Weighted Sum-Rate Maximization for Reconfigurable Intelligent Surface Aided Wireless Networks

Huayan Guo, *Member, IEEE*, Ying-Chang Liang, *Fellow, IEEE*, Jie Chen, *Student Member, IEEE*, and Erik G. Larsson, *Fellow, IEEE*

Abstract—Reconfigurable intelligent surfaces (RIS) is a promising solution to build a programmable wireless environment via steering the incident signal in fully customizable ways with reconfigurable passive elements. In this paper, we consider a RIS-aided multiuser multiple-input single-output (MISO) downlink communication system. Our objective is to maximize the weighted sum-rate (WSR) of all users by joint designing the beamforming at the access point (AP) and the phase vector of the RIS elements, while both the perfect channel state information (CSI) setup and the imperfect CSI setup are investigated. For perfect CSI setup, a low-complexity algorithm is proposed to obtain the stationary solution for the joint design problem by utilizing the fractional programming technique. Then, we resort to the stochastic successive convex approximation technique and extend the proposed algorithm to the scenario wherein the CSI is imperfect. The validity of the proposed methods is confirmed by numerical results. In particular, the proposed algorithm performs quite well when the channel uncertainty is smaller than 10%.

Index Terms—Reconfigurable intelligent surfaces (RIS), passive radio, multiple-input-multiple-output (MIMO), fractional programming, stochastic successive convex approximation.

I. INTRODUCTION

Reconfigurable intelligent surface (RIS), also known as intelligent reflection surface, is an artificial structure consisting of passive radio elements, each of which could adjust the reflection of the incident electromagnetic waves with unnatural properties [1]–[7]. Moreover, owing to the passive structure,
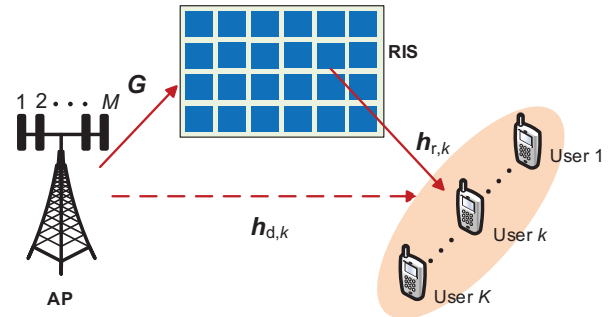
Fig. 1. The RIS-aided multiuser MISO communication system.

the power consumption is extremely low, and there is nearly no additional thermal noise added during reflecting. As a result, the RIS attracts more and more attentions in academia and industry with vast application prospect, e.g., wireless power transfer [8], [9], physical layer security [10]–[12], cognitive radio network [13], etc. Among them, one of the most promising applications is to improve the quality-of-service of users in the wireless communication system suffering from unfavorable propagation conditions [14]–[18].

In this paper, we investigate the RIS-aided *multiple-input single-output* (MISO) multiuser downlink communication system as shown in Fig. 1, in which a multi-antenna *access point* (AP) serves multiple single-antenna mobile users. The direct links between the AP and the mobile users may suffer from deep fading and shadowing, and the RIS improves the propagation conditions by providing high-quality virtual links from the AP to the users. While RIS resembles a full-duplex amplify-and-forward relay [19], [20], it forwards the RF signals via passive reflection, and thus has advantages in both energy- and cost- efficiency. The objective of this paper is to maximize the *weighted sum-rate* (WSR) of the mobile users by jointly optimizing the beamforming at the AP and the phase coefficients of the RIS elements.

A. Related Works

The system in this paper has already investigated by some early-attempt works, in which different objectives are considered while most works assume that the perfect *channel state information* (CSI) of all involved channels is available. In [21] and [22], the transmit power of the AP is minimized by decomposing the joint optimization problem into two subproblems:

one is the conventional power-minimization problem in MIMO system, and the other is for the RIS phase vector optimization. Then the phase optimization problem is solved via *semidefinite relaxation* (SDR) technique. Although this alternating optimization approach achieves quite good performance, the main shortcoming is that the proposed algorithm cannot obtain the stationary solution, and the complexity is a little high especially for large-size RIS. In [23] and [24], the energy efficiency is maximized, while employing *zero-forcing* (ZF) precoding at the AP. Since the ZF precoding completely cancels the inter-user interference, the power allocation at the AP and the phase optimization at RIS can be well decoupled. However, the ZF precoding may as well amplify the background noise, and the performance may be severely compromised when the channel is ill-conditioned. Unfortunately, the derivations in [24] are not applicable directly for other precoding schemes.

Another important issue for the RIS-aided system is the channel estimation. It is known from [21]–[24] that, to optimize the phase vector of the RIS, the system needs high-accuracy CSI about the AP-RIS channel and the RIS-user channels, respectively. However, to obtain perfect CSI is not always possible, since the RIS is passive without channel sensing capability in typical setup. This challenge has been addressed by [25] and [26] via exploiting statistical CSI. Specifically, the single user system is investigated in [25], and the average received SNR is maximized while assuming that the *line-of-sight* (LoS) component of the channel is known. In [26], multiuser system is considered in which all the users are located in the same cluster whose spatial correlation relation is known by the system. Then, the max-min fairness problem is investigated by means of large dimensional random matrix theory. However, the performance of these methods in [25] and [26] depends heavily on the channel model assumptions, as well as the objective functions investigated.

B. Contributions

In this paper, we first assume that perfect CSI is available to study the ultimate performance of the RIS-aided system. The formulated problem looks mathematically similar to the WSR maximization problem in the hybrid digital/analog precoding [27]–[29] in massive MIMO systems. Nevertheless, the main difference is that, the RIS can only control and optimize the behavior of the wireless environment, and has no capability to suppress inter-user interference. Due to that, the beamforming design at the AP and phase optimization at the RIS are deeply coupled, and the convergence speed of the alternating optimization approach is slow. Therefore, the computational complexity in each iteration step should be low and scalable to the number of RIS elements. To tackle this issue, we decompose the original problem into four disjoint blocks by utilizing the *fractional programming* (FP) technique [30]. Subsequently, low-complexity algorithm is designed based on the non-convex *block coordinate descent* (BCD) method [31].

Then, we address the imperfect CSI issue. Specifically, we assume that the AP may perfectly know the combined channel for beamforming design, since this knowledge can be obtained via conventional channel estimation method and protocol, and

the antenna number of the AP may be not huge in the femtocell network [14]. However, the system only has partial knowledge about the channels related to the RIS phase optimization. Hence, we design the phase of the RIS to maximize the average WSR for the incoming channel realizations. This problem formulation is more general than those in [25] and [26], since it is independent to the structure of channel model assumptions. Besides, although similar problem formulation can be found in the hybrid precoding problem [32]–[35], the coupled optimization variables here make this problem much more complicated to be solved. Fortunately, we show that the proposed algorithm for perfect CSI cases can be extended to the imperfect CSI setup, by utilizing the recently proposed stochastic *successive convex approximation* (SCA) technique [36], [37].

The main contributions of this work are summarized as follows:

- Firstly, this paper is one of the early attempts to study the WSR maximization problem for the RIS-aided multiuser downlink MISO system, and both the perfect and imperfect CSI setups are investigated.
- Secondly, for perfect CSI setup, a BCD based method is proposed to carry out the stationary solution for the joint beamforming design and RIS phase optimization problem. The complexity of the proposed algorithm is much lower than the conventional approach.
- Finally, the proposed algorithm is extended to the imperfect CSI cases. Numerical results verify that the proposed algorithm may perform well, when the channel uncertainty is smaller than 10%.

C. Organization and Notations

The rest of the paper is organized as follows. Section II outlines the system model and formulates the joint optimization problem. In Section III, conventional alternating optimization approach is presented for the joint optimization problem under perfect CSI setup. Then, in Section IV, a low-complexity algorithm is proposed based on the non-convex BCD technique. Next, in Section V, the proposed algorithm is extended to the imperfect CSI setup. Simulation results are provided in Section VI to verify the effectiveness of the proposed algorithms, and Section VII concludes the paper.

The notations used in this paper are listed as follows. $\mathbb{E}[\cdot]$ denotes statistical expectation. $\mathcal{CN}(\mu, \sigma^2)$ denotes the *circularly symmetric complex Gaussian* (CSCG) distribution with mean μ and variance σ^2. $\mathbf{I}_M$ denotes the $M \times M$ identity matrix. For any general matrix $\mathbf{G}$, $g_{i,j}$ is the i-th row and j-th column element. $\mathbf{G}^*$, $\mathbf{G}^T$ and $\mathbf{G}^H$ denote the conjugate, the transpose and the conjugate transpose of $\mathbf{G}$, respectively. For any vector $\mathbf{w}$ (all vectors in this paper are column vectors), w_i is the i-th element, and $\|\mathbf{w}\|$ and $\|\mathbf{w}\|_F$ denotes the Euclidean norm and the Frobenius norm, respectively. $|x|$ denotes the absolute value of a complex number x, and $\text{Re}\{x\}$ is its real part.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. Channel Model

This paper investigates a RIS-aided multiuser MISO communication system, using modeling that substantially follows [22]. As shown in Fig. 1, the system consists of one AP equipped with M antennas, one RIS which has N reflection elements, and K single-antenna users. We assume that all the channels experience quasi-static flat-fading. The baseband equivalent channels from AP to user k, from AP to RIS, and from RIS to user k are denoted by $\mathbf{h}_{\mathrm{d},k} \in \mathbb{C}^{M \times 1}$, $\mathbf{G} \in \mathbb{C}^{N \times M}$, and $\mathbf{h}_{\mathrm{r},k} \in \mathbb{C}^{N \times 1}$, respectively. The phase-shift matrix is defined as a diagonal matrix $\boldsymbol{\Theta} = \mathrm{diag}(\theta_1, \cdots, \theta_n, \cdots, \theta_N)$, where $\theta_n = e^{j\varphi_n}$ is the phase of the n-th reflection element on RIS. The reflection operation on the n-th RIS element resembles multiplying the incident signals with θ_n, and then forwarding these composite signals as if from a point source.

Denote the transmit data symbol to user k by s_k, which is independent random variables with zero mean and unit variance. Then, the transmitted signal at the AP can be expressed as

$$\mathbf{x} = \sum_{k=1}^{K} \mathbf{w}_k s_k,$$

where $\mathbf{w}_k \in \mathbb{C}^{M \times 1}$ is the corresponding transmit beamforming vector.

