

Joint Transmit Beamforming and Reflection Optimization for Beyond Diagonal RIS Aided Multi-Cell MIMO Communication

Shuo Zheng, Shuowen Zhang

Abstract—The sixth-generation (6G) wireless networks will rely on ultra-dense multi-cell deployment to meet the high rate and connectivity demands. However, frequency reuse leads to severe inter-cell interference, particularly for cell-edge users, which limits the communication performance. To overcome this challenge, we investigate a beyond diagonal reconfigurable intelligent surface (BD-RIS) aided multi-cell multi-user downlink MIMO communication system, where a BD-RIS is deployed to enhance desired signals and suppress both intra-cell and inter-cell interference. We formulate the joint optimization problem of the transmit beamforming matrices at the BSs and the BD-RIS reflection matrix to maximize the weighted sum rate of all users, subject to the challenging unitary constraint of the BD-RIS reflection matrix and transmit power constraints at the BSs. To tackle this non-convex and difficult problem, we apply the weighted minimum mean squared error (WMMSE) method to transform the problem into an equivalent tractable form, and propose an efficient alternating optimization (AO) based algorithm to iteratively update the transmit beamforming and BD-RIS reflection using Lagrange duality theory and manifold optimization. Numerical results demonstrate the superiority of the proposed design over various benchmark schemes, and provide useful practical insights on the BD-RIS deployment strategy for multi-cell systems.

Index Terms—Beyond diagonal reconfigurable intelligent surface (BD-RIS), multi-cell multi-user communication, multiple-input multiple-output (MIMO).

I. INTRODUCTION

The evolution towards the sixth-generation (6G) wireless networks calls for ultra-dense multi-cell deployments to support ubiquitous connectivity and high-rate applications. In such scenarios, frequency reuse across neighboring cells is inevitable to meet the stringent communication requirements, which, however, also results in inter-cell interference. This is particularly critical for cell-edge users, which often experience weak desired signals while suffering from strong interference from neighboring cells. To address this problem, various interference management strategies have been investigated, among which reconfigurable intelligent surface (RIS) is a promising technology due to its capability of reconfiguring the wireless propagation environment [1]–[4]. By intelligently adjusting the phase shifts of signals impinging on its passive reflecting elements, RIS can enhance desired signal strength and suppress interference, thereby increasing the rate of multi-cell communication [5], [6].

The authors are with the Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong SAR, China (E-mails: shuo.zheng@connect.polyu.hk; shuowen.zhang@polyu.edu.hk).

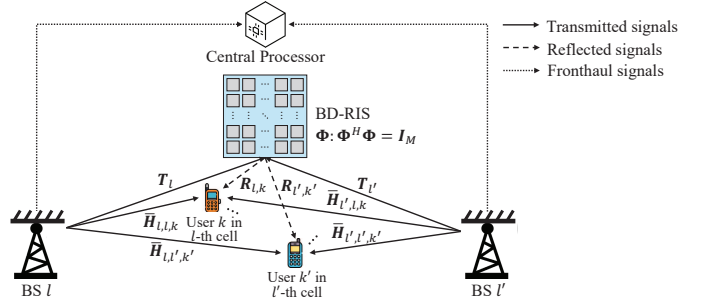


Fig. 1: Illustration of a BD-RIS aided multi-cell multi-user downlink MIMO communication system.

However, conventional (diagonal) RIS adopts a single-connected architecture that only supports diagonal reflection matrices, which restricts its beamforming flexibility and consequently limits the achievable performance gains. Recently, a novel RIS architecture, termed *beyond diagonal reconfigurable intelligent surface (BD-RIS)*, has been proposed as an advancement beyond conventional RIS [7]. Specifically, BD-RIS is capable of realizing reflection matrices beyond the diagonal structure with greater design flexibility enabled by an interconnected impedance network [8], [9]. Motivated by these advantages, several works have studied the BD-RIS reflection design in various scenarios [10]–[15]. While these works have demonstrated the potential of BD-RIS, most of them focused on single-cell systems. For multi-cell systems, prior efforts either considered assigning a dedicated BD-RIS to each base station (BS) [16]–[18], which leads to high deployment cost, or relied on frequency division across cells [19], which may cause resource underutilization. These limitations motivate our study in this work to unlock the full potential of BD-RIS in general multi-cell multi-user multiple-input multiple-output (MIMO) communication systems.

In this paper, we study a BD-RIS aided multi-cell multi-user downlink MIMO communication system, where each cell consists of a multi-antenna BS and multiple multi-antenna users and different BSs are connected via fronthaul links through a central processor. One BD-RIS is deployed to enhance the multi-cell communication performance. We aim to maximize the weighted sum rate of all users via joint optimization of the transmit beamforming matrices at all BSs and the BD-RIS reflection matrix, subject to the challenging unitary constraint of the BD-RIS reflection matrix and the transmit power constraints. This problem is non-convex and difficult

to solve due to the complex rate expression under both intra-cell and inter-cell interference as well as the couplings among optimization variables. By leveraging the weighted minimum mean squared error (WMMSE) method, we transform the problem into an equivalent form, based on which we further propose an alternating optimization (AO) based algorithm that iteratively updates the transmit beamforming matrices using Lagrange duality theory and the BD-RIS reflection matrix using manifold optimization. It is shown via numerical results that our proposed design is effective in multi-cell multi-user interference management and achieves superior rate performance over various benchmark schemes. It is also observed that under a given number of elements, deploying a large centralized BD-RIS at the cell edge yields higher rate over deployment of multiple smaller-sized distributed BD-RISs each near one BS.

