Asynchronous Distributed Beamforming for Beyond-Diagonal RIS-aided Movable Antenna Systems

Bokai Xu, Jiayi Zhang, *Senior Member, IEEE*, Ziheng Liu, Zhe Wang,
Bo Ai, *Fellow, IEEE*, and Derrick Wing Kwan Ng, *Fellow, IEEE*

Abstract—Movable antenna (MA) technology has recently attracted significant research attention as a promising solution for enhancing wireless network performance. However, conventional MAs can only effectively serve users in close proximity, resulting in restricted coverage. To overcome this limitation, in this paper, we explore a beyond-diagonal reconfigurable intelligent surface (BD-RIS)-aided MA system. First, we propose a penalty-based block coordinate descent optimization algorithm tailored to the new constraints imposed by BD-RIS-aided MA systems. Specifically, our method decouples the inherently non-convex and coupled antenna distance constraints by introducing auxiliary optimization variables. Subsequently, the resulting problem is efficiently addressed via alternating optimization, with closed-form updates for the auxiliary variables. Furthermore, recognizing the challenges posed by large-scale BD-RIS deployments, which have the potential for serving a substantial number of users, traditional centralized optimization frameworks encounter considerable difficulties, including high computational complexity, excessive communication overheads, as well as limited scalability with increasing system size. To address these limitations, we propose an efficient asynchronous alternating direction method of multipliers (AS-ADMM) scheme aimed at maximizing the sum rate. Our numerical results demonstrate that the BD-RIS-aided MA system achieves superior performance compared to both conventional fixed position antenna and BD-RIS-aided systems. Furthermore, the proposed AS-ADMM framework can achieve a trade-off between performance and computational overhead, highlighting its potential for practical implementation in large-scale wireless communication networks.

Index Terms—Movable antenna (MA), beyond diagonal reconfigurable intelligent surface (BD-RIS), beamforming.

I. INTRODUCTION

The evolution of wireless communication systems is driven by the relentless demand for higher data rates, improved reliability, and enhanced spectral efficiency (SE) [1], [2]. Over the past decades, multiple-input multiple-output (MIMO) technology has been a cornerstone of this progression, leveraging spatial diversity and multiplexing gains to significantly enhance both system capacity and reliability. However, as wireless networks scale to accommodate massive connectivity, such as in Internet-of-Things (IoT) ecosystems and dense urban environments, traditional MIMO architectures face substantial challenges. These include prohibitive computational complexity, excessive communication overhead, and limited scalability, particularly in large-scale deployments supporting hundreds or even thousands of users. Moreover, existing communication systems are unable to exploit the spatial diversity of wireless channels, i.e., the spatial degrees of freedom (DoF), within a given service, thus failing to guarantee stringent quality of service (QoS) requirements.

To address these limitations, reconfigurable intelligent surfaces (RISs) have emerged as a promising technology. Specifically, an RIS is a quasi-passive surface composed of reconfigurable elements that can proactively manipulate the wireless propagation environment by deliberately reflecting or refracting signals, thereby extending coverage and enhancing capacity [3], [4]. Traditional RIS implementation, however, are often limited by their diagonal phase-shift matrices, which restrict their ability to achieve optimal interference suppression.

In response, beyond-diagonal RIS (BD-RIS) has recently emerged as a promising technology to address this limitation by allowing non-diagonal scattering matrices, enabling higher flexibility and efficient control over the wireless channel, particularly in practical scenarios with strong mutual coupling among elements [5]–[8]. Such flexibility is crucial for achieving improved performance gains in complex propagation environments. It is widely recognized that BD-RIS can be classified into three primary architectures, balancing performance-hardware complexity trade-off: single-connected, group-connected, and fully-connected [5]. Specifically, the single-connected architecture, widely studied in prior works, e.g., [5], [9], offers simplicity in design but yields modest performance improvements. In contrast, the fully-connected architecture delivers maximal flexibility and optimization potential, although at the expense of higher complexity and stringent control requirements. Meanwhile, the group-connected architecture strikes a pragmatic balance by partitioning RIS elements into interconnected groups, achieving substantial performance gains while maintaining tractable implementation complexity. Indeed, recent research on BD-RIS centers

on three interrelated themes: accurate electromagnetic modeling [10], hardware-efficient architecture design, and joint transceiver-RIS optimization [6], [11], [12]. For instance, various studies have introduced general BD-RIS communication models that unify diverse operating modes and architectural variants, developing computationally efficient algorithms to jointly optimize transmit precoders and BD-RIS phase shift matrices for maximizing sum-rates in communication systems. However, conventional BD-RIS-enabled MIMO systems typically employ fixed-position antenna (FPA) arrays. This immobility prevents the transceiver from full exploiting the continuous spatial DoFs in ultimately constraining attainable spatial diversity and multiplexing gains.

Recently, the emerging concept of movable antennas (MAs) has been introduced as a ground-breaking approach to overcoming the inherent limitations of FPA systems [13]–[18]. Specifically, by allowing physical antenna repositioning within a constrained space, MAs introduce a novel dimension of adaptability, facilitating on-the-fly optimization of their placement in response to real-time channel conditions [19]–[21]. As a result, it can dynamically exploit spatial variations in the wireless medium, offering potential improvements in channel capacity, interference mitigation, and received signal power. Moreover, the six-dimensional movable antenna (6DMA) significantly expands upon conventional MA capabilities by enabling adjustments in both three-dimensional (3D) position and 3D rotation [22], [23]. The integration of MA, fluid antenna (FA), and 6DMA into wireless systems has demonstrated substantial performance improvements across diverse application scenarios. For example, [13] proposed a field-response channel model to evaluate MA versus FPA system performance in both deterministic and statistical channels. Their findings highlighted the MA-aided system's significant advantages in increasing receive power and reducing outage probability. Further studies in [24] and [25] explored channel estimation for MA-aided systems leveraging compressive sensing and maximum likelihood estimation (MLE) methods, respectively. However, the practical deployment of MAs also introduces various hardware challenges, such as the need for reliable and power-efficient mechanical actuation systems and adaptive channel estimation techniques that account for antenna positions [13], [24]. In this paper, we explore a BD-RIS-aided MA system. The motivation behind integrating these two technologies lies in their complementary strengths, enabling a powerful two-stage channel optimization. Specifically, MAs at the transmitter can dynamically adjust their positions to optimize the BS-RIS channel link, for instance, by maximizing the signal power incident upon the RIS. Subsequently, the BD-RIS leverages its superior and flexible signal reflection capabilities, to further effectively manipulate this enhanced incident signal. This synergy significantly improves the RIS-user link quality, resulting in stronger desired signals and more effective a suppression of interference for multiple users. Furthermore, despite the individual merits of BD-RIS and MA technologies, their joint deployment in large-scale MIMO systems also presents a new set of optimization challenges. First, to mitigate antenna coupling, MA introduces intricate constraints on the minimum separation distance. As a result,

the feasible set of MA positions is obtained by excluding invalid areas on the antenna panel that violate this constraint, leading to a more complex solution space than those in conventional wireless resource allocation problems. Moreover, the non-convex nature of these distance constraints, coupled with their interdependence, prevents them from being address individually.