The signal received at user k is expressed as

$$y_k = \underbrace{\mathbf{h}_{\mathrm{d},k}^{\mathrm{H}}\mathbf{x}}_{\text{Direct link}} + \underbrace{\mathbf{h}_{\mathrm{r},k}^{\mathrm{H}}\boldsymbol{\Theta}\mathbf{G}\mathbf{x}}_{\text{RIS-aided link}} + u_k$$
$$= \left(\mathbf{h}_{\mathrm{d},k}^{\mathrm{H}} + \mathbf{h}_{\mathrm{r},k}^{\mathrm{H}}\boldsymbol{\Theta}\mathbf{G}\right) \sum_{k=1}^{K} \mathbf{w}_k s_k + u_k,$$

where $u_k \sim \mathcal{CN}(0, \sigma_0^2)$ denotes the *additive white Gaussian noise* (AWGN) at the k-th user receiver. To make above expression more tractable, we further define $\boldsymbol{\theta} = [\theta_1, \cdots, \theta_N]^{\mathrm{H}}$ and $\mathbf{H}_{\mathrm{r},k} = \mathrm{diag}(\mathbf{h}_{\mathrm{r},k}^{\mathrm{H}})\mathbf{G} \in \mathbb{C}^{N \times M}$, and then the received signal y_k is equivalently represented as

$$y_k = \left(\mathbf{h}_{\mathrm{d},k}^{\mathrm{H}} + \boldsymbol{\theta}^{\mathrm{H}}\mathbf{H}_{\mathrm{r},k}\right) \sum_{k=1}^{K} \mathbf{w}_k s_k + u_k. \qquad (1)$$

The k-th user treats all the signals from other users (i.e., $s_1, \cdots, s_{k-1}, s_{k+1}, \cdots, s_K$) as interference. Hence, the decoding SINR of s_k at user k is

$$\gamma_k = \frac{\left|(\mathbf{h}_{\mathrm{d},k}^{\mathrm{H}} + \boldsymbol{\theta}^{\mathrm{H}}\mathbf{H}_{\mathrm{r},k})\mathbf{w}_k\right|^2}{\sum_{i=1,i\neq k}^{K} \left|(\mathbf{h}_{\mathrm{d},k}^{\mathrm{H}} + \boldsymbol{\theta}^{\mathrm{H}}\mathbf{H}_{\mathrm{r},k})\mathbf{w}_i\right|^2 + \sigma_0^2}. \qquad (2)$$

In addition, the transmit power constraint of AP is

$$\sum_{k=1}^{K} \|\mathbf{w}_k\|^2 \leq P_{\mathrm{T}}. \qquad (3)$$

B. Discussion on Channel Estimation

Generally, to optimize the phase vector $\boldsymbol{\theta}$, the RIS-aided system should estimate $\mathbf{h}_{\mathrm{d},k}$, $\mathbf{G}$, and $\mathbf{h}_{\mathrm{r},k}$, respectively. However, to obtain the high-accurate CSI is a key challenging problem, since the dimension of $\mathbf{h}_{\mathrm{d},k}$ and $\mathbf{G}$ grows linearly with N. In existing literature, there are three main methods to estimate the CSI:

- *Brute-Force Method*: In [3], a brute-force method is proposed, in which the CSI with respect to each RIS element is estimated sequentially by the AP while turning off other elements. The main drawback of this method is that the training overhead is huge (proportional to N). Then, in [38], this method is modified by grouping the adjacent elements to reduce the training overhead.
- *Compressive-Sensing Method*: In [39], a compressive-sensing based method is proposed by exploiting the low-rank property of the RIS-aided link to further reduce the training overhead.
- *Semi-passive RIS*: In [40], a semi-passive structure is suggested by integrating active elements on the RIS which have the channel estimation capability. Then, it is shown that the training overhead may become negligible by exploiting deep learning and compressive sensing tools.

C. Problem Formulation

In this paper, our objective is to maximize the WSR of all the users by jointly designing the transmit beamforming at the AP and the phase vector at RIS, subject to the transmit power constraint in (3). In addition, two setups are investigated for different assumptions on the CSI.

1) Perfect CSI: We first consider an ideal setup in which the CSI of all channels involved is perfectly known. The algorithms proposed for this setup may serve as a benchmark to study the ultimate performance of the system, as well as providing training labels for the machine learning based designs [41], [42].

Let $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_K] \in \mathbb{C}^{M \times K}$. The WSR maximization problem is formulated as

$$\mathcal{P}(\mathrm{A}) \quad \max_{\mathbf{W},\boldsymbol{\theta}} \quad f_{\mathrm{A}}(\mathbf{W}, \boldsymbol{\theta}) = \sum_{k=1}^{K} \omega_k \log(1 + \gamma_k)$$

$$\text{s.t.} \quad |\theta_n| = 1, \quad \forall n = 1, \cdots, N, \qquad (4\mathrm{a})$$

$$\sum_{k=1}^{K} \|\mathbf{w}_k\|^2 \leq P_{\mathrm{T}}, \qquad (4\mathrm{b})$$

where the weight ω_k is used to represent the priority of user k. Since the optimal solution is irrelevant to the base of the logarithm function, we use the natural logarithm throughout the paper.

Despite the conciseness of $\mathcal{P}(\mathrm{A})$, the joint beamforming and phase optimization problem is generally much more difficult than the power minimization problem in [22], and the ZF transmission based design in [24], since the optimization variables $\mathbf{W}$ and $\boldsymbol{\theta}$ are deeply coupled in the non-convex objective function. In addition, as N is usually large in practice, we
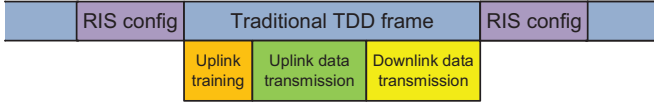
Fig. 2. The TDD based transmission frame structure for the RIS-aided communication system.

prefer an algorithm with lower complexity which is scalable to N, while the complexity of the SDR technique adopted by [22] is $\mathcal{O}(N^6)$ which is a little high.

2) Imperfect CSI: We consider the TDD based transmission frame structure for the RIS-aided communication system as illustrated in Fig. 2. Specially, one time slot for RIS configuration is inserted between two traditional TDD transmission frames. After $\boldsymbol{\theta}$ is configured, the rest system design is totally the same as the traditional communication system.

As such, to design $\mathbf{W}$, the AP only requires the knowledge of the combined channel information as follows:

$$\mathbf{h}_k = \mathbf{h}_{\mathrm{d},k} + \mathbf{H}_{\mathrm{r},k}^{\mathrm{H}}\boldsymbol{\theta}, \tag{5}$$

whose dimension is M (irrelevant to N). We still assume that $\mathbf{h}_k$ for all users are perfectly known. This is reasonable as the antenna number M at the AP in a femtocell network is usually not large, e.g., the AP in the smart-home application scenario is usually equipped with 2 to 8 transmit antennas.

Therefore, in this setup, the key task is optimizing $\boldsymbol{\theta}$. However, in order to configure $\boldsymbol{\theta}$, the channel coefficients $\mathbf{h}_{\mathrm{d},k}$, $\mathbf{G}$, and $\mathbf{h}_{\mathrm{r},k}$ should be estimated separately. According to the references in Section II-B, these channel coefficients are difficult to be estimated in high accuracy with short training pilots. In addition, the pilots and data symbols for channel estimation all come from the previous uplink-transmission slots, which also introduces estimation error. Hence, denote the imperfect channel estimations by $\hat{\mathbf{h}}_{\mathrm{d},k}$, $\hat{\mathbf{G}}$ and $\hat{\mathbf{h}}_{\mathrm{r},k}$, respectively. The estimation error could be expressed by [43]–[45]:

$$\mathbf{z}_{\mathrm{d},k} = \mathbf{h}_{\mathrm{d},k} - \hat{\mathbf{h}}_{\mathrm{d},k},$$
$$\mathbf{Z}_{\mathrm{G}} = \mathbf{G} - \hat{\mathbf{G}},$$
$$\mathbf{z}_{\mathrm{r},k} = \mathbf{h}_{\mathrm{r},k} - \hat{\mathbf{h}}_{\mathrm{r},k}.$$

If minimum *mean square error* (MSE) estimation is applied, the error $\mathbf{z}_{\mathrm{d},k}$, $\mathbf{Z}_{\mathrm{G}}$, and $\mathbf{z}_{\mathrm{r},k}$ are uncorrelated with the estimated channel coefficients $\hat{\mathbf{h}}_{\mathrm{d},k}$, $\hat{\mathbf{G}}$ and $\hat{\mathbf{h}}_{\mathrm{r},k}$ [43]–[45]. Then, the true channel coefficients $\mathbf{h}_{\mathrm{d},k}$, $\mathbf{G}$ and $\mathbf{h}_{\mathrm{r},k}$ in the incoming data transmission frame can be modeled as a realization from the sample space $\mathcal{F} \triangleq \{\mathbf{h}_{\mathrm{d},k}(\xi), \mathbf{G}(\xi), \mathbf{h}_{\mathrm{r},k}(\xi), \forall k, \forall \xi\}$ dominated by the knowledge of the imperfect CSI ($\hat{\mathbf{h}}_{\mathrm{d},k}$, $\hat{\mathbf{G}}$ and $\hat{\mathbf{h}}_{\mathrm{r},k}$) and the distribution of the channel estimation error ($\mathbf{z}_{\mathrm{d},k}$, $\mathbf{Z}_{\mathrm{G}}$, and $\mathbf{z}_{\mathrm{r},k}$), where ξ denotes the index of the random realizations drawn from $\mathcal{F}$.[1]

In this setup, the optimization problem is formulated as

---

[1]If the CSI is perfectly estimated, $\mathbf{h}_{\mathrm{d},k}(\xi)$, $\mathbf{G}(\xi)$, and $\mathbf{h}_{\mathrm{r},k}(\xi)$ keep constants for different ξ.

maximizing the expectation of the achievable WSR:

$$\mathcal{P}(\mathrm{B}) \quad \max_{\boldsymbol{\theta}} \quad f_{\mathrm{B}}(\boldsymbol{\theta}) = \mathbb{E}_\xi\left[\max_{\boldsymbol{W}(\xi)} f_{\mathrm{A}}(\mathbf{W}(\xi), \boldsymbol{\theta}; \xi)\right]$$

$$\mathbf{s.t.} \quad |\theta_n| = 1, \quad \forall n = 1, \cdots, N, \tag{6a}$$

$$\sum_{k=1}^K \|\mathbf{w}_k(\xi)\|^2 \leq P_{\mathrm{T}}, \quad \forall \xi. \tag{6b}$$

Compared with existing works, the advantage of this setup is that, the RIS configuration only requires very small change from tradition wireless system. However, $\mathcal{P}(\mathrm{B})$ is the stochastic optimization problem with inner-layer variable $\boldsymbol{W}(\xi)$ and outer-layer variable $\boldsymbol{\theta}$, and both the inner-layer and outer-layer subproblems are non-convex with no closed-form solutions. In addition, the objective function contains expectation operator, and the probability density function of the sample space $\mathcal{F}$ usually very complicated with no closed-form expression as well. Therefore, designing algorithm to solve $\mathcal{P}(\mathrm{B})$ is a really challenging task.

III. ALTERNATING OPTIMIZATION FOR THE PERFECT CSI SETUP

The alternating optimization approach is the two-block version of the BCD method, the basic idea of which is to decompose the optimization variables into several blocks, and then each block is updated following some specific rules while fixing the remaining blocks at their last updated values [46]. In existing work [22] and [24], the alternating optimization approach has been commonly used to address the joint optimization problem in the RIS-aided system. In particular, the joint optimization problem is decomposed into two subproblems: one is the conventional beamforming design problem at the AP, and the other is the phase optimization problem given optimized beamforming vectors.

A. Algorithm Description

When $\boldsymbol{\theta}$ is fixed, the subproblem to optimize $\mathbf{W}$ is reduced to the WSR maximization problem for the conventional multiuser MISO system. This problem has been studied extensively in the literature, and one famous method to obtain the stationary solution is the WMMSE algorithm with the following iterative updating rule [47]:

$$\chi_k = \left(\sum_{i=1}^K \left|\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_i\right|^2 + \sigma_0^2\right)^{-1} \mathbf{h}_k^{\mathrm{H}}\mathbf{w}_k, \tag{7a}$$

$$\kappa_k = \left(1 - \chi_k^*\mathbf{h}_k^{\mathrm{H}}\mathbf{w}_k\right)^{-1}, \tag{7b}$$

$$\mathbf{w}_k = \omega_k \chi_k \kappa_k \left(\lambda\mathbf{I}_M + \sum_{i=1}^K \omega_i|\chi_i|^2\kappa_i\mathbf{h}_i\mathbf{h}_i^{\mathrm{H}}\right)^{-1} \mathbf{h}_k, \tag{7c}$$

where $\lambda \geq 0$ is the optimal dual variable for the transmit power constraint.