II. SYSTEM MODEL

We consider a multi-cell multi-user downlink MIMO communication system with $L \geq 1$ cells. In each cell, one BS equipped with $N_t \geq 1$ transmit antennas aims to communicate with $K \geq 1$ users, each of which is equipped with $N_r \geq 1$ receive antennas. Different BSs are connected via fronthaul links through a central processor. An M -element fully-connected BD-RIS operating in the reflective mode is deployed to enhance the multi-cell multi-user communication performance, as illustrated in Fig. 1. Let $\Phi \in \mathbb{C}^{M \times M}$ denote the BD-RIS reflection matrix, which is subjected to the constraint $\Phi^H \Phi = \mathbf{I}_M$ due to the lossless property of the BD-RIS.

Let $\mathbf{s}_{l,k} \in \mathbb{C}^{N_s \times 1}$ denote the information symbol vector for the k -th user in the l -th cell, where N_s denotes the maximum number of data streams among all users. The non-zero elements in $\mathbf{s}_{l,k}$ are independent of each other and follow the circularly symmetric complex Gaussian (CSCG) distribution with zero mean and unit variance. Considering linear beamforming at the BS transmitters, the transmitted signal vector at each l -th cell, $\mathbf{x}_l \in \mathbb{C}^{N_t \times 1}$, is modeled as $\mathbf{x}_l = \sum_{k=1}^K \mathbf{F}_{l,k} \mathbf{s}_{l,k}$, where $\mathbf{F}_{l,k} \in \mathbb{C}^{N_t \times N_s}$ denotes the transmit beamforming matrix for the k -th user in the l -th cell. Let P_l denote the transmit power budget of the l -th BS, which yields $\sum_{k=1}^K \|\mathbf{F}_{l,k}\|_F^2 \leq P_l$.

The signals sent from each l -th cell will generally arrive at users in all cells through both the direct channel and a reflected channel via the BD-RIS. Let $\bar{\mathbf{H}}_{l',l,k} \in \mathbb{C}^{N_r \times N_t}$, $\mathbf{T}_{l'} \in \mathbb{C}^{M \times N_t}$, and $\mathbf{R}_{l,k} \in \mathbb{C}^{N_r \times M}$ denote the direct channel from the l' -th BS to the k -th user in the l -th cell, the channel from the l' -th BS to the BD-RIS, and the channel from the BD-RIS to the k -th user in the l -th cell, respectively. We assume that these channels are perfectly known at the BSs via channel estimation techniques [20]–[22] and fronthaul links connecting the BSs. The effective channel from the l' -th BS to the k -th user in the l -th cell is given by

$$\mathbf{H}_{l',l,k} = \bar{\mathbf{H}}_{l',l,k} + \mathbf{R}_{l,k} \Phi \mathbf{T}_{l'}. \quad (1)$$

Note that each user receives its desired signals while suffering from intra-cell and inter-cell interference. The received signal

vector at the k -th user in the l -th cell is thus given by

$$\mathbf{y}_{l,k} = \mathbf{H}_{l,l,k} \mathbf{x}_l + \sum_{\substack{l'=1 \\ l' \neq l}}^L \mathbf{H}_{l',l,k} \mathbf{x}_{l'} + \mathbf{n}_{l,k} \quad (2)$$

$$= \mathbf{H}_{l,l,k} \mathbf{F}_{l,k} \mathbf{s}_{l,k} + \sum_{(l',k') \neq (l,k)} \mathbf{H}_{l',l,k} \mathbf{F}_{l',k'} \mathbf{s}_{l',k'} + \mathbf{n}_{l,k}, \quad (3)$$

where $\mathbf{n}_{l,k} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_{N_r})$ denotes the CSCG noise vector at the k -th user in the l -th cell with σ^2 denoting the average noise power at each receive antenna. The interference-plus-noise covariance matrix at the receiver of the k -th user in the l -th cell is thus expressed as

$$\Upsilon_{l,k} = \sum_{(l',k') \neq (l,k)} \mathbf{H}_{l',l,k} \mathbf{F}_{l',k'} \mathbf{F}_{l',k'}^H \mathbf{H}_{l',l,k}^H + \sigma^2 \mathbf{I}_{N_r}. \quad (4)$$

The achievable rate of the k -th user in the l -th cell in bits per second per Hertz (bps/Hz) is thus given by

$$R_{l,k}(\mathbf{F}, \Phi) = \log_2 \det(\mathbf{I}_{N_r} + \mathbf{H}_{l,l,k} \mathbf{F}_{l,k} \mathbf{F}_{l,k}^H \mathbf{H}_{l,l,k}^H \Upsilon_{l,k}^{-1}), \quad (5)$$

where $\mathbf{F} \triangleq \{\mathbf{F}_{l,k}, \forall (l,k)\}$ denotes the collection of transmit beamforming matrices at all BSs. It is worth noting that the achievable rate for all users are dependent on the transmit beamforming matrices for all users in all cells as well as the common BD-RIS reflection matrix, which motivates our investigation on their joint optimization to maximize the multi-cell multi-user communication performance.

