel estimation error or enhances the quality of channel estimation, which in turn increases both UL and DL SE for both STARS- and cRIS-enabled systems as evident from the figure. However, the STARS-enabled system attains a significantly higher sum SE compared to the cRIS-assisted system. Notably, in the downlink, initially, for low p_{Train} , the channel estimation error is less significant in STARS. However, as p_{Train} increases,

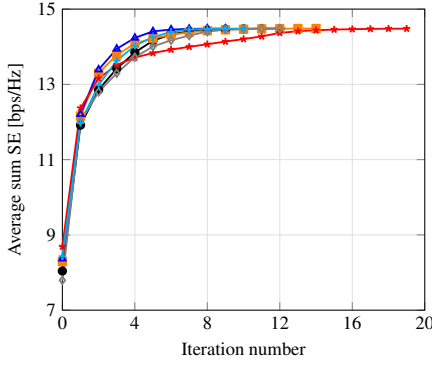


Fig. 4. Impact of random initial points ($\theta_r^{(0)}, \theta_t^{(0)}$) on the convergence of the proposed ProGrAM.

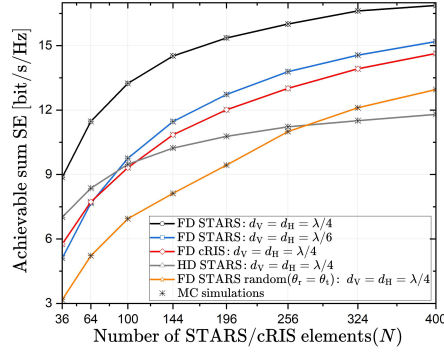


Fig. 5. Impact of the number of STARS elements on the average sum SE.

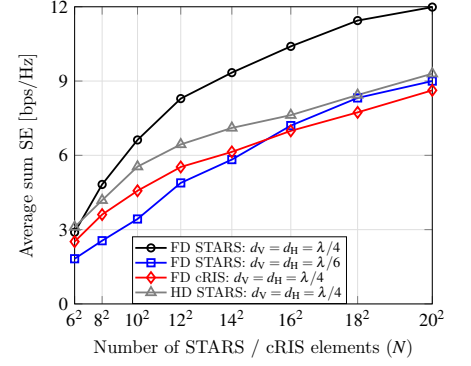


Fig. 6. Impact of the BS transmit power on the FD system performance for fixed $p_u = 15$ dBm.

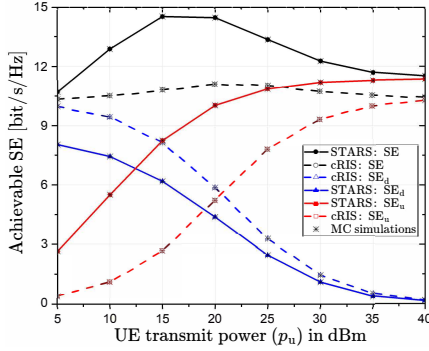


Fig. 7. Impact of UEs' transmit power on the FD system performance for fixed $p_b = 30$ dBm.

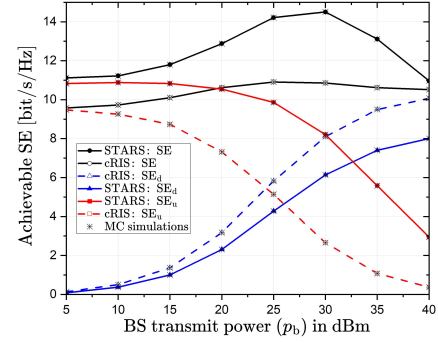


Fig. 8. Impact of random realizations of the users' positions based on the second simulation setup

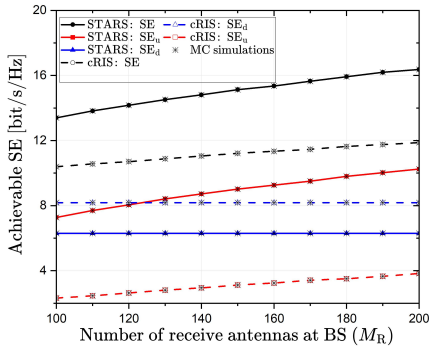


Fig. 9. Impact of the number of receive antennas at the BS on the FD system performance at fixed $M_T = 128$.

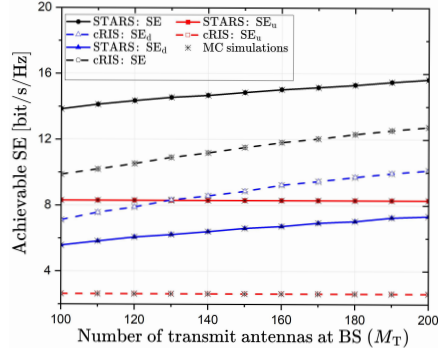


Fig. 10. Impact of the number of transmit antennas at the BS on the FD system performance at fixed $M_R = 128$.

the channel estimation error becomes more significant in STARS because this term is included in the interference, which is higher in the downlink of STARS as explained before. This justifies the following behaviour, where the downlink of the cRIS system is initially inferior to the downlink of the STARS. However, as the pilot power increases beyond a certain threshold, the downlink of the cRIS shows superior performance and the gap widens with further increase in the pilot power.

In Fig. 12, we show the impact of increasing the number of UEs on the SE of STARS- and cRIS-enabled systems. Note that in this figure, we assume $K_r = K_t = K/2$. For both STARS- and cRIS-aided systems, it is clear from the figure

that an increasing value of K decreases the DL SE due to an increase in the DL interference. On the other hand, the UL SE first increases and then saturates for large K . However, more interestingly, the sum SE for both the systems first increases and then decreases, indicating that the system cannot serve a massive number of UEs with high SE without invoking optimal power control at the UEs.

VIII. CONCLUSION

This work has considered a novel multi-user STARS-enabled mMIMO deployment in an FD two-way communication network under the practical settings of correlated fading and imperfect CSI. We have modeled both the uplink and

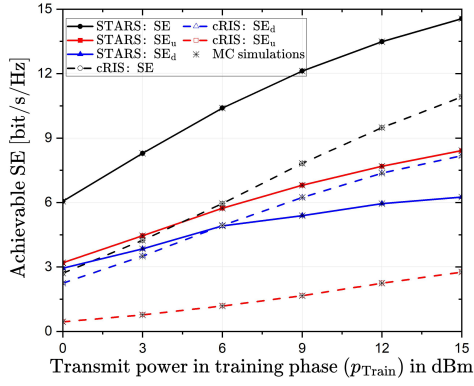


Fig. 11. Impact of transmit power in the UL and DL training phase on the FD system performance.

downlink channels of the considered system, and also derived the MMSE estimates of the effective channels. Next, we have derived closed-form expressions for the average uplink and downlink SE. Furthermore, we have devised a numerically efficient projected gradient ascent method to obtain optimal STARS PBM structure to maximize the sum SE. The superiority of the STARS-enabled FD system over the corresponding cRIS-enabled FD system and STARS-enabled HD system was established via extensive numerical results.