Then one can focus on the phase optimization subproblem. For ease of representation, define the effective channels for the direct link and the RIS link as follows

$$\mathbf{a}_{i,k} = \mathbf{H}_{\mathrm{r},k}\mathbf{w}_i, \tag{8a}$$

$$b_{i,k} = \mathbf{h}_{\mathrm{d},k}^{\mathrm{H}}\mathbf{w}_i, \tag{8b}$$

respectively. Then the phase optimization subproblem is represented as

$$\mathcal{P}(\mathrm{C}) \quad \max_{\boldsymbol{\theta}} \quad f_{\mathrm{C}}(\boldsymbol{\theta})$$
$$\mathbf{s.t.} \quad |\theta_n| = 1, \quad \forall n = 1, \cdots, N,$$

where

$$f_{\mathrm{C}}(\boldsymbol{\theta}) = \sum_{k=1}^{K} \omega_k \log\left(1 + \frac{\left|\boldsymbol{\theta}^{\mathrm{H}}\mathbf{a}_{k,k} + b_{k,k}\right|^2}{\sum_{i \neq k} \left|\boldsymbol{\theta}^{\mathrm{H}}\mathbf{a}_{i,k} + b_{i,k}\right|^2 + \sigma_0^2}\right).$$

One can see that, $f_{\mathrm{C}}(\boldsymbol{\theta})$ is continuous and differentiable, and the constraint sets of $\boldsymbol{\theta}$ forms a complex circle manifold. Thus, the stationary solution of $\mathcal{P}(\mathrm{C})$ can be obtained via the *Riemannian conjugate gradient* (RCG) algorithm [48], which has been widely applied for the analog precoder design in hybrid precoding problem [27] and has also shown good performance in the single-user RIS-aided MISO system [49]. Conceptually, the RCG algorithm has three key steps in each iteration:

1) Compute Riemannian Gradient: The Riemannian gradient is the orthogonal projection of the Euclidean gradient ∇f_{C} onto the complex circle:

$$\mathrm{grad} f_{\mathrm{C}} = \nabla f_{\mathrm{C}} - \mathrm{Re}\left\{\nabla f_{\mathrm{C}} \circ \boldsymbol{\theta}^*\right\} \circ \boldsymbol{\theta},$$

where the Euclidean gradient is

$$\nabla f_{\mathrm{C}} = \sum_{k=1}^{K} 2\omega_k \mathbf{A}_k,$$

with parameters

$$\mathbf{A}_k = \frac{\sum_i \mathbf{a}_{i,k}\mathbf{a}_{i,k}^{\mathrm{H}}\boldsymbol{\theta} + \sum_i \mathbf{a}_{i,k}b_{i,k}^*}{\sum_i \left|\boldsymbol{\theta}^{\mathrm{H}}\mathbf{a}_{i,k} + b_{i,k}\right|^2 + \sigma_0^2}$$
$$- \frac{\sum_{i \neq k} \mathbf{a}_{i,k}\mathbf{a}_{i,k}^{\mathrm{H}}\boldsymbol{\theta} + \sum_{i \neq k} \mathbf{a}_{i,k}b_{i,k}^*}{\sum_{i \neq k} \left|\boldsymbol{\theta}^{\mathrm{H}}\mathbf{a}_{i,k} + b_{i,k}\right|^2 + \sigma_0^2}.$$

2) Search Direction: One may find the tangent vector conjugate to $\mathrm{grad} f_{\mathrm{C}}$ as the search direction:

$$\mathbf{d} = -\mathrm{grad} f_{\mathrm{C}} + \tau_1 \mathcal{T}(\bar{\mathbf{d}}),$$

where $\mathcal{T}(\cdot)$ is the vector transport function defined as

$$\mathcal{T}(\mathbf{d}) = \bar{\mathbf{d}} - \mathrm{Re}\left\{\mathbf{d} \circ \boldsymbol{\theta}^*\right\} \circ \boldsymbol{\theta},$$

τ_1 is the conjugate gradient update parameter, and $\bar{\mathbf{d}}$ is the previous search direction.

3) Retraction: Project the tangent vector back to the complex circle manifold

$$\boldsymbol{\theta}_n \leftarrow \frac{(\boldsymbol{\theta} + \tau_2 \mathbf{d})_n}{|(\boldsymbol{\theta} + \tau_2 \mathbf{d})_n|},$$

where τ_2 is the Armijo step size.

B. Discussion

The alternating optimization approach in Section III-A is actually a multi-stage iterative optimization algorithm. The outer loop involves two subproblems for optimizing $\mathbf{W}$ and $\boldsymbol{\theta}$, respectively, and each subproblem still requires iterative updating method to solve.

Specifically, the WMMSE algorithm requires the matrix inverse operations in all the three updating steps with complexity $\mathcal{O}(KM^3)$. In addition, in (7c), one dimensional search (usually bi-search) for λ is needed. Hence, the complexity of the WMMSE algorithm is $\mathcal{O}(I_\lambda I_{\mathrm{w}} KM^3)$, where I_λ and I_{w} are the iteration numbers of searching λ and the three-step updating loop, respectively. Besides, the complexity of the RCG algorithm is dominated by computing the Euclidean gradient, which is $\mathcal{O}(K^2 N^2)$. The retraction step also requires iteratively searching τ_2, but fortunately the complexity is only $\mathcal{O}(K^2 N)$ and can be ignored when N is large. Therefore, the total complexity of the alternating optimization approach is $\mathcal{O}\left(I_{\mathrm{O}}\left(I_\lambda I_{\mathrm{w}} KM^3 + I_{\mathrm{R}} K^2 N^2\right)\right)$, where I_{O} and I_{R} denote the iteration times of the outer loop, and the iteration times of the inner RCG algorithm, respectively.

The idea of alternatingly updating the variables is quite straightforward, and this method generally has good performance as verified by the simulations in [22] and [24]. However, this approach has two main drawbacks for the WSR problem in this paper:

- For the WSR maximization problem in the RIS-aided system, the improvement of the objective function is mainly obtained by suppressing inter-user interference and power allocation among users, both of which are beyond the capabilities of the RIS. Thus, the beamforming design $\mathbf{W}$ and the phase vector $\boldsymbol{\theta}$ are deeply coupled. As a result, the alternating optimization converges slowly, and its complexity becomes unacceptable when both the two subproblems also require iterative method to solve. In addition, for the iterative method on the phase optimization subproblem, the precision of the output $\boldsymbol{\theta}$ should be high enough (requiring large I_{R}) to prevent the alternating optimization approach stopping at an uninteresting point.
- If the two subproblems are solved independently, it is difficult to extend the algorithm to the imperfect CSI setup, where the optimization of $\mathbf{W}$ and $\boldsymbol{\theta}$ requires a coordinated design.

Therefore, a new algorithm designing with lower complexity and better extendibility is still necessary for $\mathcal{P}(\mathrm{A})$.

IV. LOW-COMPLEXITY BCD FOR THE PERFECT CSI SETUP

In this section, we design new algorithm for $\mathcal{P}(\mathrm{A})$. To be specific, we first apply the closed-form FP approach in [30] to equivalently translate the sum-of-logarithms-of-ratio problem $\mathcal{P}(\mathrm{A})$ into a more tractable form $\mathcal{P}(\mathrm{A1})$. Then, we decompose $\mathcal{P}(\mathrm{A1})$ into four disjoint blocks. Non-convex BCD method [31] is exploited to carry out the stationary solution for $\mathcal{P}(\mathrm{A1})$. Specially, low-complexity updating rules for block $\mathbf{W}$ and block $\boldsymbol{\theta}$ based on the the prox-linear BCD update rule [46] and the SCA method [31], respectively.

A. Closed-Form FP Approach

The closed-form FP approach was proposed in [30] to deal with the sum-of-logarithms-of-ratio problem as follows:

$$\max_{\mathbf{x}} \quad \sum_{k=1}^{K} \log(1 + \frac{|A_k(\mathbf{x})|^2}{B_k(\mathbf{x}) - |A_k(\mathbf{x})|^2}),$$

where $B_k(\mathbf{x}) > |A_k(\mathbf{x})|^2$ for all k. Conceptually, the closed-form FP approach has two key steps:

1) Lagrangian Dual Transform: By introducing an auxiliary variable α_k, the logarithm function can be tackled based on the following equation:

$$\log(1 + \gamma_k) = \max_{\alpha_k \geq 0} \log(1 + \alpha_k) - \alpha_k + \frac{(1 + \alpha_k)\gamma_k}{1 + \gamma_k}. \quad (9)$$

Then, the original problem is equivalently transformed to

$$\max_{\mathbf{x},\boldsymbol{\alpha}} \quad \sum_{k=1}^{K} \left(\log(1 + \alpha_k) - \alpha_k + (1 + \alpha_k)\frac{|A_k(\mathbf{x})|^2}{B_k(\mathbf{x})} \right)$$

$$\text{s.t.} \quad \alpha_k \geq 0, \quad \forall k = 1, \cdots, K, \quad (10)$$

where $\boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_K]^{\mathrm{T}}$.

2) Quadratic Transform: Given $\boldsymbol{\alpha}$, one may focus on the following sum-of-ratios problem

$$\max_{\mathbf{x}} \quad \sum_{k=1}^{K} \frac{|A_k(\mathbf{x})|^2}{B_k(\mathbf{x})}.$$

The key idea is introducing auxiliary variables $\boldsymbol{\beta} = [\beta_1, \cdots, \beta_K]^{\mathrm{T}}$, and then the above problem is equivalently translated to

$$\max_{\mathbf{x},\boldsymbol{\beta}} \quad \sum_{k=1}^{K} \left(2\mathrm{Re}\left\{ \beta_k^* A_k(\mathbf{x}) \right\} - |\beta_k|^2 B_k(\mathbf{x}) \right).$$

The equivalence can be verified by substituting $\beta_k = \frac{A_k(\mathbf{x})}{B_k(\mathbf{x})}$ into above problem.

B. Non-Convex BCD

Applying the closed-form FP approach introduced above, problem $\mathcal{P}(\mathrm{A})$ is equivalent to following problem

$$\mathcal{P}(\mathrm{A1}) \quad \max_{\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{W},\boldsymbol{\theta}} \quad f_{\mathrm{A1}}(\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{W},\boldsymbol{\theta})$$

$$\text{s.t.} \quad (4a), (4b), (10),$$

where the new objective function is

$$f_{\mathrm{A1}}(\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{W},\boldsymbol{\theta}) = \sum_{k=1}^{K} \omega_k \left(\log(1 + \alpha_k) - \alpha_k \right)$$

$$+ \sum_{k=1}^{K} 2\sqrt{\omega_k(1 + \alpha_k)} \mathrm{Re}\left\{ \beta_k^*(\mathbf{h}_{\mathrm{d},k}^{\mathrm{H}} + \boldsymbol{\theta}^{\mathrm{H}}\mathbf{H}_{\mathrm{r},k})\mathbf{w}_k \right\}$$

$$- \sum_{k=1}^{K} |\beta_k|^2 \left(\sum_{i=1}^{K} \left| (\mathbf{h}_{\mathrm{d},k}^{\mathrm{H}} + \boldsymbol{\theta}^{\mathrm{H}}\mathbf{H}_{\mathrm{r},k})\mathbf{w}_i \right|^2 + \sigma_0^2 \right).$$

In this paper, we adopt the BCD method [46] to decompose the optimization variables $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\mathbf{W}$, and $\boldsymbol{\theta}$, and aim to get a stationary solution for $\mathcal{P}(\mathrm{A1})$.

The BCD is an iterative method, where $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\mathbf{W}$, and $\boldsymbol{\theta}$ are cyclically updated. To be specific, denote by $\bar{\boldsymbol{\alpha}}$, $\bar{\boldsymbol{\beta}}$, $\bar{\mathbf{W}}$, and $\bar{\boldsymbol{\theta}}$ the temporal optimization results in last iteration. Then, it is easy to carry out

$$\alpha_k = \frac{\bar{\zeta}_k^2 + \bar{\zeta}_k\sqrt{\bar{\zeta}_k^2 + 4}}{2}, \quad (11)$$

$$\beta_k = \frac{\sqrt{\omega_k(1 + \bar{\alpha}_k)}(\mathbf{h}_{\mathrm{d},k}^{\mathrm{H}} + \bar{\boldsymbol{\theta}}^{\mathrm{H}}\mathbf{H}_{\mathrm{r},k})\bar{\mathbf{w}}_k}{\sum_{i=1}^{K} \left| (\mathbf{h}_{\mathrm{d},k}^{\mathrm{H}} + \bar{\boldsymbol{\theta}}^{\mathrm{H}}\mathbf{H}_{\mathrm{r},k})\bar{\mathbf{w}}_i \right|^2 + \sigma_0^2}, \quad (12)$$

where $\bar{\zeta}_k = \frac{1}{\sqrt{\omega_k}}\mathrm{Re}\left\{ \bar{\beta}_k^* \bar{\mathbf{h}}_k^{\mathrm{H}}\bar{\mathbf{w}}_k \right\}$ and $\bar{\mathbf{h}}_k$ the combined channel:

$$\bar{\mathbf{h}}_k = \mathbf{h}_{\mathrm{d},k} + \mathbf{H}_{\mathrm{r},k}^{\mathrm{H}}\bar{\boldsymbol{\theta}}.$$

Remark that, in [30], it is suggested updating α_k by the temporal SINR which is not BCD. Nevertheless, its convergence is established as well [30, Appendix A].