III. PROBLEM FORMULATION

We aim to maximize the weighted sum rate of all users by jointly optimizing the BS transmit beamforming matrices in $\mathbf{F}$ and the BD-RIS reflection matrix Φ , subject to the unitary constraint on the BD-RIS reflection matrix and the transmit power constraint at each BS. The optimization problem is formulated as

$$(P1) \max_{\mathbf{F}, \Phi} \sum_{l=1}^L \sum_{k=1}^K \alpha_{l,k} R_{l,k}(\mathbf{F}, \Phi) \quad (6)$$

$$\text{s.t.} \quad \Phi^H \Phi = \mathbf{I}_M \quad (7)$$

$$\sum_{k=1}^K \|\mathbf{F}_{l,k}\|_F^2 \leq P_l, \quad l = 1, \dots, L, \quad (8)$$

where $\alpha_{l,k}$ denotes the weight coefficient for user k in cell l .

It is worth noting that (P1) is a non-convex optimization problem due to the non-concave objective function over the transmit beamforming matrices $\mathbf{F}_{l,k}$'s and the BD-RIS reflection matrix Φ , and the non-convex unitary constraint in (7). Moreover, (P1) is particularly difficult to solve since the BD-RIS reflection matrix affects both the desired signal vectors as well as the intra-cell and inter-cell interference signal vectors, and are also coupled with the transmit beamforming matrices in different cells. The complex interference in the multi-cell system and multi-antenna nature of the user receivers also make the achievable rate expressions in (5) challenging to deal with.

In the following, we apply the WMMSE method to equivalently simplify the rate expressions, and develop an AO-based algorithm for finding a high-quality suboptimal solution to (P1).

IV. PROPOSED SOLUTION TO (P1)

In this section, we first equivalently transform (P1) into a more tractable form via the WMMSE method. Then, we propose an AO-based algorithm for the equivalent problem where the transmit beamforming matrices and the BD-RIS reflection matrix are iteratively updated via convex optimization or manifold optimization.

A. WMMSE-based Equivalent Transformation of (P1)

First, let $\mathbf{U}_{l,k}^H \in \mathbb{C}^{N_s \times N_r}$ denote the receiver decoding matrix at the k -th user of the l -th cell. The decoded information symbol vector at the k -th user in the l -th cell is given by $\hat{\mathbf{s}}_{l,k} = \mathbf{U}_{l,k}^H \mathbf{y}_{l,k}$. The corresponding mean squared error (MSE) matrix is thus given by

$$\mathbf{E}_{l,k} \triangleq \mathbb{E}[(\hat{\mathbf{s}}_{l,k} - \mathbf{s}_{l,k})(\hat{\mathbf{s}}_{l,k} - \mathbf{s}_{l,k})^H] \quad (9)$$

$$= (\mathbf{U}_{l,k}^H \mathbf{H}_{l,l,k} \mathbf{F}_{l,k} - \mathbf{I}_{N_s})(\mathbf{U}_{l,k}^H \mathbf{H}_{l,l,k} \mathbf{F}_{l,k} - \mathbf{I}_{N_s})^H + \sigma^2 \mathbf{U}_{l,k}^H \mathbf{U}_{l,k} + \sum_{(l',k') \neq (l,k)} \mathbf{U}_{l,k}^H \mathbf{H}_{l',l,k} \mathbf{F}_{l',k'} \mathbf{F}_{l',k'}^H \mathbf{H}_{l',l,k}^H \mathbf{U}_{l,k}. \quad (10)$$

Then, we introduce a set of auxiliary matrices denoted by $\mathbf{W}_{l,k} \succeq \mathbf{0}$, $\forall (l,k)$. We further define $\mathbf{W} \triangleq \{\mathbf{W}_{l,k}, \forall (l,k)\}$ and $\mathbf{U} \triangleq \{\mathbf{U}_{l,k}, \forall (l,k)\}$. Based on this, (P1) can be shown to be equivalent to the following problem [23]:

$$(P2) \quad \min_{\substack{\Phi: (7), \mathbf{F}: (8), \mathbf{U} \\ \mathbf{W}: \mathbf{W}_{l,k} \succeq \mathbf{0}, \forall l, \forall k}} \sum_{l=1}^L \sum_{k=1}^K \alpha_{l,k} (\text{tr}(\mathbf{W}_{l,k} \mathbf{E}_{l,k}) - \log_2 \det(\mathbf{W}_{l,k})). \quad (11)$$

Note that (P2) is in a more tractable form than (P1) due to the absence of the interference-plus-noise covariance matrix inversion. In the following, we propose an AO-based algorithm to iteratively update the decoding matrices in $\mathbf{U}$, the auxiliary matrices in $\mathbf{W}$, as well as the transmit beamforming matrices in $\mathbf{F}$ and the BD-RIS reflection matrix Φ with the other variables being fixed at each time.

B. Optimal Solution of $\mathbf{U}$ or $\mathbf{W}$ for (P2)

With given $\mathbf{F}$, $\mathbf{W}$, and Φ , the optimal solution of each $\mathbf{U}_{l,k}$ can be obtained according to the first-order optimality as

$$\mathbf{U}_{l,k}^* = \mathbf{J}_{l,k}^{-1} \mathbf{H}_{l,l,k} \mathbf{F}_{l,k}, \quad l = 1, \dots, L, k = 1, \dots, K, \quad (12)$$

where $\mathbf{J}_{l,k} \triangleq \mathbf{\Upsilon}_{l,k} + \mathbf{H}_{l,l,k} \mathbf{F}_{l,k} \mathbf{F}_{l,k}^H \mathbf{H}_{l,l,k}^H$.

