Reconfigurable Intelligent Surface Based Uplink MU-MIMO Symbiotic Radio System

Jinlin Hu, *Graduate Student Member, IEEE*, Ying-Chang Liang, *Fellow, IEEE*, Yiyang Pei, *Senior Member, IEEE*, Sumei Sun, *Fellow, IEEE*, and Ruolun Liu, *Member, IEEE*

***Abstract*— In this paper, we investigate a novel uplink *reconfigurable intelligent surface* (RIS) based multi-user multi-input multi-output symbiotic radio system. It indicates that each RIS, as an *Internet-of-Things* (IoT) device, enhances the primary transmission from a nearby user to the *base station* (BS), and simultaneously transmits its own information to the BS by backscattering modulation. By embedding environmental sensors on the RISs, the proposed system enables the IoT transmission of locally collected environmental data to the BS while assisting the primary communications from the users to the BS. We consider both the case of perfect and imperfect *channel state information* (CSI), and design the active beamforming at the BS and the passive beamforming at the RISs jointly to maximize the weighted sum-rate of both the primary and IoT transmissions. For the perfect CSI case, we propose an algorithm based on the *block coordinate descent* (BCD) method to solve the problem. We also propose another algorithm with a similar framework to reduce the computational complexity. For the imperfect CSI case, an algorithm based on BCD and the online successive convex approximation technique is proposed. Simulation results show that the proposed system achieves significant performance gain over a number of baseline schemes for both the perfect and imperfect CSI cases. Furthermore, when the channel estimation error is small, the performance loss due to imperfect CSI is insignificant.**

***Index Terms*— Symbiotic radio (SR), reconfigurable intelligent surfaces (RIS), multi-user multi-input multi-output (MU-MIMO) system, block coordinate descent method (BCD).**

Manuscript received 25 July 2021; revised 16 January 2022 and 25 April 2022; accepted 23 July 2022. Date of publication 5 August 2022; date of current version 9 January 2023. This work was supported in part by the National Natural Science Foundation of China under Grant U1801261, Grant 61631005, and Grant 61571100; in part by the National Key Research and Development Program of China under Grant 2018YFB1801105; in part by the Key Areas of Research and Development Program of Guangdong Province, China, under Grant 2018B010114001; in part by the Macau Science and Technology Development Fund (FDCT), Macau, SAR, under Grant 0009/2020/A1; in part by the Fundamental Research Funds for the Central Universities under Grant ZYGX2019Z022; in part by the Program of Introducing Talents of Discipline to Universities under Grant B20064; and in part by the National Research Foundation, Singapore, under its the AI Singapore Program under Garnt AISG2-RP-2020-019. An earlier version of this paper was presented in part at the 2021 IEEE Globecom [DOI: 10.1109/GLOBECOM46510.2021.9685222]. The associate editor coordinating the review of this article and approving it for publication was M. C. Gursoy. *(Corresponding author: Ying-Chang Liang.)*

Jinlin Hu is with the Center for Intelligent Networking and Communications (CINC), University of Electronic Science and Technology of China (UESTC), Chengdu 611731, China (e-mail: hujinlin@std.uestc.edu.c).

Ying-Chang Liang is with the CINC, UESTC, Chengdu 611731, China, and also with the Yangtze Delta Region Institute (Huzhou), University of Electronic Science and Technology of China, Huzhou 313001, China (e-mail: liangyc@ieee.org).

Yiyang Pei is with the Singapore Institute of Technology, Singapore 138683 (e-mail: yiyang.pei@singaporetech.edu.sg).

Sumei Sun is with the Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore 119613 (e-mail: sunsm@i2r.a-star.edu.sg).

Ruolun Liu is with the Mechanical, Electrical, & Information Engineering School, Shandong University at Weihai, Weihai, Shandong 264209, China (e-mail: ruolun.liu@sdu.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2022.3194910.

Digital Object Identifier 10.1109/TWC.2022.3194910

I. Introduction

WITH the development of digitalization, enhanced devices that require increased bandwidth are increasingly adopted. This trend contributes to the exponential growth of global mobile traffic. Specifically, the International Telecommunication Union has estimated that the global mobile traffic has grown in the range of 10-100 times from 2020 to 2030 [2]. To overcome the challenges of high throughput and super-massive access, novel wireless communication technologies with high energy and spectrum efficiency have aroused significant attention worldwide.

Recently, *reconfigurable intelligent surface* (RIS), also known as intelligent reflecting surface, has been regarded as one of the promising solutions to address the mentioned challenges. RIS is an array of nearly passive reflecting elements, and each element can independently draw a phase shift onto the incident electromagnetic waves [3]. With the passive reflection components, RIS can enhance the transmission of the other primary systems with low power consumption [4]. Due to its advantages, RIS has attracted extensive attention [5]–[10]. [5] introduces RIS into the downlink transmission of a cellular network to assist multiple users. The phase-shift matrix (i.e., passive beamforming) at the RIS is optimized to maximize energy efficiency. [6] optimizes both the passive beamforming at the RIS and the transmit beamforming at the *base station* (BS) (i.e., active beamforming) to maximize the received power at the user. [8] proves that the joint active and passive beamforming is effective and crucial for achieving better system performance. Typically, RIS has a large number of reflection elements, which imposes a significant challenge for online transmission. Thus [10] develops a physics-based model and a scalable optimization framework to reduce the complexity of reconfiguring RIS elements online.

Another emerging promising wireless technology is *symbiotic radio* (SR). In an SR system, secondary transmitters, i.e., *backscatter devices* (BDs), transmit their information to the corresponding receivers over the *radio frequency* (RF) signals received from a primary transmitter. As there is no need for a dedicated RF source, the power consumption of the secondary transmission can be reduced significantly [11]. In return, the secondary system provides multi-path diversity to the primary system, yielding mutual benefits between the two spectrum sharing systems. Therefore, the SR system can achieve high energy and spectrum efficiency, making it a good candidate for *Internet-of-Things* (IoT) transmission. Many studies have been conducted to explore the different SR applications [12]–[15]. [12] considers a basic SR system with two different practical setups, where a BD is applied to enhance the communication from a transmitter to a receiver and simultaneously transmits its own information to the receiver. [13] proposes a full-duplex SR system, where a full-duplex transmitter not only transmits primary signals to the BDs and the receivers, but also receives the backscatter signals from the BDs simultaneously. [14] develops another SR system with a full-duplex BD, where the BD absorbs a fraction of the incident signal to decode messages and simultaneously transmits its own message to a receiver. [15] proposes a backscatter *non-orthogonal multiple access* (NOMA) system, where a transmitter transmits information to one far-away cellular receiver and a nearby receiver, and the symbiotic BD simultaneously transmits its information over the incident signals to the nearby receiver.

However, all of the symbiotic BDs involved in the above literature are equipped with only one antenna, which cannot provide significant performance gain due to the double fading effect and reflection loss involved in the backscatter links. To further exploit the potential of SR systems, the researchers of [16] and [17] have introduced the RIS as a BD into downlink SR systems. Unlike the above studies on RISs, the symbiotic RISs can not only relay the incident signals but also transmit their own information to the receivers. Due to the large number of reflection elements deployed at the RIS, the backscatter link can be enhanced significantly. Thus, while the IoT transmission is enabled, the primary communication can also achieve significant performance gain [16]. Specifically, in [16], RIS, as an IoT device, transmits messages to an *IoT receiver* (IR) by reflecting radio and simultaneously assists the primary transmission from a transmitter to a primary receiver (PR). [17] combines the IR and the PR in [16] into one cooperative receiver, which can jointly decode the primary and IoT signals. In addition, [17] further considers a downlink scenario with multiple receivers and RISs. Specifically, each RIS is applied to assist the primary transmission from a transmitter to a nearby receiver and simultaneously transmits its own messages to the receiver. According to [16] and [17], the RIS-based SR system not only enables the downlink transmission from the RIS to the receiver, but also significantly enhances the primary communication from the transmitter to the receiver.

Motivated by the above works, we propose a novel RIS-based uplink *multi-user multi-input multi-output* (MU-MIMO) SR system in this paper. As shown in Fig. 1, it comprises a multi-antenna BS, multiple RISs, and multiple single-antenna users. Each RIS is deployed to enhance the communication from a nearby user to the BS. At the same time, each RIS transmits its own information to the BS through backscattering modulation.

The proposed RIS-based uplink MU-MIMO SR system is of practical interest, e.g., enabling the spectrum- and energy-efficient IoT transmissions for smart cities. Specifically, we consider that each RIS is deployed close to nearby potential users. Besides, various environmental sensors are embedded on the RISs, which can collect environmental data of the surroundings, such as temperature, humidity, and noise level. With the proposed system, such local environment information can be transmitted to the BS and eventually aggregated at the application server for centralized processing. The results of data processing, such as noise monitoring, pollution warning, and pollution warning, can in return bring convenience to the users near the RISs.

The main contributions of this paper are summarized as follows

- To the best of our knowledge, this is the first time the RIS-based uplink MU-MIMO SR system has been investigated. This system can enhance the primary communication from the users to the BS and enable the RISs to transmit their own data, e.g., the locally collected environmental data, via IoT transmission to the BS.
- We first consider the case when all the CSI is perfectly available at the BS. Besides, we study the joint active and passive beamforming to maximize the *weighted sum-rate* (WSR) of the primary and IoT transmissions subject to the phase constraints at the RISs.
- The formulated problem is non-convex with many complications such as coupling variables and fractional expressions. To overcome the difficulty, we first decompose the original problem into several disjoint blocks by the *fractional programming* (FP) technique [18]. Then, an algorithm based on the *block coordinate descent* (BCD) method [19] is proposed.
- The proposed algorithm introduces multiple auxiliary variables and repeatedly applies several techniques, which leads to high complexity in each iteration. To address it, we propose another BCD-based algorithm to reduce the optimization complexity in each iteration.
- Furthermore, we formulate another optimization problem that the CSI between the RISs and the users and between the RISs and the BS are imperfect. The formulation is more general than those in [20] and [21] as it does not rely on the assumption of channel models.
- To solve the challenging problem of imperfect CSI, we propose another BCD-based algorithm. The stochastic *success convex approximation* (SCA) technique [22] is applied, which is different from the conventional SCA method used in [19], [23], and [24].
- Simulation results show that the proposed system can achieve significant performance gain compared with different baseline schemes. The simulation results also demonstrate that the proposed system incurs only marginal performance loss when the CSI error is small.

Noted that different from the above conventional works on RIS [5]–[9], the RISs in our work transmit their information. The introduction of the secondary transmission from the RISs makes the passive and active beamforming more challenging to manage the interference between them so that both systems can mutually benefit. Unlike [16] and [17], which consider the downlink scenarios in RIS-based SR systems, we consider the uplink scenario. The scenario represents a more practical implementation of an environment-monitoring application for future smart cities. The uplink scenario has different rate expressions from the downlink one, leading to a different problem formulation and beamforming design. Besides, [16] and [17] consider perfect CSI. In contrast, in this paper, we take the channel estimation error into consideration and formulate two different problems including both the perfect and the imperfect CSI cases. This makes our system model more practical in this regard. It is worth mentioning that although both [17] and this paper adopt the BCD framework, the specific algorithms in [17] cannot be applied for solving the optimization problems proposed in this paper. This is because we have to deal with both perfect and imperfect CSI in this paper. To overcome the difficulty, we adopt completely different algorithms to optimize $\mathbf{W}$ and $\mathbf{V}$, utilizing techniques such as prox-linear BCD, SCA, and Armijo search. Then we extend the method for perfect CSI to the imperfect CSI case.

The rest of this paper is organized as follows. Section II presents the system model of the proposed system and the formulation of the WSR maximization problem. Section III proposes a BCD method to solve the formulated problem with perfect CSI. Section IV proposes another BCD method with low complexity under the assumption of perfect CSI. Section V extends the BCD method to the imperfect CSI case. Section VI provides simulation results to verify the effectiveness of the proposed algorithms. Section VII concludes the paper.

These notations are used in this paper. The scalar, vector, and matrix are lowercase, bold lowercase, and bold uppercase, i.e., a, $\mathbf{a}$, and $\mathbf{A}$, respectively. $(\cdot)^*$, $(\cdot)^{\mathrm{H}}$, $\mathrm{Tr}\,(\cdot)$, $\mathbb{E}(\cdot)$, $\angle(\cdot)$, $\mathrm{Re}(\cdot)$, and $\mathrm{Im}(\cdot)$ denote conjugation, conjugate transpose, trace, expectation, phase, real dimension, and imaginary dimension, respectively. $\mathcal{CN}\left(\mu, \sigma^2\right)$ denotes the distribution of a *circularly symmetric complex Gaussian* (CSCG) random variable with mean μ and variance σ^2.

II. System Model and Problem Formulation

A. System Model

As shown in Fig. 1, we consider the uplink of a RIS-based MU-MIMO SR communication system in which multiple RISs are deployed to assist the transmission from users to the BS and simultaneously act as IoT devices to transmit their information to the BS using backscatter modulation. We assume that each RIS serves one nearby user at a time. Different nearby users can be served by the associated RIS in a *time division multiple address* (TDMA) manner.

Specifically, M-antenna BS, K single-antenna users, and K RISs are considered. RIS i is equipped with N_i reflection elements ($N_i > 1$). For simplicity, we denote the user that is near RIS i as user i. By inducing phase shift on the incident signal at each reflection element and steering it to the BS, RIS i assists the primary communication from user i to BS, and at the same time, it also modulates its information

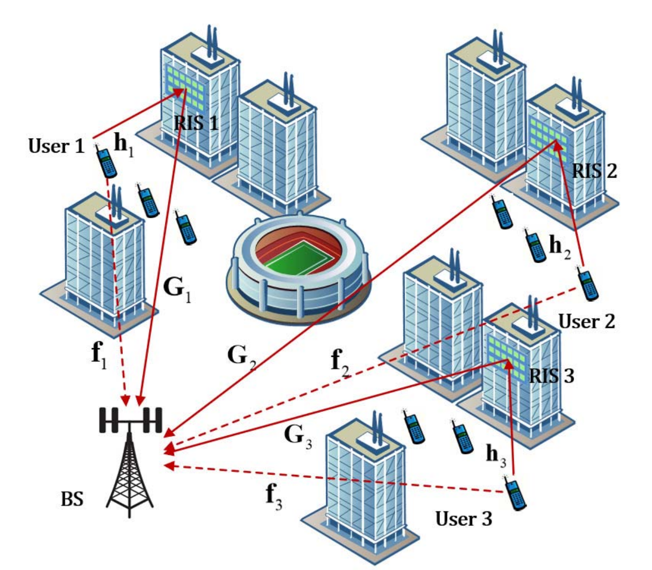

Fig. 1. RIS-based uplink MU-MIMO SR communication system.

onto the incident signals. Therefore, the IoT communication between RIS i and the BS is enabled, and no dedicated RF source at RIS i is required during this process. We consider that each RIS is equipped with a receiving module, so it can receive the control information of the BS to reconfigure the reflection elements. We assume that RISs are deployed sufficiently far away, e.g., on different streets. As a result, the interfering links from the RISs and other far users are very weak due to the high penetration and diffraction losses introduced by buildings and trees along the transmission paths. Besides, due to the double fading effect, the RIS should be deployed near the user to achieve better system performance [25]. As a result, the interference links of each user via the faraway RISs will be severely attenuated. Moreover, Once the phase-shift matrices of RISs are optimized, each RIS will form beam pairs with the corresponding user and the BS [25]. As an example, the beams of RIS i will point to user i and the BS. Therefore, the backscatter link from user i to BS via RIS i is enhanced, and the interfering links from other faraway users are further restrained. Therefore, in this paper, only the link between RIS i and user i is considered.