C. Prox-linear Update for $\mathbf{W}$

One may update $\mathbf{W}$ by solving following problem:

$$\mathbf{W} = \arg\max_{\mathbf{W}} \quad f_{\mathrm{A2}}(\mathbf{W}),$$

$$\text{s.t.} \quad \sum \|\mathbf{w}_k\|^2 \leq P_{\mathrm{T}},$$

where $f_{\mathrm{A2}}(\mathbf{W}) = f_{\mathrm{A1}}(\bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\beta}}, \mathbf{W}, \bar{\boldsymbol{\theta}})$, and thus have:

$$\boldsymbol{w}_k = \sqrt{\omega_k(1 + \bar{\alpha}_k)}\bar{\beta}_k \left(\lambda\mathbf{I}_M + \sum_{i=1}^{K} |\bar{\beta}_i|^2 \bar{\mathbf{h}}_i\bar{\mathbf{h}}_i^{\mathrm{H}} \right)^{-1} \bar{\mathbf{h}}_k,$$

where $\lambda \geq 0$ is the optimal dual variable for the transmit power constraint. However, the matrix inverse operation is expensive, and to obtain a high-accurate $\mathbf{W}$, the iteration numbers for searching λ is usually high.

To eliminate the one dimensional search of λ as well as the expensive matrix inverse operation, we apply the prox-linear BCD update rule as follows [46]:

$$\mathbf{W} = \arg\min_{\mathbf{W}} \sum_{k=1}^{K} \left(\mathrm{Re}\left\{ \mathbf{g}_k^{\mathrm{H}}(\mathbf{w}_k - \hat{\mathbf{w}}_k) \right\} + \frac{L}{2}\|\mathbf{w}_k - \hat{\mathbf{w}}_k\|^2 \right)$$

$$\text{s.t.} \quad \sum \|\mathbf{w}_k\|^2 \leq P_{\mathrm{T}},$$

where $L > 0$, the gradient is denoted by

$$\mathbf{g}_k = -\frac{\partial f_{\mathrm{A2}}}{\partial \mathbf{w}_k}\bigg|_{\mathbf{w}_k = \hat{\mathbf{w}}_k}$$

$$= -2\sqrt{\omega_k(1 + \bar{\alpha}_k)}\bar{\beta}_k\bar{\mathbf{h}}_k + 2\sum_{i=1}^{K} |\bar{\beta}_i|^2 \bar{\mathbf{h}}_i\bar{\mathbf{h}}_i^{\mathrm{H}}\hat{\mathbf{w}}_k,$$

$\hat{\mathbf{w}}_k = \bar{\mathbf{w}}_k + \epsilon(\bar{\mathbf{w}}_k - \ddot{\mathbf{w}}_k)$ is the extrapolated point, $\ddot{\mathbf{w}}_k$ is the value of $\mathbf{w}_k$ before it was updated to $\bar{\mathbf{w}}_k$, and $\epsilon \geq 0$ is the extrapolation weight. Then we have a simple update rule:

$$\boldsymbol{w}_k = \frac{1}{L - 2\lambda}\left(L\hat{\mathbf{w}}_k - \mathbf{g}_k \right), \quad (13)$$

$$\lambda = \frac{L}{2} - \frac{1}{2P_{\mathrm{T}}}\sum_{k=1}^{k} \|L\hat{\mathbf{w}}_k - \mathbf{g}_k\|^2. \quad (14)$$

One can see that, the complexity to update $\boldsymbol{W}$ is reduced to $\mathcal{O}(KM^2)$, and no iteration is required.

We set $L = 2\left\|\sum_{i=1}^{K}|\bar{\beta}_i|^2\bar{\mathbf{h}}_i\bar{\mathbf{h}}_i^{\mathrm{H}}\right\|_{\mathrm{F}}$, which is the Lipschitz constant of the gradient $\mathbf{g}_k$. Then, the extrapolation weight is taken by

$$\epsilon = \min\left(\frac{d-1}{\bar{d}}, 0.9999\sqrt{\frac{\bar{L}}{L}}\right),$$

where $\bar{d}$ and $\bar{L}$ are the values adopted in previous iteration, and d is recursively defined by $d = \frac{1}{2}\left(1 + \sqrt{1+4\bar{d}^2}\right)$ with initial value 1. Since $f_{\mathrm{A2}}(\mathbf{W})$ is strongly convex which satisfies the KL property [46], the convergence of the prox-linear update rule is established (see [46, Lemma 2.2]).

D. Successive Convex Approximation for Updating $\boldsymbol{\theta}$

In conventional BCD method, $\boldsymbol{\theta}$ is updated according to $\boldsymbol{\theta} = \arg\max_{\boldsymbol{\theta}} f_{\mathrm{A1}}(\bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\beta}}, \bar{\mathbf{W}}, \boldsymbol{\theta})$. After dropping irrelevant constant terms, this updating rule is represented as

$$\boldsymbol{\theta} = \arg\min_{\boldsymbol{\theta}} f_{\mathrm{A3}}(\boldsymbol{\theta}) \triangleq \boldsymbol{\theta}^{\mathrm{H}}\boldsymbol{U}\boldsymbol{\theta} - 2\mathrm{Re}\left\{\boldsymbol{\theta}^{\mathrm{H}}\boldsymbol{\nu}\right\}$$
$$\text{s.t.} \quad |\theta_n| = 1, \quad \forall n = 1, \cdots, N.$$

where $\boldsymbol{U}$ and $\boldsymbol{\nu}$ are

$$\boldsymbol{U} = \sum_{k=1}^{K}\left|\bar{\beta}_k\right|^2\sum_{i=1}^{K}\bar{\mathbf{a}}_{i,k}\bar{\mathbf{a}}_{i,k}^{\mathrm{H}}, \tag{15a}$$

$$\boldsymbol{\nu} = \sum_{k=1}^{K}\left(\sqrt{\omega_k(1+\bar{\alpha}_k)}\bar{\beta}_k^*\bar{\mathbf{a}}_{k,k} - |\beta_k|^2\sum_{i=1}^{K}\bar{b}_{i,k}^*\bar{\mathbf{a}}_{i,k}\right), \tag{15b}$$

with $\bar{\mathbf{a}}_{i,k} = \mathbf{H}_{r,k}\bar{\mathbf{w}}_i$, and $\bar{b}_{i,k} = \mathbf{h}_{d,k}^{\mathrm{H}}\bar{\mathbf{w}}_i$. We further replace θ_n by φ_n, where $\theta_n = e^{\jmath\varphi_n}$ and $\varphi_n \in \mathbb{R}$, and then the update rule is recast to

$$\boldsymbol{\varphi} = \arg\min_{\boldsymbol{\varphi}\in\mathbb{R}^N} f_{\mathrm{A4}}(\boldsymbol{\varphi}) \triangleq (e^{\jmath\boldsymbol{\varphi}})^{\mathrm{H}}\boldsymbol{U}e^{\jmath\boldsymbol{\varphi}} - 2\mathrm{Re}\left\{\boldsymbol{\nu}^{\mathrm{H}}e^{\jmath\boldsymbol{\varphi}}\right\}$$

where $\boldsymbol{\varphi} = [\varphi_1, \cdots, \varphi_N]^{\mathrm{T}}$.

However, $f_{\mathrm{A4}}(\boldsymbol{\varphi})$ is non-convex, and its really hard to solve the optimal solution. Fortunately, it is pointed out in [31] that, we only need to solve the following surrogate problem by exploiting the SCA technique, and the BCD method will still converge to a stationary solution. In particular, denote the surrogate function for $f_{\mathrm{A4}}(\boldsymbol{\varphi})$ by $f_{\mathrm{A5}}(\boldsymbol{\varphi}, \bar{\boldsymbol{\varphi}})$, and the output $\boldsymbol{\varphi}$ is obtained from

$$\boldsymbol{\varphi} = \arg\min_{\boldsymbol{\varphi}\in\mathbb{R}^N} f_{\mathrm{A5}}(\boldsymbol{\varphi}, \bar{\boldsymbol{\varphi}}).$$

It is known that $f_{\mathrm{A4}}(\boldsymbol{\varphi})$ continuously differentiable, and thus we need $f_{\mathrm{A5}}(\boldsymbol{\varphi}, \bar{\boldsymbol{\varphi}})$ satisfies following two constraint [31, Proposition 1]:

$$f_{\mathrm{A5}}(\bar{\boldsymbol{\varphi}}, \bar{\boldsymbol{\varphi}}) = f_{\mathrm{A4}}(\bar{\boldsymbol{\varphi}}), \tag{16a}$$
$$f_{\mathrm{A5}}(\boldsymbol{\varphi}, \bar{\boldsymbol{\varphi}}) \geq f_{\mathrm{A4}}(\boldsymbol{\varphi}). \tag{16b}$$

In this paper, we adopt the surrogate function constructed by the second order Taylor expansion [50]:

$$f_{\mathrm{A5}}(\boldsymbol{\varphi}, \bar{\boldsymbol{\varphi}}) = f_{\mathrm{A4}}(\bar{\boldsymbol{\varphi}}) + \nabla f_{\mathrm{A4}}(\bar{\boldsymbol{\varphi}})^{\mathrm{T}}(\boldsymbol{\varphi} - \bar{\boldsymbol{\varphi}}) + \frac{\kappa}{2}\|\boldsymbol{\varphi} - \bar{\boldsymbol{\varphi}}\|^2,$$

Algorithm 1 $[\boldsymbol{U}, \boldsymbol{\nu}] = \mathrm{ParfunA}(\boldsymbol{\varphi}, \bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\beta}}, \bar{\mathbf{W}})$.

1: Update $\boldsymbol{\beta}$ by (12);
2: Update $\boldsymbol{W}$ by (13);
3: Update $\boldsymbol{\alpha}$ by (11);
4: Update $\boldsymbol{\beta}$ based on (12);
5: Update $\boldsymbol{U}$ and $\boldsymbol{\nu}$ by (15a) and (15b), respectively.

---

where $\nabla f_{\mathrm{A4}}(\bar{\boldsymbol{\varphi}}) = 2\mathrm{Re}\left\{-\jmath\bar{\boldsymbol{\theta}}^* \circ \left(\boldsymbol{U}\bar{\boldsymbol{\theta}} - \boldsymbol{\nu}\right)\right\}$ is the gradient, and κ is chosen to satisfy (16b). Finally, $\boldsymbol{\varphi}$ is updated by the minimum value of $f_{\mathrm{A5}}(\boldsymbol{\varphi}, \bar{\boldsymbol{\varphi}})$:

$$\boldsymbol{\varphi} = \bar{\boldsymbol{\varphi}} - \frac{\nabla f_{\mathrm{A4}}(\bar{\boldsymbol{\varphi}})}{\kappa}. \tag{17}$$

E. Algorithm Development

The block selection rule for BCD method is designed as follows

$$\cdots\cdots\boldsymbol{\alpha} \to \boldsymbol{\beta} \to \boldsymbol{\varphi} \to \boldsymbol{\beta} \to \mathbf{W} \to \boldsymbol{\alpha}\cdots\cdots \tag{18}$$

The stationary solution for $\mathcal{P}(\mathrm{A})$ can be carried out by simply setting κ as the Lipschitz constant of ∇f_{A4} [31]. It is known that, the complexity to update $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\mathbf{W}$ are $\mathcal{O}(KNM)$, $\mathcal{O}(KNM)$, and $\mathcal{O}(KM^2)$, respectively. Besides, the complexity of update $\boldsymbol{\varphi}$ is dominated by the parameter $\mathbf{U}$ in (15a), which is $\mathcal{O}(K^2N^2)$. Therefore, the total complexity of the proposed BCD method is $\mathcal{O}\left(I_{\mathrm{O}}\left(2KNM + KM^2 + K^2N^2\right)\right)$, where I_{O} is the number of iterations.