With given $\mathbf{F}$, $\mathbf{U}$, and Φ , the optimal solution of each $\mathbf{W}_{l,k}$ can be obtained by the first-order optimality condition as

$$\mathbf{W}_{l,k}^* = \mathbf{E}_{l,k}^{-1}, \quad l = 1, \dots, L, k = 1, \dots, K. \quad (13)$$

C. Optimal Solution of $\mathbf{F}$ for (P2)

With given Φ , $\mathbf{W}$, and $\mathbf{U}$, (P2) reduces to the following problem:

$$(P2-F) \quad \min_{\mathbf{F}: (8)} \sum_{l=1}^L \sum_{k=1}^K \alpha_{l,k} \text{tr}(\mathbf{W}_{l,k} \mathbf{E}_{l,k}). \quad (14)$$

Let $\mathbf{F}_l \triangleq \{\mathbf{F}_{l,k}, \forall k\}$ denote the collection of transmit beamforming matrices at the l -th BS. We then have

$\sum_{l=1}^L \sum_{k=1}^K \alpha_{l,k} \text{tr}(\mathbf{W}_{l,k} \mathbf{E}_{l,k}) = \sum_{l=1}^L f_l(\mathbf{F}_l) + \text{constants}$, where $f_l(\mathbf{F}_l)$ is only dependent on the transmit beamforming matrices in $\mathbf{F}_l$ for the l -th BS and is given by

$$f_l(\mathbf{F}_l) = \sum_{k=1}^K \left[\alpha_{l,k} \text{tr}(\mathbf{W}_{l,k} (\mathbf{U}_{l,k}^H \mathbf{H}_{l,l,k} \mathbf{F}_{l,k} - \mathbf{I}_{N_s}) \times (\mathbf{U}_{l,k}^H \mathbf{H}_{l,l,k} \mathbf{F}_{l,k} - \mathbf{I}_{N_s})^H) + \sum_{(l',k') \neq (l,k)} \alpha_{l',k'} \text{tr}(\mathbf{W}_{l',k'} \mathbf{U}_{l',k'}^H \times \mathbf{H}_{l,l',k'} \mathbf{F}_{l,k} \mathbf{F}_{l,k}^H \mathbf{H}_{l,l',k'}^H \mathbf{U}_{l',k'}) \right]. \quad (15)$$

By further noting that $\mathbf{F}_l$'s are not coupled in the constraints of (P2-F), (P2-F) can be equivalently solved by solving the following problem for $l = 1, \dots, L$:

$$(P2-F_l) \quad \min_{\mathbf{F}_l: \sum_{k=1}^K \|\mathbf{F}_{l,k}\|_F^2 \leq P_l} f_l(\mathbf{F}_l). \quad (16)$$

Problem (P2-F_l) is convex and can be solved via the interior-point method or existing software such as CVX.

It is worth noting that the optimal solution can also be derived in semi-closed form as follows via the Lagrange duality theory. Let $\mu_l \geq 0$ denote the dual variable associated with the transmit power constraint of (P2-F_l). The Lagrangian of (P2-F_l) is given by $\mathcal{L}(\mathbf{F}_l, \mu_l) = f_l(\mathbf{F}_l) + \mu_l (\sum_{k=1}^K \|\mathbf{F}_{l,k}\|_F^2 - P_l)$. Based on the first-order optimality condition, the optimal solution to each $\mathbf{F}_{l,k}$ can be derived in semi-closed form as

$$\mathbf{F}_{l,k}^* = \alpha_{l,k} \left(\sum_{l'=1}^L \sum_{k'=1}^K \alpha_{l',k'} \mathbf{H}_{l,l',k'}^H \mathbf{U}_{l',k'} \mathbf{W}_{l',k'} \mathbf{U}_{l',k'}^H \mathbf{H}_{l,l',k'} + \mu_l^* \mathbf{I}_{N_t} \right)^{-1} \mathbf{H}_{l,l,k}^H \mathbf{U}_{l,k} \mathbf{W}_{l,k}, \quad l = 1, \dots, L, k = 1, \dots, K, \quad (17)$$

where each μ_l^* satisfies the complement slackness condition $\mu_l^* (\sum_{k=1}^K \|\mathbf{F}_{l,k}^*\|_F^2 - P_l) = 0$. Specifically, if $\sum_{l'=1}^L \sum_{k'=1}^K \alpha_{l',k'} \mathbf{H}_{l,l',k'}^H \mathbf{U}_{l',k'} \mathbf{W}_{l',k'} \mathbf{U}_{l',k'}^H \mathbf{H}_{l,l',k'}$ is invertible and $\mathbf{F}_{l,k}^*$ in (17) with $\mu_l^* = 0$ satisfies the transmit power constraint, we have $\mu_l^* = 0$. Otherwise, we need to find $\mu_l^* > 0$ such that $\sum_{k=1}^K \|\mathbf{F}_{l,k}(\mu_l^*)\|_F^2 = P_l$. Denote $\mathbf{Q}_l \triangleq \sum_{l'=1}^L \sum_{k'=1}^K \alpha_{l',k'} \mathbf{H}_{l,l',k'}^H \mathbf{U}_{l',k'} \mathbf{W}_{l',k'} \mathbf{U}_{l',k'}^H \mathbf{H}_{l,l',k'}$ and $\mathbf{D}_l \mathbf{\Lambda}_l \mathbf{D}_l^H = \mathbf{Q}_l$ as its eigenvalue decomposition (EVD). We further denote $\mathbf{C}_l = \sum_{k=1}^K \alpha_{l,k}^2 \mathbf{D}_l^H \mathbf{H}_{l,l,k}^H \mathbf{U}_{l,k} \mathbf{W}_{l,k} \mathbf{W}_{l,k}^H \mathbf{U}_{l,k}^H \mathbf{H}_{l,l,k} \mathbf{D}_l$. Then, the complement slackness condition can be equivalently written as

$$\text{tr}((\mathbf{\Lambda}_l + \mu_l^* \mathbf{I}_{N_t})^{-2} \mathbf{C}_l) = \sum_{n_t=1}^{N_t} \frac{C_{l n_t, n_t}}{(\Lambda_{l n_t, n_t} + \mu_l^*)^2} = P_l. \quad (18)$$