When the symbol rate of IoT signals transmitted by RISs is lower than that of the primary signal transmitted by users, better collaboration between the primary and IoT system can be obtained [26], leading to better performance. Therefore, we assume that every time RIS i transmits an IoT symbol to the BS, Q primary symbols are transmitted from user i. Denoting the symbol period of the user and the RIS as T_s and T_c, respectively, we have $T_c = QT_s$.

Denote the IoT signal of RIS i as c_i. We consider c_i uses *Binary phase-shift keying* (BPSK) modulation, i.e., $c_i = \{-1, 1\}$. Then the reflection matrix at RIS i can be expressed as $\mathbf{R}_i = \mathrm{diag}(r_{i1}, r_{i2}, \ldots, r_{iN_i}) = c_i\alpha_i\mathrm{diag}(v_{i1}, v_{i2}, \ldots, v_{iN_i}) \in \mathbb{C}^{N_i \times N_i}$, where r_{in} is the reflection coefficient of the n-th element at RIS i, v_{in} is the phase shift of the n-th element at RIS i, respectively, and

α_i ($|\alpha_i| \leq 1$) is the reflection efficiency of RIS i. For the ease of notation, we denote the phase-shift matrix at RIS i as $\mathbf{V}_i = \alpha_i \text{diag}(v_{i1}, v_{i2}, \ldots, v_{iN_i}) \in \mathbb{C}^{N_i \times N_i}$. Similar to the passive RIS designs in [27]–[29], we assume that $v_{in} = e^{j\theta_{in}}$ where $\theta_{in} \in [0, 2\pi)$. In this case, the maximum amplitude and the phase shift of the reflection coefficient r_{in} are $|\alpha_i|$ and 2π, respectively.

Let $\mathbf{f}_i \in \mathbb{C}^{M\times 1}$ denote the direct channel from user i to the BS, $\mathbf{h}_i \in \mathbb{C}^{N_i \times 1}$ denote the channel from user i to RIS i, and $\mathbf{G}_i \in \mathbb{C}^{N_i \times M}$ denote the channel from RIS i to the BS. For the ease of notation, and denote the backscatter channel from user i to the BS via RIS i as $\mathbf{b}_i$, i.e., $\mathbf{b}_i = \mathbf{G}_i^H \mathbf{V}_i \mathbf{h}_i \in \mathbb{C}^{M\times 1}$. This means when RIS i sends '+1', it uses $\mathbf{V}_i$ to reflect signals when RIS i sends '-1', it uses $-\mathbf{V}_i$ to reflect signals. In addition, we let $s_i(n)$ denote the n-th primary signal transmitted by user i during one symbol period of c_i. It is assumed that $s_i(n) \sim \mathcal{CN}(0, 1)$. Then the n-th uplink received signals at the BS during a period of T_c is

$$\mathbf{y}(n) = \sum_{i=1}^{K} \sqrt{P_i}(\mathbf{f}_i + c_i \mathbf{b}_i)s_i(n) + \mathbf{z}, \tag{1}$$

where $\mathbf{y}(n) = [y_1(n), y_2(n), \ldots, y_M(n)] \in \mathbb{C}^{M\times 1}$, $\mathbf{z} = [z_1, z_2, \ldots, z_M] \in \mathbb{C}^{M\times 1}$, and P_i denotes the transmit power of user i. Here, $y_l(n)$ denotes the n-th received signal at the l-th antenna of the BS, and $z_l \sim \mathcal{CN}(0, \sigma^2)$ denotes the noise at the l-th antenna of the BS.

To decode $s_i(n)$ and c_i for $i = 1, \ldots, K$, *successive interference cancellation* (SIC) technique is applied. Specifically, the BS first demodulates a set of $s_i(n)$ and the corresponding c_i with the best channel condition.[1] Then it cancels the demodulated symbols from (1) and demodulates the next set of $s_i(n)$ and c_i. The process will repeat until all $s_i(n)$ and c_i are demodulated. Without loss of generality, we use smaller index to denote the set of $s_i(n)$ and c_i with better channel condition. It means that the set of $s_1(n)$ and c_1 has the best channel condition and will be demodulated first, whereas the set of $s_K(n)$ and c_K has the worst channel condition and they will be demodulated last. Assuming perfect cancellation,[2] when BS decodes the set of $s_k(n)$ and c_k, the intermediate signal after receive beamforming is

$$y_k(n) = \sum_{i=k}^{K} \sqrt{P_i}\mathbf{w}_k^H(\mathbf{f}_i + c_i \mathbf{b}_i)s_i(n) + z_k, \tag{2}$$

where $\mathbf{w}_k \in \mathbb{C}^{M\times 1}$ is the receive beamforming vector for decoding $s_k(n)$ and c_k, and $z_k = \mathbf{w}_k^H \mathbf{z} \in \mathbb{C}$. It is seen that $s_i(n)$ and c_i for $i < k$ have been demodulated and canceled. However, in (2), $s_k(n)$ and c_k are coupled together, which makes the demodulation difficult. To address the problem, we adopt the cooperative *maximum-likelihood* (ML) detector proposed in [26], which can decode $s_k(n)$ and c_k jointly to attain better performance. The specific demodulation process is described as follows.

According to [16], when BS jointly decodes $s_k(n)$ and c_k with the ML detector, the backscatter link can be treated as an extra multi-path component for decoding $s_k(n)$. If c_k can be decoded perfectly, the conditional *signal-to-interference-plus-noise ratio* (SINR) for $s_k(n)$ is given by

$$\gamma_{\mathrm{s},k}(c_k) = \frac{\left|\sqrt{P_k}\mathbf{w}_k^H(\mathbf{f}_k + c_k \mathbf{b}_k)\right|^2}{\sum\limits_{i=k+1}^{K} \left|\sqrt{P_i}\mathbf{w}_k^H \mathbf{f}_i\right|^2 + \sum\limits_{i=k+1}^{K} \left|\sqrt{P_i}\mathbf{w}_k^H \mathbf{b}_i\right|^2 + \sigma_k^2}, \tag{3}$$

where $\sigma_k^2 = E\left\{\left|\mathbf{w}_k^H \mathbf{z}\right|^2\right\} = \mathbf{w}_k^H E\left\{\mathbf{z}\mathbf{z}^H\right\}\mathbf{w}_k = \sigma^2 \|\mathbf{w}_k\|_2^2$.

According to [11], when Q is large, the perfect decode of c_k is possible, thus the achievable rate of $s_k(n)$ can be approximated by taking an expectation over c_k, which is given by (4). Note that if Q is small, (4) should be viewed as the upper bound of $R_{\mathrm{s},k}$.

$$R_{\mathrm{s},k} \approx \mathbb{E}_{c_k}\left\{\log_2(1 + \gamma_{\mathrm{s},k}(c_k))\right\}. \tag{4}$$

To decode c_k, the *maximal ratio combing* (MRC) technique can be adopted since c_k covers Q symbol periods of the primary signal $s_k(n)$. In this case, according to [16], when $Q \gg 1$, the approximated SINR for c_k can be expressed as

$$\gamma_{\mathrm{c},k} \approx \frac{Q\left|\sqrt{P_k}\mathbf{w}_k^H \mathbf{b}_k\right|^2}{\sum\limits_{i=k+1}^{K} \left|\sqrt{P_i}\mathbf{w}_k^H \mathbf{f}_i\right|^2 + \sum\limits_{i=k+1}^{K} \left|\sqrt{P_i}\mathbf{w}_k^H \mathbf{b}_i\right|^2 + \sigma_k^2}. \tag{5}$$

With the MRC technique, although $\gamma_{\mathrm{c},k}$ is increased by Q times as shown in (5), the symbol rate of c_k is decreased by $\frac{1}{Q}$. Therefore, the achievable rate of c_k can be approximated as

$$R_{\mathrm{c},k} \approx \frac{1}{Q}\log_2\left(1 + \frac{1}{\Gamma}\gamma_{\mathrm{c},k}\right), \tag{6}$$

where Γ is the SINR gap to capacity [32], which is introduced to represent the capacity reduction due to the lower-order BPSK modulation used.

Remark 1: It is seen that the received SINR of c_k (5) increases as Q increases because of spreading gain. However, it is seen that the achievable rate of c_k (6) decreases as Q increases. Therefore, there is a trade-off between the data rate and reliability of transmitting $c_k(n)$ when we set the Q value. The design of Q is an interesting and complex topic in SR systems. We refer the readers to [33] for more insights on the selection of Q values. In [33], the authors have derived the expression of the minimum Q that enables the primary transmission to gain benefits from the existence of the secondary transmission and analyzed the trade-off between the data rate and reliability in SR for arbitrary K.

B. Frame Structure and Channel Estimation

Based on [16], we develop a new frame structure as shown in Fig. 2 to support the RIS-based MU-MIMO SR system,

[1] As a suboptimal example, we use $\tau_i = \left|\sqrt{P_i}\mathbf{f}_i\right|^2 + \left|\sqrt{P_i}\mathbf{b}_i\right|^2$ to denote the quality of channel condition of each set.

[2] Note that decoding error might be inevitable during the procedure of SIC [30], thus only imperfect SIC can be performed. As the imperfect-SIC interference is a complicated function of multiple factors and the characteristics of error propagation, it is difficult to model the impact of imperfect SIC in our system setup [31]. Therefore we still adopt the perfect-SIC assumption in this paper, which is similar to [12]–[17]. The result of this paper can be viewed as the upper bound of the imperfect-SIC case.

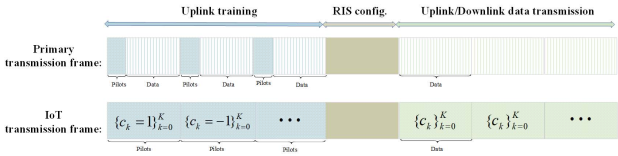

Fig. 2. The transmission frame for RIS-based MU-MIMO SR system.

where the users and RISs jointly send pilots to estimate channels.

In the uplink training slot, each set of RIS and the user and RIS takes turns sending pilots, and then the relevant channels are estimated. Specifically, the combined channel $\mathbf{h}_{\text{eq},k} = \mathbf{f}_k + c_k\mathbf{b}_k$ where $\mathbf{h}_{\text{eq},k} \in \mathbb{C}^{M\times 1}$ is first estimated via the conventional channel estimation method and protocol [34]. According to **Remark 2**, with the knowledge of $\mathbf{h}_{\text{eq},k}$, $\mathbf{f}_k$ and $\mathbf{b}_k$ can also be estimated based on the IoT transmission enabled by the SR system. Next, in order to configure the phase-shift matrix at RIS k, the CSI between the BS and RIS k ($\mathbf{G}_k$) and that between RIS k and user k ($\mathbf{h}_k$) should be estimated separately, e.g., using the techniques introduced in [8], [35], or [36]. Note that it is hard to estimate the cascaded channel with high accuracy because the RIS, unlike the BS or the users, has limited signal processing capability [17]. Besides, the number of RIS elements is usually large, which also increases the difficulty of channel estimation [8].[3]

In the RIS configuration slot, with the knowledge of CSI, the received beamforming matrix $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_K] \in \mathbb{C}^{M\times K}$ and the phase-shift matrix $\mathbf{V} = [\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_K]$ can be optimized jointly using the proposed algorithms. Then, each RIS is reconfigured accordingly. After that, with the optimized $\mathbf{W}^*$ and $\mathbf{V}^*$, the uplink and downlink data transmission starts.[4]

Remark 2: As shown in Fig. 2, we let RIS k send '$+1$' and '-1'. When RIS k sends '$+$', i.e., $c_k = 1$, we have $\mathbf{h}_{\text{eq},k} = \mathbf{f}_k + \mathbf{b}_k$. Similarly, when RIS k sends '-1', i.e., $c_k = -1$, we have $\mathbf{h}_{\text{eq},k} = \mathbf{f}_k - \mathbf{b}_k$. Then, with a simple calculation, we can obtain $\mathbf{f}_k$ and $\mathbf{b}_k$ separately based on the knowledge of $\mathbf{h}_{\text{eq},k}$.

C. Problem Formulation

This paper aims to maximize the WSR of all the primary and IoT signals by jointly designing the receive beamforming vector at the BS and the phase-shift matrix at each RIS. Two optimization problems are formulated, one for the case of perfect CSI and the other for the case of imperfect CSI.

[3]It is worth mentioning that if we adopt only the compressed sensing methods introduced in [35] and [36], we can still obtain the combined channels from the estimated cascaded channels. However, as the channel estimation time is limited, it is difficult for the compressed sensing method to estimate the cascaded channels in high accuracy with short training pilots [37]. In this case, the combined channels will be affected by the error in the estimated cascaded channels or it may take a long time for channel estimation to reach higher accuracy. Therefore, we decide to use different algorithms to estimate the combined channels and the cascaded channels and thus assume differences in the estimation accuracy of different channels. Based on this, different CSI models are considered for different channels, which is shown in Section II-C.2.

[4]Note that, similar to [16], in the downlink transmission slot, training pilots are still needed to make the downlink users know the knowledge of optimized $\mathbf{W}^*$ and $\mathbf{V}^*$, which are omitted in Fig. 2.

1) Perfect CSI: First, we assume that the CSI of all the channels is perfectly known at the BS. This assumption is similar to [5]–[9]. The result obtained under the assumption of perfect CSI can be considered as an upper bound to the case when such CSI information is only known imperfectly. We have $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_K]$ and $\mathbf{V} = [\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_K]$, thus the WSR maximization problem can be formulated as

$$\underline{\textbf{P1:}} \quad \max_{\mathbf{W},\mathbf{V}} \quad f_{\text{a}}(\mathbf{W},\mathbf{V}) = \sum_{k=1}^{K} \rho_{\text{s},k} R_{\text{s},k} + \sum_{k=1}^{K} \rho_{\text{c},k} R_{\text{c},k}$$
$$\text{s.t.} \quad |v_{kn}| = 1, 1 \le k \le K, 1 \le n \le N_k, \tag{7}$$
$$\|\mathbf{w}_k\|_2^2 \le 1, 1 \le k \le K, \tag{8}$$

where $\rho_{\text{s},k}$ and $\rho_{\text{c},k}$ are the weighting factors for $s_k(n)$ and c_k, respectively.

As shown in (4) and (6), fractional terms are involved in $R_{\text{s},k}$ and $R_{\text{c},k}$, which makes the objective function non-convex. Besides, the coupling effect of $\mathbf{W}$ and $\mathbf{V}$ as well as the mathematical operations of expectation involved in $R_{\text{s},k}$ further complicate the optimization. Therefore, it is challenging to find a global optimal solution. Besides, although P1 could be solved by the semi-definite relaxation technique, the computational complexity for optimizing $\mathbf{V}_k$ reaches $\mathcal{O}(N_k^6)$ [37], which is unacceptable as N_k is usually large in reality. Therefore, our objective is to find a low-complexity suboptimal algorithm that can solve P1 effectively.