On the other hand, traditional centralized optimization frameworks, which rely on solving high-dimensional, non-convex problems, are computationally prohibitive and lack the scalability required for such systems. Fortunately, distributed optimization methods [26], [27] have emerged as a scalable and efficient alternative to centralized approaches in wireless communications, overcoming the computational bottlenecks associated with solving high-dimensional non-convex problems. Unlike centralized methods that require aggregating global channel state information (CSI), distributed frameworks enable parallel processing, making them particularly suitable for large-scale systems. Prominent techniques in this domain include dual decomposition and alternating direction method of multipliers (ADMM) variants, which have proven effective for tasks such as multi-cell coordinated beamforming [28], [29]. Among these, the asynchronous ADMM (AS-ADMM) framework stands out due to its unique advantage: it allows immediate updates upon receiving partial server responses, thus enhancing both fault tolerance and computational efficiency by eliminating synchronization delays [30]. Crucially, AS-ADMM also maintains theoretical convergence guarantees even for non-convex problems under mild conditions. In particular, some asynchronous distributed algorithms employ scaled gradient projection or primal decomposition [31], [32], but their application to BD-RIS-aided MA communication systems remains still a largely open issue. To the best of our knowledge, this work is an initial endeavor to establish an asynchronous distributed optimization framework for BD-RIS-aided MA communication systems.

This paper addresses the aforementioned challenges by proposing an AS-ADMM framework, tailored to MA-assisted BD-RIS-aided MIMO systems. Our proposed approach maximizes the sum rate while effectively ensuring scalability and computational efficiency. The primary contributions of our work are summarized below:

- First, we transform the intractable BD-RIS-aided MA sum-rate maximization problem into a tractable formulation by jointly applying the Lagrangian dual and quadratic transform methods. Subsequently, we develop an efficient alternating optimization algorithm with closed-form updates for all auxiliary variables. To handle tightly coupled MA positions, we employ a geometry-based iterative optimization approach that alternately fixes all but one antenna position to derive a suboptimal yet effective solution. The proposed algorithm avoids conservative approximations of the feasible set, ensuring no performance loss due to its reduced solution space.
- To independently optimize the beamformer for each user, we exploit the above optimization strategy as a foundation to develop a novel distributed framework based on asynchronous ADMM algorithm. The proposed distributed
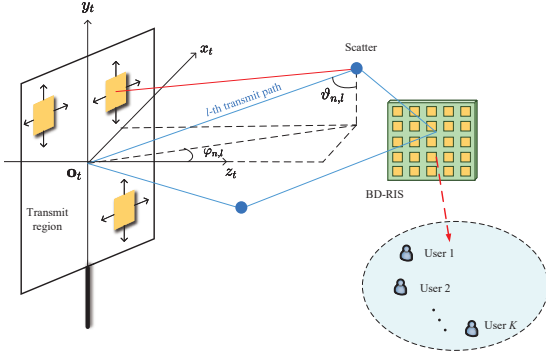
Fig. 1: A BD-RIS-aided movable antenna communication system.

framework efficiently addresses the reformulated problem via semi-closed-form solutions for both the master and worker updates. As a result, the AS-ADMM framework substantially lowers per-iteration computational complexity, rendering it well-suited for large-scale systems.

- Numerical results confirm that MA-enhanced BD-RIS systems achieve superior performance to FPA-based systems. Specifically, MAs can effectively reduce user channel correlation, thus mitigating multi-user interference. Additionally, dynamic antenna repositioning helps harness multipath effects associated with channel fading for achieving channel gain. Furthermore, our proposed asynchronous distributed beamforming optimization algorithm based on AS-ADMM can closely approach the performance of synchronous algorithms, while significantly improving computational efficiency.

The remainder of this paper is organized as follows. Section II introduces the BD-RIS-aided MA system model and problem formulation. Section III describes the centralized beamforming algorithm in BD-RIS-aided MA systems. In Section IV, we extend the centralized beamforming design to a distributed framework. Section V analyzes the convergence of the proposed algorithm. In Section VI, we provide simulation results to validate the effectiveness of the proposed optimization framework. Finally, in Section VII, we draw conclusions for this paper.

Notations: Boldface lowercase $\mathbf{a}$ and uppercase letters $\mathbf{A}$ represent vectors and matrices, respectively. The operations $\mathbf{A}^*$, $\mathbf{A}^T$, and $\mathbf{A}^H$ denote the conjugate, transpose, and conjugate transpose (Hermitian) of the matrix $\mathbf{A}$, respectively. The l_2-norm of a vector $\mathbf{a}$ is denoted by $\|\mathbf{a}\|_2$, while the Frobenius norm for matrices is represented as $\|\mathbf{A}\|_F$. The sets of $P \times Q$ complex and real matrices are denoted by $\mathbb{C}^{P \times Q}$ and $\mathbb{R}^{P \times Q}$, respectively. The operator $\mathrm{Re}[\cdot]$ refers to the real part of a complex number or variable, while $\mathbb{E}[\cdot]$ represents the expectation operator. The symbols $\nabla_{\mathbf{a}}(\cdot)$ and $\mathrm{Proj}(\cdot)$ indicate the gradient operator with respect to the variable $\mathbf{a}$ and projection operator, respectively. $\mathcal{CN}\left(0, \sigma^2\right)$ denotes the complex normal distribution with zero mean and variance σ^2. $\odot$ and $\otimes$ represent Hadamard product and Kronecker product, respectively. $\mathrm{diag}(\mathbf{a})$ denotes a diagonal matrix whose i-th diagonal element corresponds to the i-th entry of vector $\mathbf{a}$.

II. SYSTEM MODEL AND PROBLEM FORMULATION

As shown in Fig. 1, we consider a BD-RIS-aided multiuser downlink MISO system comprising a base station (BS) equipped with M movable antennas indexed as $\mathcal{M} = \{1, \cdots, M\}$, K single-antenna users indexed as $\mathcal{K} = \{1, \cdots, K\}$, and a BD-RIS with N passive reflecting elements. We assume that the direct links between the BS and users are blocked, leaving only BS-RIS-user propagation paths[1]. The effectiveness of MA technology hinges on the dynamic repositioning of a small number of antennas [14], [34], to establish favorable propagation conditions.

