RIS-Aided Multiuser MIMO-OFDM With Linear Precoding and Iterative Detection: Analysis and Optimization

Mingyang Yue®, Lei Liu®, Member, IEEE, and Xiaojun Yuan®, Senior Member, IEEE

Abstract—In this paper, we consider a reconfigurable intelligent surface (RIS) aided uplink multiuser multi-input multi-output (MIMO) orthogonal frequency division multiplexing (OFDM) system, where the receiver is assumed to conduct low-complexity iterative detection. We aim to minimize the total transmit power by jointly designing the precoder of the transmitter and the passive beamforming of the RIS. This problem can be tackled from the perspective of information theory. But this information-theoretic approach may involve prohibitively high complexity since the number of rate constraints that specify the capacity region of the uplink multiuser channel is exponential in the number of users. To avoid this difficulty, we formulate the design problem of the iterative receiver under the constraints of a maximal iteration number and target bit error rates of users. To tackle this challenging problem, we propose a groupwise successive interference cancellation (SIC) optimization approach, where the signals of users are decoded and canceled in a group-by-group manner. We present a heuristic user grouping strategy, and resort to the alternating optimization technique to iteratively solve the precoding and passive beamforming subproblems. Specifically, for the precoding subproblem, we employ fractional programming to convert it to a convex problem; for the passive beamforming subproblem, we adopt successive convex approximation to deal with the unit-modulus constraints of the RIS. We show that the proposed groupwise SIC approach has significant advantages in both performance and computational complexity, as compared with the counterpart approaches.

Index Terms—Reconfigure intelligent surface, MIMO-OFDM, precoding, passive beamforming, iterative detection.

I. INTRODUCTION

RECONFIGURABLE intelligent surface (RIS) has been regarded as a key enabling technology for sixth-generation (6G) wireless communications [1], [2]. RIS is composed of a large number of low-cost passive reflecting elements, where a controller is equipped to adjust the reflection coefficient of each element in real time. By effectively designing the passive beamforming of the elements, RIS can reconfigure the wireless propagation environment by intelligently manipulating the reflection direction of the incident electromagnetic wave. In addition, RIS also has the advantages of flexible deployment, low energy consumption, and low noise [3]. Extensive research has been conducted on the design of RIS to improve the system performance in terms of energy and spectral efficiency [4], [5], [6], [7], [8] and channel capacity [9], [10], [11].

The optimization of RIS to improve the performance of multiuser systems has been investigated, e.g., in [4], [6], [12], [13], [14], [15], [16], [17], [18], and [19]. Particularly, the authors in [12] considered a RIS-aided downlink multiuser system, and designed the active and passive beamforming to maximize the minimal user rate. In [13], the authors aimed to minimize the total transmit power under the user rate constraints for the RIS-aided downlink multiuser multiple-input-single-output (MISO) system. In [14], a difference-of-convex algorithm was proposed to jointly design active and passive beamforming by minimizing the total transmit power of all users. In addition, the (weighted) sum rate of the multiuser system has been studied in [15] and [16], and has been extended to other scenarios, such as millimeter wave [17], [18] and multi-RIS-aided cooperative transmission [19]. Furthermore, the energy efficiency, defined as the sum rate normalized by the energy consumption, is investigated for RIS-aided downlink systems [5] and multiple-input-multiple-output (MIMO) uplink systems [4], [6].

The above works are designed from the perspective of information-theoretic performance metrics. In other words, these works assumed an ideal receiver that can achieve the capacity of the system. However, it is so far unclear how to design a practical low-complexity receiver to achieve the capacity of the RIS-aided multiuser MIMO channel. In fact, a low-complexity receiver, such as the zero-forcing (ZF) receiver [20] or the maximum ratio receiver [21], may perform
In contrast to the existing works, we consider the system optimization for the low-complexity iterative LMMSE receiver. We establish the state evolution to characterize the system performance, and formulate the joint optimization problem for the iterative receiver under the constraints of target BERs and a maximal iteration number.

- We propose a groupwise SIC approach to solve the challenging optimization problem. We employ a heuristic but efficient user grouping strategy to avoid the exhaustive search over all possible user groupings. For a fixed user grouping, we tackle the problem by alternately solving the precoding and passive beamforming sub-problems. We employ the fractional programming method to convert the precoding sub-problem to a convex problem. We further show that the passive beamforming sub-problem is a feasibility-check problem, which can be solved by skilfully expanding the feasible region and applying successive convex approximation.

- We show that the complexity of the groupwise SIC approach is much lower than that of the information-theoretic approach. We also show the groupwise SIC approach substantially outperforms the information-theoretic approach and other counterparts, especially for a relatively small allowed iteration number of the receiver.

The rest of this paper is organized as follows. Section II presents the channel model and the transceiver for a RIS-aided uplink MIMO-OFDM system. In Section III, the state evolution is described, and the optimization problem for the iterative receiver is formulated. In Section IV the groupwise SIC approach is described. Numerical results are provided in Section V. Section VI concludes the paper.

**Notions:** We use a bold symbol lowercase letter and a bold symbol capital letter to denote a vector and a matrix, respectively. The Frobenius norm, trace, transpose, conjugate transpose, and inverse of a matrix are denoted by $\| \cdot \|_F$, $\text{tr}(\cdot)$, $(\cdot)^\dagger$, $(\cdot)^\text{H}$ and $(\cdot)^{-1}$, respectively; $[\mathbf{A}]^2$ denotes $\mathbf{A}\mathbf{A}^\text{H}$; $\text{diag}(\alpha)$ forms a diagonal matrix with the diagonal elements in $\alpha$; $\| \cdot \|_2$ denotes $\ell_2$-norm of a vector; $\{ \cdot \}$ denotes the modulus of a complex number; the complex conjugate of a vector is denoted by $(\cdot)^*$; $\otimes$ denotes the Kronecker product; $\odot$ denotes the Hadamard product; $* \!$ denotes the circular convolution; $\nabla$ denotes the gradient operator; $\mathcal{O}$ is the big-O notation. The complex Gaussian distribution of $x$ with mean $\mathbf{m}$ and covariance matrix $\Sigma$ is denoted by $x \sim \mathcal{CN}(\mathbf{m}, \Sigma)$. 

**II. System Model**

**A. Channel Model**

As shown in Fig. 1, we consider an uplink RIS-aided multiuser MIMO-OFDM system, where a RIS with $N$ reflecting elements is deployed to enhance the communication between $K$ single-antenna users and an $M$-antenna base station (BS). Let $J$ denote the number of subcarriers of each OFDM symbol, and these $J$ subcarriers are shared by all the users. We assume frequency-selective fading channels, and the user-BS, user-RIS, and RIS-BS links contain $L_{ub}$, $L_{ur}$ and $L_{rb}$ taps in the impulse response, respectively [35]. Each user adopts a cyclic prefix (CP) of length $L_{cp}$ with $L_{cp} \geq \max\{L_{ub}, L_{ur} + L_{rb}\}$.
where each block of the RIS remain constant over the subcarriers. The frequency selectivity of the RIS can be ignored, i.e., the reflecting coefficients occupied bandwidth is large enough. In this paper, for ease of exposition, given by

\[ k \]

where perfect channel state information (CSI) is available at the BS are unit-modulus, i.e., \( |\theta_n| = 1 \), \( \forall n \). We also assume that \( \theta_n \) is constant within the considered frequency band.\(^1\) Besides, the phase can be adjusted independently in \( [0, 2\pi] \), and all the reflecting coefficients are unit-modulus, i.e., \( |\theta_n| = 1 \), \( \forall n \). We also assume that perfect channel state information (CSI) is available at the BS side. In practice, the CSI can be acquired by existing channel estimation methods; see, e.g., [37] and [38].