2) Imperfect CSI: In this case, we adopt the imperfect CSI setup introduced in [8]. Specifically, as mentioned in Section II-B, it is difficult to estimate the cascaded channels separately with high accuracy. Thus the CSI of $\mathbf{G}_k$ and $\mathbf{h}_k$ is assumed to be imperfect. And the CSI of $\mathbf{f}_k$ and $\mathbf{b}_k$ is still assumed to be perfect for simplifying the model.

Similar to [8], to measure the impact of channel estimation errors of $\mathbf{G}_k$ and $\mathbf{h}_k$, we let $\hat{\mathbf{h}}_k \in \mathbb{C}^{N_k\times 1}$ and $\hat{\mathbf{G}}_k \in \mathbb{C}^{N_k\times M}$ denote the estimation of $\mathbf{h}_k$ and $\mathbf{G}_k$, respectively, and let $\Delta\mathbf{h}_k = \mathbf{h}_k - \hat{\mathbf{h}}_k \in \mathbb{C}^{N_k\times 1}$ and $\Delta\mathbf{G}_k = \mathbf{G}_k - \hat{\mathbf{G}}_k \in \mathbb{C}^{N_k\times M}$ denote the channel estimation errors of $\mathbf{h}_k$ and $\mathbf{G}_k$, respectively. According to [38] and [39], the distribution of $\Delta\hat{\mathbf{h}}_k$ and $\Delta\hat{\mathbf{G}}_k$ are independent of the estimated channels $\hat{\mathbf{h}}_k$ and $\hat{\mathbf{G}}_k$ when minimum mean square error estimation is applied. Therefore, $\Delta\mathbf{h}_k$ and $\Delta\mathbf{G}_k$ can be modeled as the realizations from $\mathcal{F} = \{\Delta\mathbf{h}_k(\varsigma), \Delta\mathbf{G}_k(\varsigma), \forall k, \forall\varsigma\}$, where ς denotes the index of random realizations.

Under the above assumption of frame structure and imperfect CSI model, the WSR maximization problem can be expressed as

$$\underline{\textbf{P2:}} \quad \max_{\mathbf{V}} \quad f_{\text{b}}(\mathbf{V}) = \mathbb{E}_{\varsigma}\left\{\max_{\mathbf{W}} f_{\text{a}}(\mathbf{W},\mathbf{V};\varsigma)\right\}$$
$$\text{s.t.} \quad |v_{kn}| = 1, 1 \le k \le K, 1 \le n \le N_k, \tag{9}$$
$$\|\mathbf{w}_k\|_2^2 \le 1, 1 \le k \le K, \ \forall\varsigma. \tag{10}$$

P2 is a stochastic optimization problem with inner-layer variable $\mathbf{W}$ and outer-layer variable $\mathbf{V}$, and both the inner-layer and outer-layer subproblems are non-convex with no closed-form solutions. Specifically, in the inner optimization concerning $\mathbf{W}$, the objective function is non-convex and the operation of mathematical expectation is involved. In the outer optimization concerning $\mathbf{V}$, another expectation operator is involved in the objective function, and even worse, the probability density function of $\mathcal{F}$ is usually very complicated and has no closed-form expression. Therefore, it is challenging to find an algorithm to solve P2.

III. Block Coordinate Descent Method With Perfect CSI

In this section, we solve P1 with the BCD method. BCD method is widely used for minimizing a continuous function of multiple block variables. The basic idea of BCD is to minimize the objective function concerning one block of variables while the other blocks are held fixed. To solve P1, we first transform the non-convex objective function to a form that is suitable for the BCD method based on several different fractional programming techniques such as Lagrange dual transformation and quadratic transformation [18]. Then we decompose the original problem into two subproblems with the BCD method and solve them one by one.

A. Fractional Programming

Firstly, as message c_k adopts the BPSK modulation scheme, the expectation operation involved in $f_{\rm a}(\mathbf{W},\mathbf{V})$ can be replaced with the statistical average. Next, with the Lagrangian dual transform [18], the logarithm in $f_{\rm a}(\mathbf{W},\mathbf{V})$ is eliminated, and P1 is equivalently written as

$$\underline{\textbf{P1-a:}} \quad \max_{\mathbf{W},\mathbf{V},\boldsymbol{\beta}} \quad f_{\rm r}(\mathbf{W},\mathbf{V},\boldsymbol{\beta}) \quad \text{s.t.} \quad (7),(8),$$

where $\boldsymbol{\beta}=\{\beta_{{\rm s},kj},\beta_{{\rm c},k}\}_{k=1,2,\dots,K,}^{j=1,2,}$ is a set of auxiliary variables, and $f_{\rm r}(\mathbf{W},\mathbf{V},\boldsymbol{\beta})$ is given by

$$\begin{aligned} &f_{\rm r}(\mathbf{W},\mathbf{V},\boldsymbol{\beta})\\ &=\sum_{k=1}^{K}\sum_{j=1}^{2}\frac{\rho_{{\rm s},k}}{2}\log_2\left(1+\beta_{{\rm s},kj}\right)-\sum_{k=1}^{K}\sum_{j=1}^{2}\frac{\rho_{{\rm s},k}}{2}\beta_{{\rm s},kj}\\ &\quad+\sum_{k=1}^{K}\frac{\rho_{{\rm c},k}}{Q}\log_2\left(1+\beta_{{\rm c},k}\right)-\sum_{k=1}^{K}\frac{\rho_{{\rm c},k}}{Q}\beta_{{\rm c},k}\\ &\quad+\sum_{k=1}^{K}\sum_{j=1}^{2}\frac{\rho_{{\rm s},k}}{2}\left(1+\beta_{{\rm s},kj}\right)\frac{\left|\sqrt{P_k}\mathbf{w}_k^H\left(\mathbf{f}_k+\tilde{c}_j\mathbf{b}_k\right)\right|^2}{\left|\sqrt{P_k}\mathbf{w}_k^H\left(\mathbf{f}_k+\tilde{c}_j\mathbf{b}_k\right)\right|^2+n_k}\\ &\quad+\sum_{k=1}^{K}\frac{\rho_{{\rm c},k}}{\Gamma}\left(1+\beta_{{\rm c},k}\right)\frac{\left|\mathbf{b}_k^H\mathbf{w}_k\right|^2}{\frac{Q}{\Gamma}\left|\mathbf{b}_k^H\mathbf{w}_k\right|^2+n_k}, \end{aligned} \tag{11}$$

where $\tilde{c}_1=1$, $\tilde{c}_2=-1$, and $n_k=\sum_{i=k+1}^{K}\left|\sqrt{P_i}\mathbf{w}_k^H\mathbf{f}_i\right|^2+\sum_{i=k+1}^{K}\left|\sqrt{P_i}\mathbf{w}_k^H\mathbf{b}_i\right|^2+\sigma_k^2$. As $f_{\rm r}(\mathbf{W},\mathbf{V},\boldsymbol{\beta})$ is concave for each $\beta_{{\rm s},kj}$ or $\beta_{{\rm c},k}$, by setting $\frac{\partial f_r(\mathbf{W},\mathbf{V},\boldsymbol{\beta})}{\partial\beta_{{\rm c},k}}=0$ or $\frac{\partial f_r(\mathbf{W},\mathbf{V},\boldsymbol{\beta})}{\partial\beta_{{\rm s},kj}}=0$, the optimal $\boldsymbol{\beta}$ is:

$$\boldsymbol{\beta}^o=\left\{\begin{array}{l}\beta_{{\rm c},k}^o=\frac{Q}{\Gamma}\frac{\left|\sqrt{P_k}\mathbf{w}_k^H\mathbf{b}_k\right|^2}{n_k},\\ \beta_{{\rm s},kj}^o=\frac{\left|\sqrt{P_k}\mathbf{w}_k^H(\mathbf{f}_k+\tilde{c}_j\mathbf{b}_k)\right|^2}{n_k}.\end{array}\right\}_{\substack{j=1,2,\\ k=1,2,\dots,K.}} \tag{12}$$

As discussed above, we have eliminated both the expectation and logarithm operator involved in P1 and equivalently transformed P1 into P1-a. Next, based on the technique of quadratic transformation [18], we will apply the BCD method to optimize $\mathbf{W}$ and $\mathbf{V}$ successively.

B. The Optimization of W

In this subsection, the optimization of $\mathbf{W}$ is considered with the given value of $\boldsymbol{\beta}$ and $\mathbf{V}$.

It is seen that $\mathbf{W}$ is only involved in the last two non-convex fractional terms of $f_{\rm r}(\mathbf{W},\mathbf{V},\boldsymbol{\beta})$. Therefore, we apply the quadratic transformation to deal with the fractional terms, then the equivalent optimization problem of P1-a for given values of $\boldsymbol{\beta}$ and $\mathbf{V}$ is written as

$$\underline{\textbf{P1-b:}} \quad \max_{\mathbf{W},\boldsymbol{\xi}} \quad f_{\rm ql}(\mathbf{W},\boldsymbol{\xi}) \quad \text{s.t.} \quad (8),$$

where $\boldsymbol{\xi}=\{\xi_{{\rm s},jk},\xi_{{\rm c},k}\}_{k=1,2,\dots,K,}^{j=1,2,}$ is another set of auxiliary variables, and $f_{\rm ql}(\mathbf{W},\boldsymbol{\xi})$ is given by

$$\begin{aligned} &f_{\rm ql}(\mathbf{W},\boldsymbol{\xi})\\ &=\mathrm{const}(\boldsymbol{\beta})+\sum_{k=1}^{K}[-|\xi_{{\rm c},k}|^2(\frac{Q\left|\sqrt{P_k}\mathbf{w}_k^H\mathbf{b}_k\right|^2}{\Gamma}+n_k)\\ &\quad+2\mathrm{Re}\{\xi_{{\rm c},k}^*\sqrt{\frac{P_k\rho_{{\rm c},k}\left(1+\beta_{{\rm c},k}\right)}{\Gamma}}\mathbf{w}_k^H\mathbf{b}_k\}]\\ &\quad+\sum_{k=1}^{K}\sum_{j=1}^{2}[2\mathrm{Re}\{\xi_{{\rm s},kj}^*\sqrt{\frac{P_k\rho_{{\rm s},k}\left(1+\beta_{{\rm s},kj}\right)}{2}}\mathbf{w}_k^H\left(\mathbf{f}_k+\tilde{c}_j\mathbf{b}_k\right)\}\\ &\quad-|\xi_{{\rm s},kj}|^2(\left|\sqrt{P_k}\mathbf{w}_k^H(\mathbf{f}_k+\tilde{c}_j\mathbf{b}_k)\right|^2+n_k)]. \end{aligned} \tag{13}$$

It can be proved that P1-b is convex for each $\xi_{{\rm s},kj}$ or $\xi_{{\rm c},k}$. Thus, by setting $\frac{\partial f_{\rm ql}(\mathbf{W},\boldsymbol{\xi})}{\partial\xi_{{\rm c},k}}=0$ and $\frac{\partial f_{\rm ql}(\mathbf{W},\boldsymbol{\xi})}{\partial\xi_{{\rm s},kj}}=0$, the optimal solution of $\boldsymbol{\xi}$ can be obtained as

$$\boldsymbol{\xi}^o=\left\{\begin{array}{l}\xi_{{\rm c},k}^o=\frac{\sqrt{\frac{P_k\rho_{{\rm c},k}\left(1+\beta_{{\rm c},k}\right)}{\Gamma}}\mathbf{w}_k^H\mathbf{b}_k}{(Q/\Gamma)\left|\sqrt{P_k}\mathbf{w}_k^H\mathbf{b}_k\right|^2+n_k},\\ \xi_{{\rm s},kj}^o=\frac{\sqrt{\frac{P_k\rho_{{\rm s},k}\left(1+\beta_{{\rm s},kj}\right)}{2}}\mathbf{w}_k^H\left(\mathbf{f}_k+\tilde{c}_j\mathbf{b}_k\right)}{\left|\sqrt{P_k}\mathbf{w}_k^H(\mathbf{f}_k+\tilde{c}_j\mathbf{b}_k)\right|^2+n_k}.\end{array}\right\}_{\substack{j=1,2,\\ k=1,2,\dots,K.}} \tag{14}$$

As P1-b is also convex for each $\mathbf{w}_k$, $\mathbf{W}$ can be obtained with the optimal value of $\boldsymbol{\xi}$ as

$$\mathbf{w}_k^o=\begin{cases}\mathbf{B}_{{\rm w},k}^{-1}\mathbf{a}_{{\rm w},k}, & \text{if } \left\|\mathbf{B}_{{\rm w},k}^{-1}\mathbf{a}_{{\rm w},k}\right\|\le 1\\ \mathbf{B}_{{\rm w},k}^{-1}\mathbf{a}_{{\rm w},k}/\left\|\mathbf{B}_{{\rm w},k}^{-1}\mathbf{a}_{{\rm w},k}\right\|, & \text{otherwise}\end{cases} \tag{15}$$

where $\mathbf{a}_{{\rm w},k}=\sum_{j=1}^{2}\xi_{{\rm s},kj}^*\sqrt{\frac{P_k\rho_{{\rm s},k}(1+\beta_{{\rm s},kj})}{2}}\left(\mathbf{f}_k+\tilde{c}_j\mathbf{b}_k\right)+\xi_{{\rm c},k}^*\sqrt{\frac{P_k\rho_{{\rm c},k}(1+\beta_{{\rm s},kj})}{\Gamma}}\mathbf{b}_k$, $\mathbf{B}_{{\rm w},k}=\sum_{j=1}^{2}|\xi_{{\rm s},kj}|^2[P_k(\mathbf{f}_k+\tilde{c}_j\mathbf{b}_k)(\mathbf{f}_k+\tilde{c}_j\mathbf{b}_k)^H+\delta_k]+|\xi_{{\rm c},k}|^2[(\frac{QP_k\mathbf{b}_k\mathbf{b}_k^H}{\Gamma})+\delta_k]$, where $\delta_k=\sum_{i=k+1}^{K}P_i(\mathbf{f}_i\mathbf{f}_i^H+\mathbf{b}_i\mathbf{b}_i^H)+\sigma^2 I_M$.

As discussed above, with the given value of $\boldsymbol{\beta}$ and $\mathbf{V}$, the optimal value of $\boldsymbol{\xi}$ is obtained according to (14). With the optimal value of $\boldsymbol{\xi}$, the optimal value of $\mathbf{W}$ can be obtained according to (15). In the next subsection, we deal with the optimization of $\mathbf{V}$.

C. Successive Convex Approximation Method for V

In this subsection, $\mathbf{V}$ is optimized with the given value of $\boldsymbol{\beta}$ and $\mathbf{W}$. We apply the SCA method [19] to deal with the non-convexity. The specific process is discussed as follows.