Specifically, let us define $\tilde{\mathbf{r}} = [\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_M] \in \mathbb{R}^{2 \times M}$ as the positions of MAs at the BS, where $\mathbf{r}_m = [x_m, y_m]^T$, $m \in \mathcal{M}$, represents the two-dimensional (2D) coordinates of the m-th MA. The array response vector of the MA array can be given by

$$
\begin{aligned}
&\mathbf{a}\left(\theta_{n,l}, \phi_{n,l}, \tilde{\mathbf{r}}\right) \\
&= \left[e^{j\frac{2\pi}{\lambda}(\phi_{n,l}x_1 + \theta_{n,l}y_1)}, \cdots, e^{j\frac{2\pi}{\lambda}(\phi_{n,l}x_M + \theta_{n,l}y_M)}\right]^T,
\end{aligned} \quad (1)
$$

where variables $\theta_{n,l} \triangleq \cos\left(\vartheta_{n,l}\right)$ and $\phi_{n,l} \triangleq \sin\left(\varphi_{n,l}\right)\sin\left(\vartheta_{n,l}\right)$ represent the direction of arrivals (DoAs) from the n-th element of the BD-RIS via the l-th path. Then, the channel[2] between the BS and the n-th element of the BD-RIS is expressed as [35]

$$
\mathbf{g}_n\left(\tilde{\mathbf{r}}\right) = \sqrt{\frac{1}{L}} \sum_{l=1}^{L} \rho_{n,l} \mathbf{a}\left(\theta_{n,l}, \phi_{n,l}, \tilde{\mathbf{r}}\right) \in \mathbb{C}^{M \times 1}, \quad (2)
$$

where $\rho_{n,l} \in \mathbb{C}$ represents the complex gain of the l-th path in the channel for the n-th BD-RIS element and L denotes the total number of resolvable spatial channel paths. Therefore, the channel of the transmit MAs for all N BD-RIS elements is given by

$$
\mathbf{G}(\tilde{\mathbf{r}}) \triangleq [\mathbf{g}_1\left(\tilde{\mathbf{r}}\right), \mathbf{g}_2\left(\tilde{\mathbf{r}}\right), \ldots, \mathbf{g}_N\left(\tilde{\mathbf{r}}\right)]^T \in \mathbb{C}^{N \times M}. \quad (3)
$$

Remark 1. *The effectiveness of MA technology primarily hinges on the dynamic repositioning of a small number of antennas [14], [34], to establish favorable propagation conditions. It is worth noting that MAs operating within a confined region $\mathcal{C}$ primarily influence the channel's small-scale fading by exploiting multipath effects.*

Then, we define the channel between user k and BD-RIS as $\mathbf{h}_k \in \mathbb{C}^{N \times 1}$, $k \in \mathcal{K}$. Let $\mathbf{s} \triangleq [s_1, \ldots, s_K]^T \in \mathbb{C}^{K \times 1}$ represent the transmit symbol vector, with $s_k \sim \mathcal{CN}(0, 1)$, $\mathbb{E}\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}_K$. At the BS, these symbols are precoded adopting the matrix $\mathbf{W} \triangleq [\mathbf{w}_1, \ldots, \mathbf{w}_K] \in \mathbb{C}^{M \times K}$, where $\mathbf{w}_k \in \mathbb{C}^{M \times 1}$ is the precoding vector for user k, $k \in \mathcal{K}$. The passive beamforming matrix of the BD-RIS is denoted as $\boldsymbol{\Theta} \in \mathbb{C}^{N \times N}$. This paper considers all three BD-RIS categories, i.e., single-, group-, and fully-connected. Specifically, for a single-connected RIS, the

---

[1]For simplicity, we omit the consideration of direct links, although our proposed algorithms are still applicable when direct links exist, as commonly assumed in literature, e.g., [5], [11], [33].

[2]CSI is derived from channel estimation algorithms [24], [25]. For simplicity, we assume perfect CSI in this work.

reconfigurable impedance network isolates each port, thereby satisfying the following constraint:

$$\mathcal{X}_1 = \left\{ \boldsymbol{\Theta} \mid \boldsymbol{\Theta} = \mathrm{diag}\left(e^{j\alpha_1}, e^{j\alpha_2}, \cdots, e^{j\alpha_N} \right) \right\}, \quad (4)$$

where $\alpha_n \in [0, 2\pi)$ represents the phase shift angle. Furthermore, the fully-connected BD-RIS comprises a reconfigurable impedance network, where each port is interconnected with the others via a reconfigurable reactance. As a result, the scattering matrix $\boldsymbol{\Theta}$ of the fully-connected BD-RIS must satisfy the following constraint

$$\mathcal{X}_2 = \left\{ \boldsymbol{\Theta} \mid \boldsymbol{\Theta} = \boldsymbol{\Theta}^T, \boldsymbol{\Theta}\boldsymbol{\Theta}^H = \mathbf{I} \right\}. \quad (5)$$

For the group-connected architecture, the N BD-RIS elements are divided into G groups, each containing N_g elements. In particular, elements within the same group are fully-connected with each other but disconnected from elements in other groups. As a result, the scattering matrix adopts the structure of a block diagonal matrix, i.e.,

$$\mathcal{X}_3 = \left\{ \boldsymbol{\Theta} \mid \boldsymbol{\Theta} = \mathrm{diag}\left(\boldsymbol{\Theta}_1, \cdots, \boldsymbol{\Theta}_G \right), \boldsymbol{\Theta}_g^H \boldsymbol{\Theta}_g = \mathbf{I}, \boldsymbol{\Theta}_g = \boldsymbol{\Theta}_g^T, \forall g \right\}, \quad (6)$$

where $\boldsymbol{\Theta}_g, g \in \{1, \cdots, G\}$ are complex-valued symmetric unitary matrices.