APPENDIX A PROOF OF THEOREM 1

Using the fact that $\mathbf{x}^H \mathbf{y} = \text{Tr}(\mathbf{y} \mathbf{x}^H)$, (16) yields

$$\begin{aligned} S_{uk} &= p_{uk} |\mathbb{E}\{\mathbf{v}_k^H \tilde{\mathbf{u}}_k\}|^2 = p_{uk} |\text{Tr}(\mathbb{E}\{\tilde{\mathbf{u}}_k \mathbf{v}_k^H\})|^2 \\ &= p_{uk} |\text{Tr}(\mathbb{E}\{\tilde{\mathbf{u}}_k \hat{\mathbf{u}}_k^H\})|^2 = p_{uk} |\text{Tr}(\mathbb{E}\{\tilde{\mathbf{u}}_k \hat{\mathbf{r}}_k^H \tilde{\mathbf{Q}}_k \tilde{\mathbf{R}}_k\})|^2 \\ &= p_{uk} |\text{Tr}(\mathbb{E}\{\tilde{\mathbf{u}}_k \hat{\mathbf{u}}_k^H\} \tilde{\mathbf{Q}}_k \tilde{\mathbf{R}}_k)|^2 = p_{uk} |\text{Tr}(\tilde{\Psi}_k)|^2 \\ &= p_{uk} \text{Tr}^2(\tilde{\Psi}_k). \end{aligned} \quad (39)$$

Note that the last equality holds since $\tilde{\Psi}_k$ is Hermitian. The first term in (17) leads to

$$\begin{aligned} \text{var}\{\sqrt{p_{uk}} \mathbf{v}_k^H \tilde{\mathbf{u}}_k\} &= p_{uk} \mathbb{E}\{|\mathbf{v}_k^H \tilde{\mathbf{u}}_k - \mathbb{E}\{\mathbf{v}_k^H \tilde{\mathbf{u}}_k\}|^2\} \\ &= p_{uk} [\mathbb{E}\{|\mathbf{v}_k^H \tilde{\mathbf{u}}_k|^2\} - |\mathbb{E}\{\mathbf{v}_k^H \tilde{\mathbf{u}}_k\}|^2] \\ &= p_{uk} [\mathbb{E}\{|\hat{\mathbf{u}}_k^H \tilde{\mathbf{u}}_k|^2\} - |\mathbb{E}\{\hat{\mathbf{u}}_k^H \tilde{\mathbf{u}}_k\}|^2] \\ &= p_{uk} [\mathbb{E}\{|\hat{\mathbf{u}}_k^H \tilde{\mathbf{u}}_k + \hat{\mathbf{u}}_k^H \tilde{\mathbf{e}}_k|^2\} - |\mathbb{E}\{\hat{\mathbf{u}}_k^H \tilde{\mathbf{u}}_k\}|^2] \\ &= p_{uk} \mathbb{E}\{|\hat{\mathbf{u}}_k^H \tilde{\mathbf{e}}_k|^2\} = p_{uk} \mathbb{E}\{\text{Tr}(\hat{\mathbf{u}}_k^H \tilde{\mathbf{e}}_k \tilde{\mathbf{e}}_k^H \hat{\mathbf{u}}_k)\} \\ &= p_{uk} \mathbb{E}\{\text{Tr}(\tilde{\mathbf{e}}_k \tilde{\mathbf{e}}_k^H \hat{\mathbf{u}}_k \hat{\mathbf{u}}_k^H)\} = p_{uk} \text{Tr}(\{\tilde{\mathbf{R}}_k - \tilde{\Psi}_k\} \tilde{\Psi}_k) \\ &= p_{uk} [\text{Tr}(\tilde{\Psi}_k \tilde{\mathbf{R}}_k) - \text{Tr}(\tilde{\Psi}_k^2)]. \end{aligned} \quad (40)$$

Next, by using the independence between $\mathbf{v}_k^H$ and $\tilde{\mathbf{u}}_i$, the second term in (17) yields

$$\begin{aligned} \sum_{i \in \mathcal{K} \setminus \{k\}} p_{ui} \mathbb{E}\{|\mathbf{v}_k^H \tilde{\mathbf{u}}_i|^2\} &= \sum_{i \in \mathcal{K} \setminus \{k\}} p_{ui} \mathbb{E}\{|\hat{\mathbf{u}}_k^H \tilde{\mathbf{u}}_i|^2\} \\ &= \sum_{i \in \mathcal{K} \setminus \{k\}} p_{ui} \mathbb{E}\{\text{Tr}(\hat{\mathbf{u}}_k^H \tilde{\mathbf{u}}_i \tilde{\mathbf{u}}_i^H \hat{\mathbf{u}}_k)\} = \sum_{i \in \mathcal{K} \setminus \{k\}} p_{ui} \mathbb{E}\{\text{Tr}(\hat{\mathbf{u}}_k \hat{\mathbf{u}}_k^H \tilde{\mathbf{u}}_i \tilde{\mathbf{u}}_i^H)\} \\ &= \sum_{i \in \mathcal{K} \setminus \{k\}} p_{ui} \text{Tr}(\mathbb{E}\{\hat{\mathbf{u}}_k \hat{\mathbf{u}}_k^H\} \mathbb{E}\{\tilde{\mathbf{u}}_i \tilde{\mathbf{u}}_i^H\}) = \sum_{i \in \mathcal{K} \setminus \{k\}} p_{ui} \text{Tr}(\tilde{\Psi}_k \tilde{\mathbf{R}}_i). \end{aligned} \quad (41)$$

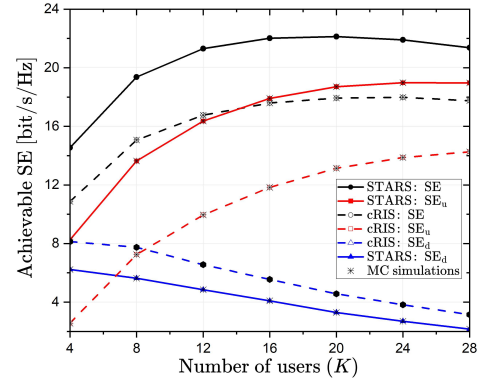


Fig. 12. Impact of the number of users on the FD system performance.