Since $\sum_{n_t=1}^{N_t} \frac{C_{l n_t, n_t}}{(\Lambda_{l n_t, n_t} + \mu_l^*)^2}$ is a decreasing function over $\mu_l^* > 0$, μ_l^* can be efficiently obtained by using the bi-section method.

D. Optimization of Φ for (P2)

With given $\mathbf{F}$, $\mathbf{W}$, and $\mathbf{U}$, (P2) is reduced to the following problem:

$$(P2-\Phi) \quad \min_{\Phi: (7)} \sum_{l=1}^L \sum_{k=1}^K \alpha_{l,k} \text{tr}(\mathbf{W}_{l,k} \mathbf{E}_{l,k}). \quad (19)$$

By extracting the terms related to Φ , we can re-express the objective in a more compact form as $\sum_{l=1}^L \sum_{k=1}^K \alpha_{l,k} \text{tr}(\mathbf{W}_{l,k} \mathbf{E}_{l,k}) = \text{tr}(\mathbf{A}_1 \Phi \mathbf{A}_2 \Phi^H) + 2\Re\{\text{tr}((\mathbf{B}_1 + \mathbf{B}_2)\Phi)\} + \text{constants}$, where

$$\mathbf{A}_1 = \sum_{l=1}^L \sum_{k=1}^K \alpha_{l,k} \mathbf{R}_{l,k}^H \mathbf{U}_{l,k} \mathbf{W}_{l,k} \mathbf{U}_{l,k}^H \mathbf{R}_{l,k} \quad (20)$$

$$\mathbf{A}_2 = \sum_{l'=1}^L \sum_{k'=1}^K \mathbf{T}_{l'} \mathbf{F}_{l',k'} \mathbf{F}_{l',k'}^H \mathbf{T}_{l'}^H \quad (21)$$

$$\mathbf{B}_1 = \sum_{l=1}^L \sum_{k=1}^K \sum_{l'=1}^L \sum_{k'=1}^K \alpha_{l,k} \mathbf{T}_{l'} \mathbf{F}_{l',k'} \mathbf{F}_{l',k'}^H \bar{\mathbf{H}}_{l',l,k}^H \times \mathbf{U}_{l,k} \mathbf{W}_{l,k} \mathbf{U}_{l,k}^H \mathbf{R}_{l,k} \quad (22)$$

$$\mathbf{B}_2 = - \sum_{l=1}^L \sum_{k=1}^K \alpha_{l,k} \mathbf{T}_l \mathbf{F}_{l,k} \mathbf{W}_{l,k} \mathbf{U}_{l,k}^H \mathbf{R}_{l,k}. \quad (23)$$

Then, (P2'- Φ) can be equivalently rewritten as

$$(P2'-\Phi) \min_{\Phi: (7)} \text{tr}(\mathbf{A}_1 \Phi \mathbf{A}_2 \Phi^H) + 2\Re\{\text{tr}((\mathbf{B}_1 + \mathbf{B}_2)\Phi)\}. \quad (24)$$

Note that (P2'- Φ) is still non-convex due to the unitary constraint in (7). However, notice that Φ belongs to an $M \times M$ complex Stiefel manifold represented by $\text{St}_M = \{\Phi \in \mathbb{C}^{M \times M} : \Phi^H \Phi = \mathbf{I}_M\}$. Therefore, to tackle (P2'- Φ), we can employ a manifold-based algorithm to iteratively update Φ [24]. Specifically, we first calculate the Euclidean gradient as

$$\nabla = \frac{\partial g(\Phi)}{\partial \Phi^*} = \mathbf{A}_1 \Phi \mathbf{A}_2 + \mathbf{B}_1^H + \mathbf{B}_2^H. \quad (25)$$

Based on this, the Riemannian gradient can be obtained as

$$\tilde{\nabla} = \nabla - \Phi \nabla^H \Phi. \quad (26)$$

Then, we translate the Riemannian gradient to identity for further calculating the rotation matrix:

$$\hat{\nabla} = \tilde{\nabla} \Phi^H = \nabla \Phi^H - \Phi \nabla^H. \quad (27)$$

Based on this, the rotation matrix can be obtained as follows, which provides the geodesic motion along the steepest descent direction:

$$\mathbf{P} = \exp(-\xi \hat{\nabla}), \quad (28)$$

with ξ denoting the step length that can be tuned based on the Armijo rule. Finally, Φ can be updated as follows:

$$\Phi = \mathbf{P} \Phi. \quad (29)$$

By repeating these procedures and iteratively optimizing Φ , we can obtain a high-quality suboptimal solution to Problem (P2'- Φ). According to [24], the manifold-based algorithm for (P2'- Φ) is guaranteed to converge. The computational complexity of the manifold-based algorithm is given by $\mathcal{O}(I_{\text{MF}} I_{\xi} M^3)$, where I_{MF} and I_{ξ} denote the numbers of iterations for the geodesic steepest descent and step size adjustment, respectively.