The precoded symbols are then up-converted to the radio frequency (RF) domain via M RF chains. After propagation through the RIS-aided channels, the signals are corrupted by additive white Gaussian noise (AWGN). As a result, the received signal at user k can be written as

$$y_k = \mathbf{h}_k^H \boldsymbol{\Theta} \mathbf{G}\left(\tilde{\mathbf{r}} \right) \sum_{k \in \mathcal{K}} \mathbf{w}_k s_k + n_k, \quad (7)$$

where $n_k \sim \mathcal{CN}\left(0, \sigma_k^2 \right)$ denotes the Gaussian noise and σ_k^2 is the corresponding noise power. Furthermore, the signal-to-interference-plus-noise ratio (SINR) for each user can be expressed as

$$\gamma_k = \frac{\left| \tilde{\mathbf{h}}_k\left(\tilde{\mathbf{r}} \right)^H \mathbf{w}_k \right|^2}{\sum_{i \in \mathcal{K}, i \neq k} \left| \tilde{\mathbf{h}}_k\left(\tilde{\mathbf{r}} \right)^H \mathbf{w}_i \right|^2 + \sigma_k^2} \quad (8)$$

and we define $\tilde{\mathbf{h}}_k\left(\tilde{\mathbf{r}} \right) \triangleq \left(\mathbf{h}_k^H \boldsymbol{\Theta} \mathbf{G}(\tilde{\mathbf{r}}) \right)^H$. Accordingly, the optimization problem is mathematically formulated as

$$\max_{\{\mathbf{W}, \tilde{\mathbf{r}}, \boldsymbol{\Theta}\}} \sum_{k \in \mathcal{K}} \log_2\left(1 + \gamma_k \right)$$
$$\text{s.t. } \mathrm{C}_1 : \sum_{k \in \mathcal{K}} \|\mathbf{w}_k\|_2^2 \leq P_{\max},$$
$$\mathrm{C}_2 : \mathbf{r}_m \in \mathcal{C}, \quad \forall m \in \mathcal{M}, \quad (9)$$
$$\mathrm{C}_3 : \|\mathbf{r}_m - \mathbf{r}_l\|_2 \geq D, \quad \forall m, l \in \mathcal{M}, \quad m \neq l,$$
$$\mathrm{C}_4 : \boldsymbol{\Theta} \in \mathcal{X}_i, \forall i \in \{1, 2, 3\},$$

where $\mathcal{C}$ represents the MA transmit region and D is the minimum antenna separation.

In general, it is highly challenging to acquire the globally optimal solution to problem (9), since the objective function is non-convex with tightly coupled optimization variables. Furthermore, the introduction of the MA's position variables $\{\mathbf{r}_m\}_{m \in \mathcal{M}}$ exacerbates the complexity, as the resulting coupled non-convex constraints C_3 further amplify the problem.

To address this challenge, the following subsections first transform problem (9) into a more tractable block optimization using fractional programming theory, followed by an iterative approach to handle each block.

III. Centralized Beamforming Algorithm

A. Equivalent Transformation

To begin with, we address the original sum rate maximization problem in a centralized manner. Specifically, we first transform problem (9) into a tractable form by applying the Lagrangian dual transform and the quadratic transform [36], [37]. Based on this, we have the following Proposition 1.

Proposition 1. *By introducing auxiliary optimization variables $\boldsymbol{\iota} = [\iota_1, \cdots, \iota_K]^T \in \mathbb{R}^K$ and $\boldsymbol{\tau} = [\tau_1, \cdots, \tau_K]^T \in \mathbb{C}^K$, the original problem in (9) can be equivalently transformed into*

$$\min_{\mathbf{W}, \tilde{\mathbf{r}}, \boldsymbol{\Theta}, \boldsymbol{\iota}, \boldsymbol{\tau}} f_a\left(\mathbf{W}, \tilde{\mathbf{r}}, \boldsymbol{\Theta}, \boldsymbol{\iota}, \boldsymbol{\tau} \right)$$
$$\text{s.t. } \mathrm{C}_1, \mathrm{C}_2, \mathrm{C}_3, \mathrm{C}_4, \quad (10)$$

where

$$f_a\left(\mathbf{W}, \tilde{\mathbf{r}}, \boldsymbol{\Theta}, \boldsymbol{\iota}, \boldsymbol{\tau} \right) = \sum_{k \in \mathcal{K}} \Bigg(-\log(1 + \iota_k) + \iota_k$$
$$-2\sqrt{1 + \iota_k} \Re\left\{ \tilde{\mathbf{h}}_k(\tilde{\mathbf{r}})^H \mathbf{w}_k \tau_k^* \right\} + |\tau_k|^2 \Big(\sum_{i \in \mathcal{K}} |\tilde{\mathbf{h}}_k(\tilde{\mathbf{r}})^H \mathbf{w}_i|^2 + \sigma_k^2 \Big) \Bigg). \quad (11)$$

Proof: Please see Appendix A. ∎

After introducing two auxiliary variables $\boldsymbol{\iota}$ and $\boldsymbol{\tau}$, we address the equivalent problem (10) using the block coordinate descent (BCD) method. This approach iteratively optimizes each block while keeping the others fixed. Below, we detail the optimization process for each block.

B. Auxiliary Vectors: Blocks $\boldsymbol{\iota}$ and $\boldsymbol{\tau}$

Given fixed $(\mathbf{W}, \tilde{\mathbf{r}}, \boldsymbol{\Theta})$, the subproblem with respect to ι_k and τ_k becomes an unconstrained convex optimization problem. Its solution can be derived by setting $\partial f_a\left(\mathbf{W}, \tilde{\mathbf{r}}, \boldsymbol{\Theta}, \boldsymbol{\iota}, \boldsymbol{\tau} \right)/\partial \iota_k = 0$ and $\partial f_a\left(\mathbf{W}, \tilde{\mathbf{r}}, \boldsymbol{\Theta}, \boldsymbol{\iota}, \boldsymbol{\tau} \right)/\partial \tau_k = 0, \forall k \in \mathcal{K}$. The optimal solutions for the auxiliary variables ι_k and $\tau_k, \forall k \in \mathcal{K}$, are then given by

$$\iota_k = \frac{\left| \tilde{\mathbf{h}}_k\left(\tilde{\mathbf{r}} \right)^H \mathbf{w}_k \right|^2}{\sum_{i \in \mathcal{K}, i \neq k} \left| \tilde{\mathbf{h}}_k\left(\tilde{\mathbf{r}} \right)^H \mathbf{w}_i \right|^2 + \sigma_k^2} \quad (12)$$

and

$$\tau_k = \frac{\sqrt{1 + \iota_k} \tilde{\mathbf{h}}_k\left(\tilde{\mathbf{r}} \right)^H \mathbf{w}_k}{\sum_{i \in \mathcal{K}} \left| \tilde{\mathbf{h}}_k\left(\tilde{\mathbf{r}} \right)^H \mathbf{w}_i \right|^2 + \sigma_k^2}, \quad (13)$$

respectively.