Similarly, the third term in (17) is given by

$$\begin{aligned} &\frac{\beta p_b}{K} \sum_{j \in \mathcal{K}} \mathbb{E}\{|\mathbf{v}_k^H (\tilde{\mathbf{G}} \Theta_r \mathbf{G} + \mathbf{G}_b) \mathbf{f}_j|^2\} \\ &= \frac{\beta p_b}{K} \sum_{j \in \mathcal{K}} [\mathbb{E}\{|\mathbf{v}_k^H \tilde{\mathbf{G}} \Theta_r \mathbf{G} \mathbf{f}_j|^2\} + \mathbb{E}\{|\mathbf{v}_k^H \mathbf{G}_b \mathbf{f}_j|^2\}] \\ &= \frac{\beta p_b}{K} \sum_{j \in \mathcal{K}} [\delta_g \tilde{\delta}_g \text{Tr}(\tilde{\Psi}_k \tilde{\mathbf{R}}_b) \text{Tr}(\mathbf{A}_r \Theta_r^H) \text{Tr}(\mathbf{R}_b \Psi_j) \\ &\quad + \sigma_L^2 \text{Tr}(\tilde{\Psi}_k \tilde{\mathbf{R}}_b) \text{Tr}(\Psi_j \mathbf{R}_b)]. \end{aligned} \quad (42)$$

Moreover, the normalization term β is given by

$$\begin{aligned} \beta &= \frac{K}{\sum_{j \in \mathcal{K}} \mathbb{E}\{\mathbf{f}_j^H \mathbf{f}_j\}} = \frac{K}{\sum_{j \in \mathcal{K}} \mathbb{E}\{\hat{\mathbf{u}}_j \hat{\mathbf{u}}_j^H\}} \\ &= \frac{K}{\sum_{j \in \mathcal{K}} \text{Tr}(\Psi_j)} = \frac{K}{\text{Tr}(\Psi_{\text{sum}})}. \end{aligned} \quad (43)$$

The last term in (17) is given by

$$\sigma_b^2 \mathbb{E}\{|\mathbf{v}_k^H|^2\} = \sigma_b^2 \mathbb{E}\{|\hat{\mathbf{b}}_k^H|^2\} = \sigma_b^2 \text{Tr}(\tilde{\Psi}_k). \quad (44)$$

Using (39) – (44), closed-form expressions for S_{uk} and I_{uk} can directly be written as (18) and (19), respectively. This completes the proof.

APPENDIX B PROOF OF THEOREM 2

Since the BS uses MRT/conjugate beamforming based on the estimated CSI, we have $\mathbf{f}_k = \hat{\mathbf{u}}_k^H$. Therefore, (21) yields

$$\begin{aligned} S_{dk} &= \frac{p_b}{\text{Tr}(\Psi_{\text{sum}})} |\mathbb{E}\{\mathbf{u}_k \mathbf{f}_k\}|^2 = \frac{p_b}{\text{Tr}(\Psi_{\text{sum}})} |\mathbb{E}\{\text{Tr}(\mathbf{f}_k \mathbf{u}_k)\}|^2 \\ &= \frac{p_b}{\text{Tr}(\Psi_{\text{sum}})} |\mathbb{E}\{\text{Tr}(\hat{\mathbf{u}}_k^H \mathbf{u}_k)\}|^2 = \frac{p_b}{\text{Tr}(\Psi_{\text{sum}})} |\text{Tr}(\mathbb{E}\{\mathbf{R}_k \mathbf{Q}_k \mathbf{r}_k^H \mathbf{u}_k\})|^2 \\ &= \frac{p_b}{\text{Tr}(\Psi_{\text{sum}})} |\text{Tr}(\mathbf{R}_k \mathbf{Q}_k \mathbb{E}\{\mathbf{u}_k^H \mathbf{u}_k\})|^2 = \frac{p_b}{\text{Tr}(\Psi_{\text{sum}})} \text{Tr}^2(\Psi_k). \end{aligned} \quad (45)$$

Next, following the arguments in Appendix A, closed-form expressions for the first two terms in (22) can

$$I_{uk} = \underbrace{\text{Tr}(\tilde{\Psi}_k \tilde{\mathbf{R}}_{\text{sum}})}_{I_{uk1}} + \underbrace{\frac{p_b}{\text{Tr}(\Psi_{\text{sum}})} \text{Tr}(\tilde{\Psi}_k \tilde{\mathbf{R}}_b) \text{Tr}(\mathbf{R}_b \Psi_{\text{sum}})}_{I_{uk2}} \{ \delta_g \tilde{\delta}_g \text{Tr}(\mathbf{A}_r \Theta_r^H) + \sigma_L^2 \} - \underbrace{p_{uk} \text{Tr}(\tilde{\Psi}_k^2)}_{I_{uk3}} + \underbrace{\sigma_b^2 \text{Tr}(\tilde{\Psi}_k)}_{I_{uk4}}. \quad (48)$$

be expressed as $\frac{p_b}{\text{Tr}(\Psi_{\text{sum}})} [\text{Tr}(\Psi_k \mathbf{R}_k) - \text{Tr}(\Psi_k^2)]$ and $\frac{p_b}{\text{Tr}(\Psi_{\text{sum}})} \sum_{i \in \mathcal{K} \setminus \{k\}} \text{Tr}(\mathbf{R}_k \Psi_i)$. The third term in (22) yields,

$$\begin{aligned} & \sum_{j \in \mathcal{K}} p_{uj} (\sigma_{kj}^2 + \mathbb{E}\{|\mathbf{h}_k \Theta_{wk} \tilde{\mathbf{h}}_j|^2\}) \\ &= \sum_{j \in \mathcal{K}} p_{uj} [\sigma_{kj}^2 + \mathbb{E}\{\text{Tr}(\mathbf{h}_k \Theta_{wk} \tilde{\mathbf{h}}_j \tilde{\mathbf{h}}_j^H \Theta_{wk}^H \mathbf{h}_k^H)\}] \\ &= \sum_{j \in \mathcal{K}} p_{uj} [\sigma_{kj}^2 + \mathbb{E}\{\text{Tr}(\mathbf{h}_k^H \mathbf{h}_k \Theta_{wk} \tilde{\mathbf{h}}_j \tilde{\mathbf{h}}_j^H \Theta_{wk}^H)\}] \\ &= \sum_{j \in \mathcal{K}} p_{uj} [\sigma_{kj}^2 + \text{Tr}(\mathbb{E}\{\mathbf{h}_k^H \mathbf{h}_k\} \Theta_{wk} \mathbb{E}\{\tilde{\mathbf{h}}_j \tilde{\mathbf{h}}_j^H\} \Theta_{wk}^H)] \\ &= \sum_{j \in \mathcal{K}} p_{uj} [\sigma_{kj}^2 + \delta_{h_k} \tilde{\delta}_{h_j} \text{Tr}(\mathbb{E}\{\mathbf{R}_s^{1/2} \mathbf{c}_k^H \mathbf{c}_k \mathbf{R}_s^{1/2}\} \\ & \quad \times \Theta_{wk} \mathbb{E}\{\mathbf{R}_s^{1/2} \tilde{\mathbf{c}}_j^H \mathbf{c}_j \mathbf{R}_s^{1/2}\} \Theta_{wk}^H)] \\ &= \sum_{j \in \mathcal{K}} p_{uj} [\sigma_{kj}^2 + \delta_{h_k} \tilde{\delta}_{h_j} \text{Tr}(\mathbf{R}_s \Theta_{wk} \mathbf{R}_s \Theta_{wk}^H)] \\ &= \sum_{j \in \mathcal{K}} p_{uj} [\sigma_{kj}^2 + \delta_{h_k} \tilde{\delta}_{h_j} \text{Tr}(\mathbf{A}_{wk} \Theta_{wk}^H)]. \quad (46) \end{aligned}$$

Therefore, using the fact that $z_{dk} \sim \mathcal{CN}(0, \sigma_d^2)$, (45), (46), and the arguments above, closed-form expressions for S_{dk} and I_{dk} are respectively given by (23) and (24). This completes the proof.