However, the improvement of f_{A3} by updating $\boldsymbol{\varphi}$ is even much smaller than that of f_{C} in the alternating optimization approach in Section III. So the convergence speed is much slower, and the algorithm complexity may not decrease. In the next, we show that the convergence of the proposed BCD algorithm can be accelerated by chosen a proper search step size in (17).

Consider following problem for $\boldsymbol{\varphi}$

$$\mathcal{P}(\mathrm{A6}) \min_{\boldsymbol{\varphi}} f_{\mathrm{A6}}(\boldsymbol{\varphi}) \triangleq (e^{\jmath\boldsymbol{\varphi}})^{\mathrm{H}}\boldsymbol{U}e^{\jmath\boldsymbol{\varphi}} - 2\mathrm{Re}\left\{\boldsymbol{\nu}^{\mathrm{H}}e^{\jmath\boldsymbol{\varphi}}\right\}$$

where $[\boldsymbol{U}, \boldsymbol{\nu}] = \mathrm{ParfunA}(\boldsymbol{\varphi}, \bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\beta}}, \bar{\mathbf{W}})$ which is summarized in Algorithm 1. According to the block selection rule in (18), every stationary solution of $\mathcal{P}(\mathrm{A6})$ is the critical point of the BCD method, which is as well the stationary solution of $\mathcal{P}(\mathrm{A})$.

To solve $\mathcal{P}(\mathrm{A6})$, let's construct function as follows

$$h_{\mathrm{A6}}(\boldsymbol{\varphi}, \bar{\boldsymbol{\varphi}}) = f_{\mathrm{A6}}(\bar{\boldsymbol{\varphi}}) + \nabla f_{\mathrm{A4}}(\bar{\boldsymbol{\varphi}})^{\mathrm{T}}(\boldsymbol{\varphi} - \bar{\boldsymbol{\varphi}}) + \frac{\kappa}{2}\|\boldsymbol{\varphi} - \bar{\boldsymbol{\varphi}}\|^2, \tag{19}$$

where the parameters $\boldsymbol{U}$ and $\boldsymbol{\nu}$ in $\nabla f_{\mathrm{A4}}(\bar{\boldsymbol{\varphi}})$ are determined by $[\boldsymbol{U}, \boldsymbol{\nu}] = \mathrm{ParfunA}(\bar{\boldsymbol{\varphi}}, \bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\beta}}, \bar{\mathbf{W}})$. Then, one can verify that, for sufficient large κ, we always have

$$h_{\mathrm{A6}}(\bar{\boldsymbol{\varphi}}, \bar{\boldsymbol{\varphi}}) = f_{\mathrm{A6}}(\bar{\boldsymbol{\varphi}}),$$
$$h_{\mathrm{A6}}(\boldsymbol{\varphi}, \bar{\boldsymbol{\varphi}}) \geq f_{\mathrm{A6}}(\boldsymbol{\varphi}).$$

It is known that, both f_{A5} and f_{A6} are continuously differentiable functions, and the constraints in (16a) and (16b) are satisfied. Therefore, $h_{\mathrm{A6}}(\boldsymbol{\varphi}, \bar{\boldsymbol{\varphi}})$ is the SCA surrogate function of $f_{\mathrm{A6}}(\boldsymbol{\varphi})$, and we shall have the gradient of $f_{\mathrm{A6}}(\boldsymbol{\varphi})$:

$$\nabla f_{\mathrm{A6}}(\boldsymbol{\varphi}) = \nabla f_{\mathrm{A4}}(\boldsymbol{\varphi})$$
$$= 2\mathrm{Re}\left\{-\jmath e^{-\jmath\boldsymbol{\varphi}} \circ \left(\boldsymbol{U}e^{\jmath\boldsymbol{\varphi}} - \boldsymbol{\nu}\right)\right\}, \tag{20}$$

Algorithm 2 Low-complexity BCD for the perfect CSI setup.

1: Initialize $\boldsymbol{W}^{(0)}$ and $\boldsymbol{\theta}^{(0)}$ to feasible values.
2: Initialize $\boldsymbol{\alpha}^{(0)}$ and $\boldsymbol{\beta}^{(0)}$ by (11) and (12), and set $t = 0$.
   Repeat
3: Design κ using (21) according to $\boldsymbol{\varphi}^{(t)}$, $\boldsymbol{\alpha}^{(t)}$, $\boldsymbol{\beta}^{(t)}$, $\mathbf{W}^{(t)}$;
4: Set $t = t + 1$;
5: Update $\boldsymbol{\varphi}^{(t)}$, $\boldsymbol{\alpha}^{(t)}$, $\boldsymbol{\beta}^{(t)}$, $\mathbf{W}^{(t)}$.
   Until The value of the objective function f_{A1} converges.

Algorithm 3 $[\boldsymbol{U}, \boldsymbol{\nu}] = \mathrm{ParfunB}(\boldsymbol{\varphi}, \boldsymbol{\alpha}^{(0)}, \boldsymbol{\beta}^{(0)}, \mathbf{W}^{(0)}; \xi)$.

Input: Channel realizations $\mathbf{h}_{\mathrm{d},k}(\xi)$, $\mathbf{G}(\xi)$, $\mathbf{h}_{\mathrm{r},k}(\xi)$ for all k;
1: Set $t = 0$.
   Repeat
2: Design κ using (21) according to $\boldsymbol{\varphi}$, $\boldsymbol{\alpha}^{(t)}$, $\boldsymbol{\beta}^{(t)}$, $\mathbf{W}^{(t)}$;
3: Set $t = t + 1$;
4: Update $\boldsymbol{\alpha}^{(t)}$, $\boldsymbol{\beta}^{(t)}$, $\mathbf{W}^{(t)}$.
   Until The value of the objective function f_{A1} converges.
5: Update $\boldsymbol{U}$ and $\boldsymbol{\nu}$ by (15a) and (15b), respectively.

---

since both f_{A5} and f_{A6} are continuously differentiable functions (see [31, Proposition 1]).

From (20), the update rule in (17) has the same formation as the gradient projection algorithm for $\mathcal{P}(\mathrm{A6})$. We design the step size κ can be determined by the Armijo rule [51]:

$$f_{\mathrm{A6}}(\bar{\boldsymbol{\varphi}}) - f_{\mathrm{A6}}(\boldsymbol{\varphi}) \geq \zeta\kappa\|\nabla f_{\mathrm{A4}}(\bar{\boldsymbol{\varphi}})\|_2^2, \qquad (21)$$

where $0 < \zeta < 0.5$, κ is the largest element in $\{\kappa_0^{-j}\}_{j=0,1,\cdots}$ and $\kappa_0 > 1$.

The proposed BCD algorithm above is summarized in Algorithm 2. Denote by I_{A} the iteration number of the Armijo search. The complexity of the proposed algorithm is $\mathcal{O}\left(I_{\mathrm{O}}\left(I_{\mathrm{A}}\left(2KNM + KM^2\right) + K^2N^2\right)\right)$. We will show in simulation that, the proposed algorithm has nearly the same I_{O} as that of the alternating optimization approach in Section III. Hence, the complexity of Algorithm 2 with respect to N is reduced by I_{R} times, since no iterative updating is required in the block with respect to $\boldsymbol{\theta}$.

V. EXTEND THE NON-CONVEX BCD FOR THE IMPERFECT CSI SETUP

In this section, we extend the non-convex BCD method in Algorithm 2 to solve $\mathcal{P}(\mathrm{B})$. Specifically, after applying the closed-form FP approach in Section IV-A, $\mathcal{P}(\mathrm{B})$ is equivalently transformed as follows:

$$\mathcal{P}(\mathrm{B1}) \quad \min_{\boldsymbol{\theta}} \quad f_{\mathrm{B1}}(\boldsymbol{\theta}) = \mathbb{E}_\xi\left[g(\boldsymbol{\theta};\xi)\right]$$
$$\mathbf{s.t.} \quad |\theta_n| = 1, \quad \forall n = 1, \cdots, N,$$

where

$$g(\boldsymbol{\theta};\xi) = \min_{\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{W}} -f_{\mathrm{A1}}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{W}, \boldsymbol{\theta}; \xi)$$
$$\mathbf{s.t.} \quad \sum_{k=1}^K \|\mathbf{w}_k\|^2 \leq P_{\mathrm{T}},$$
$$\alpha_k \geq 0, \quad \forall k = 1, \cdots, K.$$

A. Solve The Inner-Layer Subproblem

Exploiting the BCD approach in Section IV, we may obtain a stationary solution for $g(\boldsymbol{\theta};\xi)$, and have following approximated function:

$$\hat{g}(\boldsymbol{\theta};\xi) = \boldsymbol{\theta}^{\mathrm{H}}\boldsymbol{U}\boldsymbol{\theta} - 2\mathrm{Re}\left\{\boldsymbol{\nu}^{\mathrm{H}}\boldsymbol{\theta}\right\},$$

where $[\boldsymbol{U}, \boldsymbol{\nu}] = \mathrm{ParfunB}(\boldsymbol{\varphi}, \boldsymbol{\alpha}^{(0)}, \boldsymbol{\beta}^{(0)}, \mathbf{W}^{(0)}; \xi)$ which is summarized in Algorithm 3, and $\boldsymbol{\varphi} = [\varphi_1, \cdots, \varphi_N]^{\mathrm{T}}$ with $\theta_n = e^{j\varphi_n}$. One can also verify that, $\hat{g}(\boldsymbol{\theta};\xi)$ is continuously differentiable, which has the unique and finite output given

fixed $\boldsymbol{\alpha}^{(0)}$, $\boldsymbol{\beta}^{(0)}$, $\mathbf{W}^{(0)}$, and stop criterion. Then $\mathcal{P}(\mathrm{B1})$ can be approximately solved by

$$\mathcal{P}(\mathrm{B2}) \quad \min_{\boldsymbol{\varphi}} f_{\mathrm{B2}}(\boldsymbol{\varphi}) = \mathbb{E}_\xi\left[\hat{g}(\boldsymbol{\varphi};\xi)\right]$$

where

$$\hat{g}(\boldsymbol{\varphi};\xi) = (e^{j\boldsymbol{\varphi}})^{\mathrm{H}}\boldsymbol{U}e^{j\boldsymbol{\varphi}} - 2\mathrm{Re}\left\{\boldsymbol{\nu}^{\mathrm{H}}e^{j\boldsymbol{\varphi}}\right\}.$$

Since we obtain the stationary solution for the inner-layer subproblem, the stationary solution of $\mathcal{P}(\mathrm{B2})$ is also the stationary solution of $\mathcal{P}(\mathrm{B1})$.

B. Stochastic Successive Convex Approximation for $\mathcal{P}(B2)$

A classical approach to deal with the expectation operation in $f_{\mathrm{B2}}(\boldsymbol{\varphi})$ is the sample average approximation method. To be specific, at the r-th iteration, a new realization ξ_r is obtained and $\boldsymbol{\varphi}$ is updated by:

$$\boldsymbol{\varphi}_r = \arg\min_{\boldsymbol{\varphi}} f_{\mathrm{B},r}(\boldsymbol{\varphi}) \triangleq \frac{1}{r}\sum_{i=1}^r \hat{g}(\boldsymbol{\varphi};\xi_i). \qquad (22)$$

1) Stochastic SCA: However, $\hat{g}(\boldsymbol{\varphi};\xi_i)$ is still a non-convex function of $\boldsymbol{\varphi}$, and the parameters $\boldsymbol{U}$ and $\boldsymbol{\nu}$ are obtained by iterative operations. In [36], a stochastic optimization version of SCA is proposed to deal with this kind of problem, in which $\boldsymbol{\varphi}$ is updated by:

$$\boldsymbol{\varphi}_r = \arg\min_{\boldsymbol{\varphi}} h_r(\boldsymbol{\varphi}) \triangleq \frac{1}{r}\sum_{i=1}^r \hat{h}_i(\boldsymbol{\varphi}, \boldsymbol{\varphi}_{i-1}; \xi_i), \qquad (23)$$

where $\hat{h}_i(\boldsymbol{\varphi}, \boldsymbol{\varphi}_{i-1}; \xi_i)$ is an approximation of $\hat{g}(\boldsymbol{\varphi};\xi_i)$ around the output of $\{i-1\}$-th iteration.