With (1), the time-domain channel matrix of the direct link between user \( k \) and the BS is a block-circulant matrix given by

\[
H_{ub}^k = \begin{bmatrix}
H_{ub}^k(1) & H_{ub}^k(J) & \cdots & H_{ub}^k(2)
\end{bmatrix} & \in \mathbb{C}^{JM \times J}, (2)
\]

where each block \( H_{ub}^k(j) = [h_{ub,k}(j), h_{ub,k}(j+1), \ldots, h_{ub,k}(j+M-1)]^T \in \mathbb{C}^{M \times 1} \) is the channel at the \( j \)-th tap. Similarly, the time-domain channel matrix between the \( j \)-th user and the \( n \)-th RIS element \( H_{ur}^{n,k} \in \mathbb{C}^{J \times J} \) is a circulant matrix, and the time-domain channel matrix between the \( n \)-th RIS element and the BS \( H_{rb}^n \in \mathbb{C}^{JM \times J} \) is a block-circulant matrix. Let \( r = [r^T(1), \ldots, r^T(J)]^T \in \mathbb{C}^{JM \times 1} \) and \( n = [n^T(1), \ldots, n^T(J)]^T \in \mathbb{C}^{JM \times 1} \), where \( r(j) = [r_1(j), \ldots, r_M(j)]^T \) and \( n(j) = [n_1(j), \ldots, n_M(j)]^T \).

Then, the received baseband signal is given by

\[
r = \sum_{k=1}^{K} \left( \sum_{n=1}^{N} \theta_n H_{rb}^n H_{ur}^{n,k} \right) s_k + n. \tag{3}
\]

The frequency-domain channel \( G_{rb}^k \) corresponding to \( H_{rb}^n \) is a block-diagonal matrix given by

\[
G_{rb}^k = \begin{bmatrix}
G_{rb}^k(1) & 0 & \cdots & 0
0 & G_{rb}^k(2) & \cdots & 0
\vdots & \vdots & \ddots & \vdots
0 & 0 & \cdots & G_{rb}^k(J)
\end{bmatrix} & \in \mathbb{C}^{JM \times J} \tag{4a}
\]

with its \( j \)-th diagonal block calculated as

\[
G_{rb}^k(j) = \frac{1}{J} \sum_{l=1}^{J} H_{rb}^k(l) \exp(-i2\pi j(l-1)/J), \tag{4b}
\]

where \( i = \sqrt{-1} \). A concise form of (4) is given by

\[
G_{rb}^k = F_M H_{rb}^k F_N^H, \quad H_{rb}^k = F_M G_{rb}^k F_N, \tag{5}
\]

where \( F \in \mathbb{C}^{J \times J} \) is the \( J \)-by-\( J \) unitary discrete Fourier transform (DFT) matrix, \( F_M = F \otimes I_M \) with \( I_M \) being the \( M \times M \) identity matrix. Similar to \( G_{rb}^k \), \( G_n \in \mathbb{C}^{JM \times J} \) is a block-diagonal matrix with block-size \( M \)-by-1, and \( G_{ur}^{n,k} \in \mathbb{C}^{J \times J} \) is a diagonal matrix. After applying the DFT at the BS, we obtain the frequency-domain received signal as

\[
r' = F_M r = \sum_{k=1}^{K} \left( \sum_{n=1}^{N} \theta_n G_{rb}^k G_{ur}^{n,k} \right) F s_k + \eta, \tag{6}
\]

where \( \eta = F_M n \sim \mathcal{CN}(0, \sigma^2 I) \) is an AWGN.

**B. Transmitter Structure**

The transmitter is illustrated in the upper part of Fig. 2. For each user \( k \), its data is first encoded by forward-error-correction codes, such as the low-density parity-check (LDPC) code, with coding rate \( R_k \), and then permuted by an interleaver to obtain \( b_k = [b_{k,1}, \ldots, b_{k,J}]^T \), where \( b_{k,j} \in \{0, 1\}^{Q_k} \) and \( Q_k \) is the number of bits per modulated symbol of user \( k \). For simplicity, we assume that all the users adopt the same type of modulation, implying \( Q_1 = \ldots = Q_K = Q \). The modulated vector \( x_k \in \mathbb{C}^{J \times 1} \) is generated by mapping each \( b_{k,j} \) to a discrete constellation \( \mathcal{X}_k = \{c_{k,1}, \ldots, c_{k,Q} \} \) with zero mean and unit variance. Then, \( x_k \) is linearly precoded by \( [24] \)

\[
s_k = F^H W_k F x_k, \tag{7}
\]

where \( W_k = \text{diag}\{W_k(1,1), \ldots, W_k(J,J)\} \) is used for power allocation among the subcarriers. Substituting (7)
into (6), we obtain

\[ r' = \sum_{k=1}^{K} \left( G_k^{ab} + \sum_{n=1}^{N} \theta_n G_n^{th} G_{k,n}^{ur} \right) W_k F x_k + \eta \tag{8a} \]

\[ = \sum_{k=1}^{K} G_k(\theta) W_k F x_k + \eta, \tag{8b} \]

where \( G_k(\theta) = G_k^{ab} + \sum_{n=1}^{N} \theta_n G_n^{th} G_{k,n}^{ur} \) is the equivalent channel matrix, and \( \theta = [\theta_1, \ldots, \theta_N]^T \).

C. Receiver Structure

The existing works on RIS mostly aim to improve the system performance by assuming an ideal receiver with capacity-achieving performance. Such a capacity-achieving receiver usually involves prohibitively high complexity. Instead, we here consider a low-complexity iterative receiver [24], which is meaningful for practical implementation.

As illustrated in the lower part of Fig. 2, the iterative receiver consists of an elementary signal estimator (ESE) that handles the linear constraint and the coding constraint, respectively. The ESE and the DEC are executed iteratively.

1) LMMSE-ESE Module: The ESE carries out the linear minimum mean-square error (LMMSE) estimation based on the channel input \( r' \) and the messages from the DEC \( \{x_{pri,k}, v_k\} \). The signal \( x_k \) is approximated by a Gaussian distribution \( CN(0, \rho^2 I) \). Given \( r' \), the a posteriori covariance matrix and a posterior mean of \( x_k \) are expressed as [24]

\[ V_{post,k} = v_k I - v_k^2 A_k^H V^{-1} A_k, \tag{9a} \]

\[ x_{post,k} = x_{pri,k} + v_k A_k^H V^{-1} \left( r' - \sum_{k'=1}^{K} A_k^c x_{pri,k} \right), \tag{9b} \]

where

\[ A_k = G_k(\theta) W_k F, \tag{9c} \]

\[ V = \sum_{k'=1}^{K} v_{pri,k'} A_k^c A_k^{c*} + \sigma^2 I. \tag{9d} \]

Following [39], we calculate the extrinsic variance and mean as the output of the ESE:

\[ v_{ext,k} = \left( v_{post,k} - v_k^{-1} \right)^{-1}, \tag{10a} \]

\[ x_{ext,k} = v_{ext,k} \left( v_{post,k} x_{post,k} - v_k^{-1} x_{pri,k} \right), \tag{10b} \]

where \( v_{post,k} = \frac{1}{2} \text{tr} \{ V_{post,k} \} \). In (10), \( x_{ext,k} \) is modeled as an AWGN observation of \( x_k \):

\[ x_{ext,k} = x_k + \xi_k, \tag{11} \]

where \( \xi_k \sim \mathcal{CN}(0, \rho_k^2 I) \) with \( \rho_k = v_{ext,k}^{-1} \) being the SINR of the effective channel in (11).

2) DEC Module: The DEC module is composed of a posterior probability (APP) decoders, de-interleavers/interleavers, and soft demodulators/modulators for all the users. Given \( x_{ext,k} \), the soft demodulator calculates the log-likelihood ratio (LLR) of each \( b_{k,j} \) as

\[ \lambda_{k,j}(q) = \ln \frac{p(b_{k,j} = 0|x_{ext,k}, j)}{p(b_{k,j} = 1|x_{ext,k}, j)}, \tag{12} \]

where \( q = 1, 2, \ldots, Q \).