APPENDIX C PROOF OF THEOREM 3

Using (18), we have

$$\begin{aligned} \nabla_{\theta_m} S_{uk} &= 2p_{uk} \text{Tr}(\tilde{\Psi}_k) \text{Tr}\{\nabla_{\theta_m}(\tilde{\mathbf{R}}_k \tilde{\mathbf{Q}}_k \tilde{\mathbf{R}}_k)\} \\ &= 2p_{uk} \text{Tr}(\tilde{\Psi}_k) \text{Tr}\{\nabla_{\theta_m}(\tilde{\mathbf{R}}_k) \tilde{\mathbf{Q}}_k \tilde{\mathbf{R}}_k \\ & \quad - \tilde{\mathbf{R}}_k \tilde{\mathbf{Q}}_k \nabla_{\theta_m}(\tilde{\mathbf{R}}_k) \tilde{\mathbf{Q}}_k \tilde{\mathbf{R}}_k + \tilde{\mathbf{R}}_k \tilde{\mathbf{Q}}_k \nabla_{\theta_m}(\tilde{\mathbf{R}}_k)\} \\ &= 2\tilde{\delta}_g \tilde{\delta}_{h_k} p_{uk} \text{Tr}(\tilde{\Psi}_k) \text{Tr}\{\tilde{\mathbf{R}}_b \tilde{\mathbf{C}}_k \nabla_{\theta_m}(\mathbf{A}_r \Theta_r^H)\} \\ &= \begin{cases} \tilde{\nu}_k \text{diag}(\mathbf{A}_r), & \text{if } m = r \\ \mathbf{0}, & \text{otherwise} \end{cases}, \quad (47) \end{aligned}$$

where $\tilde{\mathbf{C}}_k \triangleq \tilde{\mathbf{Q}}_k \tilde{\mathbf{R}}_k - \tilde{\mathbf{Q}}_k \tilde{\mathbf{R}}_k^2 \tilde{\mathbf{Q}}_k + \tilde{\mathbf{R}}_k \tilde{\mathbf{Q}}_k$, and $\tilde{\nu}_k \triangleq 2\tilde{\delta}_g \tilde{\delta}_{h_k} p_{uk} \text{Tr}(\tilde{\Psi}_k) \text{Tr}(\tilde{\mathbf{R}}_b \tilde{\mathbf{C}}_k)$. Analogously, closed-form expressions for $\nabla_{\theta_r} S_{dk}$ and $\nabla_{\theta_t} S_{dk}$ are given by (32), and (33), respectively.

Next, using (19), let I_{uk} is given by (48). Therefore, the gradient of the first term in (48) is given by

$$\begin{aligned} \nabla_{\theta_m} I_{uk1} &= \text{Tr}\{\nabla_{\theta_m}(\tilde{\Psi}_k) \tilde{\mathbf{R}}_{\text{sum}}\} + \sum_{j \in \mathcal{K}} p_{uj} \text{Tr}\{\tilde{\Psi}_k \nabla_{\theta_m}(\tilde{\mathbf{R}}_j)\} \\ &= \text{Tr}\{\tilde{\mathbf{B}}_{k1} \nabla_{\theta_m}(\tilde{\mathbf{R}}_k)\} + \sum_{j \in \mathcal{K}} p_{uj} \text{Tr}\{\tilde{\Psi}_k \nabla_{\theta_m}(\tilde{\mathbf{R}}_j)\}, \quad (49) \end{aligned}$$

where $\tilde{\mathbf{B}}_{k1} \triangleq \tilde{\mathbf{Q}}_k \tilde{\mathbf{R}}_k \tilde{\mathbf{R}}_{\text{sum}} - \tilde{\mathbf{Q}}_k \tilde{\mathbf{R}}_k \tilde{\mathbf{R}}_{\text{sum}} \tilde{\mathbf{R}}_k \tilde{\mathbf{Q}}_k + \tilde{\mathbf{R}}_{\text{sum}} \tilde{\mathbf{R}}_k \tilde{\mathbf{Q}}_k$. The gradient of the second term in (48) is given by