Similar to the conventional SCA constraints in (16a) and (16b), if $\hat{h}_i(\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2; \xi_i)$ is continuously differentiable strongly convex with uniformly bounded second order derivatives, we still require following two constraints to guarantee the convergence of the stochastic SCA:

$$\hat{h}_i(\boldsymbol{\varphi}, \boldsymbol{\varphi}; \xi_i) = \hat{g}(\boldsymbol{\varphi}; \xi_i), \qquad (24a)$$
$$\hat{h}_i(\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2; \xi_i) \geq \hat{g}(\boldsymbol{\varphi}_1; \xi_i). \qquad (24b)$$

2) Design Surrogate Function: Fortunately, the surrogate function designed for the perfect CSI setup in (19) also satisfies above requirement. Thus we have

$$\hat{h}_i(\boldsymbol{\varphi}, \boldsymbol{\varphi}_{i-1}; \xi_i) = \hat{g}(\boldsymbol{\varphi}_{i-1}; \xi_i) + \frac{\kappa_i}{2}\|\boldsymbol{\varphi} - \boldsymbol{\varphi}_{i-1}\|^2$$
$$+ \nabla\hat{g}(\boldsymbol{\varphi}_{i-1}; \xi_i)^{\mathrm{T}}(\boldsymbol{\varphi} - \boldsymbol{\varphi}_{i-1}), \qquad (25)$$

where the gradient is

$$\nabla \hat{g}(\boldsymbol{\varphi}_{i-1}; \xi_i) = 2\mathrm{Re}\left\{ -\jmath e^{-\jmath\boldsymbol{\varphi}_{i-1}} \circ \left(\bar{U}e^{\jmath\boldsymbol{\varphi}_{i-1}} - \bar{\nu} \right) \right\}, \quad (26a)$$

$$[\bar{U}, \bar{\nu}] = \mathrm{ParfunB}(\boldsymbol{\varphi}_{i-1}, \boldsymbol{\alpha}^{(0)}, \boldsymbol{\beta}^{(0)}, \mathbf{W}^{(0)}; \xi_i), \quad (26b)$$

and $\kappa_i > 0$ should be properly chosen to satisfy (24b). One can see that, given $\boldsymbol{\varphi}_{i-1}$ and channel realizations about ξ_i, the surrogate function $\hat{h}_i(\boldsymbol{\varphi}, \boldsymbol{\varphi}_{i-1}; \xi_i)$ in (25) has closed-form expression, and the gradient of $h_r(\boldsymbol{\varphi})$ is $\frac{1}{r}\sum_{i=1}^{r} \nabla\hat{g}(\boldsymbol{\varphi}_{i-1}; \xi_i)$.

3) Gradient Projection for Updating $\boldsymbol{\varphi}_r$: The main drawback of the sample average approximation in (22) is that, the upadte of $\boldsymbol{\varphi}_r$ is related to all $\hat{g}(\boldsymbol{\varphi}; \xi_i)$ for $i = 1, \cdots, r$. In [37, Theorem 1], a better recursive approximation is proposed which can be exploited to further simplify the update rule. To be specific, $f_{\mathrm{B},r}(\boldsymbol{\varphi})$ can be approximated recursively by h_r as follows:

$$h_r = (1 - \delta_r)h_{r-1} + \delta_r \hat{g}(\boldsymbol{\varphi}; \xi_r), \quad (27)$$

where $h_0 = 0$, and we set $\delta_r = r^{-0.501}$. Then, replacing $\frac{1}{r}$ with δ_r, the gradient of $f_{\mathrm{B},r}(\boldsymbol{\varphi})$ is approximated by

$$\mathbf{h}_r = (1 - \delta_r)\mathbf{h}_{r-1} + \delta_r \nabla\hat{g}(\boldsymbol{\varphi}; \xi_r), \quad (28)$$

with $\mathbf{h}_0 = \mathbf{0}$. Combining (25), the surrogate function for h_r can be also expressed by a recursive formula as

$$\bar{h}_r(\boldsymbol{\varphi}) = (1 - \delta_r)h_{r-1} + \delta_r \hat{g}(\boldsymbol{\varphi}_{r-1}; \xi_r)$$
$$+ \mathbf{h}_r^{\mathrm{T}}(\boldsymbol{\varphi} - \boldsymbol{\varphi}_{r-1}) + \frac{\kappa_r}{2}\|\boldsymbol{\varphi} - \boldsymbol{\varphi}_{r-1}\|^2.$$

Finally, we have following update rule for $\boldsymbol{\varphi}_r$:

$$\boldsymbol{\varphi}_r = \boldsymbol{\varphi}_{r-1} - \frac{\mathbf{h}_r}{\kappa_r}, \quad (29)$$

which is similar to the gradient projection updating in (17), and thus κ_r can be determined by the Armijo rule.

C. Algorithm Development

Putting all above together, when a new realization ξ_r is obtained, we first calculate the gradient $\nabla\hat{g}(\boldsymbol{\varphi}_{r-1}; \xi_r)$ according to (26a) and (26b). Then, $\mathbf{h}_r$ is updated according to (28). Substituting $\mathbf{h}_r$ into (29), we have $\boldsymbol{\varphi}_r$.

The proposed stochastic SCA approach is summarized in Algorithm 4. Denote by I_{A} the iteration number of the Armijo search, by I_{S} the iteration number in Algorithm 3. The complexity of the proposed algorithm is $\mathcal{O}\left(I_{\mathrm{O}}\left(I_{\mathrm{A}}I_{\mathrm{S}}\left(2KNM + KM^2\right) + K^2N^2\right)\right)$. Hence, compared with Algorithm 2, Algorithm 4 cost I_{S} times complexity with respect to M to obtain a stationary beamforming solution in every outer-loop iteration. However, the complexity with respect to N is still about $\mathcal{O}\left(I_{\mathrm{O}}K^2N^2\right)$.

VI. NUMERICAL RESULTS

A. Simulation Scenario

In this section, numerical examples are provided to validate the effectiveness of the proposed algorithms. We consider a RIS-aided femtocell network illustrated in Fig. 3, in which one AP equipped with 4 antennas, and 4 single-antenna users ($K = 4$) uniformly and randomly distributed in a circle centered at $(200\ \mathrm{m}, 30\ \mathrm{m})$ with radius 10 m. The RIS is applied to

---

Algorithm 4 Stochastic SCA for the imperfect CSI setup.

1: Initialize $\boldsymbol{\varphi}_0$ and set $r = 0$.
   Repeat
2: Set $r = r+1$ and obtain new channel realizations $\mathbf{h}_{\mathrm{d},k}(\xi_r)$, $\mathbf{G}(\xi_r)$, $\mathbf{h}_{\mathrm{r},k}(\xi_r)$ for all k.
3: Calculate the gradient $\nabla\hat{g}(\boldsymbol{\varphi}_{r-1}; \xi_r)$ by (26a) and (26b).
4: Update $\mathbf{h}_r$ by (28).
5: Search κ_r in (29) by Armijo rule, and then update $\boldsymbol{\varphi}_r$.
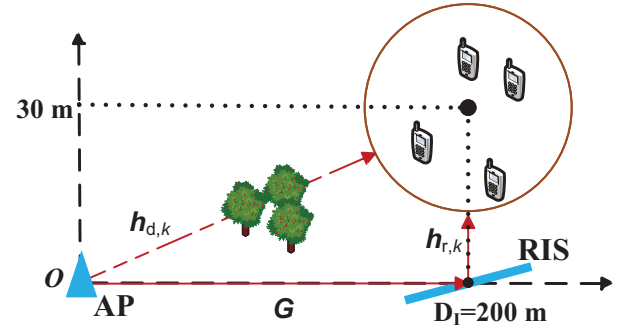   **Until** The value of h_r in (27) converges.

Fig. 3. The simulated RIS-aided K-user MISO communication scenario comprising of one M-antenna AP and one N-element RIS.

TABLE I
SIMULATION PARAMETERS

Parameters	Values
AP location	(0m, 0m)
Path-loss for $\mathbf{G}$ and $\mathbf{h}_{\mathrm{r},k}$ (dB)	$35.6 + 22.0\lg d$
Path-loss for $\mathbf{h}_{\mathrm{d},k}$ (dB)	$32.6 + 36.7\lg d$
Transmission bandwidth	180 kHz
Noise power spectral density	-170 dBm/Hz

provide high-quality link between the AP and users, and we assume that the LOS component is contained by the channel between AP and RIS, and channel between RIS and each user. The system parameters are summarized in Table I, which are almost the same as those in [24]. In particular, the path-loss is set according to the 3GPP propagation environment [52, Table B.1.2.1-1].

We assume the direct link channel $\mathbf{h}_{\mathrm{d},k}$ follows Rayleigh fading, while the RIS-aided channels follow Rician fading. Same as [25] and [39], we further assume that the antenna elements form a half-wavelength uniform linear array configuration at the AP and the RIS, and thus the channels $\mathbf{G}$ and $\mathbf{h}_{\mathrm{r},k}$ are modeled by

$$\mathbf{G} = L_1\left(\sqrt{\frac{\varepsilon}{\varepsilon+1}}\mathbf{a}_N(\vartheta)\mathbf{a}_M(\psi)^{\mathrm{H}} + \sqrt{\frac{1}{\varepsilon+1}}\bar{\mathbf{G}}\right), \quad (30)$$

$$\mathbf{h}_{\mathrm{r},k} = L_{2,k}\left(\sqrt{\frac{\varepsilon}{\varepsilon+1}}\mathbf{a}_N(\varsigma_k) + \sqrt{\frac{1}{\varepsilon+1}}\bar{\mathbf{h}}_{\mathrm{r},k}\right), \quad (31)$$

where L_1 and $L_{2,k}$ denote the corresponding path-losses, ε is the Rician factor and we set $\varepsilon = 10$, $\mathbf{a}$ is the steering vector, ϑ, ψ and ς_k are the angular parameters, and $\bar{\mathbf{G}}$ and $\bar{\mathbf{h}}_{\mathrm{r},k}$ denote the NLOS components whose elements are chosen from $\mathcal{CN}(0, 1)$.

Based on above assumption, only the small-scale fading variables $\mathbf{h}_{d,k}$, $\bar{\mathbf{G}}$, and $\bar{\mathbf{h}}_{r,k}$ need to be estimated in every frame. Denote x as one element in above variables, and $\hat{x}$ is the corresponding estimate value. We assume that the estimate error $x - \hat{x}$ follows zero mean complex Gaussian distribution, and all these elements have the same normalized MSE:

$$\varrho = \frac{\mathbb{E}\left[|x - \hat{x}|^2\right]}{\mathbb{E}\left[|\hat{x}|^2\right]}.$$

To better understand the channel conditions of the direct link and the RIS-aided link, we provide a simple example here. Consider a reference point at (200m, 30m). According to Table I, the direct-link path-loss is about 117.23 dB, meanwhile, the path-loss of channel $\mathbf{G}$ and channel $\mathbf{h}_r$ are 86.22 dB and 68.10 dB, respectively, so the path-loss of the RIS-aided link ($N = 1$) is 154.32 dB, which is much larger than that of the direct link (about 37 dB). Therefore, the direct link cannot be ignored, and extremely large N is required to achieve performance gain, if the surface phase vector $\boldsymbol{\theta}$ is not properly designed. In the next, we will show that, by utilizing the proposed joint optimization algorithms, significant performance gain can be achieved.

We evaluate the performance of the proposed algorithms with the following 3 baselines:

- **Baseline 1** (Without RIS): Let $N = 0$, and then $\mathcal{P}(A)$ is solved by the WMMSE.
- **Baseline 2** (Random Phase): $\boldsymbol{\theta}$ is initialized by random value, and then $\mathbf{W}$ is optimized by WMMSE.
- **Baseline 3** (Upper Bound): The KKT conditions are necessary conditions for a solution to be optimal. Thus one may run Algorithm 2 sufficient times (e.g., 100 times) with random initializations, and then the maximum output might approximate the optimal solution well.