Similar to Section III-B, the technique of quadratic transformation is applied to deal with the non-convex terms. Thus, with the given value of $\boldsymbol{\beta}$ and $\mathbf{W}$, P1-a is equivalently transformed into

$$\underline{\textbf{P1-c:}} \quad \max_{\mathbf{V},\boldsymbol{\varepsilon}} \quad f_{\mathrm{q2}}(\mathbf{V},\boldsymbol{\varepsilon}) \quad \text{s.t.} \quad (7),$$

where $\boldsymbol{\varepsilon} = \{\varepsilon_{\mathrm{s},jk}, \varepsilon_{\mathrm{c},k}\}_{k=1,2,\ldots,K,}^{j=1,2,}$ is a new set of auxiliary variables, and $f_{\mathrm{q2}}(\mathbf{V},\boldsymbol{\varepsilon})$ is given by

$$\begin{aligned}
&f_{\mathrm{q2}}(\mathbf{V},\boldsymbol{\varepsilon}) \\
&= \mathrm{const}(\boldsymbol{\beta}) + \sum_{k=1}^{K}[-|\varepsilon_{\mathrm{c},k}|^2(\frac{Q\left|\mathbf{p}_{kk}^H\mathbf{v}_k\right|^2}{\Gamma} + n_k) \\
&\quad +2\mathrm{Re}\{\varepsilon_{\mathrm{c},k}^*\sqrt{\frac{\rho_{\mathrm{c},k}\left(1+\beta_{\mathrm{c},k}\right)}{\Gamma}}\mathbf{p}_{kk}^H\mathbf{v}_k)\}] \\
&\quad +\sum_{k=1}^{K}\sum_{j=1}^{2}[2\mathrm{Re}\{\varepsilon_{\mathrm{s},kj}^*\sqrt{\frac{\rho_{\mathrm{s},k}\left(1+\beta_{\mathrm{s},kj}\right)}{2}}(B_{kk}+\tilde{c}_j\mathbf{p}_{kk}^H\mathbf{v}_k)\} \\
&\quad -|\varepsilon_{\mathrm{s},kj}|^2(\left|B_{kk}+\tilde{c}_j\mathbf{p}_{kk}^H\mathbf{v}_k\right|^2 + n_k)], \qquad (16)
\end{aligned}$$

where $B_{ki} = \sqrt{P_k}\mathbf{w}_k^H\mathbf{f}_i$, $\mathbf{p}_{ki}^H = \sqrt{\alpha_i}\sqrt{P_k}\mathbf{w}_k^H\mathbf{G}_i^H\mathrm{diag}(\mathbf{h}_i)$, $n_k = \sum_{i=k+1}^{K}\left|\mathbf{p}_{ki}^H\mathbf{v}_i\right|^2 + \sum_{i=k+1}^{K}|B_{ki}|^2 + \sigma_k^2$, and $\mathbf{v}_i = [v_{i1}, v_{i2}, \ldots, v_{iN_i}]^H \in \mathbb{C}^{N_i\times 1}$ denotes the passive beamforming vector at RIS i.

Similarly, by setting $\frac{\partial f_{\mathrm{q2}}(\mathbf{V},\boldsymbol{\varepsilon})}{\partial \varepsilon_{\mathrm{c},k}} = 0$ and $\frac{\partial f_{\mathrm{q2}}(\mathbf{V},\boldsymbol{\varepsilon})}{\partial \varepsilon_{\mathrm{s},kj}} = 0$, the optimal value of $\boldsymbol{\varepsilon}$ can be obtained as

$$\boldsymbol{\varepsilon}^o = \left\{\begin{array}{l} \varepsilon_{\mathrm{c},k}^o = \frac{\sqrt{\frac{\rho_{\mathrm{c},k}\left(1+\beta_{\mathrm{c},k}\right)}{\Gamma}}\mathbf{p}_{kk}^H\mathbf{v}_k}{\frac{Q}{\Gamma}\left|\mathbf{p}_{kk}^H\mathbf{v}_k\right|^2+n_k}, \\ \varepsilon_{\mathrm{s},kj}^o = \frac{\sqrt{\frac{\rho_{\mathrm{s},k}\left(1+\beta_{\mathrm{s},kj}\right)}{2}}B_{kk}+\tilde{c}_j\mathbf{p}_{kk}^H\mathbf{v}_k}{\left|B_{kk}+\tilde{c}_j\mathbf{p}_{kk}^H\mathbf{v}_k\right|^2+n_k}. \end{array}\right\}_{k=1,2,\ldots,K.}^{j=1,2,} \qquad (17)$$

Next, with the given value of $\boldsymbol{\varepsilon}$, P1-c can be equivalently transformed to

$$\underline{\textbf{P1-c-V:}} \quad \min_{\mathbf{V}} \quad -f_{\mathrm{q2}}(\mathbf{V}) \quad \text{s.t.} \quad (7).$$

Similar to [37], it is easily proved that $f_{\mathrm{q2}}(\mathbf{V})$ can be transformed into a much simpler form by combining the linear terms and the quadratic terms of $\mathbf{v}_k$, which is given by

$$f_{\mathrm{q2}}(\mathbf{V}) = -\sum_{k=1}^{K}\mathbf{v}_k^H\mathbf{U}_k\mathbf{v}_k + \sum_{k=1}^{K}2\mathrm{Re}\{\mathbf{d}_k^H\mathbf{v}_k\} + \mathrm{const}, \quad (18)$$

where

$$\begin{aligned}
\mathbf{U}_k &= \sum_{j=1}^{2}|\varepsilon_{\mathrm{s},kj}|^2\mathbf{p}_{kk}\mathbf{p}_{kk}^H + \frac{Q|\varepsilon_{\mathrm{c},k}|^2\mathbf{p}_{kk}\mathbf{p}_{kk}^H}{\Gamma} \\
&\quad + \sum_{i=1}^{k-1}\mathbf{p}_{ik}\mathbf{p}_{ik}^H(|\varepsilon_{\mathrm{c},i}|^2 + \sum_{j=1}^{2}|\varepsilon_{\mathrm{s},ij}|^2), \qquad (19) \\
\mathbf{d}_k &= \sum_{j=1}^{2}\tilde{c}_j(\varepsilon_{\mathrm{s},kj}\sqrt{\frac{\rho_{\mathrm{s},k}\left(1+\beta_{\mathrm{s},kj}\right)}{2}}\mathbf{p}_{kk} - |\varepsilon_{\mathrm{s},kj}|^2B_{kk}\mathbf{p}_{kk}) \\
&\quad + \varepsilon_{\mathrm{c},k}\sqrt{\frac{\rho_{\mathrm{c},k}\left(1+\beta_{\mathrm{c},k}\right)}{\Gamma}}\mathbf{p}_{kk}. \qquad (20)
\end{aligned}$$

It is seen that the optimization of $\mathbf{v}_k$ is independent of each other. Therefore, by replacing $\mathbf{v}_k$ with $\boldsymbol{\theta}_k = [e^{j\theta_{k1}}, e^{j\theta_{k2}}, \ldots, e^{j\theta_{kN_k}}]^H \in \mathbb{C}^{N_k\times 1}$, P1-c-V can be transformed to K unconstrained subproblems, and the k-th subproblem given by

$$\underline{\textbf{P1-d-(}k\textbf{):}} \quad \min_{\boldsymbol{\theta}_k\in\mathbb{C}^{N_k}} \quad f_{\mathrm{q3},k}(\boldsymbol{\theta}_k),$$

where $f_{\mathrm{q3},k}(\boldsymbol{\theta}_k) \triangleq (e^{j\boldsymbol{\theta}_k})^H\mathbf{U}_k e^{j\boldsymbol{\theta}_k} - 2\mathrm{Re}\left\{\mathbf{d}_k^H e^{j\boldsymbol{\theta}_k}\right\}$.

As $f_{\mathrm{q3},k}(\boldsymbol{\theta}_k)$ is still non-convex, we adopt the SCA method [19] to optimize $\boldsymbol{\theta}_k$. According to Proposition 1 of [19], as $f_{\mathrm{q3},k}(\boldsymbol{\theta}_k)$ is continuously differentiable, we can construct $f_{\mathrm{q4},k}(\boldsymbol{\theta}_k,\bar{\boldsymbol{\theta}}_k)$ as the approximate function of $f_{\mathrm{q3},k}(\boldsymbol{\theta}_k)$, as long as the following conditions are satisfied:

$$f_{\mathrm{q4},k}(\bar{\boldsymbol{\theta}}_k,\bar{\boldsymbol{\theta}}_k) = f_{\mathrm{q3},k}(\bar{\boldsymbol{\theta}}_k), \qquad (21)$$

$$f_{\mathrm{q4},k}(\boldsymbol{\theta}_k,\bar{\boldsymbol{\theta}}_k) \geq f_{\mathrm{q3},k}(\boldsymbol{\theta}_k), \qquad (22)$$

where $\bar{\boldsymbol{\theta}}_k$ is the value of $\boldsymbol{\theta}_k$ at the previous iteration.

According to (21) and (22), by applying the second-order Taylor expansion, we can obtain an approximate function of $f_{\mathrm{q3},k}(\boldsymbol{\theta}_k)$ as $f_{\mathrm{q4},k}(\boldsymbol{\theta}_k,\bar{\boldsymbol{\theta}}_k) = f_{\mathrm{q3},k}(\bar{\boldsymbol{\theta}}_k) + \nabla f_{\mathrm{q3},k}(\bar{\boldsymbol{\theta}}_k)^H(\boldsymbol{\theta}_k-\bar{\boldsymbol{\theta}}_k) + \frac{\mu_k}{2}\left\|\boldsymbol{\theta}_k-\bar{\boldsymbol{\theta}}_k\right\|^2$, where $\nabla f_{\mathrm{q3},k}(\bar{\boldsymbol{\theta}}_k) = 2\mathrm{Re}\left\{-j\bar{\boldsymbol{\theta}}_k^*\circ(\mathbf{U}_k\bar{\boldsymbol{\theta}}_k-\mathbf{d}_k)\right\}$, and the value of μ_k should satisfy (22).[5] With the approximate function $f_{\mathrm{q4},k}(\boldsymbol{\theta}_k,\bar{\boldsymbol{\theta}}_k)$, we can formulate the surrogate problem of P1-d-(k) as

$$\underline{\textbf{P1-e-(}k\textbf{):}} \quad \min_{\boldsymbol{\theta}_k\in\mathbb{C}^{N_k}} \quad f_{\mathrm{q4},k}(\boldsymbol{\theta}_k,\bar{\boldsymbol{\theta}}_k).$$

According to [19], we can still obtain the stationary solution of P1-d-(k) by solving P1-e-(k). As P1-e-(k) is convex for $\boldsymbol{\theta}_k$, we can obtain the update rule of $\boldsymbol{\theta}_k$ as

$$\boldsymbol{\theta}_k = \bar{\boldsymbol{\theta}}_k - \frac{\nabla f_{\mathrm{q3},k}(\bar{\boldsymbol{\theta}}_k)}{\mu_k}. \qquad (23)$$

Besides, we can accelerate the convergence of the proposed BCD method by carefully choosing a proper search step $\frac{1}{\mu_k}$ [8]. Specifically, we can design μ_k based on the Armijo rule [41]:

$$f_{\mathrm{q3},k}(\bar{\boldsymbol{\theta}}_k) - f_{\mathrm{q3},k}(\boldsymbol{\theta}_k) \geq \upsilon_k\frac{1}{\mu_k}\left\|\nabla f_{\mathrm{q3},k}(\bar{\boldsymbol{\theta}}_k)\right\|_2^2, \qquad (24)$$

where $0 < \upsilon_k < 0.5$, and $\mu_k = \min\{\frac{1}{\mu_0^j}\}_{j=0,1,2,\ldots}$, i.e., $j = \min_j\left\{\frac{1}{\mu_0^j}\right\}$, where $\mu_0 < 1$.

D. Discussion

The proposed BCD method is summarized in **Algorithm 1** and described as follows. First, $\mathbf{W}$ and $\mathbf{V}$ are initialized with $\mathbf{W}^{(0)}$ and $\mathbf{V}^{(0)}$, respectively. Then they are updated iteratively. In the t-th iteration, $\mathbf{W}^{(t)}$ is updated with the values of $\mathbf{W}^{(t-1)}$ and $\mathbf{V}^{(t-1)}$. While updating $\mathbf{W}^{(t)}$, a set of auxiliary variables $\boldsymbol{\beta}^{(t)}$, $\boldsymbol{\xi}^{(t)}$, and $\boldsymbol{\varepsilon}^{(t)}$ are also updated. The set of operations for updating $\boldsymbol{\beta}^{(t)}$, $\boldsymbol{\xi}^{(t)}$, $\mathbf{W}^{(t)}$, and $\boldsymbol{\varepsilon}^{(t)}$ is grouped into **functionA**

[5]We adopt the second-order Taylor expansion to ensure the convexity of the surrogate function, and use constraints (21) and (22) to ensure that the surrogate function is the global upper bound of the original function. In this case, according to [40], the SCA method for solving P1-d-(k) is equivalent to the *Majorize-Minimization* (MM) method.

as shown in **Algorithm 2**. With update $\mathbf{W}^{(t)}$, the phase-shift matrix for each RIS is updated one by one. Note that once the updated value of the phase-shift matrix $\mathbf{V}_k^{(t)}$ for RIS k is obtained, the beamforming matrix $\mathbf{W}^{(t)}$ is updated again. Once all the phase-shift matrices are updated, the values of $\mathbf{W}^{(t)}$ and $\mathbf{V}^{(t)} = [\mathbf{V}_1^{(t)}, \mathbf{V}_2^{(t)}, \ldots, \mathbf{V}_K^{(t)}]$ are fed into the $(t+1)$-th iteration. Then the above-described process repeats.

Algorithm 1 BCD Method for the Perfect CSI Case

1: Initialize $\mathbf{V}^{(0)}$ and $\mathbf{W}^{(0)}$ to feasible values; Let $t = 0$;
2: **repeat**
3: $\quad t = t + 1$;
4: $\quad [\tilde{\boldsymbol{\beta}}^{(t)}, \tilde{\boldsymbol{\xi}}^{(t)}, \tilde{\mathbf{W}}^{(t)}, \tilde{\boldsymbol{\varepsilon}}^{(t)}] = \textbf{functionA}(\mathbf{V}^{(t-1)},\ \mathbf{W}^{(t-1)})$.
5: $\quad$ **for** k=1,2,…,K **do**
6: $\qquad$ Update $\mathbf{U}_k$ and $\mathbf{d}_k$ according to (19) and (20) for all k;
7: $\qquad$ Design $\mu_k^{(t)}$ with Armijo rule shown in (24), and update $\boldsymbol{\theta}_k^{(t)}$ with (23).
8: $\qquad$ Update $\mathbf{V}_k^{(t)}$ with $\boldsymbol{\theta}_k^{(t)}$ and let $\tilde{\mathbf{V}}^{(t)} = [\mathbf{V}_1^{(t)}, \mathbf{V}_2^{(t)}, \ldots, \mathbf{V}_k^{(t)}, \mathbf{V}_{k+1}^{(t-1)}, \ldots, \mathbf{V}_K^{(t-1)}]$;
9: $\qquad [\tilde{\boldsymbol{\beta}}^{(t)}, \tilde{\boldsymbol{\xi}}^{(t)}, \tilde{\mathbf{W}}^{(t)}, \tilde{\boldsymbol{\varepsilon}}^{(t)}] = \textbf{functionA}(\tilde{\mathbf{V}}^{(t)}, \tilde{\mathbf{W}}^{(t)})$.
10: $\quad$ **end for**
11: $\quad$ Let $\mathbf{W}^{(t)} = \tilde{\mathbf{W}}^{(t)},\ \mathbf{V}^{(t)} = \tilde{\mathbf{V}}^{(t)}$.
12: **until** The value of function $f_{\mathrm{a}}(\mathbf{W}, \mathbf{V})$ in P1 converges.