This article has been accepted for publication in IEEE Transactions on Wireless Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TWC.2025.3613556

5

C. Transmit Precoder: Block $\mathbf{W}$

With a given set $(\check{\mathbf{r}}, \boldsymbol{\Theta}, \boldsymbol{\iota}, \boldsymbol{\tau})$, the sub-problem with respect to $\mathbf{W}$ is given by

$$\min_{\mathbf{W}} \sum_{k \in \mathcal{K}} \left(-2\Re\{\tilde{\tau}_k \tilde{\mathbf{h}}_k(\tilde{\mathbf{r}})^H \mathbf{w}_k\} + |\tau_k|^2 \sum_{i \in \mathcal{K}} |\tilde{\mathbf{h}}_k(\tilde{\mathbf{r}})^H \mathbf{w}_i|^2 \right)$$
$$\text{s.t. } \mathrm{C}_1, \tag{14}$$

where $\tilde{\tau}_k = \sqrt{1 + \iota_k} \tau_k$. Since both the objective function and the constraint of problem (14) are convex, classical optimization methods can be applied to find its optimal solution. To this end, we adopt the Lagrange multiplier method via introducing a non-negative multiplier $\lambda \geq 0$ for the power constraint C_1. By checking the first-order optimality condition, the optimal precoder $\mathbf{w}_k$ is obtained as

$$\mathbf{w}_k^\star = \left(\sum_{i \in \mathcal{K}} \overline{\mathbf{h}}_i(\tilde{\mathbf{r}}) \overline{\mathbf{h}}_i(\tilde{\mathbf{r}})^H + \lambda^\star \mathbf{I}_N \right)^{-1} \times \sqrt{1 + \iota_k} \mathbf{h}_k(\tilde{\mathbf{r}}), \tag{15}$$

where $\lambda^\star$ can be efficiently determined via a simple bisection search, and $\overline{\mathbf{h}}_k(\tilde{\mathbf{r}}) \triangleq \tau_k \tilde{\mathbf{h}}_k(\tilde{\mathbf{r}}), \forall k \in \mathcal{K}$.

D. BD-RIS Matrix: Block $\boldsymbol{\Theta}$

When $(\check{\mathbf{r}}, \mathbf{W}, \boldsymbol{\iota}, \boldsymbol{\tau})$ are determined, the sub-problem with respect to $\boldsymbol{\Theta}$ is written as

$$\min_{\boldsymbol{\Theta}} \sum_{k \in \mathcal{K}} \left(-2\Re\{\tilde{\tau}_k^* \mathbf{h}_k^H \boldsymbol{\Theta} \mathbf{G}(\tilde{\mathbf{r}}) \mathbf{w}_k\} + |\tau_k|^2 \sum_{i \in \mathcal{K}} |\mathbf{h}_k^H \boldsymbol{\Theta} \mathbf{G}(\tilde{\mathbf{r}}) \mathbf{w}_i|^2 \right)$$
$$\text{s.t. } \mathrm{C}_4. \tag{16}$$

We propose a symmetric unity projection method for handling problem (16), which consists of the following two steps. First, we relax the non-convex constraints into a convex set and solve the resulting relaxed problem. Second, we project the relaxed solution back onto the feasible points within the non-convex sets C_4.

Relaxed solution of (16): To derive a practice solution while maintaining a nearly optimal performance, we propose the following solution based on the gradient decent approach. The relaxed problem becomes

$$\min_{\boldsymbol{\Theta}} \sum_{k \in \mathcal{K}} \left(-2\Re\{\tilde{\tau}_k^* \mathbf{h}_k^H \boldsymbol{\Theta} \mathbf{G}(\tilde{\mathbf{r}}) \mathbf{w}_k\} + |\tau_k|^2 \sum_{i \in \mathcal{K}} |\mathbf{h}_k^H \boldsymbol{\Theta} \mathbf{G}(\tilde{\mathbf{r}}) \mathbf{w}_i|^2 \right)$$
$$\text{s.t. } \boldsymbol{\Theta} \in \mathcal{M}, \tag{17}$$

where $\mathcal{M}$ is a convex sphere set satisfying $\{\boldsymbol{\Theta} \mid \|\boldsymbol{\Theta}\|_F^2 \leq N\}$. The gradient of the objective function is computed as

$$\nabla_{\boldsymbol{\Theta}} f_a(\boldsymbol{\Theta})$$
$$= \sum_{k \in \mathcal{K}} \left(-2\tilde{\tau}_k \mathbf{h}_k \mathbf{w}_k^H \mathbf{G}(\tilde{\mathbf{r}})^H + 2|\tau_k|^2 \sum_{i \in \mathcal{K}} \mathbf{h}_k \mathbf{h}_k^H \boldsymbol{\Theta} \mathbf{G}(\tilde{\mathbf{r}}) \mathbf{w}_i \mathbf{w}_i^H \mathbf{G}(\tilde{\mathbf{r}})^H \right) \tag{18}$$

and set it equals to $\mathbf{0}$ and project to $\mathcal{M}$. Consequently, the low-complexity solution to (17) is $\boldsymbol{\Theta} = \frac{\sqrt{N}}{\|\hat{\boldsymbol{\Theta}}\|_F} \hat{\boldsymbol{\Theta}}$, where $\hat{\boldsymbol{\Theta}}$ is given by

$$\hat{\boldsymbol{\Theta}} = \mathrm{mat}\left(\left(\left(\sum_{i \in \mathcal{K}} \mathbf{G}(\tilde{\mathbf{r}}) \mathbf{w}_i \mathbf{w}_i^H \mathbf{G}(\tilde{\mathbf{r}})^H \right)^T \otimes \left(\sum_{k \in \mathcal{K}} |\tau_k|^2 \mathbf{h}_k \mathbf{h}_k^H \right) \right)^{-1} \right.$$
$$\left. \mathrm{vec}\left(\sum_{k \in \mathcal{K}} \tilde{\tau}_k \mathbf{h}_k \mathbf{w}_k^H \mathbf{G}(\tilde{\mathbf{r}})^H \right) \right). \tag{19}$$

While the relaxed solutions provide valuable insights, they generally fail to satisfy the constraints of any feasible set $\mathcal{X}_i$ for $i \in \{1, 2, 3\}$. Therefore, we first develop a projection method to map the relaxed solution onto the feasible set $\mathcal{X}_2$, ensuring compliance with the symmetric unitary requirements. Subsequently, we generalize this approach to handle cases for both $\mathcal{X}_1$ and $\mathcal{X}_3$.