B. Weighted Sum Rate Analyses

In this subsection, we assume that the RIS is deployed at (200m, 0m), and the users' locations are fixed once randomly generated.[2] Then, for fairness comparison, the weights are first chosen inversely proportional to the direct-link path-loss, and then normalized by $\sum \omega_k = 1$. All the simulation curves have been averaged over 10^3 independent realizations of channel small scale fading.

Fig. 4(a) illustrates the WSR of different schemes with respect to the transmit power P_T when $N = 100$. It is seen that, if the phase vector is not optimized, the performance gain by deploying RIS is negligible as expected. However, the joint beamforming and phase optimization schemes may achieve about 4 dB gain. In addition, in perfect CSI setup, the proposed Algorithm 2 and the alternating optimization approach have almost the same performance. We conjecture that this is due to both algorithms are initialized by the same point, and then they both converge to the same stationary solution with high probability. In imperfect CSI setup, one can see that, the proposed algorithm may still achieve about 3 dB

[2]In the simulation, the user locations are (205.65m, 34.48m), (193.47m, 30.24m), (198.30m, 22.40m), and (207.00m, 24.28m).
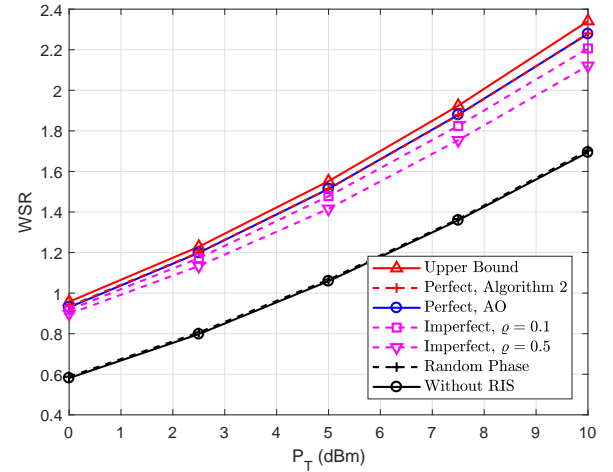
(a) P_T vs WSR
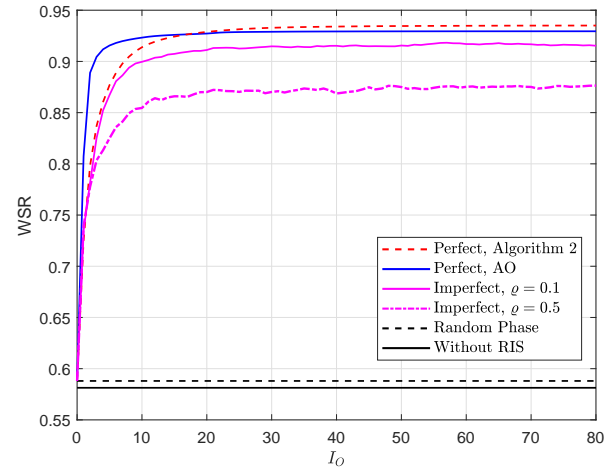
(b) Convergence behavior when $P_T = 0$ dBm

Fig. 4. The WSR versus transmit power when $N = 100$.

gain when $\varrho \leq 0.5$. Besides, the performance loss increases as P_T increases.

Next, in Fig. 4(b), we fix the transmit power $P_T = 0$ dBm and show the convergence behaviors of all the proposed algorithms. In perfect CSI setup, the convergence speed of the proposed algorithm is slightly slower than the alternating optimization approach, which the performance is sightly better. In addition, as we have analyzed in Section IV-E, in each iteration, the proposed algorithm has much lower complexity. In imperfect CSI setup, one can see that, as ϱ increases, the channel becomes more uncertain, and the proposed algorithm needs more steps to get converged.

Fig. 5 compares the WSR with the size N of RIS, while the transmit power of AP is fixed to $P_T = 5$ dBm. The random-phase scheme still has only a small gain, meanwhile all the schemes optimized $\boldsymbol{\theta}$ achieve remarkable performance gain as N increases. Besides, we observe that, the performance of the proposed algorithm at $N = 200$ is similar to that at $P_T = 8$ dBm in Fig. 4(a). This observation implies that, different from [22, Proposition 2], the RIS phase design could not achieve
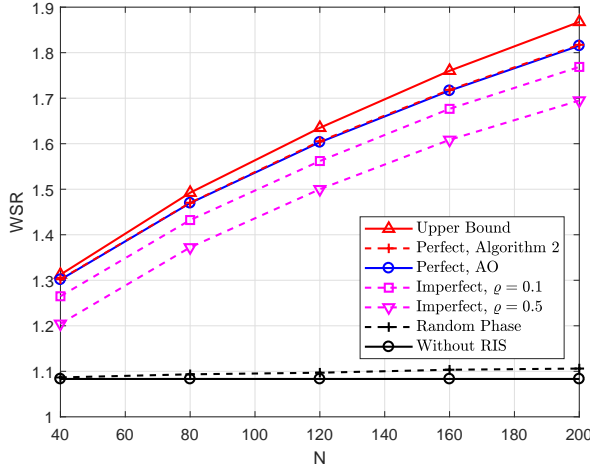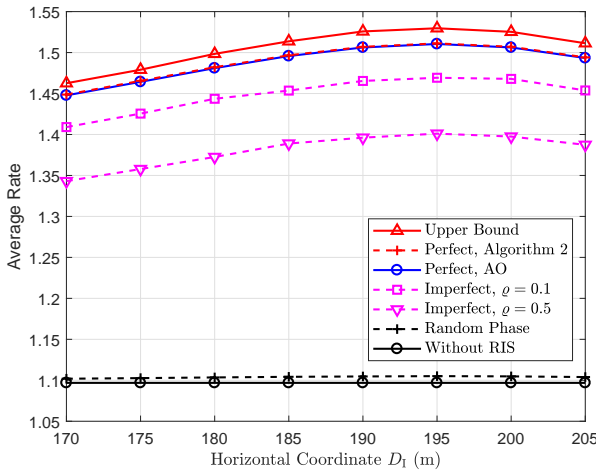
Fig. 5.   WSR versus N, when $P_{\mathrm{T}} = 5$ dBm.

Fig. 7.   CDF curves for random user locations.

Fig. 6.   The average rate versus the horizontal coordinate of RIS, when $P_{\mathrm{T}} = 5$ dBm and $N = 100$.

the "squared gain" here, since the aperture gain of the RIS is relatively small.

C. RIS Deployment and User Locations

In this subsection, we discuss on the impact of the RIS deployment locations and users' locations, and the horizontal coordinate of RIS is denoted by D_{I}. The weights are set to be equal to $1/K$, so the objective function becomes average rate. We generate 100 snapshots for randomly located users. Then, for each snapshot, we further generate 100 channel realizations with independent small-scale fading.

Fig. 6 illustrates the average rate of users when $P_{\mathrm{T}} = 5$ dBm and $N = 100$, while moving the RIS from (170m, 0m) to (205m, 0m). It is seen that, when D_{I} increases from 200 m to 205 m, the average rate decreases, since the path-losses of channel $\mathbf{G}$ and channel $\mathbf{h}_{\mathrm{r}}$ both increase. However, when decreasing D_{I} from 200 m to 170 m, the average rate first increases, and then decreases, while the optimal location is $D_{\mathrm{I}} = 195$ m. This is because, the path-loss of the RIS-aided
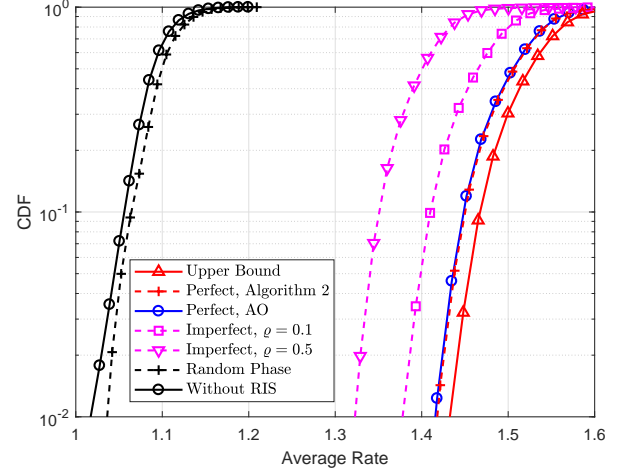
link is the product of the path-losses of $\mathbf{G}$ and $\mathbf{h}_{\mathrm{r}}$. Therefore, although the summation of the transmission distance decreases, the propagation condition might not necessarily become better.

Finally, Fig. 7 plots the *cumulative distribution function* (CDF) of the average rate over different snapshots by deploying RIS at (195m, 0m). It is seen that, the performance gains of all the proposed schemes are stable over the CDF curves, and also keep consistent with their counterparts in Fig. 6. Therefore, we conclude that, with high probability, the performance of the proposed algorithms will be good irrespective of user locations.
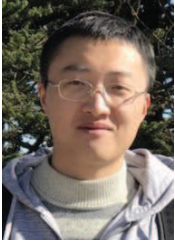
VII. CONCLUSION

In this paper, we investigate the RIS-aided multiuser downlink MISO system. Specifically, a joint transmit beamforming design and RIS phase optimization problem is formulated to maximize the WSR under the AP transmit power constraint. The perfect CSI setup is firstly addressed, and a low-complexity algorithm is designed to carry out stationary solution for the joint design problem by utilizing the recently proposed FP technique. The proposed algorithm is then leveraged to the imperfect CSI setup, and the average WSR is maximized by resorting to the stochastic SCA technique. Extensive simulation results demonstrated that the proposed joint design schemes achieve significant performance gain compared to the benchmarks by deploying a RIS with 100 passive elements. In addition, it is also shown that the performance degradation of the proposed algorithm is very small, when the channel estimation uncertainty is smaller than 10%.

REFERENCES

[1] E. Basar, M. D. Renzo, J. D. Rosny, M. Debbah, M. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116 753–116 773, 2019.
[2] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *arXiv preprint arXiv:1905.00152*, 2019.

[3] Q.-U.-A. Nadeem, A. Kammoun, A. Chaaban, M. Debbah, and M.-S. Alouini, "Intelligent reflecting surface assisted multi-user MISO communication," *arXiv preprint arXiv:1906.02360*, 2019.

[4] Y.-C. Liang, R. Long, Q. Zhang, J. Chen, H. V. Cheng, and H. Guo, "Large intelligent surface/antennas (LISA): Making reflective radios smart," *J. Commun. Inf. Netw.*, vol. 4, no. 2, pp. 40–50, June 2019.

[5] X. Tan, Z. Sun, D. Koutsonikolas, and J. M. Jornet, "Enabling indoor mobile millimeter-wave networks based on smart reflect-arrays," in *Proc. IEEE INFOCOM*, Apr. 2018, pp. 270–278.

[6] F. Liu, O. Tsilipakos, A. Pitilakis, A. C. Tasolamprou, M. S. Mirmoosa, N. V. Kantartzis, D.-H. Kwon, M. Kafesaki, C. M. Soukoulis, and S. A. Tretyakov, "Intelligent metasurfaces with continuously tunable local surface impedance for multiple reconfigurable functions," *Physical Rev. Appl.*, vol. 11, no. 4, p. 044024, 2019.

[7] L. Li, T. J. Cui, W. Ji, S. Liu, J. Ding, X. Wan, Y. B. Li, M. Jiang, C.-W. Qiu, and S. Zhang, "Electromagnetic reprogrammable coding-metasurface holograms," *Nature Commun.*, vol. 8, no. 1, p. 197, 2017.

[8] D. Mishra and H. Johansson, "Channel estimation and low-complexity beamforming design for passive intelligent surface assisted MISO wireless energy transfer," in *Proc. IEEE ICASSP*, 2019, pp. 4659–4663.

[9] Q. Wu and R. Zhang, "Weighted sum power maximization for intelligent reflecting surface aided SWIPT," *arXiv preprint arXiv:1907.05558*, 2019.