After de-interleaving, APP decoding, and interleaving, the a posteriori LLR for each \( b_{k,j} \) is obtained, denoted by \( \gamma_{k,j} \). Then we update the mean and variance of each \( x_{k,j} \):

\[ \tilde{x}_{k,j} = E \{ x_{k,j} | \gamma_{k,j} \} = \sum_{c_k \in \mathcal{A}_k} c_k p \left( x_{k,j} = c_k | \gamma_{k,j} \right), \tag{13a} \]

\[ \bar{v}_k = \frac{1}{J} \sum_{j=1}^{J} \sum_{c_k \in \mathcal{A}_k} |c_k - \tilde{x}_{k,j}|^2 p \left( x_{k,j} = c_k | \gamma_{k,j} \right). \tag{13b} \]

Similarly to (10a) and (10b), we update the a priori mean and variance of \( x_k \) for the ESE by

\[ v_k = \left( v_{ext,k}^2 - v_k^{-1} \right)^{-1}, \tag{14a} \]

\[ x_{pri,k} = v_k \left( v_k^{-1} x_k - v_{ext,k} x_{ext,k} \right). \tag{14b} \]

The remainder of this paper is devoted to the optimization of the system performance over the power allocation matrix \( \{W\}_k \) and the RIS phase shifts \( \Theta \). This optimization can be done based on the well-known capacity region of the multiple access channel (MAC), with the details given in Appendix A. However, there are two issues with this information-theoretic approach. First, the complexity of the information-theoretic approach is exponential in the number of users (i.e., \( K \)), since the capacity region of a \( K \)-user MAC channel involves \( 2^K - 1 \) rate constraints. Second, a practical iterative receiver may perform far away from the capacity due to implementation limitations. For example, with a limited computational power, the iterative receiver may strictly limit its iteration number in exchanging messages between the ESE and the DEC. In this case, the optimization result based on information theory may be not a good choice for the practical system. For these reasons, we propose a state evolution based optimization approach with lower complexity and better performance in the following section.
III. PERFORMANCE ANALYSIS

In this section, we describe the state evolution to characterize the performance of the iterative receiver, and then formulate the joint precoding and passive beamforming optimization problem. First, we provide the transfer functions of the DEC and the LMMSE-ESE, based on which the state evolution is established. Then, we formulate the joint optimization problem to reduce the total transmit power under the constraints of user BER and maximal iteration number.

A. State Evolution (SE)

The SE is a semi-analytical method for the performance evaluation of an iterative receiver. Specifically, we use the output SINR \( \rho = [\rho_1, \ldots, \rho_K]^T \) to characterize the performance of the LMMSE-ESE, and the output variance \( v = [v_1, \ldots, v_K]^T \) to characterize the performance of the DEC, while the output of a module is the input of the other module. As shown in Fig. 3, we track the performance by the following recursion: Start with \( v(0) = [1, \ldots, 1]^T \),

\[
\rho(t) = \phi(v(t-1)), \\
v(t) = \psi(\rho(t)),
\]

where the superscript \( t = 1, 2, \ldots, T \) is the iteration number with \( T \) being the maximum iteration number, the MIMO function \( \rho = \phi(v) \) is the transfer function of the LMMSE-ESE, and the function \( v = \psi(\rho) \) is the transfer function of the DEC, consisting of \( K \) separable single-input single-output (SISO) functions \( v_k = \psi_k(\rho_k), k = 1, \ldots, K \). We assume that the channel code of each user is identical to each other, and thus \( \psi(\cdot) = \psi_1(\cdot) = \ldots = \psi_K(\cdot) \).

We first describe the ESE transfer function \( \phi(v) \). Without loss of generality, denote \( \phi(v) = [\phi_1(v), \ldots, \phi_K(v)]^T \). From (9), (10) and \( \rho_k = v_{ext,k} \) under (11), \( \phi_k(v) \) can be expressed as

\[
\rho_k = \phi_k(v) = \frac{T_k}{1 - v_k T_k},
\]

where

\[
T_k = \frac{1}{J} \text{tr} \left\{ A_k^H \left( \sum_{k'=1}^K v_{k'}^2 [G_k^H(\theta)W_k]_k^2 + \sigma^2 I \right) A_k \right\}^{-1}.
\]

Hence, \( \phi_k(\cdot) \) is a function of \( \{W_k'\} \) and \( \theta \).

We now consider the DEC transfer function \( \psi(\rho) \). Since each user’s data are decoded separately at the DEC, we can express \( \psi(\rho) \) as \( \psi(\rho) = [\psi(\rho_1), \ldots, \psi(\rho_K)]^T \), where \( \psi(\cdot) \) is a monotone increasing function of the SINR \( \rho_k \). It is difficult to obtain the analytical expression of the DEC transfer function \( \psi(\cdot) \), which can be numerically obtained by local Monte Carlo decoding by taking (11) as the input [24].

Denote by \( P_{e,k} \) the bit error probability of user \( k \). Note that \( P_{e,k} = \xi(v_k) \) is a monotone increasing function that maps the output of DEC \( v_k \) to the bit error probability \( P_{e,k} \). We assume that \( v_k \) can be obtained by simulations similarly to \( \psi(\cdot) \). Thus, the output bit error probability at the last iteration is given by \( \xi(v_k^{(T)}) \). Given a target BER performance \( P_{\text{tar},k} \), the required \( v_k \) can be calculated by \( v_{\text{tar},k} = \xi^{-1}(P_{\text{tar},k}) \). We now consider the DEC transfer function \( \phi_k(v) \), above the inverse of the DEC transfer function \( \psi(\cdot) \), i.e.,

\[
\phi_k(v) > \psi^{-1}(v_k), \quad \forall v_k \in \mathcal{L}, \quad \forall k.
\]

Fig. 4 illustrates the path condition (17) in a two-user system. We see that there generally exist multiple potential paths that satisfy the path condition.

B. Problem Formulation

We aim to jointly optimize \( \{W_k\}, \theta \) and path \( \mathcal{L} \) to minimize the total transmit power for the iterative receiver under the constraints of target BER and the maximum number of iterations. This problem can be formulated as

\[
\begin{align}
\mathcal{P}_{\text{He}} : \min_{\{W_k\}, \theta, \mathcal{L}} & \sum_{k=1}^K \sum_{j=1}^J |W_k(j, j)|^2 \\
\text{s.t.} \quad & \phi_k(v) > \psi^{-1}(v_k), \quad \forall v_k \in \mathcal{L}, \\
& T \leq T_{\text{max}}, \\
& |\theta_n| = 1, \forall n,
\end{align}
\]

where (19b) is the target BER constraint from (18), (19c) the constraint of the maximal allowed iteration number \( T_{\text{max}} \), and (19d) the unit-modulus constraint of the RIS elements. In practice, \( T_{\text{max}} \) is usually set to a small integer to reduce the computational complexity of the receiver.

It is generally difficult to find the globally optimal solution to \( \mathcal{P}_{\text{He}} \). On one hand, the unit-modulus constraint (19d) is non-convex, and so is the constraint (19b) as seen from (16). Therefore, \( \mathcal{P}_{\text{He}} \) is a non-convex optimization program that is difficult to solve. On the other hand, the optimization result of \( \{W_k\} \) and \( \theta \) highly depends on the choice of the path \( \mathcal{L} \), where \( \mathcal{L} \) is a curve starting from \( (1, \ldots, 1) \) and ending at \( (v_{\text{tar},1}, \ldots, v_{\text{tar},K}) \). It is computationally infeasible to search
over all the possible paths in the $K$-dimension space [25], [26]. In [26], the authors proposed a diagonal path as shown in Fig. 4(c), where the input variances of the ESE are set to equal to each other. Thus, the corresponding path is the diagonal of the unit square (or equivalently, a two-dimensional unit cube) formed by the initial point $(1, 1)$ and the point $(0, 0)$.

In this section, we present a groupwise SIC approach for problem $\mathcal{P}_{\text{SIC}}$. We start with the description of the groupwise SIC approach. Based on that, $\mathcal{P}_{\text{SIC}}$ is simplified and then solved by alternately optimizing precoding and passive beamforming.