Symmetric Unitary Projection [6]: Having established a feasible solution to problem (17), we now address problem (16) by projecting $\boldsymbol{\Theta}$ onto $\mathcal{X}_2$. To facilitate this projection, we define the symmetric unitary projection for a square matrix $\mathbf{A}_1$ as follows:

$$\mathrm{Proj}_{\mathrm{sym}}(\mathbf{A}_1) \triangleq \frac{1}{2}(\mathbf{A}_1 + \mathbf{A}_1^T) = \arg\min_{\mathbf{B} = \mathbf{B}^T} \|\mathbf{A}_1 - \mathbf{B}\|_F^2. \tag{20}$$

Furthermore, for a square matrix $\mathbf{A}_2$, we define the unitary projection as

$$\mathrm{Proj}_{\mathrm{uni}}(\mathbf{A}_2) \triangleq \mathbf{U}\mathbf{V}^H = \arg\min_{\mathbf{B}\mathbf{B}^H = \mathbf{I}} \|\mathbf{A}_2 - \mathbf{B}\|_F^2, \tag{21}$$

where $\mathbf{U}, \mathbf{V}$ are unitary matrices obtained via the singular value decomposition (SVD) of $\mathbf{A}_2$, i.e., $\mathbf{A}_2 = \mathbf{U}\mathbf{S}\mathbf{V}^H$, and $\mathbf{S}$ is a diagonal matrix.

Suppose that $\mathrm{rank}(\mathrm{Proj}_{\mathrm{sym}}(\mathbf{A})) = R$, with its singular value decomposition given by $\mathrm{Proj}_{\mathrm{sym}}(\mathbf{A}) = \mathbf{U}\mathbf{S}\mathbf{V}^H$. We partition the matrices $\mathbf{U}$ and $\mathbf{V}$ as $\mathbf{U} = [\mathbf{U}_R, \mathbf{U}_{N-R}]$ and $\mathbf{V} = [\mathbf{V}_R, \mathbf{V}_{N-R}]$, respectively. The symmetric unitary projection is thus then defined as

$$\mathrm{Proj}_{\mathrm{symuni}}(\mathbf{A}) = \mathrm{Proj}_{\mathrm{uni}}(\mathrm{Proj}_{\mathrm{sym}}(\mathbf{A})) \triangleq \hat{\mathbf{U}}\mathbf{V}^H, \tag{22}$$

where $\hat{\mathbf{U}} \triangleq [\mathbf{U}_R, \mathbf{V}_{N-R}^*]$. Note that if $\mathrm{Proj}_{\mathrm{sym}}(\mathbf{A})$ is rank-deficient, $\hat{\mathbf{U}}\mathbf{V}^H$ represents one of the unitary projections for $\mathrm{Proj}_{\mathrm{sym}}(\mathbf{A})$. The proposed symmetric unitary projection indeed yields the closest point projection, as formalized in the following Proposition 2.

Proposition 2. *For a square matrix $\mathbf{A} \in \mathbb{C}^{N \times N}$, we have*

$$\mathrm{Proj}_{\mathrm{symuni}}(\mathbf{A}) = \arg\min_{\mathbf{B} \in \mathcal{X}_2} \|\mathbf{A} - \mathbf{B}\|_F^2. \tag{23}$$

Proof: See Appendix B. ∎

Using Proposition 2, we derive a closed-form for the passive beamforming solution in problem (16) when $\boldsymbol{\Theta} \in \mathcal{X}_2$, expressed as

$$\boldsymbol{\Theta} = \mathrm{Proj}_{\mathrm{symuni}}(\hat{\boldsymbol{\Theta}}). \tag{24}$$

Next, we extend the projection operation to $\boldsymbol{\Theta} \in \mathcal{X}_1$ and $\boldsymbol{\Theta} \in \mathcal{X}_3$. For the group-connected case, we first block-diagonalize it and then apply symmetric unitary projection in each block. The block-diagonalization operation for a given square matrix $\mathbf{A}$ is defined as

$$\mathrm{blkdiag}(\mathbf{A}) = \mathrm{diag}\{\mathbf{1}, \ldots, \mathbf{1}\} \odot \mathbf{A}, \tag{25}$$

where $\mathbf{1}$ is an N_g-dimensional square matrix filled with ones. Let $\mathrm{diag}\{\mathbf{X}_1, \ldots, \mathbf{X}_G\} = \mathrm{blkdiag}(\hat{\boldsymbol{\Theta}})$. The solution for the passive beamforming matrix satisfying the constraint set $\mathcal{X}_3$ is

$$\boldsymbol{\Theta} = \mathrm{diag}\{\mathrm{Proj}_{\mathrm{symuni}}(\mathbf{X}_1), \ldots, \mathrm{Proj}_{\mathrm{symuni}}(\mathbf{X}_G)\}. \tag{26}$$

For a single-connected RIS satisfying $\boldsymbol{\Theta} \in \mathcal{X}_3$, which is a special case of the group-connected RIS with the number

of user groups equals to the number of RIS elements (i.e., $G = N$), equation (26) simplifies as

$$\boldsymbol{\Theta} = \mathrm{diag}\{e^{j\angle x_1}, \ldots, e^{j\angle x_N}\}, \tag{27}$$

where $\mathbf{X}_n$, simplifies to a scalar $x_n, \forall n \in \{1, \ldots, N\}$. j is the imaginary unit, and $\angle x_n$ represents the corresponding phase angle.

E. Antenna Position Optimization: Block $\tilde{\mathbf{r}}$

This subsection focuses on optimizing the antenna position $\tilde{\mathbf{r}}$ with fixed $(\mathbf{W}, \boldsymbol{\Theta}, \boldsymbol{\iota}, \boldsymbol{\tau})$. This subproblem is given by

$$\min_{\tilde{\mathbf{r}}} \sum_{k \in \mathcal{K}} \left(-2\Re\{\tilde{\tau}_k^* \tilde{\mathbf{h}}(\tilde{\mathbf{r}})^H \mathbf{w}_k\} + |\tau_k|^2 \sum_{i \in \mathcal{K}} |\tilde{\mathbf{h}}(\tilde{\mathbf{r}})^H \mathbf{w}_i|^2 \right)$$
$$\text{s.t. } C_2, C_3. \tag{28}$$

To address the non-convex constraint C_3 in (28), Using variable splitting, we define auxiliary variables $\{\mathbf{b}_m\}_{m=1}^M = \{\mathbf{r}_m\}_{m=1}^M$. By substituting $\{\mathbf{r}_m\}_{m=1}^M$ in (28) with $\{\mathbf{b}_m\}_{m=1}^M$ and augmenting a penalty term $\varrho \sum_{m=1}^M \|\mathbf{r}_m - \mathbf{b}_m\|_2^2$, problem (28) is reformulated as two subproblems.