$$\begin{aligned} \nabla_{\theta_m} I_{uk2} &= \nabla_{\theta_m} \frac{p_b}{\text{Tr}(\Psi_{\text{sum}})} \\ & \quad \times \text{Tr}(\tilde{\Psi}_k \tilde{\mathbf{R}}_b) \text{Tr}(\mathbf{R}_b \Psi_{\text{sum}}) \{ \delta_g \tilde{\delta}_g \text{Tr}(\mathbf{A}_r \Theta_r^H) + \sigma_L^2 \} \\ &= \nabla_{\theta_m} \frac{p_b \delta_g \tilde{\delta}_g}{\text{Tr}(\Psi_{\text{sum}})} \text{Tr}(\tilde{\Psi}_k \tilde{\mathbf{R}}_b) \text{Tr}(\mathbf{R}_b \Psi_{\text{sum}}) \text{Tr}(\mathbf{A}_r \Theta_r^H) \\ & \quad + \nabla_{\theta_m} \frac{p_b \sigma_L^2}{\text{Tr}(\Psi_{\text{sum}})} \text{Tr}(\tilde{\Psi}_k \tilde{\mathbf{R}}_b) \text{Tr}(\mathbf{R}_b \Psi_{\text{sum}}) \\ &= \frac{p_b \delta_g \tilde{\delta}_g}{\text{Tr}^2(\Psi_{\text{sum}})} \left[\text{Tr}(\Psi_{\text{sum}}) \nabla_{\theta_m} \left\{ \text{Tr}(\tilde{\Psi}_k \tilde{\mathbf{R}}_b) \text{Tr}(\mathbf{R}_b \Psi_{\text{sum}}) \right. \right. \\ & \quad \times \text{Tr}(\mathbf{A}_r \Theta_r^H) \left. \right\} - \text{Tr}(\tilde{\Psi}_k \tilde{\mathbf{R}}_b) \text{Tr}(\mathbf{R}_b \Psi_{\text{sum}}) \text{Tr}(\mathbf{A}_r \Theta_r^H) \\ & \quad \times \nabla_{\theta_m} \text{Tr}(\Psi_{\text{sum}}) \left. \right] + \frac{p_b \sigma_L^2}{\text{Tr}^2(\Psi_{\text{sum}})} \left[\text{Tr}(\Psi_{\text{sum}}) \right. \\ & \quad \times \nabla_{\theta_m} \left\{ \text{Tr}(\tilde{\Psi}_k \tilde{\mathbf{R}}_b) \text{Tr}(\mathbf{R}_b \Psi_{\text{sum}}) \right\} - \text{Tr}(\tilde{\Psi}_k \tilde{\mathbf{R}}_b) \\ & \quad \times \text{Tr}(\mathbf{R}_b \Psi_{\text{sum}}) \nabla_{\theta_m} \text{Tr}(\Psi_{\text{sum}}) \left. \right] \\ &= \frac{p_b \delta_g \tilde{\delta}_g}{\text{Tr}^2(\Psi_{\text{sum}})} \left[\text{Tr}(\Psi_{\text{sum}}) \text{Tr}(\mathbf{R}_b \Psi_{\text{sum}}) \text{Tr}(\mathbf{A}_r \Theta_r^H) \right. \\ & \quad \times \text{Tr}\{(\nabla_{\theta_m} \tilde{\Psi}_k) \tilde{\mathbf{R}}_b\} + \text{Tr}(\Psi_{\text{sum}}) \text{Tr}(\tilde{\Psi}_k \tilde{\mathbf{R}}_b) \\ & \quad \times \text{Tr}(\mathbf{A}_r \Theta_r^H) \text{Tr}\{\mathbf{R}_b (\sum_{j \in \mathcal{K}} \nabla_{\theta_m} \Psi_j)\} + \text{Tr}(\Psi_{\text{sum}}) \\ & \quad \times \text{Tr}(\tilde{\Psi}_k \tilde{\mathbf{R}}_b) \text{Tr}(\mathbf{R}_b \Psi_{\text{sum}}) \nabla_{\theta_m} \text{Tr}(\mathbf{A}_r \Theta_r^H) - \text{Tr}(\tilde{\Psi}_k \tilde{\mathbf{R}}_b) \\ & \quad \times \text{Tr}(\mathbf{R}_b \Psi_{\text{sum}}) \text{Tr}(\mathbf{A}_r \Theta_r^H) \nabla_{\theta_m} \text{Tr}\{\sum_{j \in \mathcal{K}} \nabla_{\theta_m} \Psi_j\} \left. \right] \\ & \quad + \frac{p_b \sigma_L^2}{\text{Tr}^2(\Psi_{\text{sum}})} \left[\text{Tr}(\Psi_{\text{sum}}) \text{Tr}(\mathbf{R}_b \Psi_{\text{sum}}) \text{Tr}\{(\nabla_{\theta_m} \tilde{\Psi}_k) \tilde{\mathbf{R}}_b\} \right. \\ & \quad + \text{Tr}(\Psi_{\text{sum}}) \text{Tr}(\tilde{\Psi}_k \tilde{\mathbf{R}}_b) \text{Tr}\{\mathbf{R}_b (\sum_{j \in \mathcal{K}} \nabla_{\theta_m} \Psi_j)\} \\ & \quad - \text{Tr}(\tilde{\Psi}_k \tilde{\mathbf{R}}_b) \text{Tr}(\mathbf{R}_b \Psi_{\text{sum}}) \nabla_{\theta_m} \text{Tr}\{\sum_{j \in \mathcal{K}} \nabla_{\theta_m} \Psi_j\} \left. \right] \\ &= \text{Tr}\{\tilde{\mathbf{B}}_{k2} (\nabla_{\theta_m} \tilde{\mathbf{R}}_k)\} + \sum_{j \in \mathcal{K}} \text{Tr}\{\tilde{\mathbf{L}}_{kj} (\nabla_{\theta_m} \mathbf{R}_j)\} \\ & \quad + \tilde{\chi}_k \nabla_{\theta_m} \text{Tr}(\mathbf{A}_r \Theta_r^H), \quad (50) \end{aligned}$$

where $\varpi \triangleq \frac{p_b}{\text{Tr}(\Psi_{\text{sum}})} \text{Tr}(\mathbf{R}_b \Psi_{\text{sum}}) \{ \delta_g \tilde{\delta}_g \text{Tr}(\mathbf{A}_r \Theta_r^H) + \sigma_L^2 \}$, $\tilde{\mathbf{B}}_{k2} \triangleq \frac{p_b}{\text{Tr}(\Psi_{\text{sum}})} \text{Tr}(\tilde{\Psi}_k \tilde{\mathbf{R}}_b) \{ \delta_g \tilde{\delta}_g \text{Tr}(\mathbf{A}_r \Theta_r^H) + \sigma_L^2 \} [\mathbf{R}_b -$

$$I_{dk} = \underbrace{\text{Tr}(\mathbf{R}_k \Psi_{\text{sum}})}_{I_{dk1}} + \underbrace{\frac{\text{Tr}(\Psi_{\text{sum}})}{p_b} \sum_{j \in \mathcal{K}} \left[p_{uj} \{ \sigma_{kj}^2 + \delta_{hk} \tilde{\delta}_{hj} \text{Tr}(\mathbf{A}_{w_k} \Theta_{w_k}^H) \} \right]}_{I_{dk2}} - \underbrace{\text{Tr}(\Psi_k^2)}_{I_{dk3}} + \underbrace{\frac{\sigma_d^2}{p_b} \text{Tr}(\Psi_{\text{sum}})}_{I_{dk4}}. \quad (54)$$

$\frac{\text{Tr}(\mathbf{R}_b \Psi_{\text{sum}})}{\text{Tr}(\Psi_{\text{sum}})]$, $\tilde{\chi}_k \triangleq \frac{p_b \delta_g \tilde{\delta}_g}{\text{Tr}(\Psi_{\text{sum}})} \text{Tr}(\tilde{\Psi}_k \tilde{\mathbf{R}}_b) \text{Tr}(\mathbf{R}_b \Psi_{\text{sum}})$, $\tilde{\mathbf{B}}_{k2} \triangleq \varpi(\tilde{\mathbf{Q}}_k \tilde{\mathbf{R}}_k \tilde{\mathbf{R}}_b - \tilde{\mathbf{Q}}_k \tilde{\mathbf{R}}_k \tilde{\mathbf{R}}_b \tilde{\mathbf{Q}}_k + \tilde{\mathbf{R}}_b \tilde{\mathbf{R}}_k \tilde{\mathbf{Q}}_k)$, and $\tilde{\mathbf{L}}_{kj} \triangleq \mathbf{Q}_j \mathbf{R}_j \tilde{\Xi}_k - \mathbf{Q}_j \mathbf{R}_j \tilde{\Xi}_k \mathbf{R}_j \mathbf{Q}_j + \tilde{\Xi}_k \mathbf{R}_j \mathbf{Q}_j$. Similarly, the gradient of the third term in (48) can be obtained as