### A. Groupwise SIC

We first describe the groupwise SIC approach [29], a.k.a, sequential group detection [27], [28]. In the groupwise SIC, the users are divided into several groups, and are decoded and canceled successively in a group-by-group manner. Following this idea, we divide the users into $N_G = T_{\text{max}}$ groups, and the users in group $t$ are decoded (i.e., to achieve the target BER) at the $t$th iteration. Denote by $K = \{1, \ldots, K\}$ the total user set, and by $G_t = \{k_{t,1}, \ldots, k_{t,J_t}\}$ the user set of group $t$. As shown in Fig. 5, the users are decoded in the order from $G_1$ to $G_{N_G}$, and the interference from the decoded groups is assumed to be perfectly canceled. The path $\mathcal{L}$ of the groupwise SIC in a 2-user system is illustrated in Fig. 4(b). The maximal iteration number is $T_{\text{max}} = 2$, and thus the users are divided into 2 groups. The first group only includes user 2, and the second group only includes user 1. At the first iteration, i.e., from the initial point to point $v^{(1)}$, the first group is decoded and then canceled. At the second iteration, i.e., from $v^{(1)}$ to the target point, the second group is decoded.

The optimal user grouping is generally difficult to determine. We will discuss how to find a sub-optimal user grouping strategy later in Section IV-D. Here we focus on the joint optimization of $\{W\}$ and $\theta$ for fixed user grouping $\{G_t\}$. Given $\{G_t\}$, $\mathcal{P}_{\text{SIC}}$ is reduced to

$$
\mathcal{P}_{\text{SIC}} : \min_{\{W_k\}, \theta} \sum_{k=1}^{K} \sum_{j=1}^{J} |W_k(j,j)|^2
$$

s.t.

$$
\phi_k^j \geq \rho_{\text{tar}}, \forall k,\forall \phi_k^j,
$$

$$
|\theta_n| = 1, \forall n.
$$

where

$$
\rho_{\text{tar}} = \psi^{-1}(v_{\text{tar}}),
$$

$$
\phi_k^j = \frac{\tau_k^j}{1 - \tau_k^j},
$$

$$
\tau_k^j = \frac{1}{J} \text{tr} \left( A_k^H \left( \sum_{k' \in \{G_t\}} [\mathbf{G}_{k'}(\theta)W_{k'}]^2 + \sigma^2 \mathbf{I} \right) A_k \right)^{-1}.
$$

Eq. (20b) is the user BER constraint, and (20c) is the RIS unit-modulus constraint. Different from (16b), for user $k$ in group
We employ the fractional programming (FP) [34] to replace the fractions and matrix inversions involved in \(\phi^P\). Then, by introducing auxiliary variables \(\tau^P_k\) and \(\phi^P\) is non-convex and is hard to solve directly. In this paper, to obtain an approximate solution, we use the alternating optimization (AO) method to find a sub-optimal choice of \(\{W_k\}\) and \(\theta\).

**B. Optimization of \(\{W_k\}\) Given \(\theta\)**

Given \(\theta\), \(P_{\text{SIC}}^P\) in (20) is reduced to

\[
P_{1,1}^P: \min_{\{W_k\}} \sum_{k=1}^K \sum_{j=1}^J |W_k(j,j)|^2 \quad (21a)
\]

\[
s.t. \quad \phi^P_k \geq \rho_{\text{tar},k}, \forall k. \quad (21b)
\]

However, \(P_{1,1}^P\) is a non-convex optimization problem due to the fractions and matrix inversions involved in \(\phi^P\). We employ the fractional programming (FP) [34] to replace (21b) with convex constraints. For user \(k\) in group \(t\), let \(w_k = [W_k(1,1), \ldots, W_k(j,j)]^T\), \(a_k = G_k w_k\) and \(B_k = \sum_{k' \in \mathcal{G}^P_t} \alpha^P_{kk'} [G_k W_k^T] + \sigma^2 I\), and we rewrite \(\tau^P_k\) in (20) as

\[
\tau^P_k = \frac{1}{J} a_k^H B_k^{-1} a_k. \quad (22)
\]

With (20c) and (22), constraint (21b) can be rewritten as

\[
a_k^H B_k^{-1} a_k \geq \frac{J \rho_{\text{tar},k}}{1 + \rho_{\text{tar},k}}, \quad \forall k. \quad (23)
\]

Then, by introducing auxiliary variables \(\{y_k \in \mathbb{C}^{J \times 1}\}\), we can convert \(P_{1,1}^P\) to

\[
P_{1,2}^P: \min_{\{W_k, y_k\}} \sum_{k=1}^K \sum_{j=1}^J |W_k(j,j)|^2 \quad (24a)
\]

\[
s.t. \quad 2\text{Re}\{a_k^H a_k\} - y_k^H B_k y_k \geq \frac{J \rho_{\text{tar},k}}{1 + \rho_{\text{tar},k}}, \quad \forall k. \quad (24b)
\]

The equivalence between \(P_{1,1}^P\) and \(P_{1,2}^P\) is obtained by [34, Theorem 2]. Then, we solve \(P_{1,2}^P\) by alternately optimizing \(\{y_k\}\) and \(\{W_k\}\) as follows.

1) Given \(\{W_k\}\), the optimal solution of \(y_k\) is

\[
y_k = B_k^{-1} a_k. \quad (25)
\]

Note that (24b) is equivalent to (23) by substituting (25) into (24b).

2) Given \(\{y_k\}\), constraint (24b) is convex [34]. Thus, \(P_{1,2}^P\) is convex and can be solved through convex optimization tools such as the interior-point method [40].

**C. Optimization of \(\theta\) Given \(\{W_k\}\)**

Given \(\{W_k\}\), \(P_{\text{SIC}}^P\) in (19) is reduced to

\[
P_{2,1}^P: \min_\theta \sum_{k=1}^K \sum_{j=1}^J |W_k(j,j)|^2 \quad (26a)
\]

\[
s.t. \quad \phi^P_k \geq \rho_{\text{tar},k}, \quad \forall k, \quad |\theta_n| = 1, \forall n. \quad (26b)
\]

In fact, \(P_{2,1}^P\) is a feasibility-check problem since the objective function is invariant to the variable \(\theta\). To improve the optimization performance, we try to increase the minimum gap between \(\phi^P_k\) and \(\rho_{\text{tar},k}\) by optimizing \(\theta\), which provides a wider feasible region of \(\{W_k\}\) for \(P_{2,1}^P\) at the next iteration and allows for further power reduction. Thus, we reformulate the problem as

\[
P_{2,2}^P: \max_\theta \min_k \phi^P_k(\theta) - \rho_{\text{tar},k} \quad (27a)
\]

\[
s.t. \quad |\theta_n| = 1, \forall n, \quad (27b)
\]

where \(\phi^P_k\) is rewritten as \(\phi^P_k(\theta)\) to indicate that it is a function of \(\theta\). To deal with the non-convex unit-modulus constraint (27b), we replace \(\theta_n\) by \(\beta_n\), where \(\theta_n = e^{j\beta_n}\) and \(\beta_n \in \mathbb{R}, \forall n\). Let \(\beta = [\beta_1, \ldots, \beta_N]^T\), \(P_{2,2}^P\) is recast to

\[
P_{2,3}^P: \max_{\beta} \min_k \phi^P_k(\beta) - \rho_{\text{tar},k}. \quad (28)
\]

Then, by exploiting successive convex approximation (SCA), we solve a serial of surrogate problems to obtain a sub-optimal solution to \(P_{2,3}^P\) [41].

**Lemma 1:** The surrogate function

\[
l_k(\beta, \bar{\beta}) = \frac{-\beta^H \rho_{\text{tar},k} \beta}{2} + C_k(\beta) \quad (29)
\]

satisfies

\[
l_k(\bar{\beta}, \bar{\beta}) = \phi^P_k(\bar{\beta}) - \rho_{\text{tar},k}, \quad (30a)
\]

\[
l_k(\bar{\beta}, \beta) \leq \phi^P_k(\beta) - \rho_{\text{tar},k}, \quad (30b)
\]

where \(\kappa_k\) is a constant no less than the Lipschitz constant of \(\nabla l_k(\bar{\beta}, \bar{\beta})\).

\[
\alpha_k(\beta) = \frac{1}{J(1 - \tau^P_k(\beta))}, \quad (31a)
\]

\[
C_k(\beta) = \phi^P_k(\beta) - J \alpha_k(\beta) \tau^P_k(\beta) + \alpha_k(\bar{\beta}) \tau^P_k(\bar{\beta}) - \rho_{\text{tar},k}, \quad (31b)
\]

\[
l_k(\beta, \bar{\beta}) = \frac{2}{J} \text{Re}\{y_k(\beta)^H a_k(\beta)\} - \frac{1}{J} y_k(\beta)^H B_k(\beta) y_k(\bar{\beta}), \quad (31c)
\]

and \(y_k\) is given by (25).