Algorithm 1 Proposed Algorithm for (P1)

- 1: **repeat**
 - 2: Update $\mathbf{U}$, $\mathbf{W}$, and $\mathbf{F}$ according to (12), (13), and (17), respectively.
 - 3: **repeat**
 - 4: Calculate ∇ , $\tilde{\nabla}$, $\hat{\nabla}$, and $\mathbf{P}$ according to (25), (26), (27), and (28), respectively.
 - 5: Update Φ according to (29).
 - 6: **until** convergence.
 - 7: **until** convergence.
-

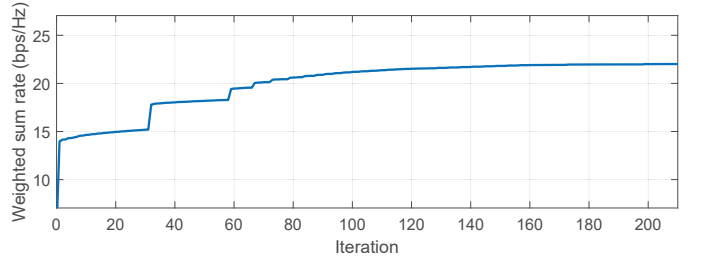


Fig. 2: Convergence behavior of Algorithm 1.

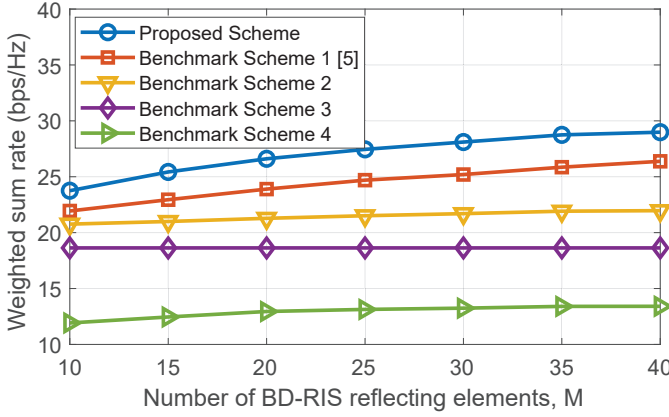
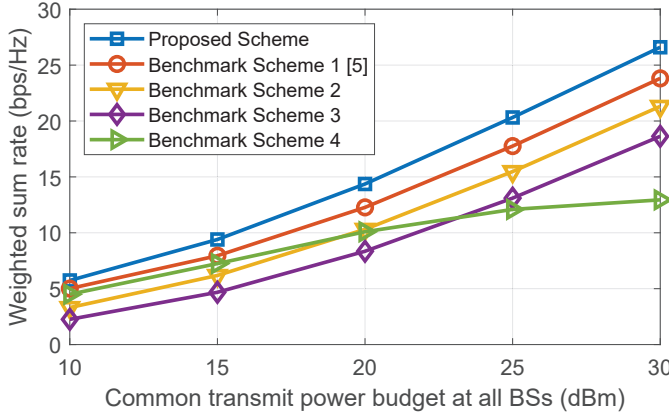
E. Overall Algorithm for (P1)

To summarize, the overall algorithm for Problem (P1) via solving (P2) is summarized in Algorithm 1. Since the objective function of (P2) is lower-bounded and each update of optimization variables leads to a non-increasing objective value, the overall algorithm is guaranteed to converge. By noting that $N_s \leq \min\{N_t, N_r\}$, the worst-case computational complexity for Algorithm 1 can be shown to be $\mathcal{O}(I_{\text{AO}}(L^2 K^2 (N_t N_r^2 + N_t N_r M + N_r M^2) + LK(N_r^3 + N_t^2 N_r + N_r^2 M + N_t^2 M) + LN_t^3 + I_{\text{MF}} I_{\xi} M^3))$, where I_{AO} denotes the total number of AO iterations.

V. NUMERICAL RESULTS

In this section, we provide numerical results to evaluate the performance of the proposed joint transmit beamforming and BD-RIS reflection design. We set $L = 2$, $K = 2$, $N_t = 4$, $N_r = 2$, $N_s = 2$, $M = 20$, $P_l = P = 30$ dBm, $l = 1, \dots, L$, and $\sigma^2 = -104$ dBm unless otherwise specified. The coordinates of the two BSs and the BD-RIS are set as $[0, 0]$ m, $[600, 0]$ m, and $[300, 0]$ m, respectively. Two users in each cell are randomly and uniformly distributed in a disk of radius 20 m centered at $[280, 0]$ m; or $[320, 0]$ m, respectively. The direct BS-user channels are modeled under Rayleigh fading with path loss exponent 3.75. The channels related to the BD-RIS are modeled under Rician fading with path loss exponent 2.2 and Rician factor 3. The path loss at reference distance 1 m is set as 30 dB. We set the weight coefficient for all users as $\alpha_{l,k} = 1$.

Firstly, we show the convergence behavior of Algorithm 1 in Fig. 2. It is observed that the proposed algorithm converges fastly within 200 iterations, and the converged weighted sum rate is significantly higher than that at the initial point. This is due to the optimal solutions of $\mathbf{U}$, $\mathbf{W}$, and $\mathbf{F}$ or high-quality suboptimal solution of Φ obtained in each iteration.