$$\begin{aligned} \nabla_{\theta_m} I_{uk3} &= p_{uk} \nabla_{\theta_m} \text{Tr}(\tilde{\Psi}_k^2) = p_{uk} \text{Tr} \{ 2 \tilde{\Psi}_k (\nabla_{\theta_m} \tilde{\Psi}_k) \} \\ &= \text{Tr} \{ \tilde{\mathbf{B}}_{k3} \nabla_{\theta_m} (\tilde{\mathbf{R}}_k) \}, \end{aligned} \quad (51)$$

where $\tilde{\mathbf{B}}_{k3} = 2p_{uk}(\tilde{\mathbf{Q}}_k \tilde{\mathbf{R}}_k \tilde{\Psi}_k - \tilde{\mathbf{Q}}_k \tilde{\mathbf{R}}_k \tilde{\Psi}_k \tilde{\mathbf{R}}_k \tilde{\mathbf{Q}}_k + \tilde{\Psi}_k \tilde{\mathbf{R}}_k \tilde{\mathbf{Q}}_k)$. Moreover, the derivative of the last term in (48) can be computed as

$$\nabla_{\theta_m} I_{uk4} = \sigma_b^2 \nabla_{\theta_m} \text{Tr}(\tilde{\Psi}_k) = \sigma_b^2 \text{Tr} \{ \tilde{\mathbf{C}}_k \nabla_{\theta_m} (\tilde{\mathbf{R}}_k) \}. \quad (52)$$

Using (49) – (52) with some algebraic manipulations yields

$$\begin{aligned} \nabla_{\theta_m} I_{uk} &= \tilde{\delta}_g \tilde{\delta}_{h_k} \text{Tr} \{ \tilde{\mathbf{R}}_b (\tilde{\mathbf{B}}_{k1} + \tilde{\mathbf{B}}_{k2} - \tilde{\mathbf{B}}_{k3} + \sigma_u^2 \tilde{\mathbf{C}}_k) \nabla_{\theta_m} (\mathbf{A}_r \Theta_r^H) \} \\ &+ \tilde{\delta}_g \sum_{j \in \mathcal{K}} \tilde{\delta}_{h_j} p_{uj} \text{Tr} \{ \tilde{\mathbf{R}}_b \tilde{\Psi}_k \nabla_{\theta_m} (\mathbf{A}_r \Theta_r^H) \} \\ &+ \delta_g \sum_{j \in \mathcal{K}} \delta_{h_j} \text{Tr} \{ \mathbf{R}_b \tilde{\mathbf{L}}_{kj} \nabla_{\theta_m} (\mathbf{A}_{w_j} \Theta_{w_j}^H) \} \\ &+ \tilde{\chi}_k \nabla_{\theta_m} \text{Tr}(\mathbf{A}_r \Theta_r^H). \end{aligned} \quad (53)$$

Using (53), the closed-form expressions for $\nabla_{\theta_r} I_{uk}$ and $\nabla_{\theta_t} I_{uk}$ are respectively given by (34) and (35).

Turning our attention to the DL signaling, from (26) we write I_{dk} as given in (54). Therefore, the gradient of the first term in (54) is given by

$$\begin{aligned} \nabla_{\theta_m} I_{dk1} &= \nabla_{\theta_m} \text{Tr}(\Psi_{\text{sum}} \mathbf{R}_k) \\ &= \sum_{j \in \mathcal{K}} \left[\text{Tr} \{ \nabla_{\theta_m} (\Psi_j) \mathbf{R}_k \} + \text{Tr}(\Psi_j \nabla_{\theta_m} (\mathbf{R}_k)) \right] \\ &= \sum_{j \in \mathcal{K}} \text{Tr} \{ \mathbf{L}_{kj} \nabla_{\theta_m} (\mathbf{R}_j) \} + \text{Tr} \{ \Psi_{\text{sum}} \nabla_{\theta_m} (\mathbf{R}_k) \}, \end{aligned} \quad (55)$$

where $\mathbf{L}_{kj} = \mathbf{Q}_j \mathbf{R}_j \mathbf{R}_k - \mathbf{Q}_j \mathbf{R}_j \mathbf{R}_k \mathbf{R}_j \mathbf{Q}_j + \mathbf{R}_k \mathbf{R}_j \mathbf{Q}_j$. The gradient of the second term in (54) can be obtained as

$$\begin{aligned} \nabla_{\theta_m} I_{dk2} &= \frac{1}{p_b} \sum_{j \in \mathcal{K}} p_{uj} \sigma_{kj}^2 \text{Tr}(\nabla_{\theta_m} \Psi_{\text{sum}}) \\ &+ \frac{\delta_{h_k}}{p_b} \sum_{j \in \mathcal{K}} p_{uj} \tilde{\delta}_{h_j} \nabla_{\theta_m} \text{Tr}(\Psi_{\text{sum}}) \text{Tr}(\mathbf{A}_{w_k} \Theta_{w_k}^H) \\ &= \frac{1}{p_b} \sum_{j \in \mathcal{K}} p_{uj} \left[\sigma_{kj}^2 + \delta_{hk} \tilde{\delta}_{h_j} \text{Tr}(\mathbf{A}_{w_k} \Theta_{w_k}^H) \right] \\ &\times \left\{ \sum_{i \in \mathcal{K}} \text{Tr} \{ \mathbf{C}_i \nabla_{\theta_m} (\mathbf{R}_i) \} \right\} \\ &+ \frac{\delta_{h_k}}{p_b} \text{Tr}(\Psi_{\text{sum}}) \sum_{j \in \mathcal{K}} p_{uj} \tilde{\delta}_{h_j} \nabla_{\theta_m} \text{Tr}(\mathbf{A}_{w_k} \Theta_{w_k}^H). \end{aligned} \quad (56)$$

Next, the gradient of the third term in (54) is given by

$$\nabla_{\theta_m} I_{dk3} = \nabla_{\theta_m} \text{Tr}(\Psi_k^2) = \text{Tr} \{ \mathbf{B}_k \nabla_{\theta_m} (\mathbf{R}_k) \}, \quad (57)$$

where $\mathbf{B}_k = 2(\mathbf{Q}_k \mathbf{R}_k \Psi_k - \mathbf{Q}_k \mathbf{R}_k \Psi_k \mathbf{R}_k \mathbf{Q}_k + \Psi_k \mathbf{R}_k \mathbf{Q}_k)$. Additionally, the gradient of the last term in (54) is obtained as

$$\begin{aligned} \nabla_{\theta_m} I_{dk4} &= \frac{\sigma_d^2}{p_b} \text{Tr} \left(\sum_{i \in \mathcal{K}} \nabla_{\theta_m} \Psi_i \right) \\ &= \frac{\sigma_d^2}{p_b} \sum_{i \in \mathcal{K}} \text{Tr} \{ \mathbf{C}_i \nabla_{\theta_m} (\mathbf{R}_i) \}. \end{aligned} \quad (58)$$