The proof of Lemma 1 is given in Appendix B. Then, the surrogate problem is given by

\[
\max_{\beta} \min_k l_k(\beta, \bar{\beta}), \quad (32)
\]

where \(\bar{\beta}\) is the optimization result of the previous surrogate problem. As pointed in [41], by solving a series of surrogate problems with the surrogate functions satisfying (30), we can obtain a stationary point of \(P_{2,3}^P\). Note that \(l_k(\beta, \bar{\beta})\) is a concave function, and thus (32) can be easily solved by following [40].
D. User Grouping

It is computationally involved to find the optimal user grouping with the minimum transmit power. Take the exhaustive search for example, there are total \((N_G)^K\) user groupings, and we need to solve the multiuser MIMO-OFDM power minimization problem for each user grouping. The complexity is prohibitively high in practice when \(K\) is large. In this paper, we provide a heuristic low-complexity user grouping strategy based on the sum rate maximization with a fixed transmit power for each user.

We first evenly assign the users into \(N_G\) groups as follows. Following [27], we order the users based on their channel conditions, where the users with relatively good channel conditions are decoded first, and the channel condition is characterized by the achievable rate with groupwise SIC. For each group \(t\), denote by \(K_t = K - \sum_{t'=1}^{t-1} G_{t'}\) the set consisting of users that are not in groups 1 to \(t-1\), and set \(K_1 = K\).

Then, the achievable rate of user \(k\) in group \(t\) is given by [22]

\[
R_{t,k} = \log \det \left( I + \frac{1}{\sigma^2} \sum_{k' \in K_t} [G_{k'} W_{k'}]^2 \right) - \log \det \left( I + \frac{1}{\sigma^2} \sum_{k' \in K_t, k' \neq k} [G_{k'} W_{k'}]^2 \right)
\]  

(33)

Clearly, the above rate \(R_{t,k}\) of user \(k\) is obtained by assuming that the users in the first \(t-1\) groups are already decoded and canceled from the received signal, by following the groupwise SIC strategy. We order the undecoded users by their achievable rates, and assign \(T_G = \frac{K}{N_G}\) users with the largest \(R_{t,k}\) into \(G_t\). We sequentially determine the initial user grouping from \(G_1\) to \(G_{N_G}\). The achievable sum rate is given by

\[
R_{\text{sum}} = \sum_t \sum_{k \in G_t} R_{t,k}.
\]  

(34)

Then, we adjust the user grouping to further increase the sum rate by reassigning each user to the preceding group or the subsequent group. Specifically, for user \(k\) assigned to group \(t\), we calculate the sum rate after reassigning it to the preceding group as

\[
P_{\text{pre}, t} = \begin{cases} 
R_{\text{sum}}, & t = 1, \\
\sum_{t' = 1}^{t} \sum_{k' \in G_{t'}} R_{t',k'}, & 1 < t < N_G,
\end{cases}
\]  

(35)

where \(\{G_{t'}^{\text{pre}, k}\}_{t' = 1}^{N_G}\) is obtained from \(\{G_{t'}\}_{t' = 1}^{N_G}\) by reassigning user \(k\) to the preceding group \(t-1\) for \(1 < t < N_G\). Similarly, the sum rate after reassigning user \(k\) to the subsequent group is

\[
P_{\text{pre}, t} = \begin{cases} 
R_{\text{sum}}, & t = 1, \\
\sum_{t' = 1}^{N_G} \sum_{k' \in G_{t'}} R_{t',k'}, & 1 \leq t < N_G,
\end{cases}
\]  

(36)

where \(\{G_{t'}^{\text{sub}, k}\}_{t' = 1}^{N_G}\) is obtained from \(\{G_{t'}\}_{t' = 1}^{N_G}\) by reassigning user \(k\) to the subsequent group \(t+1\) for \(1 \leq t < N_G\). The overall user grouping strategy is summarized in Algorithm 1.

We see that the sum rate is non-decreasing by adjusting the user grouping. The reassigning operation is repeated for all the users until the sum rate \(R_{\text{sum}}\) does not increase anymore.

Algorithm 1 User Grouping Algorithm

\begin{itemize}
  \item [\textbf{Input}:] \(r^\prime, \theta, \{G_{k}^{\text{sub}}\}, \{G_{k}^{\text{pre}}\}, \{G_{k,n}^{\text{sub}}\}, \{W_{k}\}, N_G\).
  \end{itemize}

\begin{algorithmic}[1]
  \STATE \textbf{for} \(t = 1, \ldots, N_G\) \textbf{do}
  \STATE \hspace{1em} Order the unassigned users by (33), and assign \(T_G\) users with the largest \(R_{t,k}\) into \(G_t\).
  \STATE \textbf{end}
  \STATE \hspace{1em} Calculate the sum rate \(R_{\text{sum}}\) by (34);
  \STATE \textbf{while} \(R_{\text{sum}}\) is increasing \textbf{do}
  \STATE \hspace{1em} \textbf{for} \(k = 1, \ldots, K\) \textbf{do}
  \STATE \hspace{2em} Calculate \(P_{\text{pre}, k}^{\text{sum}}\) and \(P_{\text{sub}, k}^{\text{sum}}\);
  \STATE \hspace{2em} \textbf{if} \(P_{\text{pre}, k}^{\text{sum}} > R_{\text{sum}}\) and \(P_{\text{pre}, k}^{\text{sum}} \geq P_{\text{sub}, k}^{\text{sum}}\) \textbf{then}
  \STATE \hspace{3em} \(G_t = G_t^{\text{pre}, k}\), \(\forall t\) and \(R_{\text{sum}} = R_{\text{pre}, k}^{\text{sum}}\);
  \STATE \hspace{2em} \textbf{end}
  \STATE \hspace{2em} \textbf{if} \(P_{\text{sub}, k}^{\text{sum}} > R_{\text{sum}}\) and \(P_{\text{sum}}^{\text{sub}} > P_{\text{pre}, k}^{\text{sum}}\) \textbf{then}
  \STATE \hspace{3em} \(G_t = G_t^{\text{sub}, k}\), \(\forall t\) and \(R_{\text{sum}} = R_{\text{sub}, k}^{\text{sum}}\);
  \STATE \hspace{2em} \textbf{end}
  \STATE \textbf{end}
  \STATE \textbf{end}
  \STATE \textbf{Output:} \(\{G_t\}\).
\end{algorithmic}

Since the sum rate is non-decreasing, the convergence of the above process is guaranteed.

E. Convergence and Complexity

The overall groupwise SIC approach is summarized in Algorithm 2. The convergence of the algorithm is guaranteed, since the objective function is non-increasing during the iteration of AO and is also lower-bounded by zero. The algorithm stops when the change of the objective function is less than a predetermined threshold. Using the interior-point method [42] for \(P_{1,2}\) and \(P_{2,3}\), the complexity of one iteration of Algorithm 2 is \(O((KJ + 1.5)^{3.5} + T_1N^{3.5})\), where \(T_1\) is the number of successive convex approximation iterations. Compared with the information-theoretic approach in Appendix A, whose complexity is \(O((KJ + 2K)^{3.5} + (2K)^{3.5}N)\), the groupwise SIC algorithm has a much lower complexity as the number of users increases.