(a) Weighted sum rate vs. M .

(b) Weighted sum rate vs. power.

Fig. 3: Weighted sum rate versus number of BD-RIS elements or BS transmit power.

Then, we consider the following benchmark schemes for performance comparison:

- **Benchmark scheme 1:** We consider a **conventional diagonal RIS** aided multi-cell multi-user downlink system, where Φ is a diagonal matrix and optimized to maximize the weighted sum rate using the majorization-minimization based algorithm in [5].
- **Benchmark scheme 2:** We consider **random BD-RIS reflection**. Specifically, we randomly generate 100 BD-RIS reflection matrices and select the one with the best performance. In each realization, we randomly generate a complex matrix $S \in \mathbb{C}^{M \times M}$ where the real and imaginary parts of each entry is randomly generated under uniform distribution in $[0, 1]$; then, we perform QR decomposition of S and obtain the Q matrix as the BD-RIS reflection matrix, which is guaranteed to be unitary.
- **Benchmark scheme 3:** We consider a multi-cell multi-user downlink system **without BD-RIS**, where the transmit beamforming matrices are optimized using the same approach as in Section IV.
- **Benchmark scheme 4:** We consider a **non-cooperative** multi-cell multi-user downlink system, where each BS independently designs its transmit beamforming to

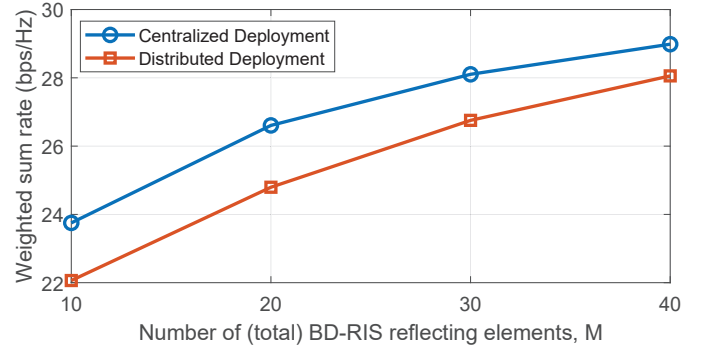


Fig. 4: Weighted sum rate with different BD-RIS deployment.

maximize the weighted sum rate of users in its cell. The BD-RIS aids the communication within one cell at a time in a round-robin manner to maximize the weighted sum rate of users in the corresponding cell.

In Fig. 3(a), we show the weighted sum rate versus the number of BD-RIS reflecting elements. It is observed that our proposed design outperforms all benchmark schemes, with an average 2.5947 bps/Hz rate gain compared with its conventional diagonal RIS aided counterpart. Moreover, as the number of BD-RIS reflecting elements increases, the performance gain of the proposed design over systems without BD-RIS, random BD-RIS reflection, and non-cooperative design increases, which demonstrates the effectiveness of the proposed BD-RIS reflection design together with the transmit beamforming design. In Fig. 3(b), we show the weighted sum rate versus the common BS transmit power budget. It is observed that our proposed design outperforms all benchmark schemes. Particularly, the non-cooperative design encounters a rate floor as the transmit power increases, which is due to its lack of inter-cell interference management ability. In contrast, our proposed design yields satisfactory interference management capability due to the judicious joint optimization of the BD-RIS reflection and transmit beamforming.

Finally, note that throughout this paper, we consider a multi-cell system aided by a single BD-RIS under a *centralized* deployment strategy. In Fig. 4, we consider another *distributed* deployment strategy of the BD-RIS and evaluate its performance in comparison with the centralized one. Specifically, we propose to deploy two distributed BD-RISs each with $M/2$ elements and located at $[5, 0]$ m or $[595, 0]$ m near one BS. The BS transmit beamforming and BD-RIS reflection under the distributed deployment are designed in a similar manner as that proposed in the paper. Note that compared to the centralized deployment where the BD-RIS is located at the cell edge, the distributed deployment strategy can enable smaller path loss for the channels between each BS and the users in its corresponding cell, due to the short distance between each BS and its adjacent BD-RIS. However, the overall reflection gain and interference management capability at each smaller-sized distributed BD-RIS is lower than that at the centralized BD-RIS. Thus, it is unclear which strategy yields better rate performance. In Fig. 4, it is observed that the centralized deployment strategy outperforms the distributed one under the

considered setup, while the performance gain decreases as the total number of reflecting elements increases. This indicates that as M increases, the signal processing capability of each distributed BD-RIS enhances, such that its gap to the large centralized BD-RIS decreases under the smaller path loss.

VI. CONCLUSIONS

This paper aimed to unlock the full potential of BD-RIS in a multi-cell multi-user downlink MIMO system via joint optimization of the transmit beamforming matrices at the BSs and the BD-RIS reflection matrix towards weighted sum rate maximization among all users. Despite the non-convexity of the problem and complex rate expressions in log-determinant forms, we proposed an AO-based algorithm which iteratively updates the transmit beamforming matrices via Lagrange duality theory and the BD-RIS reflection matrix via manifold optimization, thanks to the equivalent problem transformation via the WMMSE method. Numerical results showed that the proposed design outperforms various benchmark schemes in terms of weighted sum rate due to the judicious beamforming and reflection co-design. Moreover, it was unveiled that deploying one centralized BD-RIS at the cell edge achieves better performance than deploying multiple smaller-sized distributed BD-RISs each near one BS.