Algorithm 2 $[\boldsymbol{\beta}, \boldsymbol{\xi}, \mathbf{W}, \boldsymbol{\varepsilon}] = \textbf{functionA}(\mathbf{V}, \bar{\mathbf{W}})$

Determine $\boldsymbol{\beta}$ with (12);
Update $\boldsymbol{\xi}$ with (14);
Update $\mathbf{W}$ with (15);
Update $\boldsymbol{\varepsilon}$ with (17).
return $[\boldsymbol{\beta}, \boldsymbol{\xi}, \mathbf{W}, \boldsymbol{\varepsilon}]$.

We would like to discuss the convergence of **Algorithm 1** first. According to [42], the BCD-based algorithm is guaranteed to converge if the objective function is monotonically nondecreasing after each update. As discussed above, **Algorithm 1** contains **Algorithm 2**. As for **Algorithm 2**, it is easily proved that the objective function is monotonically nondecreasing, because the optimal value of the corresponding optimization variable for each step can be obtained. Except for **Algorithm 2**, only the line 7 in **Algorithm 1** may lead to the non-convergence, which updates $\boldsymbol{\theta}_k^{(t)}$ from $\boldsymbol{\theta}_k^{(t-1)}$ with (23). Based on (21) and (22), we have:

$$\begin{aligned} f_{\mathrm{q}3,k}\left(\boldsymbol{\theta}_k^{(t-1)}\right) &= f_{\mathrm{q}4,k}\left(\boldsymbol{\theta}_k^{(t-1)}, \boldsymbol{\theta}_k^{(t-1)}\right) \\ &\geq f_{\mathrm{q}4,k}\left(\boldsymbol{\theta}_k^{(t)}, \boldsymbol{\theta}_k^{(t-1)}\right) \geq f_{\mathrm{q}3,k}\left(\boldsymbol{\theta}_k^{(t)}\right). \end{aligned} \tag{25}$$

Therefore, the objective function of P1-d-(k), i.e., $f_{\mathrm{q}3,k}(\boldsymbol{\theta}_k)$, is monotonically nonincreasing with (23). Thus, when the other $\boldsymbol{\theta}_i$ $(i \neq k)$ are fixed, the objective function of P1-c-V, i.e., $-f_{\mathrm{q}2}(\mathbf{V})$ is monotonically non-decreasing with (23). Namely, $f_{\mathrm{q}2}(\mathbf{V})$ is monotonically non-decreasing with (23). It is shown that in [42], with fixed $\boldsymbol{\beta}$, $\boldsymbol{\xi}$, $\mathbf{W}$, and $\boldsymbol{\varepsilon}$, $f_{\mathrm{q}2}(\mathbf{V})$ is equivalent with the objective function of P1, i.e., $f_{\mathrm{a}}(\mathbf{W}, \mathbf{V})$. Therefore, the objective function of P1 is monotonically non-decreasing with (23). As discussed above, during the process of the BCD method for the perfect CSI case, the objective function of P1 is monotonically nondecreasing, thus the BCD method for the perfect CSI case must converge. The convergence for **Algorithm 3** and **Algorithm 4** can be proved in a similar way.

Next, we would like to discuss the complexity of **Algorithm 1** and **Algorithm 2**. For simplicity, we assume that $N_k = N$ for all k, which means all RISs have the same number of reflective elements. For **Algorithm 2**, the complexity of updating the auxiliary variables β, $\boldsymbol{\xi}$, and ε is $\mathcal{O}(2K^2M + 15KM)$, and the complexity of updating $\mathbf{W}$ is $\mathcal{O}(KM^3 + 21KM^2 + 13KM)$. Thus, the complexity of **Algorithm 2** can be approximated as $\mathcal{O}(2K^2M + KM^3 + 21KM^2)$. Note that in line 7 of **Algorithm 1**, we use the Armijo rule to determine the value of μ_k. During this process, **functionA** is called a number of times implicitly. Denote the number of times it is called as I_A. With the updated μ_k, the complexity of updating $\mathbf{V}_k^{(t)}$ is $\mathcal{O}(\frac{3}{2}KN^2 + 3N^2)$. Thus, the total complexity of **Algorithm 1** is $\mathcal{O}(I_O(\frac{3}{2}K^2N^2 + 3KN^2 + KI_A(2K^2M + KM^3 + 21KM^2))$.

As shown in Section VI, the proposed BCD method can obtain significant performance gain compared to different baseline systems. However, it has a few drawbacks. First, in the optimization of $\mathbf{W}$, the operation of matrix inversion is required, which is of high complexity. Second, in the optimization of $\mathbf{V}$, K iterations are needed. In every such iteration, I_A times Armijo searches are required. This further increases the complexity of **Algorithm 1**. To overcome these drawbacks, we propose another algorithm for solving P1 in Section IV, which can achieve comparable performance gain but with much lower complexity.

IV. Low-Complexity BCD Method With Perfect CSI

In this section, we develop a *low-complexity BCD method* (LC-BCD) to solve P1. While the algorithm framework is similar to that introduced in Section III, the LC-BCD method can eliminate the operation of matrix inversion in the optimization of $\mathbf{W}$, and get rid of the inner iteration for searching μ_k involved in the optimization of $\mathbf{V}$. Specifically, after applying the fractional programming techniques, the prox-linear BCD method [43] is applied to optimize $\mathbf{W}$, and the iterative method [7] is applied to optimize $\mathbf{V}$.

A. Prox-Linear BCD Method for Optimizing W

As discussed in Section III-B, with the fixed value of $\boldsymbol{\beta}$ and $\mathbf{V}$, P1-a is equivalently transformed to P1-b, and $\boldsymbol{\xi}$ and is updated by (14). Therefore, with the given value of $\boldsymbol{\xi}$, we aim to find a low-complexity algorithm for solving the equivalent question of P1-b, which is given by

$$\underline{\textbf{P1-b-W:}} \qquad \min_{\mathbf{W}} \quad -f_{\mathrm{q}1}(\mathbf{W}) \qquad \text{s.t.} \quad (8).$$

Denote $\mathbf{g}(\mathbf{w}_k)$ as the gradient of $-f_{\mathrm{q}1}(\mathbf{W})$ with respect to $\mathbf{w}_k$, i.e., $\mathbf{g}(\mathbf{w}_k) = -\frac{\partial f_{\mathrm{q}1}(\mathbf{W})}{\partial \mathbf{w}_k}$. With a simple deduction, we can obtain the expression of $\mathbf{g}(\mathbf{w}_k)$, which is given by

$$\begin{aligned}&\mathbf{g}(\mathbf{w}_k)\\&=-2\xi_{\mathrm{c},k}^*\sqrt{\frac{P_k\rho_{\mathrm{c},k}(1+\beta_{\mathrm{s},kj})}{\Gamma}}\mathbf{b}_k\\&\quad+2\left|\xi_{\mathrm{c},k}\right|^2(\frac{QP_k\mathbf{b}_k\mathbf{b}_k^H}{\Gamma}+\delta_k)\mathbf{w}_k\\&\quad+\sum_{j=1}^{2}[-2\xi_{\mathrm{s},kj}^*\sqrt{\frac{P_k\rho_{\mathrm{s},k}(1+\beta_{\mathrm{s},kj})}{2}}(\mathbf{f}_k+\tilde{c}_j\mathbf{b}_k)\\&\quad+2|\xi_{\mathrm{s},kj}|^2(P_k\left(\mathbf{f}_k+\tilde{c}_j\mathbf{b}_k\right)\left(\mathbf{f}_k+\tilde{c}_j\mathbf{b}_k\right)^H+\delta_k)\mathbf{w}_k].\end{aligned}\tag{26}$$

We find that $\|\mathbf{g}_k(\mathbf{x})-\mathbf{g}_k(\mathbf{z})\|_F\le L_k\|\mathbf{x}-\mathbf{z}\|_F$ for $\forall\mathbf{x},\mathbf{z}\in\mathbb{C}^{M\times1}$, where $L_k=\|2|\xi_{\mathrm{c},k}|^2(\frac{Q}{\Gamma}P_k\mathbf{b}_k\mathbf{b}_k^H+\delta_k)+\sum_{j=1}^{2}2|\xi_{\mathrm{s},kj}|^2(P_k(\mathbf{f}_k+\tilde{c}_j\mathbf{b}_k)(\mathbf{f}_k+\tilde{c}_j\mathbf{b}_k)^H+\delta_k)\|_F$. Therefore, $\mathbf{g}(\mathbf{w}_k)$ is Lipschitz continuous. In this case, with the prox-linear BCD method [43], P1-b-W can be approximated as

$$\begin{aligned}\underline{\textbf{P1-f:}}\qquad\min_{\mathbf{W}}\quad&\sum_{k=1}^{K}(\mathrm{Re}\{\mathbf{g}(\hat{\mathbf{w}}_k)^H(\mathbf{w}_k-\hat{\mathbf{w}}_k)\}\\&+\frac{L_k}{2}\|\mathbf{w}_k-\hat{\mathbf{w}}_k\|^2)\\\text{s.t.}\quad&(8).\end{aligned}$$

In P1-f, $\hat{\mathbf{w}}_k$ denotes the extrapolated point for $\mathbf{w}_k$, and $\hat{\mathbf{w}}_k=\dot{\mathbf{w}}_k+\epsilon_k\left(\dot{\mathbf{w}}_k-\ddot{\mathbf{w}}_k\right)$ [43], where $\dot{\mathbf{w}}_k$ is the current value of $\mathbf{w}_k$ and $\ddot{\mathbf{w}}_k$ is the previous value of $\mathbf{w}_k$ before it was updated to $\dot{\mathbf{w}}_k$. Besides, $\epsilon_k>0$ is the extrapolation weight for $\mathbf{w}_k$, and the update rule for ϵ_k is $\epsilon_k=\min(\frac{\bar{z}-1}{z},\frac{9999}{10000}\sqrt{\frac{\bar{L}_k}{L}})$, where $\bar{z}$ and $\bar{L}_k$ denote the value of z and L_k at the previous iteration. By setting the initial value of z as zero, then z can be updated by $z=\frac{1}{2}(1+\sqrt{1+4\bar{z}^2})$. Finally, based on the form of P1-f, we can obtain a simple update rule of $\mathbf{W}$, which is given by

$$\mathbf{w}_k=\frac{1}{L_k-2\lambda_k}\left(L_k\hat{\mathbf{w}}_k-\hat{\mathbf{g}}_k\right),\tag{27}$$

where $\lambda_k=\frac{L_k}{2}-\frac{1}{2}\|L_k\hat{\mathbf{w}}_k-\hat{\mathbf{g}}_k\|_2$.

As the matrix inversion operation is eliminated, the complexity for updating $\mathbf{W}$ with the prox-linear BCD method is significantly reduced compared with that introduced in Section III. Besides, as $f_{\mathrm{q1}}(\mathbf{W})$ satisfies the Kurdyka–Łojasiewicz property, the convergence of the prox-linear BCD method for optimizing $\mathbf{W}$ can be established with Lemma 2.8 of [43].

Next, with the given value of $\mathbf{W}$, we develop a new algorithm for optimizing $\mathbf{V}$.

B. Iterative Method for Optimizing $\mathbf{V}$

As discussed in Section III-C, with the fixed value of $\boldsymbol{\beta}$ and $\mathbf{W}$, P1-a can be equivalently transformed to P1-c, and $\boldsymbol{\varepsilon}$ is updated according to (17). Therefore, with the given value of $\boldsymbol{\varepsilon}$, we aim to find a low-complexity algorithm for optimizing $\mathbf{V}$ in this subsection.

According to the form of $f_{\mathrm{q2}}(\mathbf{V})$ as shown in (18), we adopt the iterative method introduced in [7]. We only optimize the reflection coefficient of one element at one RIS each time while keeping the other reflection elements unchanged. The specific process is discussed as follows.

First, we have: $\mathbf{v}_k^H\mathbf{U}_k\mathbf{v}_k=v_{kn}^*u_{kn,kn}v_{kn}+2\mathrm{Re}\{\sum_{j=1,j\neq n}^{N_k}v_{kn}^*u_{kn,kj}v_{kj}\}+\sum_{j=1,j\neq n}^{N_k}\sum_{i=1,i\neq n}^{N_k}v_{ki}^*u_{ki,kj}v_{kj}$, $\mathbf{d}_k^H\mathbf{v}_k=d_{kn}^*v_{kn}+\sum_{i=1,i\neq n}^{N}d_{ki}^*v_{ki}$. Therefore, when we optimize the n-th reflection coefficient at the k-th RIS, with the other reflection elements unchanged, $f_{\mathrm{q2}}(\mathbf{V})$ can be transformed to

$$f_{\mathrm{q5},kn}(v_{kn})=2\mathrm{Re}\{v_{kn}^*A_{2,kn}\}-\left|v_{kn}\right|^2A_{1,kn},\tag{28}$$

where $A_{1,kn}{=}u_{kn,kn}$, $A_{2,kn}{=}d_{kn}-\sum\limits_{j=1,j\neq n}^{N_k}u_{kn,kj}v_{kj}$. As a result, by applying the iterative method, P1-c can be divided into $\sum_{k=1}^{K}N_k$ subproblems:

$$\begin{aligned}\underline{\textbf{P1-g-(}\boldsymbol{k,n}\textbf{):}}\qquad\max_{v_{kn}}\quad&f_{\mathrm{q5},kn}(v_{kn})\\\text{s.t.}\quad&(7).\end{aligned}$$

When we optimize v_{kn}, the subproblem P1-g-(k,n) is convex with respect to v_{kn}. Thus we can obtain the expression of optimal v_{kn} with simple deduction:

$$v_{kn}^o=\exp\left(\arg\min_{\varphi_{kn}\in[0,2\pi]}\left|\varphi_{kn}-\angle A_{2,kn}\right|\right).\tag{29}$$

C. Discussion

We summarize the proposed LC-BCD method in **Algorithm 3**. In **Algorithm 3**, the complexity of updating $\boldsymbol{\beta}$, $\boldsymbol{\xi}$, and $\boldsymbol{\varepsilon}$ is the same as **Algorithm 2**, which is $\mathcal{O}(2K^2M+15KM)$. With the updated $\boldsymbol{\beta}$ and $\boldsymbol{\xi}$, the complexity of updating $\mathbf{W}$ is $\mathcal{O}(21KM^2+23KM)$. With the updated $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$, the complexity of updating $\mathbf{V}$ is $\mathcal{O}(\frac{3}{2}K^2N^2+4KN^2)$. Therefore, the total complexity of **Algorithm 3** is $\mathcal{O}(I_O(\frac{3}{2}K^2N^2+4KN^2+2K^2M+21KM^2))$.

It is seen that the complexity of the LC-BCD method is much lower compared with the complexity of **Algorithm 1**. Specifically, the term $\mathcal{O}(KM^3)$ involved in the complexity of **Algorithm 1** is eliminated, which corresponds to the matrix inversion operation. Besides, thanks to the iterative algorithm discussed in Section IV-B, the operation for Armijo search is eliminated and the complexity of **Algorithm 3** concerning $(2K^2M+21KM^2)$ is reduced by KI_A times.