V. NUMERICAL RESULTS

In this section, we evaluate the proposed scheme with simulations. The metrics are the average BER of all the users, i.e., \(\frac{1}{K} \sum_k P_{k,k}\), and the average transmit power defined as
Consider a three-dimensional coordinate system, where the BS is located at (0, 0, 10), and the receiving antennas are a uniform linear array located on the x-axis with half-wavelength antenna spacing. The RIS, located at (50, 50, 10), is a uniform planar array parallel to the x – z plane with half-wavelength element spacing. The locations of the users are randomly and uniformly distributed in the horizontal rectangular area formed by the point (60, 0, 1.5) and the point (110, 50, 1.5). Following [43], the first path of the user-RIS and RIS-BS links is modeled as line-of-sight (LoS) paths, and the remaining paths are treated as non-LoS (NLoS) paths characterized by Rayleigh fading. The power ratio of the LoS component is \( \kappa_1 = 10 \), and the total power ratio of the NLoS paths is \( \kappa_2 = 1 \), without otherwise specified. All the paths of the user-BS links are modeled as NLoS paths. Each LoS path is expressed as a 2D array steering vector \([43]\), and the entries of the NLoS paths are independently taken from the circularly symmetric complex gaussian (CSCG) distribution. The large-scale pathlosses of the user-RIS, the RIS-BS, and the user-BS channels are \( 10^{-3}d^{-2}, 10^{-3}d^{-2.5}, \) and \( 10^{-3}d^{-4} \), respectively, with \( d \) being the distance. Unless otherwise specified, the numbers of the delay taps are \( L_{\text{bs}} = 2, L_{\text{th}} = 5, \) and \( L_{\text{ab}} = 6 \), respectively. We assume that all the users use the same channel code. The target BER is set as \( 10^{-4} \), and the LDPC code in 3GPP TR 38.212 [44] for QPSK modulation, code rate = 1/2 and information length = 2112 is used for the channel code. Since each codeword of a user is transmitted over a single OFDM symbol, the number of subcarriers is \( J = 2112 \). In practice, for complexity consideration, it is not necessary to allocate a different power for every subcarrier. Thus, we downsample the frequency channel to \( J' \) subcarriers in power optimization, i.e., \( G_k(\theta) \in \mathbb{C}^{JM \times 1} \) to \( G'_k(\theta) \in \mathbb{C}^{J'M \times 1} \). Then, every \( J'/J' \) subcarriers share the same power. We set \( \sigma^2 = -105 \) dBm, and all the results are averaged more than 200 independent channel realizations. With the transceiver in Section II, the groupwise SIC approach is compared with the following baseline approaches:

1) **No-RIS approach**: Obtain \( \{W_k\} \) by solving \( P_{\text{SIC}}^{\text{SC}} \) in Section IV-B with \( \theta = 0 \).
2) **Random-phase approach**: Obtain \( \{W_k\} \) by solving \( P_{\text{SIC}}^{\text{SC}} \) with a randomly generated \( \theta \).
3) **Non-grouping approach**: The groupwise SIC approach without grouping, i.e., \( N_G = 1 \).
4) **Information-theoretic approach**: The optimization approach in Appendix A, where the joint precoding and passive beamforming optimization problem under the constraint of the capacity region is formulated. This optimization approach is inspired by the single-user optimization algorithm in [9].
5) **Diagonal-path approach**: Using the diagonal path shown in Fig. 4(c) and removing the constraint of iteration number, Problem \( P_{\text{dir}} \) degenerates to the form similar to the single user optimization problem in [39], and can be solved using the FP and SCA derived in this paper. However, the computational complexity of the diagonal-path approach is much higher than the groupwise SIC approach, since the number of constraints of the former one is much more due to the discretization of the path \( \mathcal{L} \) [26], [39]. Due to space limitations, the details of the diagonal-path approach are omitted.

We first evaluate the information-theoretic and other approaches in scenarios with a relatively small number of users (e.g., 4 users). In the following, the optimization approaches use \( J' = 16 \) unless specified otherwise, which means that every 2112/16 = 132 subcarriers share the same power. Fig. 6 shows the performance comparison between the approaches with \( T_{\text{max}} = 4 \). First, we see that the no-RIS and random-phase approaches have a significant performance loss compared with other approaches, which shows the advantage of RIS in multiuser systems. Second, the groupwise SIC approach has a similar performance to the information-theoretic approach, and outperforms the diagonal-path and non-grouping approaches, which demonstrates the efficiency of groupwise SIC and the grouping strategy. In addition, the groupwise SIC approaches with \( J' = 1 \) and \( J' = 16 \) have almost the same performance.

We now evaluate the groupwise SIC and diagonal-path approaches in the setting with more users (e.g., 4 to 16 users), where \( J' = 1 \) for all the approaches. Fig. 7 shows the performance comparison with a varying number of RIS elements \( N \) in a 12-user system. The groupwise SIC approach has a significant performance gain compared with the diagonal approach, no matter whether the maximum number of iterations is 2,
In addition, the performance of the groupwise SIC approach improves as the increasing of $T_{\text{max}}$, which provides a clear demonstration of the performance-complexity trade-off of our considered iterative detection scheme.

The performance comparison under different $T_{\text{max}}$ in a 12-user system are shown Fig. 8. First, we see that the groupwise SIC approach always outperforms the diagonal-path and the random-phase approaches. As the number of iterations increases to 12, more than 2 dB power gain can be obtained at the target BER $= 10^{-4}$. Second, with 4 iterations, a large portion of the performance gain can be obtained, which demonstrates the advantage of the groupwise SIC approach for a low-complexity receiver.

We further consider the impact of imperfect CSI on different approaches in Fig. 9. The channel of user $k$ is represented as $G_k = \hat{G}_k + \Delta G_k$, where $\hat{G}_k$ is the estimated CSI, and $\Delta G_k$ is the channel estimation error. We adopt the statistical CSI error model in [45], where $\Delta G_k$ follows the CSCG distribution and the variance is $\delta \|G_k\|_F/J$ with $\delta \in [0, 1)$ measuring the relative amount of CSI uncertainties. When $\delta$ is less than 0.02, the groupwise SIC approach outperforms the diagonal-path approaches with a 2 dB performance gain at the target BER $= 10^{-4}$. When $\delta$ is 0.05, the performance gain increases to 3 dB, which shows the robustness of groupwise SIC to the imperfect CSI.

The performance comparison under different LoS-path powers and different numbers of delay taps is shown in Fig. 10. We consider channels with the LoS-path power $\kappa_1 = 1$, 5, and 10. Two types of delay taps are considered, i.e., $L_{\text{ur}} = 2$, $L_{\text{rb}} = 5$ and $L_{\text{ub}} = 6$, denoted by “$L_{\text{ur}} = 2$”, and $L_{\text{ur}} = 4$, $L_{\text{rb}} = 10$ and $L_{\text{ub}} = 12$, denoted by “$L_{\text{ur}} = 4$”. We see that the increase of the LoS-path power and the number of delay taps lead to a performance loss of less than 1 dB for the groupwise SIC approach, since the LoS path is dominant and the spatial diversity is weakened. However, compared with the diagonal-path approach, the performance gain of groupwise SIC approach is stably about 2 dB under different LoS-path powers and different numbers of delay taps.

Fig. 11 shows the performance comparison under a varying number of users. The groupwise SIC approach shows a significant performance gain compared with the diagonal-path approach, especially when the number of users is large. The performance gain is less than 0.5 dB when $K = 4$, and increases to about 5 dB when $K = 16$. Therefore, it can be concluded that the groupwise SIC approach has advantages in both performance and computational complexity compared with its counterparts, especially when the number of iterations is relatively small.
VI. CONCLUSION

In this paper, we studied the RIS-aided multiuser MIMO-OFDM system with a specific iterative receiver. We formulated the joint optimization problem for the iterative receiver under the constraints of user BER and maximal iteration number. We proposed the low-complexity groupwise SIC approach and converted the problem to two sub-problems of precoding and passive beamforming. For precoding, we apply the FP to deal with the non-convex constraints with matrix inversion and fractions. For passive beamforming, we redesign the feasibility-check problem, and resort to the SCA to deal with the unit-modulus constraints of RIS. We also provided a heuristic and low-complexity user grouping approach. We show that the proposed groupwise SIC approach has much lower complexity than the information-theoretic approach. Numerical simulations showed that the groupwise SIC approach outperforms the information-theoretic approach and the diagonal-path approach, especially when the iteration number of the receiver is limited to a relatively small value.