1) Subproblem with respect to $\tilde{\mathbf{r}}$:

$$\min_{\tilde{\mathbf{r}}} \sum_{k \in \mathcal{K}} \left(-2\Re\{\tilde{\tau}_k^* \tilde{\mathbf{h}}(\tilde{\mathbf{r}})^H \mathbf{w}_k\} + |\tau_k|^2 \sum_{i \in \mathcal{K}} |\tilde{\mathbf{h}}(\tilde{\mathbf{r}})^H \mathbf{w}_i|^2 \right)$$
$$+ \varrho \sum_{m=1}^M \|\mathbf{r}_m - \mathbf{b}_m\|_2^2$$
$$\text{s.t. } C_2, \tag{29}$$

where $\varrho > 0$ is the penaty factor and it is known that as $\varrho \to \infty$, the optimal solutions of (29) and (28) are identical. Since the objective function (29) is differentiable with respect to $\{\mathbf{r}_m\}_{m=1}^M$, we apply the projected gradient approach [38], [39] to optimize $\{\mathbf{r}_m\}_{m=1}^M$, i.e.,

$$\mathbf{r}_m = \mathrm{Proj}_{\mathcal{C}}\{\mathbf{r}_m - \eta \nabla_{\mathbf{r}_m} f_a\}, \tag{30}$$

where η is the step-size, and the projection is given by $\mathrm{Proj}_{\mathcal{C}}(\mathbf{x}) \triangleq \min(\max(\mathbf{x}, \mathbf{x}_{\min}), \mathbf{x}_{\max})$ with $\mathbf{x}_{\min}$ and $\mathbf{x}_{\max}$ denoting the lower left and upper right coordinates of $\mathcal{C}$.

2) Subproblem with respect to $\{\mathbf{b}_m\}_{m=1}^M$: Since the auxiliary variables $\{\mathbf{b}_m\}_{m=1}^M$ appear only in the penalty term, the corresponding subproblem can be expressed as

$$\min_{\{\mathbf{b}_m\}_{m=1}^M} \sum_{m=1}^M \|\mathbf{b}_m - \mathbf{r}_m\|_2^2$$
$$\text{s.t. } \|\mathbf{b}_m - \mathbf{b}_l\|_2 \geq D, \forall m, l = 1, 2, \ldots, M, \quad m \neq l. \tag{31}$$

When $\{\mathbf{b}_l\}_{l \neq m}$ is fixed, the m-th subproblem of (31) is given by

$$\min_{\mathbf{b}_m} \|\mathbf{b}_m - \mathbf{r}_m\|_2^2$$
$$\text{s.t. } \|\mathbf{b}_m - \mathbf{b}_l\|_2 \geq D, \quad \forall l = 1, 2, \ldots, M, \quad l \neq m. \tag{32}$$

To systematically address problem (32), we first introduce the following notations for clarity. Let $\mathcal{A}$ represent the set of all l that satisfy the constraint $\|\mathbf{r}_m - \mathbf{b}_l\|_2 < D$. ϖ_l denotes the circle centered at $\mathbf{b}_l$ with radius D. The set $\mathcal{P}_l$ includes the

---

Algorithm 1 Proposed Penalty-BCD Algorithm for Handling Problem (9)

1: Initialize the channel parameters and optimization variables.
2: **repeat**
3:     Update the transmit precoder $\mathbf{W}$ based on (15).
4:     Update the $\boldsymbol{\Theta}$ based on (24), (26), or (27) utilizing the projected gradient-based methods.
5:     Update $\{\mathbf{r}_m\}_{m \in \mathcal{M}}$ according to (30) utilizing the projected gradient-based methods.
6:     **repeat**
7:         **for** $m = 1, \ldots, M$ **do**
8:             Compute $\mathbf{b}_m$ via (32).
9:         **end for**
10:     **until** the updated value satisfying the convergence threshold.
11:     Increase the penalty factor ϱ.
12: **until** the updated value satisfying the convergence threshold.

---

intersection points of ϖ_l with other circles (excluding ϖ_m), where these points must also satisfy constraint (32). Similarly, the set $\mathcal{Q}_l$ comprises the intersection points of ϖ_l with the line passing through $\mathbf{b}_l$ and $\mathbf{r}_m$, again satisfying constraint (32). We then analyze the optimal solution for $\mathbf{b}_m^*$ in problem (32) by considering three distinct cases:

(a) $\mathcal{A} = \emptyset$: In this case, the optimal $\mathbf{b}_m^*$ is $\mathbf{b}_m^* = \mathbf{r}_m$, as illustrated in Fig. 2(a).

(b) $|\mathcal{A}| = 1$: We first identify the unique $\mathbf{b}_l$ that satisfies $\|\mathbf{r}_m - \mathbf{b}_l\|_2 < D$. The optimal $\mathbf{b}_m^*$ then belongs to the set $\mathcal{P}_l \cup \mathcal{Q}_l$. As shown in Figs. 2(b) and 2(c), the optimal $\mathbf{b}_m^*$ is given by

$$\mathbf{b}_m^* = \arg \min_{\mathbf{b}_m \in \mathcal{P}_l \cup \mathcal{Q}_l} \|\mathbf{b}_m - \mathbf{r}_m\|_2, \tag{33}$$

where $\mathcal{P}_l$ and $\mathcal{Q}_l$ can be computed using geometric methods.

(c) $|\mathcal{A}| \geq 2$: As depicted in Fig. 2(d), the optimal $\mathbf{b}_m^*$ is determined by

$$\mathbf{b}_m^* = \arg \min_{\mathbf{b}_m} \|\mathbf{b}_m - \mathbf{r}_m\|_2, \tag{34}$$

subject to the relevant constraints.

Algorithm 1 provides a comprehensive outline of the entire procedure for handling problem (9).