Algorithm 3 LC-BCD Method for the Perfect CSI Case

1: : Initialize $\mathbf{W},\mathbf{V}$ to feasible values.
2: **repeat**
3: Update $\boldsymbol{\beta}$ with (12);
4: Update $\boldsymbol{\xi}$ with (14);
5: Update $\mathbf{W}$ with (15);
6: Update $\boldsymbol{\varepsilon}$ with (17);
7: Update v_{kn} with (29) for all k and n.
8: **until** the value of function $f_{\mathrm{a}}(\mathbf{W},\mathbf{V})$ in P1 converges.

V. BCD Method With Imperfect CSI

This section extends the BCD method to address the case with imperfect CSI and solve P2. It is easily found that P2 can be decomposed into two subproblems.

The outer-layer subproblem is given by

$$\underline{\textbf{P2-a:}} \quad \min_{\mathbf{V}} \quad f_{\text{b1}}(\mathbf{V}) \triangleq \mathbb{E}_{\varsigma}\left\{g(\mathbf{V};\varsigma)\right\} \quad \text{s.t.} \quad (9).$$

The inner-layer subproblem is given by

$$\underline{\textbf{P2-b:}} \quad g(\mathbf{V};\varsigma) = \min_{\mathbf{W}} -f_{\text{a}}(\mathbf{W},\mathbf{V};\varsigma) \quad \text{s.t.} \quad (10).$$

Next, we solve P2-b to optimize $\mathbf{W}$ first and then solve P2-a to optimize $\mathbf{V}$.

A. The Optimization of W

Algorithm 4 $[\boldsymbol{\beta},\boldsymbol{\xi},\mathbf{W},\boldsymbol{\varepsilon}] = \textbf{functionB}(\mathbf{V},\bar{\mathbf{W}};\varsigma)$

1: Set $j = 0$; $\mathbf{W}^{(0)} = \bar{\mathbf{W}}$.
2: **repeat**
3: $\quad j = j + 1$.
4: $\quad$ Update $\boldsymbol{\beta}^{(j)}$, $\boldsymbol{\xi}^{(j)}$, $\mathbf{W}^{(j)}$, and $\boldsymbol{\varepsilon}^{(j)}$ with given $\mathbf{V}$ and ς.
5: **until** The value of $f_{\text{a}}(\mathbf{W},\mathbf{V})$ converges.
6: **return** $[\boldsymbol{\beta}^{(j)},\boldsymbol{\xi}^{(j)},\mathbf{W}^{(j)},\boldsymbol{\varepsilon}^{(j)}]$.

In this subsection, P2-b is considered with a given $\mathbf{V}$ and a fixed sample index ς.

As P1 is equivalent to P2-b with given $\mathbf{V}$ and ς, we apply the BCD method introduced in Section III-B to solve P2-b, and a corresponding stationary solution $[\boldsymbol{\beta}(\varsigma),\boldsymbol{\xi}(\varsigma),\mathbf{W}(\varsigma),\boldsymbol{\varepsilon}(\varsigma)]$ is obtained. The specific process of which is shown in **Algorithm 4**.

Then, with the stationary solution, similar to Section III-C, $\hat{g}(\mathbf{V};\varsigma)$ can be approximated as $\hat{g}(\mathbf{V};\varsigma) = \sum_{k=1}^{K}\mathbf{v}_k^H\mathbf{U}_k(\varsigma)\mathbf{v}_k - \sum_{k=1}^{K}2\text{Re}\{\mathbf{d}_k(\varsigma)^H\mathbf{v}_k\} + \text{const}$, where $\mathbf{U}_k(\varsigma)$ and $\mathbf{d}_k(\varsigma)$ correspond to $\mathbf{U}_k$ and $\mathbf{d}_k$ shown in (19) and (20). Next, similar to P1-c-V, after dropping the irrelevant terms, P2-a can be transformed to K unconstrained subproblems. The k-th subproblem is:

$$\underline{\textbf{P2-c-(}k\textbf{):}} \quad \min_{\boldsymbol{\theta}_k} \quad f_{\text{b2}}(\boldsymbol{\theta}_k) \triangleq \mathbb{E}_{\varsigma}\left\{\hat{g}_k(\boldsymbol{\theta}_k;\varsigma)\right\},$$

where $\hat{g}_k(\boldsymbol{\theta}_k;\varsigma) = \left(e^{\jmath\boldsymbol{\theta}_k}\right)^H\mathbf{U}_k(\varsigma)e^{\jmath\boldsymbol{\theta}_k} - 2\text{Re}\left\{\mathbf{d}_k(\varsigma)^H e^{\jmath\boldsymbol{\theta}_k}\right\}$.

As discussed above, we have obtained a stationary solution for the inner-layer problem P2-b and transformed the outer-layer problem P2-a to P2-c-(k) by utilizing the solution of the P2-b. Next, we aim to find a stationary solution of P2-c-(k), and it will also be the stationary solution of P2.

B. Online SCA Method for the Optimization of V

Firstly, we apply the online version SCA method [22] to eliminate the expectation operator involved in P2-c-(k). Specifically, at the t-th iteration, a new set of channel samples (ς_t) is generated, and $f_{\text{b2}}(\boldsymbol{\theta}_k)$ can be approximated recursively by $r_k^{(t)}$ as follows

$$r_k^{(t)} = (1-\omega_t)\, r_k^{(t-1)} + \omega_t\hat{g}_k(\boldsymbol{\theta}_k^{(t-1)};\varsigma_t), \tag{30}$$

where $r_k^{(0)}$ equals zero for all k, $\omega_t \in (0,1]$ denotes the weighting factor of the current channel realization (ς_t) at the t-th iteration, and $(1-\omega_t)$ denotes the weighting factor of the previous channel realizations. The gradient of $f_{\text{b2}}(\boldsymbol{\theta}_k)$ is given by

$$\nabla r_k^{(t)} = (1-\omega_t)\,\nabla r_k^{(t-1)} + \omega_t\nabla\hat{g}_k(\boldsymbol{\theta}_k^{(t-1)};\varsigma_t), \tag{31}$$

where

$$\begin{aligned}&\nabla\hat{g}_k(\boldsymbol{\theta}_k^{(t-1)};\varsigma_t)\\ &\quad = 2\text{Re}\{-\jmath e^{-\jmath\boldsymbol{\theta}_k^{(t-1)}} \circ (\,\mathbf{U}_k(\varsigma_{t-1})e^{\jmath\boldsymbol{\theta}_k^{(t-1)}} - \,\mathbf{d}_k(\varsigma_{t-1}))\}.\end{aligned} \tag{32}$$

Besides, according to [22], the approximated function $r_k^{(t)}$ can also be expressed as $\bar{r}_k^{(t)} = (1-\omega_t)r_k^{(t-1)} + \omega_t\hat{g}_k(\boldsymbol{\theta}_k^{(t-1)};\varsigma_t) + (\nabla r_k^{(t)})^H(\boldsymbol{\theta}_k^{(t)} - \boldsymbol{\theta}_k^{(t-1)}) + \frac{\tau_k^{(t)}}{2}\|\boldsymbol{\theta}_k^{(t)} - \boldsymbol{\theta}_k^{(t-1)}\|^2$. As $\bar{r}_k^{(t)}$ is convex for $\boldsymbol{\theta}_k^{(t)}$, thus we can obtain the update rule of $\boldsymbol{\theta}_k^{(t)}$ at the t-th iteration as

$$\boldsymbol{\theta}_k^{(t)} = \boldsymbol{\theta}_k^{(t-1)} - \frac{\nabla r_k^{(t)}}{\tau_k^{(t)}}, \tag{33}$$

where $\tau_k^{(t)}$ can be determined by the Armijo rule similar to (23).

The weighting factors ω_t for all t can be regarded as sequence $\{\omega_t\}$. According to [22], to guarantee the convergence of the online SCA method, $\{\omega_t\}$ should satisfy the following constraints: $\lim_{t\to\infty}\omega_t \to 0$, $\lim_{t\to\infty}\sum_t\omega_t = \infty$, and $\lim_{t\to\infty}\sum_t(\omega_t)^2 < \infty$.

Remark 3: To satisfy these constraints, similar to [8], [44], and [45], we use the p-series as ω_t in this paper. The series converges when $p > 1$, and diverges when $p \leq 1$. Therefore, we set $0.5 < p \leq 1$ to ensure that the above convergence constraints are satisfied. In the simulation, similar to [8], we take $p = 0.501$ as an example.

C. Discussion

We summarize the BCD method for imperfect CSI as **Algorithm 5**. It is easily found that the basic framework of **Algorithm 5** is similar to **Algorithm 1** except for the following differences: First, at the beginning of each iteration, **Algorithm 5** needs to generate new channel realizations $\mathbf{h}_k(\varsigma)$ and $\mathbf{G}_k(\varsigma)$. Second, while **functionB** involved in **Algorithm 5** requires inner-loop to output stationary solution of P2-b, **functionA** involved in **Algorithm 1** requires no loop to generate the output. Third, while the update equations for $\boldsymbol{\beta}$, $\boldsymbol{\xi}$, $\mathbf{W}$, and $\boldsymbol{\varepsilon}$ involved in **functionB** are the same as those in **functionA**, the update equations for $\mathbf{V}$ in **Algorithm 5**, which are shown from line 7 to line 10, are different from **Algorithm 1**.

Next, we discuss the complexity of **Algorithm 5**. As **functionB** applies the same updated equations for $\boldsymbol{\beta}$, $\boldsymbol{\xi}$, $\mathbf{W}$, and $\boldsymbol{\varepsilon}$ as **functionA**, the complexity of **Algorithm 4** is $\mathcal{O}(I_E(2K^2M + KM^3 + 21KM^2))$, where I_E is the iteration number of the inner-loop in **functionB**. Besides, the complexity of updating $\mathbf{V}$ with **Algorithm 5** is the same as that with **Algorithm 1** [8], which is $\mathcal{O}(\frac{3}{2}KN^2 + 3N^2)$. Therefore, the total complexity of **Algorithm 5** is $\mathcal{O}(I_O(\frac{3}{2}K^2N^2 + 3KN^2 + KI_AI_E(2K^2M + KM^3 + 21KM^2))$. It is seen that **Algorithm 5** costs I_E times complexity concerning M compared to **Algorithm 1**, and the complexity for updating $\mathbf{V}$ remains unchanged.

Algorithm 5 BCD Method for the Imperfect CSI Case

Initialize $\mathbf{V}^{(0)}$ and $\mathbf{W}^{(0)}$ to feasible values; Let $t=0$;
repeat
  $t=t+1$;
  Generate new channel realizations $\mathbf{h}_k(\varsigma_t)$ and $\mathbf{G}_k(\varsigma_t)$ for all k;
  $[\tilde{\boldsymbol{\beta}}^{(t)}, \tilde{\boldsymbol{\xi}}^{(t)}, \tilde{\mathbf{W}}^{(t)}, \tilde{\varepsilon}^{(t)}] =$ **functionB**$(\mathbf{V}^{(t-1)},\ \mathbf{W}^{(t-1)}; \varsigma_t)$.
  for k=1, 2, ..., K **do**
    Update $\mathbf{U}_k(\varsigma_t)$ and $\mathbf{d}_k(\varsigma_t)$ according to (19) and (20) for all k;
    Update $\nabla \hat{g}_k(\boldsymbol{\theta}_k^{(t-1)}; \varsigma_t)$ with (32);
    Update $\nabla r_k^{(t)}$ with (31);
    Decide $\tau_k^{(t)}$ with Armijo rule, then update $\boldsymbol{\theta}_k^{(t)}$ with (33);
    Update $\mathbf{V}_k^{(t)}$ with $\boldsymbol{\theta}_k^{(t)}$ and let $\tilde{\mathbf{V}}^{(t)} = [\mathbf{V}_1^{(t)}, \mathbf{V}_2^{(t)}, \ldots, \mathbf{V}_k^{(t)}, \mathbf{V}_{k+1}^{(t-1)}, \ldots, \mathbf{V}_K^{(t-1)}]$;
    $[\tilde{\boldsymbol{\beta}}^{(t)}, \tilde{\boldsymbol{\xi}}^{(t)}, \tilde{\mathbf{W}}^{(t)}, \tilde{\varepsilon}^{(t)}] =$ **functionB**$(\tilde{\mathbf{V}}^{(t)}, \tilde{\mathbf{W}}^{(t)}; \varsigma_t)$.
  end for
  Let $\mathbf{W}^{(t)} = \tilde{\mathbf{W}}^{(t)},\ \mathbf{V}^{(t)} = \tilde{\mathbf{V}}^{(t)}$.
until the approximated function $r_k^{(t)}$ shown in (30) converges.

VI. Simulations

A. Simulation Setup

In this section, we provide simulation results to evaluate the proposed algorithms. Similar to [8], we consider a RIS-based femtocell network as shown in Fig. 3. Specifically, the BS is located at (0,0). Besides, we have 4 RISs and 4 corresponding users ($K=4$). We let L_k denote the distance between RIS k and the BS, thus the location of the RISs can be expressed as (L_1,0), (0,L_2), ($-L_3$,0), and (0,$-L_4$), respectively. In addition, we assume that the BS is equipped with 16 antennas ($M=16$), and all the RISs are equipped with the same number of reflection elements, i.e., $N_k = N$ for all k, and the same reflection efficiency, i.e., $\alpha_i = 0.8$ for all i. All users are assumed to have the same transmit power P_m, i.e., $P_k = P_m$ for all k. The weighting factors for $R_{\mathrm{s},k}$ and $R_{\mathrm{c},k}$ are assumed to be equal. As $\sum_{k=1}^{K}(\rho_{\mathrm{s},k}+\rho_{\mathrm{c},k}) = 1$, we have $\rho_{\mathrm{s},k} = \rho_{\mathrm{c},k} = \frac{1}{2K}$. The weighting sequence $\{\omega_t\}$ for the imperfect CSI setup is $\omega_t = \frac{1}{t^{0.501}}$. Besides, as the typical SINR of the IoT transmissions in the proposed system is very small due to the multi-user interference and noise, the discrete capacity of the BPSK signal can be approximated as the Gaussian capacity [46]. Thus, we have $\Gamma = 1$. Similar to [12], the period ratio Q is set as 128. The transmission bandwidth is 180 kHz. The noise power spectral density is -170 dBm/Hz. The path-loss factors are set according to the 3GPP propagation environment [47]. Specifically, the path loss for $\mathbf{G}_k$ and $\mathbf{h}_k$ is $22.0\lg d + 35.6$ dB. The path loss for $\mathbf{f}_k$ is $36.7\lg d + 32.6$ dB.