APPENDIX A

INFORMATION-THEORETIC APPROACH

A. Problem Formulation

Denote by \( K = \{1, \ldots, K\} \) the total user set. Following [22] and [46], the capacity region of the considered multiuser MIMO-OFDM transmission can be expressed as

\[
\sum_{k \in \mathcal{K}_u} Q R_k \leq \frac{1}{J + L_{cp}} \log \det \left( I + \frac{1}{\sigma^2} \sum_{k \in \mathcal{K}_u} \left| G_k(\theta) W_k \right|^2 \right),
\]

where \( Q R_k \) is the transmission rate per channel use. Then, the information-theoretic optimization problem is formulated by

\[
P_{\text{Info}} : \min_{\{W_k\}, \theta} \sum_{k=1}^{K} \sum_{j=1}^{J} |W_k(j, j)|^2
\]

s.t. \( (37) \), \( |\theta_n| = 1, \forall n \). (38b)

where (38a) is the total transmit power of all users, (37) is the capacity region constraint, and (38b) is the unit-modulus constraint of the RIS’s elements. As inspired by [9], we resort to the AO method to obtain an approximate solution to \( P_{\text{Info}} \) by optimizing \( \{W_k\} \) and each \( \theta_n \) alternately, as described in the following subsections.

B. Optimization of \( \{W_k\} \) Given \( \theta \)

Given \( \theta \), \( P_{\text{Info}} \) is reduced to

\[
P_{\text{Info}} : \min_{\{W_k\}} \sum_{k=1}^{K} \sum_{j=1}^{J} |W_k(j, j)|^2
\]

s.t. \( (37) \). (39a)

Since \( \log \det(\cdot) \) is a concave function of \( \{W_k\} \) [22], constraint (37) is convex. Therefore, \( P_{\text{Info}} \) can be solved by the existing convex optimization tools [40].

C. Optimization of \( \theta_n \) Given \( \{W_k\} \) and \( \{\theta_n, n' \neq n\} \)

Given \( \{W_k\} \) and \( \{\theta_n, n' \neq n\} \), \( P_{\text{Info}} \) is reduced to

\[
P_{\text{Info}}^{2.1} : \min_{\theta_n} \sum_{k=1}^{K} \sum_{j=1}^{J} |W_k(j, j)|^2
\]

s.t. \( (37) \), \( |\theta_n| = 1 \). (40a)

Note that \( P_{\text{Info}}^{2.1} \) is a feasibility-check problem. Similar to \( P_{\text{SIC}}^{2.1} \) in Section IV-C, by introducing an auxiliary variable \( \Delta R \), we reformulate \( P_{\text{Info}}^{2.1} \) as

\[
P_{\text{Info}}^{2.2} : \max_{\theta_n, \Delta R} \Delta R
\]

s.t. \( \sum_{k \in \mathcal{K}_u} Q(J + L_{cp})(R_k + \Delta R) \leq \log \det \left( I + \frac{1}{\sigma^2} \sum_{k \in \mathcal{K}_u} |G_k(\theta) W_k|^2 \right) \), \( \mathcal{K}_u \subseteq \mathcal{K} \), \( |\theta_n| = 1 \). (41a)

To solve this problem, we rewrite \( G_k(\theta) \) as

\[
G_k(\theta) = G_{nk} + \theta_n G_{nk}^R
\]

where \( G_{nk} = G_{nk}^b + \sum_{n'' \neq n} \theta_n' G_{nk}^b G_{k,n''}^R \). Then, the right-hand side of (41b) is rewritten as [9]

\[
C_u(\theta_n)
= \log_2 \det \left( I + \frac{1}{\sigma^2} \sum_{k \in \mathcal{K}_u} \left( 2\text{Re}\{\theta_n G_{nk}^b W_k W_k^H G_{nk}^R \} + |G_{nk}^b W_k|^2 + I \right) \right).
\]

We apply the convex relaxation by relaxing \( |\theta_n| \leq 1 \). Thus, \( P_{\text{Info}}^{2.2} \) is converted to

\[
P_{\text{Info}}^{2.3} : \max_{\theta_n, \Delta R} \Delta R
\]

s.t. \( \sum_{k \in \mathcal{K}_u} Q(J + L_{cp})(R_k + \Delta R) \leq C_u(\theta_n), \mathcal{K}_u \subseteq \mathcal{K} \), \( |\theta_n| \leq 1 \). (44b)

Since (44a)-(44c) are all convex, \( P_{\text{Info}}^{2.3} \) is a convex optimization problem that can be solved by existing convex optimization tools [40]. During the iteration of AO, if the obtained \( \theta_n \) for \( P_{\text{Info}}^{2.3} \) does not satisfy the constraint in (41b), we normalize \( \theta_n \) by \( |\theta_n| \) and then optimize \( \{W_k\} \) based on the normalized \( \theta \). As stated in [9], due to the above relaxation and normalization of \( \theta_n \), the convergence of the AO is not guaranteed.

D. Complexity

The numbers of constraints in (37) and (44b) are both about \( 2^K \). With the interior-point method [42], the complexity of one iteration is \( O((K + 2^K)^{3.5} + (2^K)^{3.5}N) \), where the complexity of solving the \( P_{\text{Info}}^{2.1} \) and \( P_{\text{Info}}^{2.3} \) is \( O((K + 2^K)^{3.5}) \) and \( O((2^K)^{3.5}) \), respectively.
\section*{Appendix B
Proof of Lemma 1}

We construct \( l_k(\beta, \beta) \) by using the following three steps.

1) Lower bound \( l_{1,k}(\beta, \beta) \) of \( \phi_k' \): Note that \( \phi_k' = \frac{\tau_k'}{\tau_k^2} \) is a convex function of \( \tau_k' \). Thus, the first order Taylor expansion of \( \phi_k'(\tau_k'(\beta)) \) at \( \tau_k^*(\beta) \) satisfies
\[
l_{1,k}(\beta, \beta) = \phi_k'(\beta) + \frac{\tau_k'(\beta) - \tau_k^*(\beta)}{1 - \tau_k^*(\beta)^2} \leq \phi_k'(\beta). \quad (45)
\]

2) Lower bound \( l_{2,k}(\beta, \beta) \) of \( \tau_k' \): With (25) in Section IV-B, we have \( y_k(\beta) = B_k^{-1}(\beta)ak(\beta) \) and \( \tau_k'(\beta) = \frac{1}{2}a_k^2(\beta)B_k^{-1}(\beta)ak(\beta) \). From [34, Theorem 2], we obtain
\[
l_{2,k}(\beta, \beta) = \frac{1}{2} Re \{ y_k^H(\beta)ak(\beta) \} - \frac{1}{2} y_k^H(\beta)B_k(y_k(\beta)) = \frac{1}{2} Re \{ y_k^H(\beta)e^{\beta} \} - \frac{1}{2} (e^{\beta}1)U_k(\beta)e^{\beta} + C_{1,k}(\beta) \leq \bar{\tau}_k^*(\beta), \quad (46a)
\]
where
\[
Y_k(\beta) = \text{diag}(y_k(\beta)), \quad u_k(\beta) = G_k^H(\beta)W_k \otimes I_M Y_k(\beta) \nonumber - \sum_{k' \in \mathbb{U}_{n,m}, \tilde{g}_i} v_{k'_k}G_k^H(\beta)Y_k(\beta)[W_{k'}]^2 \nonumber \times G_{k'_k}(\beta)[y_k(\beta)]^H, \quad (46b)
\]
\[
U_k(\beta) = \sum_{k' \in \mathbb{U}_{n,m}, \tilde{g}_i} v_{k'_k}G_k^H(\beta)W_kW_{k'_k}Y_k^H(\beta)G_{d,k}, \quad (46c)
\]
\[
C_{1,k}(\beta) = \sum_{k' \in \mathbb{U}_{n,m}, \tilde{g}_i} v_{k'_k}y_k^H(\beta)G_{k,k'_k}W_{k'}^2[W_{k'}]^2G_{d,k}^H(\beta)Y_k(\beta) \nonumber + 2 Re \{ y_k^H(\beta)G_{d,k}w_k(\beta) \} - \sigma^2 [y_k^H(\beta)]^2, \quad (46d)
\]
\[
\end{equation}