Using some algebraic manipulations on (54) – (58), the gradient of I_{dk} is given by

$$\begin{aligned} \nabla_{\theta_m} I_{dk} &= \text{Tr} \{ (\Psi_{\text{sum}} - \mathbf{B}_k) \nabla_{\theta_m} (\mathbf{R}_k) \} \\ &+ \sum_{j \in \mathcal{K}} \text{Tr} \{ (\mathbf{L}_{kj} + \chi_{k1} \mathbf{C}_j) \nabla_{\theta_m} (\mathbf{R}_j) \} + \chi_{k2} \nabla_{\theta_m} \text{Tr}(\mathbf{A}_{w_k} \Theta_{w_k}^H) \\ &= \delta_g \delta_{h_k} \text{Tr} \{ \mathbf{R}_b (\Psi_{\text{sum}} - \mathbf{B}_k) \nabla_{\theta_m} (\mathbf{A}_{w_k} \Theta_{w_k}^H) \} \\ &+ \delta_g \sum_{j \in \mathcal{K}} \delta_{h_j} \text{Tr} \{ \mathbf{R}_b (\mathbf{L}_{kj} + \chi_{k1} \mathbf{C}_j) \nabla_{\theta_m} (\mathbf{A}_{w_j} \Theta_{w_j}^H) \} \\ &+ \chi_{k2} \nabla_{\theta_m} \text{Tr}(\mathbf{A}_{w_k} \Theta_{w_k}^H), \end{aligned} \quad (59)$$

where $\Xi_k = \Psi_{\text{sum}} - 2\mathbf{T}_k$, $\chi_{k1} = \frac{1}{p_b} \sum_{j \in \mathcal{K}} p_{uj} \{ \sigma_{kj}^2 + \delta_{h_k} \tilde{\delta}_{h_j} \text{Tr}(\mathbf{A}_{w_k} \Theta_{w_k}^H) \} + \frac{\sigma_d^2}{p_b}$, and $\chi_{k2} = \frac{\delta_{h_k}}{p_b} \text{Tr}(\Psi_{\text{sum}}) \sum_{j \in \mathcal{K}} p_{uj} \tilde{\delta}_{h_j}$. From (59), closed-form expressions for $\nabla_{\theta_r} I_{dk}$ and $\nabla_{\theta_t} I_{dk}$ are respectively given by (36) and (37). This completes the proof.

REFERENCES

- [1] M. Di Renzo *et al.*, “Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead,” *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2450–2525, 2020.
- [2] Q. Wu *et al.*, “Intelligent reflecting surface-aided wireless communications: A tutorial,” *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3313–3351, 2021.
- [3] W. Mei, B. Zheng, C. You, and R. Zhang, “Intelligent reflecting surface-aided wireless networks: From single-reflection to multireflection design and optimization,” *Proc. IEEE*, vol. 110, no. 9, pp. 1380–1400, 2022.
- [4] H. Yang *et al.*, “A programmable metasurface with dynamic polarization, scattering and focusing control,” *Sci. Rep.*, vol. 6, no. 1, pp. 1–11, 2016.
- [5] C. Liaskos *et al.*, “A new wireless communication paradigm through software-controlled metasurfaces,” *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 162–169, 2018.
- [6] N. S. Perović, L.-N. Tran, M. Di Renzo, and M. F. Flanagan, “On the maximum achievable sum-rate of the RIS-aided MIMO broadcast channel,” *IEEE Trans. Signal Process.*, vol. 70, pp. 6316–6331, 2022.
- [7] C. Pan *et al.*, “Multicell MIMO communications relying on intelligent reflecting surfaces,” *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5218–5233, 2020.
- [8] 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, 2020.
- [9] V. Kumar, M. Flanagan, R. Zhang, and L.-N. Tran, “Achievable rate maximization for underlay spectrum sharing MIMO system with intelligent reflecting surface,” *IEEE Wireless Commun. Lett.*, vol. 11, no. 8, pp. 1758–1762, 2022.