We assume that $\mathbf{G}_k$ and $\mathbf{h}_k$ follow Rician fading, and that $\mathbf{f}_k$ follows Rayleigh fading. Besides, half-wavelength *uniform linear array* (ULA) is considered at the BS and *uniform planar array* (UPA) are considered at all the RISs, and thus $\mathbf{G}_k$ and $\mathbf{h}_k$ can be expressed as $\mathbf{G}_k = L_{\mathrm{G},k}(\sqrt{\frac{\nu_{\mathrm{G},k}}{\nu_{\mathrm{G},k}+1}}\mathbf{a}_N(\vartheta_k, o_k)\mathbf{a}_M(\psi_k)^{\mathrm{H}} + \sqrt{\frac{1}{\nu_{\mathrm{G},k}+1}}\overline{\mathbf{G}}_k)$ and $\mathbf{h}_k = L_{\mathrm{h},k}(\sqrt{\frac{\nu_{\mathrm{h},k}}{\nu_{\mathrm{h},k}+1}}\mathbf{a}_N(\varsigma_k, \iota_k) + \sqrt{\frac{1}{\nu_{\mathrm{h},k}+1}}\overline{\mathbf{h}}_k)$, respectively. $L_{\mathrm{G},k}$ and $L_{\mathrm{h},k}$ denote the corresponding path-losses of $\mathbf{G}_k$ and $\mathbf{h}_k$, $\nu_{\mathrm{G},k}$ and $\nu_{\mathrm{h},k}$ denote the Rician factors of $\mathbf{G}_k$ and $\mathbf{h}_k$. For simplicity, we set $\nu_{\mathrm{G},k} = \nu_{\mathrm{h},k} = 10$ for all k. ϑ_k, o_k, ψ_k, ς_k, and ι_k are the angular parameters, and $\mathbf{a}_N(\vartheta_k, o_k)$, $\mathbf{a}_M(\psi_k)$, and $\mathbf{a}_N(\varsigma_k, \iota_k)$ denote the corresponding steering vectors of them. $\overline{\mathbf{G}}_k$ and $\overline{\mathbf{h}}_k$ denote the NLOS components, and they are chosen from $\mathcal{CN}(0,1)$.

In the imperfect CSI setup, $\mathbf{f}_k$ is assumed to be perfectly acquired. Thus only the channel variables $\overline{\mathbf{G}}_k$ and $\overline{\mathbf{h}}_k$ need to be estimated according to the assumption in Section II-C.2. Similar to [8], we let x denote one of the above variables, and $\hat{x}$ denote the corresponding estimate value. We assume that the estimation error $x - \hat{x}$ follows zeros mean complex Gaussian distribution, and all these elements have the same normalized minimum mean square error:

$$\kappa = \frac{\mathbb{E}\left[|x-\hat{x}|^2\right]}{\mathbb{E}\left[|\hat{x}|^2\right]}. \tag{34}$$

To better evaluate the performance of the proposed algorithms, we introduce three baseline systems:

- **Baseline 1** (Random Phase): The passive beamforming variable $\mathbf{V}$ is initialized by a random value and kept fixed under each channel realization. In this case, only the active beamforming variable $\mathbf{W}$ is optimized with the proposed algorithms.
- **Baseline 2** (Without RIS): All RISs are removed from the simulation setup, and thus the system degenerates into a standard uplink MU-MIMO system. Then Zero-forcing algorithm is adopted for the active beamforming variable $\mathbf{W}$.
- **Baseline 3** (Relay only): All RISs are used only to enhance the primary transmissions from the users to the BS but not to transmit information. The joint beamforming scheme proposed in [8] is adopted.
- **Baseline 4** (Interference): $\mathbf{W}$ and $\mathbf{V}$ are still optimized with the proposed algorithms, i.e., without considering the interference links. However, the sum rates are calculated with the interference. Similar to $\mathbf{f}_k$, we assume that the interference links follow Rayleigh fading with the path loss as $36.7\lg d + 32.6$ dB.

B. Simulation Results

In this subsection, the locations of users and RISs are fixed once randomly generated.[6] The sum rates are obtained by averaging over 1000 independent small-scale fading realizations.

Fig. 4 shows the convergence speed of different algorithms with the transmit power $P_m = -5$ dBm and $N = 100$. For the case of perfect CSI, it is observed that the LC-BCD method can achieve almost the same WSR as the BCD method, but with a slightly more number of iterations. As discussed in Section IV-C, the complexity of the LC-BCD method in each iteration is significantly lower than the BCD method. Therefore, the overall complexity of the LC-BCD method is

[6] In this simulation, the RIS locations are (208m, 0), (0, 189m), (-211m, 0), (0, -198m), and the user locations are (195.45m, 35.29m), (-31.43m, 197.08m), (-199.66m, -24.18m), (31.05m, -196.40m).

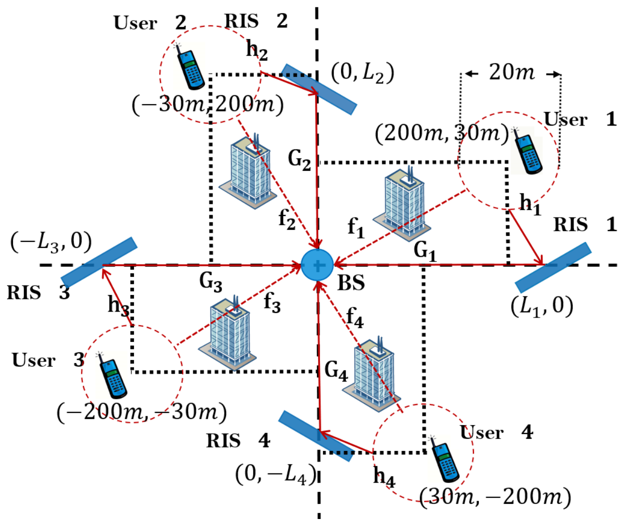

Fig. 3. The simulation scenario for the RIS-based uplink MU-MIMO SR system.

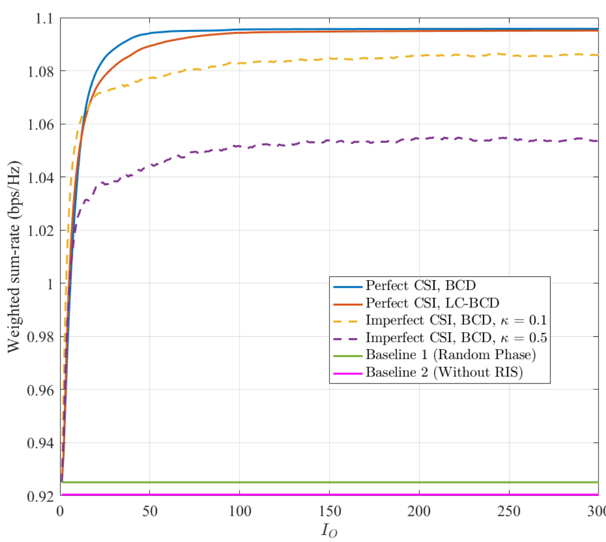

Fig. 4. Convergence speed: $P_m = -5$ dBm, $N = 100$.

much lower than the BCD method.[7] In the case of imperfect CSI, the proposed method can also converge even though there are slight fluctuations. It is also seen that when κ increases, i.e., the channel becomes more uncertain, the proposed BCD method needs more steps to converge.

Fig. 5 illustrates the sum rate of both the primary transmissions from the users and the IoT transmissions from the RISs for different algorithms concerning P_m when $N = 100$. Several observations are made from the figure. Firstly, the

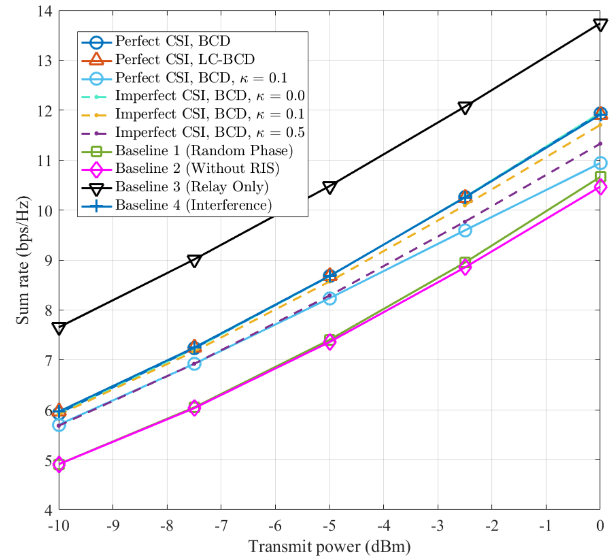

Fig. 5. Sum rate vs P_m: $N = 100$.

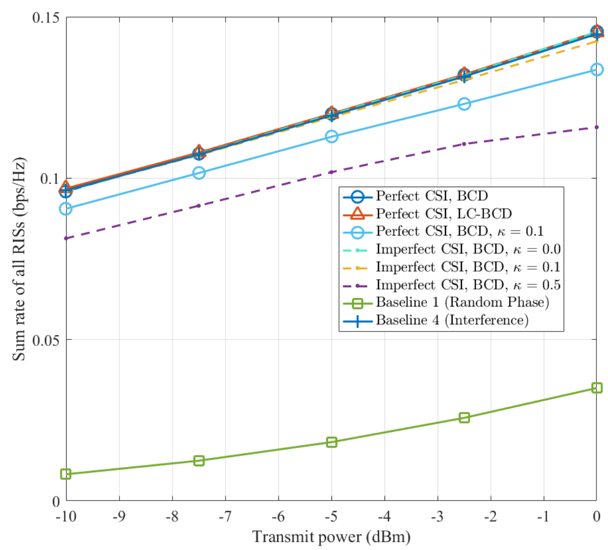

Fig. 6. Sum rate of RISs vs P_m: $N = 100$.

performance of all the schemes increases with the increase of P_m, which is as expected. Secondly, the proposed algorithms achieve significantly higher sum rates compared with **Baseline 1** and **Baseline 2**, which proves the effectiveness of jointly optimizing the active and the passive beamforming. Thirdly, the LC-BCD method achieves almost the same performance gain as the BCD method, which proves the effectiveness of the LC-BCD method. Fourthly, when the channel uncertainty equals zero ($\kappa = 0.0$), the BCD method for the imperfect CSI case achieves almost the same performance as the BCD method for the perfect CSI case, which is as expected. When the channel uncertainty is small ($\kappa = 0.1$), the BCD method for the imperfect CSI can achieve a similar performance gain as the perfect CSI. Besides, even when the channel uncertainty becomes larger ($\kappa = 0.5$), the proposed BCD method can still achieve considerable performance gain.

[7]Specifically, according to Section III-D and Section IV-C, the complexity of the BCD method is $\mathcal{O}(I_{O1}(\frac{3}{2}K^2N^2+3KN^2+KI_A(2K^2M+KM^3+21KM^2))$, and the complexity of the LC-BCD method is $\mathcal{O}(I_{O2}(\frac{3}{2}K^2N^2+4KN^2+2K^2M+21KM^2))$. As discussed in Section VI-A as an example, we have $K = 4$, $M = 16$, and $N = 100$. According to Fig. 4, $I_{O1} \approx 60$ and $I_{O2} \approx 100$. Besides, we find that the average value of I_A is 10.6. Thus the overall complexity of the BCD method is $O(1.2 \times 10^8)$, and the overall complexity of the LC-BCD method is $O(4.2 \times 10^7)$. Thus the overall complexity of the BCD method is approximately three times that of the LC-BCD method.

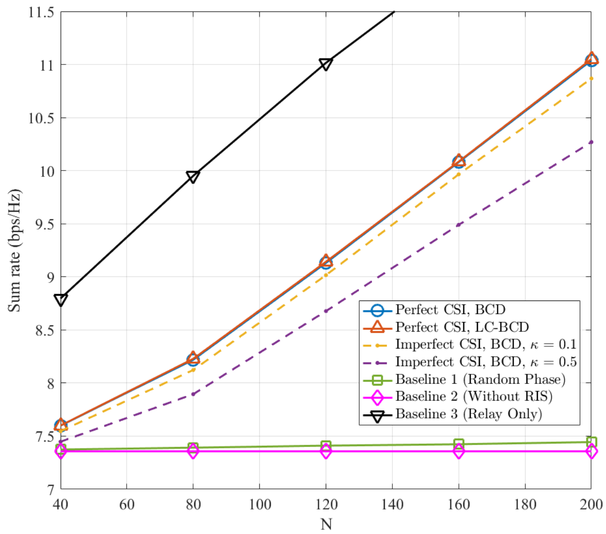

Fig. 7. Sum rate vs N: $P_m = -5$ dBm.

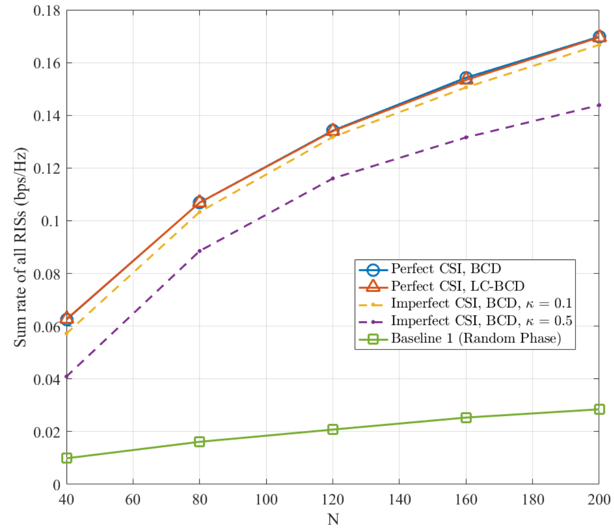

Fig. 8. Sum rate of RISs vs N: $P_m = -5$ dBm.

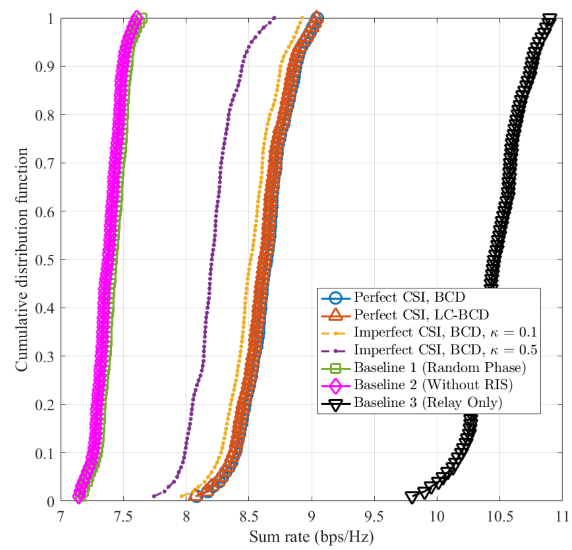

Fig. 9. CDF vs sum rate: $P_m = -5$ dBm, $N = 100$.

Fifthly, the performance of BCD assuming there is no channel uncertainty (i.e., perfect CSI, BCD, $\kappa = 0.1$) is worse than the BCD method considering imperfect CSI under the channel uncertainty of $\kappa = 0.1$. It is even worse than the imperfect CSI BCD method under the channel uncertainty of $\kappa = 0.5$. This shows the importance of robust design and the robustness of our proposed BCD method for the imperfect CSI case. In addition, we can find that **Baseline 4** achieves almost the same performance as the proposed system. This means that if the RISs are deployed far away from each other, the interference links will have a limited effect on the system performance. In this case, we can ignore the interference links and this greatly simplifies the beamforming design. Lastly, it is seen that **Baseline 3** achieves a higher rate performance than the proposed schemes in this paper. This is because RIS k in the proposed system needs to transmit sensor information c_k, and the backscatter link may change significantly with the change of the IoT symbol. Therefore, it is more demanding for the proposed system in this paper to achieve the same performance gain as **Baseline 3**, where the RISs only need to enhance the primary communication.