[10] M. Cui, G. Zhang, and R. Zhang, "Secure wireless communication via intelligent reflecting surface," *IEEE Wireless Commun. Lett.*, Early Access.

[11] J. Chen, Y.-C. Liang, Y. Pei, and H. Guo, "Intelligent reflecting surface: A programmable wireless environment for physical layer security," *IEEE Access*, vol. 7, pp. 82 599–82 612, 2019.

[12] H. Shen, W. Xu, S. Gong, Z. He, and C. Zhao, "Secrecy rate maximization for intelligent reflecting surface assisted multi-antenna communications," *IEEE Commun. Lett.*, vol. 23, no. 9, pp. 1488–1492, 2019.

[13] X. Tan, Z. Sun, J. M. Jornet, and D. Pados, "Increasing indoor spectrum sharing capacity using smart reflect-array," in *Proc. IEEE ICC*, 2016, pp. 1–6.

[14] C. Liaskos, S. Nie, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. Akyildiz, "A new wireless communication paradigm through software-controlled metasurfaces," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 162–169, 2018.

[15] E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis, and T. L. Marzetta, "Massive MIMO is a reality--what is next?: Five promising research directions for antenna arrays," *Digit. Signal Process.*, 2019.

[16] M. Di Renzo, M. Debbah, D.-T. Phan-Huy, A. Zappone, M.-S. Alouini, C. Yuen, V. Sciancalepore, G. C. Alexandropoulos, J. Hoydis, H. Gacanin *et al.*, "Smart radio environments empowered by reconfigurable AI meta-surfaces: an idea whose time has come," *EURASIP J. Wireless Commun. and Netw.*, vol. 2019, no. 1, p. 129, 2019.

[17] S. V. Hum and J. Perruisseau-Carrier, "Reconfigurable reflectarrays and array lenses for dynamic antenna beam control: A review," *IEEE Trans. Antennas Propag.*, vol. 62, no. 1, pp. 183–198, 2014.

[18] K. Ntontin, M. Di Renzo, J. Song, F. Lazarakis, J. de Rosny, D.-T. Phan-Huy, O. Simeone, R. Zhang, M. Debbah, and G. Lerosey, "Reconfigurable intelligent surfaces vs. relaying: Differences, similarities, and performance comparison," *arXiv preprint arXiv:1908.08747*, 2019.

[19] H. Q. Ngo, H. A. Suraweera, M. Matthaiou, and E. G. Larsson, "Multipair full-duplex relaying with massive arrays and linear processing," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1721–1737, 2014.

[20] E. Björnson, Ö. Ö, and E. G. Larsson, "Intelligent reflecting surface vs. decode-and-forward: How large surfaces are needed to beat relaying?" *IEEE Wireless Commun. Lett.*, Early Access.

[21] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network: Joint active and passive beamforming design," in *Proc. IEEE Globecom*, Dec. 2018, pp. 1–6.

[22] ——, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, 2019.

[23] C. Huang, A. Zappone, M. Debbah, and C. Yuen, "Achievable rate maximization by passive intelligent mirrors," in *Proc. IEEE ICASSP*, May 2018, pp. 3714–3718.

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

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

[26] Q.-U.-A. Nadeem, A. Kammoun, A. Chaaban, M. Debbah, and M.-S. Alouini, "Large intelligent surface assisted MIMO communications," *arXiv:1903.08127*, 2019.

[27] X. Yu, J. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485–500, 2016.

[28] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, 2014.

[29] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, 2016.

[30] K. Shen and W. Yu, "Fractional programming for communication systems–Part II: Uplink scheduling via matching," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2631–2644, 2018.

[31] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optimization*, vol. 23, no. 2, pp. 1126–1153, 2013.

[32] A. Liu and V. Lau, "Phase only RF precoding for massive MIMO systems with limited RF chains," *IEEE Trans. Signal Process.*, vol. 62, no. 17, pp. 4505–4515, 2014.

[33] A. Liu and V. K. N. Lau, "Impact of CSI knowledge on the codebook-based hybrid beamforming in massive MIMO," *IEEE Trans. Signal Process.*, vol. 64, no. 24, pp. 6545–6556, 2016.

[34] A. Liu, V. K. N. Lau, and M. Zhao, "Stochastic successive convex optimization for two-timescale hybrid precoding in massive MIMO," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 432–444, 2018.

[35] X. Chen, A. Liu, Y. Cai, V. K. N. Lau, and M. Zhao, "Randomized two-timescale hybrid precoding for downlink multicell massive mimo systems," *IEEE Trans. Signal Process.*, vol. 67, no. 16, pp. 4152–4167, 2019.

[36] M. Razaviyayn, M. Sanjabi, and Z.-Q. Luo, "A stochastic successive minimization method for nonsmooth nonconvex optimization with applications to transceiver design in wireless communication networks," *Mathematical Programming*, vol. 157, no. 2, pp. 515–545, 2016.

[37] A. Liu, V. K. N. Lau, and M. Zhao, "Online successive convex approximation for two-stage stochastic nonconvex optimization," *IEEE Trans. Signal Process.*, vol. 66, no. 22, pp. 5941–5955, 2018.

[38] Y. Yang, B. Zheng, S. Zhang, and R. Zhang, "Intelligent reflecting surface meets OFDM: Protocol design and rate maximization," *arXiv preprint arXiv:1906.09956*, 2019.

[39] Z. He and X. Yuan, "Cascaded channel estimation for large intelligent metasurface assisted massive MIMO," *IEEE Wireless Commun. Lett.*, Early Access.

[40] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling large intelligent surfaces with compressive sensing and deep learning," *arXiv preprint arXiv:1904.10136*, 2019.

[41] C. Huang, G. C. Alexandropoulos, C. Yuen, and M. Debbah, "Indoor signal focusing with deep learning designed reconfigurable intelligent surfaces," in *Proc. IEEE SPAWC*, 2019, pp. 1–5.

[42] C. Liaskos, A. Tsioliaridou, S. Nie, A. Pitsillides, S. Ioannidis, and I. Akyildiz, "An interpretable neural network for configuring programmable wireless environments," in *Proc. IEEE SPAWC*, 2019, pp. 1–5.

[43] W. Cheng and R. D. Murch, "Adaptive downlink multi-user MIMO wireless systems for correlated channels with imperfect CSI," *IEEE Trans. Wireless Commun.*, vol. 5, no. 9, pp. 2435–2446, 2006.

[44] Y. Taesang and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2203–2214, 2006.

[45] A. D. Dabbagh and D. J. Love, "Multiple antenna mmse based downlink precoding with quantized feedback or channel mismatch," *IEEE Trans. Commun.*, vol. 56, no. 11, pp. 1859–1868, 2008.

[46] Y. Xu and Y. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM J. Imag. Sci.*, vol. 6, no. 3, pp. 1758–1789, 2013.

[47] Q. Shi, M. Razaviyayn, Z. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, 2011.

[48] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre, "Manopt, a Matlab toolbox for optimization on manifolds," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1455–1459, 2014.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2020.2970061, IEEE Transactions on Wireless Communications

The final version of record is available at   http://dx.doi.org/10.1109/TWC.2020.2970061

13

[49] X. Yu, D. Xu, and R. Schober, "MISO wireless communication systems via intelligent reflecting surfaces : (Invited Paper)," in *Proc. IEEE ICCC*, 2019, pp. 735–740.

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

[51] D. P. Bertsekas, "Nonlinear programming," *J. Operational Research Soc.*, vol. 48, no. 3, pp. 334–334, 1997.

[52] *Further advancements for E-UTRA physical layer aspects (Release 9)*. 3GPP TS 36.814, Mar. 2010.

Jie Chen (S'17) received the B.S. degree in communication engineering from Chongqing University of Posts and Telecommunications, China, in 2016. He is currently working towards the Ph.D. degree in University of Electronic Science and Technology of China (UESTC), China. His research interests include convex optimization, resource allocation in wireless communication networks, reconfigurable intelligent surface, and short packet communications.

Huayan Guo (M'18) received the B.S. degree in electronic information science and technology from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2012, and the Ph.D. degree from Peking University, Beijing, China, in 2017. From 2017 to 2019, he was a post-doctoral fellow with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China. Currently, he is a post-doctoral fellow working with the Department of Electronic and Computer Engineering in the Hong Kong University of Science and Technology, and his research interests include passive radio communications, and machine learning for Internet-of-Things.

Ying-Chang Liang (F'11) is currently a Professor with the University of Electronic Science and Technology of China, China, where he leads the Center for Intelligent Networking and Communications and serves as the Deputy Director of the Artificial Intelligence Research Institute. He was a Professor with The University of Sydney, Australia, a Principal Scientist and Technical Advisor with the Institute for Infocomm Research, Singapore, and a Visiting Scholar with Stanford University, USA. His research interests include wireless networking and communications, cognitive radio, symbiotic networks, dynamic spectrum access, the Internet-of-Things, artificial intelligence, and machine learning techniques.

Dr. Liang has been recognized by Thomson Reuters (now Clarivate Analytics) as a Highly Cited Researcher since 2014. He received the Prestigious Engineering Achievement Award from The Institution of Engineers, Singapore, in 2007, the Outstanding Contribution Appreciation Award from the IEEE Standards Association, in 2011, and the Recognition Award from the IEEE Communications Society Technical Committee on Cognitive Networks, in 2018. He is the recipient of numerous paper awards, including the IEEE Jack Neubauer Memorial Award, in 2014, and the IEEE Communications Society APB Outstanding Paper Award, in 2012. He is a Fellow of IEEE, and a foreign member of Academia Europaea.

He is the Founding Editor-in-Chief of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS: COGNITIVE RADIO SERIES, and the Key Founder and now the Editor-in-Chief of the IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING. He is also serving as an Associate Editor-in-Chief for China Communications. He served as a Guest/Associate Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE JOURNAL OF SELECTED AREAS IN COMMUNICATIONS, the IEEE Signal Processing Magazine, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and the IEEE TRANSACTIONS ON SIGNAL AND INFORMATION PROCESSING OVER NETWORK. He was also an Associate Editor-in-Chief of the World Scientific Journal on Random Matrices: Theory and Applications. He was a Distinguished Lecturer of the IEEE Communications Society and the IEEE Vehicular Technology Society. He was the Chair of the IEEE Communications Society Technical Committee on Cognitive Networks, and served as the TPC Chair and Executive Co-Chair of the IEEE Globecom'17.

Erik G. Larsson (S'99–M'03–SM'10–F'16) received the Ph.D. degree from Uppsala University, Uppsala, Sweden, in 2002. He is currently Professor of Communication Systems at Linköping University (LiU) in Linköping, Sweden. He was with the KTH Royal Institute of Technology in Stockholm, Sweden, the George Washington University, USA, the University of Florida, USA, and Ericsson Research, Sweden. His main professional interests are within the areas of wireless communications and signal processing. He co-authored *Space-Time Block Coding for Wireless Communications* (Cambridge University Press, 2003) and *Fundamentals of Massive MIMO* (Cambridge University Press, 2016). He is co-inventor of 19 issued U.S. patents.

Currently he is an editorial board member of the *IEEE Signal Processing Magazine*, and a member of the *IEEE Transactions on Wireless Communications* steering committee. He served as chair of the IEEE Signal Processing Society SPCOM technical committee (2015–2016), chair of the *IEEE Wireless Communications Letters* steering committee (2014–2015), General respectively Technical Chair of the Asilomar SSC conference (2015, 2012), technical co-chair of the IEEE Communication Theory Workshop (2019), and member of the IEEE Signal Processing Society Awards Board (2017–2019). He was Associate Editor for, among others, the *IEEE Transactions on Communications* (2010-2014) and the *IEEE Transactions on Signal Processing* (2006-2010).

He received the IEEE Signal Processing Magazine Best Column Award twice, in 2012 and 2014, the IEEE ComSoc Stephen O. Rice Prize in Communications Theory in 2015, the IEEE ComSoc Leonard G. Abraham Prize in 2017, the IEEE ComSoc Best Tutorial Paper Award in 2018, and the IEEE ComSoc Fred W. Ellersick Prize in 2019.