3) Lower bound \( l_{3,k}(\beta, \beta) \) of \( l_{2,k}(\beta, \beta) \): We use the second order Taylor expansion as the lower bound of \( l_{2,k}(\beta, \beta) \):
\[
l_{3,k}(\beta, \beta) = l_{2,k}(\beta, \beta) + \nabla l_{2,k}(\beta, \beta)^T(\beta - \beta) - \frac{\kappa_k}{2} \| \beta - \beta \|^2 \leq l_{2,k}(\beta, \beta) \quad (47)
\]

where \( \nabla l_{2,k}(\beta, \beta) = 2 Re \{ i(\beta^* \otimes (U_k(\beta) - u_k)) \} \) is the gradient [15], and \( \kappa_k \) is a constant on less than the Lipschitz constant of \( \nabla l_{2,k}(\beta, \beta) \) [41]. Note that \( \kappa_k \) can be chosen as follows. The gradient and the Hessian matrix of \( l_{2,k}(\beta, \beta) \) are respectively given by
\[
\nabla l_{2,k}(\beta, \beta) = 2 Re \{ i(e^{\beta}^* \otimes (U_k(\beta)e^{\beta} - u_k)) \}, \quad (48a)
\]
\[
\nabla^2 l_{2,k}(\beta, \beta) = 2 Re \{ (e^{i\beta})^*\alpha_{k,1,1}, \ldots, (e^{i\beta})^*\alpha_{k,N,N}^T \}, \quad (48b)
\]
where
\[
\alpha_{k,1} = [u_{k,1} - \sum_{n \neq 1} U_{k,n}e^{i\beta n}, U_{k,1,2}e^{i\beta 2}, \ldots, U_{k,1,N}e^{i\beta N}], \quad (48c)
\]
\[
\alpha_{k,N} = [U_{k,N,1}e^{i\beta 1}, U_{k,N,2}e^{i\beta 2}, \ldots, U_{k,N,N}e^{i\beta N}], \quad \alpha_{k,N} = \text{the nth element of } u_k(\beta), \text{ and } U_{k,n,m} \text{ is the } (n,m)\text{th element of } U_k(\beta). \quad (48d)
\]

Note that \( \| \nabla l_{2,k}(\beta, \beta) \|_F \) is upper bounded by \( \| \Gamma_k \|_F \), where \( \Gamma_k = \left[ \begin{array}{c} \Gamma_{k,1,1}, \Gamma_{k,1,2}, \ldots, \Gamma_{k,N,N}^T \end{array} \right]^T \),
\[
\Gamma_{k,1} = [u_{k,1}, [u_{k,1} + \sum_{n \neq 1} [U_{k,n,1}, U_{k,n,2}, \ldots, U_{k,n,N}]]^T, \quad (49a)
\]
\[
\Gamma_{k,2} = [U_{k,2,1}, [U_{k,2,2}, \ldots, [U_{k,2,N}, \ldots, [U_{k,2,N}]]], \quad (49b)
\]
\[
\Gamma_{k,N} = [U_{k,N,1}, U_{k,N,2}, \ldots, [U_{k,N,N}, \ldots, [U_{k,N,N}]]]. \quad (49c)
\]

Thus, Lemma 1 holds by letting \( l_{2,k}(\beta, \beta) = l_k^*(\beta, \beta) \) in (50).

\section*{References}

[1] M. D. Renzo et al., “Smart radio environments empowered by reconfigurable AI meta-surfaces: An idea whose time has come,” EURASIP J. Wireless Commun. Netw., vol. 2019, no. 1, pp. 1–20, May 2019.

[2] M. Jiao et al., “Reconfigurable intelligent surfaces for wireless communications: Overview of hardware designs, channel models, and estimation techniques,” Intell. Converged Netw., vol. 3, no. 1, pp. 1–32, Mar. 2022.

[3] Q. Wu and R. Zhang, “Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network,” IEEE Commun. Mag., vol. 58, no. 1, pp. 106–112, Jan. 2020.

[4] L. You, J. Xiong, D. W. K. Ng, C. Yuen, W. Wang, and X. Gao, “Energy efficiency and spectral efficiency tradeoff in IRS-aided multiuser MIMO uplink transmission,” IEEE Trans. Signal Process., vol. 69, pp. 14017–14021, Dec. 2020.

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

[6] L. You et al., “Reconfigurable intelligent surfaces-assisted multiuser MIMO uplink transmission with partial CSI,” IEEE Trans. Wireless Commun., vol. 20, no. 9, pp. 5613–5627, Sep. 2021.

[7] S. Zhou, W. Xu, K. Wang, M. Di Renzo, and M.-S. Alouini, “Spectral and energy efficiency of IRS-assisted MISO communication with hardware impairments,” IEEE Wireless Commun. Lett., vol. 9, no. 9, pp. 1366–1369, Sep. 2020.

[8] S. Fang et al., “Energy-efficient design of IRS-NOMA networks,” IEEE Trans. Veh. Technol., vol. 69, no. 11, pp. 14088–14092, Nov. 2020.

[9] S. Zhang and R. Zhang, “Capacity characterization for intelligent reflecting surface aided MIMO communication,” IEEE J. Sel. Areas Commun., vol. 38, no. 8, pp. 1823–1838, Aug. 2020.

[10] Q. Tao, J. Wang, and C. Zhong, “Performance analysis of intelligent reflecting surface aided communication systems,” IEEE Commun. Lett., vol. 24, no. 11, pp. 2464–2468, Nov. 2020.
Lei Liu (Member, IEEE) was born in Ji’an, Jiangxi, China. He received the Ph.D. degree in communication and information systems from Xidian University in 2017. From 2014 to 2016, he was an exchange Ph.D. student with Nanyang Technological University (NTU), Singapore, supported by the State Scholarship Fund from the China Scholarship Council (CSC). From 2016 to 2017, he was a Post-Doctoral Research Fellow with the Singapore University of Technology and Design (SUTD), Singapore. From 2017 to 2019, he was a Post-Doctoral Fellow with the City University of Hong Kong (CityU), Hong Kong, where he was a Research Fellow in 2019. From 2019 to 2023, he was an Assistant Professor with the Japan Advanced Institute of Science and Technology (JAIST), Japan. He has been a tenure-track young Professor with Zhejiang University, Hangzhou, China, since February 2023. His current research interests include message passing, MIMO, multi-user systems, modern channel coding, and information theory. He is a Senior Member of the China Institute of Communications (CIC). He was the Publications Co-Chair of the 2021 IEEE Information Theory Workshop. He was an Exemplary Reviewer (fewer than 2%) of IEEE TRANSACTIONS ON COMMUNICATIONS in 2020.

Xiaojun Yuan (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the City University of Hong Kong in 2009. From 2009 to 2011, he was a Research Fellow with the Department of Electronic Engineering, City University of Hong Kong. He was a Visiting Scholar with the Department of Electrical Engineering, University of Hawaii at Manoa, in spring and summer 2009 and the same period of 2010. From 2011 to 2014, he was a Research Assistant Professor with the Institute of Network Coding, The Chinese University of Hong Kong. From 2014 to 2017, he was an Assistant Professor with the School of Information Science and Technology, ShanghaiTech University. He is currently a state-specially-recruited Professor with the Center for Intelligent Networking and Communications, University of Electronic Science and Technology of China. His research interests include signal processing, machine learning, and wireless communications, including but not limited to intelligent communications, structured signal reconstruction, Bayesian approximate inference, and distributed learning. He has published over 200 peer-reviewed research papers in the leading international journals and conferences in the related areas. He has served on a number of technical programs for international conferences. He was a co-recipient of the Best Paper Award of IEEE International Conference on Communications (ICC) in 2014 and the Best Journal Paper Award of IEEE Technical Committee on Green Communications and Computing (TCGCC) in 2017. He is an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.