Under the same setting as Fig. 5, Fig. 6 illustrates the sum rate of IoT transmissions for different algorithms with respect to P_m when $N = 100$. It is seen that the observations shown in Fig. 6 are similar to Fig. 5. Besides, it is worth mentioning that the BCD method for imperfect CSI achieves 0.1 bps/Hz IoT transmission even when $P_m = -10$ dBm and $\varrho = 0.5$. This verifies that the proposed system not only enhances the primary transmission as discussed above but also enables the IoT transmission.

Fig. 7 and Fig. 8 illustrate the sum rate of users and RISs and the sum rate of RISs for different algorithms, respectively, concerning N when $P_m = -5$ dBm. According to Fig. 7, we have several observations. Firstly, the performance of all the proposed methods increases with N. This shows that with the increase in the number of reflection elements at RISs, more energy is reflected and the beam can be better steered towards the BS. Secondly, comparing Fig. 5 and Fig. 7, we find that the proposed methods at $N = 160$ and $P_m = -5$ dBm achieve similar rate performance like that at $N = 100$ and $P_m = -3$ dBm. It shows that adjusting P_m is more effective than adjusting N. This is because, both the direct and backscatter links are enhanced when P_m increases, while only the backscatter links are enhanced when N increases. As for Fig. 8, similar observations can be made.

Next, we are interested in the impact of users' locations on the system performance. We consider that the four users are uniformly and randomly distributed in four circles with a radius of 10m and centered at (200m,30m), (−30m,200m), (−200m,−30m), (30m,−200m), respectively. We consider 100 random samples of locations for each user. Besides, under each sample of user locations, 100 channel realizations with independent small-scale fading are further generated. The simulation results in terms of the *cumulative*

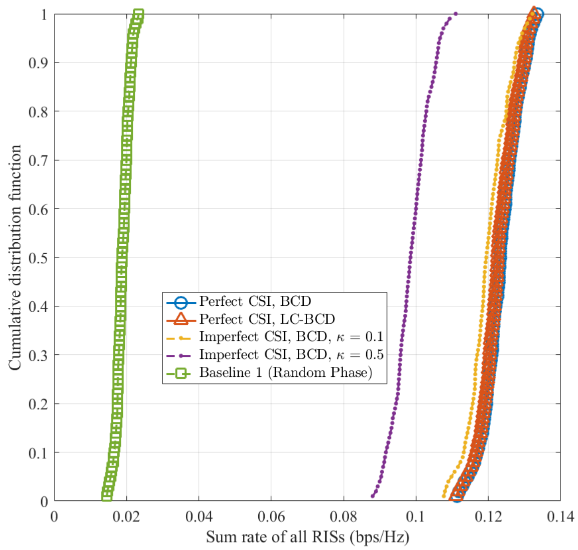

Fig. 10. CDF vs sum rate of RISs: $P_m = -5$ dBm, $N = 100$.

distribution function (CDF) of the sum rate of both users and RISs and the sum rate of RISs are illustrated in Fig. 9 and Fig. 10, respectively. It is seen that the performance gains of all the proposed schemes are stable and consistent with the case when the user locations are fixed. Therefore, we can conclude that the proposed methods can achieve stable and good performance gains compared with the baseline schemes with a high probability even when the users' locations change.

VII. Conclusion

This paper proposes a RIS-based uplink MU-MIMO SR system comprising multiple RISs and multiple PRs. Each RIS is applied to enable both the enhanced primary transmission from one nearby user and the IoT transmission from itself. The joint design of the active and passive beamforming has been investigated to maximize the WSR of the primary and IoT transmissions. Besides, both the perfect and imperfect CSI setups have been considered. For the perfect CSI setup, a BCD-based method and a corresponding low-complexity method have been proposed. Next, the BCD method has been extended to the imperfect CSI setup. Simulation results have shown that the proposed system not only enhances the primary transmission from users to the BS compared to different baseline systems, but also enables the IoT transmission from the RISs to the BS. Besides, when the channel uncertainty is small, the BCD method for the imperfect CSI setup incurs only marginal performance loss when the CSI error is small.

References

[1] J. Hu, Y. Pei, Y.-C. Liang, and S. Sun, "Reconfigurable intelligent surface based uplink massive MIMO symbiotic radio system," in *Proc. IEEE GLOBECOM*, Dec. 2021, pp. 1–6.

[2] I. Union. *IMT Traffic Estimates for the Years 2020 to 2030*. Report ITU. Accessed: Jul. 2015. [Online]. Available: https://www.itu.int

[3] Q.-U.-A. Nadeem, A. Kammoun, A. Chaaban, M. Debbah, and M.-S. Alouini, "Asymptotic max-min SINR analysis of reconfigurable intelligent surface assisted MISO systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 7748–7764, Apr. 2020.

[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, Jun. 2019.

[5] C. Huang, G. C. Alexandropoulos, A. Zappone, M. Debbah, and C. Yuen, "Energy efficient multi-user MISO communication using low resolution large intelligent surfaces," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2018, pp. 1–6.

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

[7] Q. Wu and R. Zhang, "Beamforming optimization for intelligent reflecting surface with discrete phase shifts," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 7830–7833.

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

[9] L. Yang, F. Meng, J. Zhang, M. O. Hasna, and M. D. Renzo, "On the performance of RIS-assisted dual-hop UAV communication systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 9, pp. 10385–10390, Sep. 2020.

[10] M. Najafi, V. Jamali, R. Schober, and H. V. Poor, "Physics-based modeling and scalable optimization of large intelligent reflecting surfaces," *IEEE Trans. Commun.*, vol. 69, no. 4, pp. 2673–2691, Apr. 2020.

[11] Y.-C. Liang, Q. Zhang, E. G. Larsson, and G. Y. Li, "Symbiotic radio: Cognitive backscattering communications for future wireless networks," *IEEE Trans. Cognit. Commun. Netw.*, vol. 6, no. 4, pp. 1242–1255, Dec. 2020.

[12] R. Long, Y.-C. Liang, H. Guo, G. Yang, and R. Zhang, "Symbiotic radio: A new communication paradigm for passive Internet of Things," *IEEE Internet Things J.*, vol. 7, no. 2, pp. 1350–1363, Feb. 2020.

[13] S. Xiao, H. Guo, and Y.-C. Liang, "Resource allocation for full-duplex-enabled cognitive backscatter networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 6, pp. 3222–3235, Jun. 2019.

[14] R. Long, H. Guo, L. Zhang, and Y.-C. Liang, "Full-duplex backscatter communications in symbiotic radio systems," *IEEE Access*, vol. 7, pp. 21597–21608, 2019.

[15] Q. Zhang, L. Zhang, Y.-C. Liang, and P. Y. Kam, "Backscatter-NOMA: A symbiotic system of cellular and Internet-of-Things networks," *IEEE Access*, vol. 7, pp. 20000–20013, 2019.

[16] Q. Zhang, Y.-C. Liang, and H. V. Poor, "Reconfigurable intelligent surface assisted MIMO symbiotic radio networks," *IEEE Trans. Commun.*, vol. 69, no. 7, pp. 4832–4846, Jul. 2021.

[17] J. Hu, Y.-C. Liang, and Y. Pei, "Reconfigurable intelligent surface enhanced multi-user MISO symbiotic radio system," *IEEE Trans. Commun.*, vol. 69, no. 4, pp. 2359–2371, Apr. 2021.

[18] K. Shen and W. Yu, "Fractional programming for communication systems—Part I: Power control and beamforming," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2616–2630, May 2018.

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

[20] 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, Jun. 2019.

[21] Q.-U.-U. Nadeem, A. Kammoun, A. Chaaban, M. Debbah, and M.-S. Alouini, "Asymptotic max-min SINR analysis of reconfigurable intelligent surface assisted MISO systems," 2019, *arXiv:1903.08127*.

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

[23] X. Yu, D. Xu, Y. Sun, D. W. K. Ng, and R. Schober, "Robust and secure wireless communications via intelligent reflecting surfaces," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2637–2652, Nov. 2020.

[24] Y. Liu, J. Zhao, M. Li, and Q. Wu, "Intelligent reflecting surface aided MISO uplink communication network: Feasibility and power minimization for perfect and imperfect CSI," *IEEE Trans. Commun.*, vol. 69, no. 3, pp. 1975–1989, Mar. 2020.
[25] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3313–3351, May 2021.
[26] G. Yang, Q. Zhang, and Y.-C. Liang, "Cooperative ambient backscatter communications for green Internet-of-Things," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 1116–1130, Apr. 2018.
[27] M. Riel and J.-J. Laurin, "Design of an electronically beam scanning reflectarray using aperture-coupled elements," *IEEE Trans. Antennas Propag.*, vol. 55, no. 5, pp. 1260–1266, May 2007.
[28] W. Tang *et al.*, "Programmable metasurface-based RF chain-free 8PSK wireless transmitter," *Electron. Lett.*, vol. 55, no. 7, pp. 417–420, Apr. 2019.
[29] S. Henthorn, K. L. Ford, and T. O'Farrell, "Direct antenna modulation for high-order phase shift keying," *IEEE Trans. Antennas Propag.*, vol. 68, no. 1, pp. 111–120, Jan. 2019.
[30] P. Li, R. C. de Lamare, and R. Fa, "Multiple feedback successive interference cancellation detection for multiuser MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2434–2439, Aug. 2011.
[31] X. Chen, R. Jia, and D. W. K. Ng, "On the design of massive non-orthogonal multiple access with imperfect successive interference cancellation," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 2539–2551, Mar. 2018.
[32] A. García-Armada, "SNR gap approximation for M-PSK-based bit loading," *IEEE Trans. Wireless Commun.*, vol. 5, no. 1, pp. 57–60, Jan. 2006.
[33] Q. Zhang, Y.-C. Liang, H.-C. Yang, and H. V. Poor, "Mutualistic mechanism in symbiotic radios: When can the primary and secondary transmissions be mutually beneficial?" *IEEE Trans. Wireless Commun.*, early access, Apr. 7, 2022, doi: 10.1109/TWC.2022.3163735.
[34] 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, Sep. 2018.
[35] 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, Feb. 2019.
[36] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling large intelligent surfaces with compressive sensing and deep learning," 2019, *arXiv:1904.10136*.
[37] H. Guo, Y. Liang, J. Chen, and E. G. Larsson, "Weighted sum-rate maximization for intelligent reflecting surface enhanced wireless networks," in *Proc. IEEE GLOBECOM*, Dec. 2019, pp. 1–6.
[38] C. Wang 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, Sep. 2006.
[39] T. Yoo 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, May 2006.
[40] A. Liu, V. K. N. Lau, and B. Kananian, "Stochastic successive convex approximation for non-convex constrained stochastic optimization," *IEEE Trans. Signal Process.*, vol. 67, no. 16, pp. 4189–4203, Aug. 2019.
[41] D. P. Bertsekas, "Nonlinear programming," *J. Oper. Res. Soc.*, vol. 48, no. 3, p. 334, 1997.
[42] 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, May 2018.
[43] Y. Xu and W. 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.
[44] T. T. Vu, D. T. Ngo, N. H. Tran, H. Q. Ngo, M. N. Dao, and R. H. Middleton, "Cell-free massive MIMO for wireless federated learning," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6377–6392, Oct. 2020.
[45] M.-M. Zhao, Q. Wu, M.-J. Zhao, and R. Zhang, "Two-timescale beamforming optimization for intelligent reflecting surface enhanced wireless network," in *Proc. IEEE 11th Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jun. 2020, pp. 1–5.
[46] J. M. Geist, "Capacity and cutoff rate for dense M-ary PSK constellations," in *Proc. MILCOM*, Sep. 1990, pp. 768–770.
[47] *Further Advancements for E-UTRA Physical Layer Aspects (Release 9)*, document TS 16.814, 3GPP, Mar. 2010.

Jinlin Hu (Graduate Student Member, IEEE) received the B.S. degree in communication engineering from the University of Electronic Science and Technology of China (UESTC), China, in 2020. His current research interest focuses on wireless communications.

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

He is also a Foreign Member of the Academia Europaea. He 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 was a recipient of numerous paper awards, including the IEEE Communications Society Advances in Communication Prize Paper Award in 2022, the IEEE Communications Society Stephen O. Rice Prize in 2021, and the IEEE Vehicular Technology Society Jack Neubauer Memorial Award in 2014. He was the Chair of the IEEE Communications Society Technical Committee on Cognitive Networks. He has served as the TPC Chair and the Executive Co-Chair for the IEEE Globecom'17. He was a Guest Editor/an Associate Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE JOURNAL ON 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 NETWORKS. He was an Associate Editor-in-Chief of the *Random Matrices: Theory and Applications* (World Scientific). He is also the Founding Editor-in-Chief of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS: COGNITIVE RADIO SERIES and the Key Founder and 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 was a Distinguished Lecturer of the IEEE Communications Society and the IEEE Vehicular Technology Society.

Yiyang Pei (Senior Member, IEEE) received the B.Eng. and Ph.D. degrees in electrical and electronic engineering from Nanyang Technological University, Singapore, in 2007 and 2012, respectively. From 2012 to 2016, she was a Research Scientist with the Institute for Infocomm Research, Singapore. She is currently an Associate Professor with the Singapore Institute of Technology, Singapore. Her current research interests include reconfigurable intelligent surface, dynamic spectrum access, and application of machine learning to wireless communications and networks. She was a recipient of the IEEE Communications Society Stephen O. Rice Prize Paper Award in 2021. She is an Editor of IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING.

Sumei Sun (Fellow, IEEE) is currently a Principal Scientist, the Deputy Executive Director (Research), and the Head of the Communications and Networks Department, Institute for Infocomm Research (I2R), Singapore. She is also holding a joint appointment with the Singapore Institute of Technology, and an adjunct appointment with the National University of Singapore, both as a Full Professor. Her current research interests are in next-generation wireless communications, joint sensing-communication-computing-control design, and the Industrial Internet of Things. She is the Editor-in-Chief of IEEE OPEN JOURNAL OF VEHICULAR TECHNOLOGY, the Steering Committee Chair of IEEE TRANSACTIONS ON MACHINE LEARNING IN COMMUNICATIONS AND NETWORKS, a Distinguished Speaker of the IEEE Vehicular Technology Society from 2018 to 2024, a member at large (MAL) with the IEEE Communications Society from 2021 to 2023, and a member of IEEE Vehicular Technology Society Board of Governors from 2022 to 2024.

Ruolun Liu (Member, IEEE) received the B.Eng. and M.Eng. degrees from the Jilin University of Technology, Changchun China, all in electrical engineering, and the Ph.D. degree in telecommunication and information system from Jilin University, Changchun.

From 2000 to 2001, he was a Research Scientist with the I3S Laboratory, CNRS, Nice, France, from 2002 to 2004, he was a Research Scientist with the Computer Science Information System Department, University of Limerick, Ireland. Since 2005, he has been with the Department of Electronic Engineering, Shandong University at Weihai, China, where he is currently an Associate Professor with the Faculty of Engineering. In 2009, he was the DAAD Visit Professor of the Faculty of Electrical Engineering, Helmut-Schmidt-Universität (HSU), Hamburg, Germany. His research interests include smart and reconfigurable environment, optimization methods, wide band beamforming, and computational acoustic scenes analysis.

Dr. Liu has been a member of the IEEE Signal Processing Society and Communications Society since 2008. He is also a member of ACM Special Interest Group on Multimedia Systems (SIGMM).