- [10] C. Huang *et al.*, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, 2019.
- [11] A. Mukherjee, V. Kumar, D. W. K. Ng, and L.-N. Tran, "On the energy-efficiency maximization for IRS-assisted MIMOME wiretap channels," in *IEEE Vehicular Technology Conference (VTC2022-Fall)*, 2022, pp. 1–6.
- [12] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, 2019.
- [13] V. Kumar, R. Zhang, M. D. Renzo, and L.-N. Tran, "A novel SCA-based method for beamforming optimization in IRS/RIS-assisted MU-MISO downlink," *IEEE Wireless Commun. Lett.*, vol. 12, no. 2, pp. 297–301, 2023.
- [14] J. Xu, Y. Liu, X. Mu, and O. A. Dobre, "STAR-RISs: Simultaneous transmitting and reflecting reconfigurable intelligent surfaces," *IEEE Commun. Lett.*, vol. 25, no. 9, pp. 3134–3138, 2021.
- [15] X. Mu *et al.*, "Simultaneously transmitting and reflecting (STAR) RIS aided wireless communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 5, pp. 3083–3098, 2022.
- [16] H. Niu, Z. Chu, F. Zhou, and Z. Zhu, "Simultaneous transmission and reflection reconfigurable intelligent surface assisted secrecy MISO networks," *IEEE Commun. Lett.*, vol. 25, no. 11, pp. 3498–3502, 2021.
- [17] C. Wu *et al.*, "Channel estimation for STAR-RIS-aided wireless communication," *IEEE Commun. Lett.*, vol. 26, no. 3, pp. 652–656, 2022.
- [18] H. Niu *et al.*, "Weighted sum rate optimization for STAR-RIS-assisted MIMO system," *IEEE Trans. Veh. Tech.*, vol. 71, no. 2, pp. 2122–2127, 2022.
- [19] NTT-DOCOMO. DOCOMO conducts world's first successful trial of transparent dynamic metasurface. [Online]. Available: www.nttdocomo.co.jp/english/info/mediacenter/pr/2020/011700.html
- [20] H. Alves, T. Riihonen, and H. A. Suraweera, *Full-Duplex Communications for Future Wireless Networks*. Singapore: Springer-Verlag, 2020.
- [21] G. C. Alexandropoulos, M. A. Islam, and B. Smida, "Full-duplex massive multiple-input, multiple-output architectures: Recent advances, applications, and future directions," *IEEE Veh. Technol. Mag.*, vol. 17, no. 4, pp. 83–91, 2022.
- [22] Z. Peng *et al.*, "Multiuser full-duplex two-way communications via intelligent reflecting surface," *IEEE Trans. Signal Process.*, vol. 69, pp. 837–851, 2021.
- [23] H. Shen, T. Ding, W. Xu, and C. Zhao, "Beamforming design with fast convergence for IRS-aided full-duplex communication," *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2849–2853, 2020.
- [24] 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, 2020.
- [25] S. R. Kudumala *et al.*, "Hardware impaired RIS assisted multipair FD communication with spatial correlation," *IEEE Commun. Lett.*, vol. 26, no. 9, pp. 2200–2204, 2022.
- [26] N. Deshpande, S. Dey, D. N. Amudala, and R. Budhiraja, "Spatially-correlated IRS-aided multi-user FD mMIMO systems: Analysis and optimization," *IEEE Trans. Commun.*, vol. 70, no. 6, pp. 3879–3896, 2022.
- [27] A. Papazafeiropoulos, P. Kourtessis, and I. Krikidis, "STAR-RIS assisted full-duplex systems: Impact of correlation and maximization," *IEEE Commun. Lett.*, vol. 26, no. 12, pp. 3004–3008, 2022.
- [28] F. Karim, S. K. Singh, K. Singh, and M. F. Flanagan, "STAR-RIS aided full duplex communication system: Performance analysis," in *IEEE Global Communications Conference (GLOBECOM)*, 2022, pp. 3114–3119.
- [29] P. P. Perera, V. G. Warnasooriya, D. Kudathanthirige, and H. A. Suraweera, "Sum rate maximization in STAR-RIS assisted full-duplex communication systems," in *IEEE International Conference on Communications (ICC)*, 2022, pp. 3281–3286.
- [30] T. H. Nguyen and T. T. Nguyen, "On performance of STAR-RIS-enabled multiple two-way full-duplex D2D communication systems," *IEEE Access*, vol. 10, pp. 89063–89071, 2022.
- [31] Y. Wang, P. Guan, H. Yu, and Y. Zhao, "Transmit power optimization of simultaneous transmission and reflection RIS assisted full-duplex communications," *IEEE Access*, vol. 10, pp. 61192–61200, 2022.
- [32] C. Pan *et al.*, "An overview of signal processing techniques for RIS/IRS-aided wireless systems," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 5, pp. 883–917, 2022.
- [33] B. Zheng, C. You, W. Mei, and R. Zhang, "A survey on channel estimation and practical passive beamforming design for intelligent reflecting surface aided wireless communications," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 2, pp. 1035–1071, 2022.
- [34] A. Papazafeiropoulos *et al.*, "Intelligent reflecting surface-assisted MU-MISO systems with imperfect hardware: Channel estimation and beamforming design," *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 2077–2092, 2022.
- [35] K. Zhi, C. Pan, H. Ren, and K. Wang, "Power scaling law analysis and phase shift optimization of RIS-aided massive MIMO systems with statistical CSI," *IEEE Trans. Commun.*, vol. 70, no. 5, pp. 3558–3574, 2022.
- [36] A. Papazafeiropoulos *et al.*, "Asymptotic analysis of Max-Min weighted SINR for IRS-assisted MISO systems with hardware impairments," *IEEE Wireless Commun. Lett.*, vol. 12, no. 2, pp. 192–196, 2023.
- [37] A. Papazafeiropoulos, I. Krikidis, and P. Kourtessis, "Impact of channel aging on reconfigurable intelligent surface aided massive MIMO systems with statistical CSI," *IEEE Trans. Veh. Tech.*, vol. 72, no. 1, pp. 689–703, 2023.
- [38] B. Ji *et al.*, "Several key technologies for 6G: Challenges and opportunities," *IEEE Commun. Stand. Mag.*, vol. 5, no. 2, pp. 44–51, 2021.
- [39] M. Matthaiou *et al.*, "The road to 6G: Ten physical layer challenges for communications engineers," *IEEE Commun. Mag.*, vol. 59, no. 1, pp. 64–69, 2021.
- [40] Z. Abdullah, G. Chen, S. Lambbotharan, and J. A. Chambers, "Optimization of intelligent reflecting surface assisted full-duplex relay networks," *IEEE Wireless Commun. Lett.*, vol. 10, no. 2, pp. 363–367, 2020.
- [41] Q. Nadeem *et al.*, "Intelligent reflecting surface-assisted multi-user MISO Communication: Channel estimation and beamforming design," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 661–680, 2020.
- [42] K. Zhi, C. Pan, H. Ren, K. K. Chai, and M. Elkhassan, "Active RIS versus passive RIS: Which is superior with the same power budget?" *IEEE Commun. Lett.*, vol. 26, no. 5, pp. 1150–1154, 2022.
- [43] A. Papazafeiropoulos, H. Ge, P. Kourtessis, T. Ratnarajah, S. Chatzinotas, and S. Papavassiliou, "Two-timescale design for active STAR-RIS aided massive MIMO systems," *IEEE Trans. Veh. Tech.*, vol. 73, no. 7, pp. 10118–10134, 2024.
- [44] A. Sabharwal *et al.*, "In-band full-duplex wireless: Challenges and opportunities," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1637–1652, 2014.
- [45] Y. Liu, X. Mu, R. Schober, and H. V. Poor, "Simultaneously transmitting and reflecting (STAR)-RISs: A coupled phase-shift model," in *IEEE ICC*, 2022, pp. 2840–2845.
- [46] D. Neumann, M. Joham, and W. Utschick, "Covariance matrix estimation in massive MIMO," *IEEE Signal Process. Lett.*, vol. 25, no. 6, pp. 863–867, 2018.
- [47] K. Upadhyaya and S. A. Vorobyov, "Covariance matrix estimation for massive MIMO," *IEEE Signal Process. Lett.*, vol. 25, no. 4, pp. 546–550, 2018.
- [48] A. Papazafeiropoulos, "Ergodic capacity of IRS-assisted MIMO systems with correlation and practical phase-shift modeling," *IEEE Wireless Commun. Lett.*, vol. 11, no. 2, pp. 421–425, 2022.
- [49] Z. Zhang, Z. Chen, M. Shen, and B. Xia, "Spectral and energy efficiency of multipair two-way full-duplex relay systems with massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 848–863, 2016.
- [50] A. Shojafard, K.-K. Wong, M. Di Renzo, G. Zheng, K. A. Hamdi, and J. Tang, "Massive MIMO-enabled full-duplex cellular networks," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4734–4750, 2017.
- [51] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends® in Signal Processing*, vol. 11, no. 3–4, pp. 154–655, 2017.
- [52] B. Zheng, C. You, and R. Zhang, "Efficient channel estimation for double-IRS aided multi-user MIMO system," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 3818–3832, 2021.
- [53] Z.-Q. He and X. Yuan, "Cascaded channel estimation for large intelligent metasurface assisted massive MIMO," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 210–214, 2019.
- [54] A. F. Molisch, *Wireless communications*. John Wiley & Sons, 2012.
- [55] D. Bertsekas, *Nonlinear Programming*, ser. Athena scientific optimization and computation series. Athena Scientific, 2016.
- [56] J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," *IMA J. Numerical Analysis*, vol. 8, no. 1, pp. 141–148, 1988.
- [57] J. Hoydis, S. Ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, 2013.
- [58] E. Björnson and L. Sanguinetti, "Rayleigh fading modeling and channel hardening for reconfigurable intelligent surfaces," *IEEE Wireless Commun. Lett.*, vol. 10, no. 4, pp. 830–834, 2020.