REFERENCES

- [1] 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.
- [2] 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.
- [3] —, "Intelligent reflecting surface aided multi-user communication: Capacity region and deployment strategy," *IEEE Trans. Commun.*, vol. 69, no. 9, pp. 5790–5806, Sep. 2021.
- [4] S. Zheng, B. Lv, T. Zhang, Y. Xu, G. Chen, R. Wang, and P. C. Ching, "On DoF of active RIS-assisted MIMO interference channel with arbitrary antenna configurations: When will RIS help?" *IEEE Trans. Veh. Technol.*, vol. 72, no. 12, pp. 16 828–16 833, Dec. 2023.
- [5] C. Pan, H. Ren, K. Wang, W. Xu, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Multicell MIMO communications relying on intelligent reflecting surfaces," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5218–5233, Aug. 2020.
- [6] W. Cai, R. Liu, M. Li, Y. Liu, Q. Wu, and Q. Liu, "IRS-assisted multicell multiband systems: Practical reflection model and joint beamforming design," *IEEE Trans. Commun.*, vol. 70, no. 6, pp. 3897–3911, Jun. 2022.
- [7] S. Shen, B. Clerckx, and R. Murch, "Modeling and architecture design of reconfigurable intelligent surfaces using scattering parameter network analysis," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 1229–1243, Feb. 2022.
- [8] H. Li, S. Shen, M. Nerini, and B. Clerckx, "Reconfigurable intelligent surfaces 2.0: Beyond diagonal phase shift matrices," *IEEE Commun. Mag.*, vol. 62, no. 3, pp. 102–108, 2024.
- [9] H. Li, M. Nerini, S. Shen, and B. Clerckx, "A tutorial on beyond-diagonal reconfigurable intelligent surfaces: Modeling, architectures, system design and optimization, and applications," 2025. [Online]. Available: <https://arxiv.org/abs/2505.16504>.
- [10] M. Nerini, S. Shen, and B. Clerckx, "Closed-form global optimization of beyond diagonal reconfigurable intelligent surfaces," *IEEE Trans. Wireless Commun.*, vol. 23, no. 2, pp. 1037–1051, Feb. 2024.
- [11] S. Zheng and S. Zhang, "Beyond diagonal intelligent reflecting surface aided integrated sensing and communication," *IEEE Trans. on Cogn. Commun. Netw.*, 2025, Early Access.
- [12] Y. Yuan and S. Zhang, "Beyond diagonal IRS aided OFDM: Rate maximization under frequency-dependent reflection," in *Proc. IEEE Global Commun. Conf. (GlobeCom)*, Dec. 2025.
- [13] S. Zheng and S. Zhang, "BD-IRS aided uplink ISAC exploiting prior information: SDMA or TDMA?" in *Proc. IEEE Int. Wkshps. Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2025, pp. 1–5.
- [14] X. Zhang, L. Liu, S. Zhang, W. Zhu, and H. Zhang, "Optimizing rate-CRB performance for beyond diagonal reconfigurable intelligent surface enabled ISAC," *IEEE Commun. Lett.*, 2025, Early Access.
- [15] X. Zhang, L. Liu, S. Zhang, and H. Zhang, "Beyond diagonal reconfigurable intelligent surface enabled sensing: Cramér-Rao bound optimization," *IEEE Wireless Commun. Lett.*, 2025, Early Access.
- [16] K. D. Katsanos, P. D. Lorenzo, and G. C. Alexandropoulos, "The interference broadcast channel with reconfigurable intelligent surfaces: A cooperative sum-rate maximization approach," in *Proc. IEEE Int. Wkshps. Signal Process. Adv. Wireless Commun. (SPAWC)*, Sep. 2024, pp. 551–555.
- [17] —, "Multi-RIS-empowered multiple access: A distributed sum-rate maximization approach," *IEEE J. Sel. Topics Signal Process.*, vol. 18, no. 7, pp. 1324–1338, Oct. 2024.
- [18] A. Zhang, W. U. Khan, A. K. Bashir, A. K. Dutta, A. U. Rehman, and M. M. A. Dabel, "Sum rate maximization for 6G beyond diagonal RIS-assisted multi-cell transportation systems," *IEEE Trans. Intell. Transp. Syst.*, 2025, Early Access.
- [19] A. S. de Sena, M. Rasti, N. H. Mahmood, and M. Latva-aho, "Beyond diagonal RIS for multi-band multi-cell MIMO networks: A practical frequency-dependent model and performance analysis," *IEEE Trans. Wireless Commun.*, vol. 24, no. 1, pp. 749–766, Jan. 2025.
- [20] R. Wang, S. Zhang, B. Clerckx, and L. Liu, "Low-overhead channel estimation framework for beyond diagonal reconfigurable intelligent surface assisted multi-user MIMO communication," *IEEE Trans. Signal Process.*, Mar. 2025, Early Access.
- [21] R. Wang, S. Zhang, and L. Liu, "Low-overhead channel estimation for beyond diagonal reconfigurable intelligent surface aided single-user communication," in *Proc. IEEE Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2024, pp. 305–310.
- [22] H. Li, S. Shen, Y. Zhang, and B. Clerckx, "Channel estimation and beamforming for beyond diagonal reconfigurable intelligent surfaces," *IEEE Trans. Signal Process.*, vol. 72, pp. 3318–3332, Jul. 2024.
- [23] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.
- [24] T. E. Abruđan, J. Eriksson, and V. Koivunen, "Steepest descent algorithms for optimization under unitary matrix constraint," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1134–1147, Mar. 2008.