Remark 2. *We can obtain stationary points for the subproblems concerning $\{\mathbf{r}_m\}_{m \in \mathcal{M}}$ and $\{\mathbf{b}_m\}_{m \in \mathcal{M}}$. Moreover, once a stationary point is reached for subproblem (16), Algorithm 1 is guaranteed to converge. The proposed optimization algorithm offers two key advantages. Firstly, it transforms the non-convex antenna distance constraints into a subproblem involving auxiliary variables $\{\mathbf{b}_m\}_{m \in \mathcal{M}}$, which can be optimized without resorting to any approximations. Secondly, the algorithm integrates conventional algorithms exploited in FPA systems, thereby streamlining the optimization problem associated with MA.*
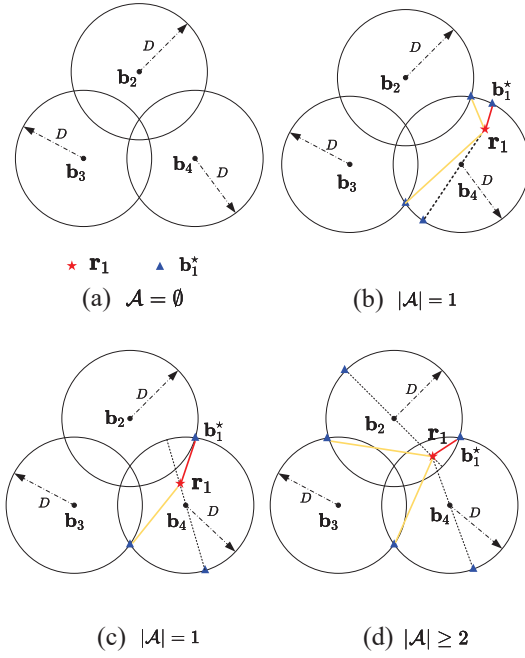
This article has been accepted for publication in IEEE Transactions on Wireless Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TWC.2025.3613556

7

Fig. 2: For $m = 1$ and $M = 4$, and the blue triangular points represent the coordinates of $\mathbf{b}_1^\star$. The orange and red lines indicate the distance between $\mathbf{r}_1$ and $\mathbf{b}_1^\star$ and the minimal distance between $\mathbf{r}_1$ and $\mathbf{b}_1^\star$, respectively. Illustrations: (a) Case a; (b) and (c) Case b; (d) Case c.

IV. Distributed Beamforming Algorithm

In this section, the focus now shifts toward distributed beamforming algorithms. Unlike the centralized approach in Section III, distributed implementations eliminate the need for aggregating all CSI at a single point, thereby reducing backhaul information exchange and enhancing robustness in time-varying scenarios. We first present the distributed ADMM [40] for solving problem (9). Furthermore, to achieve more efficient collaboration between master and workers, we further extend the synchronous ADMM algorithm to an asynchronous ADMM algorithm, inspired by [32], [37], [41].

A. S-ADMM Algorithm

We model the distributed system such that the user devices act as worker nodes, with the BS serving as the master node. The optimization process begins at the BS, which determines optimal antenna positions, followed by individual users computing their respective beamforming matrices and RIS phase configurations. We reformulate problem (9) in consensus form as

$$
\min_{\left\{\mathscr{B},\{\mathscr{A}_k\}_{k\in\mathcal{K}}\right\}} \sum_{k\in\mathcal{K}} f_k(\mathscr{A}_k)
$$
$$
\text{s.t. } \mathrm{C}_5 : \mathbf{W}_k = \mathbf{W}, \forall k \in \mathcal{K}, \quad (35)
$$
$$
\mathrm{C}_6 : \boldsymbol{\Theta}_k = \boldsymbol{\Theta}, \forall k \in \mathcal{K},
$$

where $\mathscr{A}_k = \{\mathbf{W}_k, \boldsymbol{\Theta}_k, \iota_k, \tau_k\}$ and $\mathscr{B} = \{\mathbf{W}, \boldsymbol{\Theta}, \tilde{\mathbf{r}} \mid (\mathrm{C}_1, \mathrm{C}_2, \mathrm{C}_3, \mathrm{C}_4)\}$. The objective function $f_k(\mathscr{A}_k)$ can be defined as

$$
f_k(\mathscr{A}_k) = -\log(1 + \iota_k) + \iota_k - 2\sqrt{1 + \iota_k}\Re\{\tilde{\mathbf{h}}_k(\tilde{\mathbf{r}})^H \mathbf{w}_k \tau_k^*\}
$$
$$
+ |\tau_k|^2 \left(\sum_{i\in\mathcal{K}} |\tilde{\mathbf{h}}_k(\tilde{\mathbf{r}})^H \mathbf{w}_i|^2 + \sigma_k^2\right). \quad (36)
$$

Based on the ADMM framework, the augmented Lagrangian function for problem (35) incorporates the original objective function along with the dual and penalty terms as follows [40]

$$
\mathcal{L}\left(\{\mathscr{A}_k, \mathbf{Z}_k, \mathbf{U}_k\}_{k\in\mathcal{K}}, \mathscr{B}\right)
$$
$$
= \sum_{k\in\mathcal{K}} f_k(\mathscr{A}_k) - \sum_{k\in\mathcal{K}} 2\Re\left\{\mathrm{Tr}\left(\mathbf{Z}_k^H(\mathbf{W}_k - \mathbf{W})\right)\right\}
$$
$$
+ \sum_{k\in\mathcal{K}} \rho \|\mathbf{W}_k - \mathbf{W}\|_F^2 - \sum_{k\in\mathcal{K}} 2\Re\left\{\mathrm{Tr}\left(\mathbf{U}_k^H(\boldsymbol{\Theta}_k - \boldsymbol{\Theta})\right)\right\}
$$
$$
+ \sum_{k\in\mathcal{K}} \rho \|\boldsymbol{\Theta}_k - \boldsymbol{\Theta}\|_F^2, \quad (37)
$$

where $\{\mathbf{Z}_k, \mathbf{U}_k\}$ represent dual variables for the consensus constraints, with $\rho > 0$ as the penalty parameter.

The algorithm updates local variables first in each iteration, keeping global variables fixed. First, problem (35) can be solved distributedly at each user k using the ADMM

$$
\min_{\mathscr{A}_k} f_k(\mathscr{A}_k) - 2\Re\left\{\mathrm{Tr}\left(\mathbf{Z}_k^H(\mathbf{W}_k - \mathbf{W})\right)\right\} + \rho \|\mathbf{W}_k - \mathbf{W}\|_F^2
$$
$$
- 2\Re\left\{\mathrm{Tr}\left(\mathbf{U}_k^H(\boldsymbol{\Theta}_k - \boldsymbol{\Theta})\right)\right\} + \rho \|\boldsymbol{\Theta}_k - \boldsymbol{\Theta}\|_F^2. \quad (38)
$$

We solve the k-th subproblem (38) using an efficient alternating optimization (AO) procedure. Although jointly non-convex in $\mathscr{A}_k$ due to strong variable coupling in the objective function, the problem is convex with respect to each individual variable when others are fixed. The $\mathbf{W}_k$ is updated according to

$$
\mathbf{W}_k = \arg\min_{\mathbf{W}_k} \mathrm{Tr}\left(\mathbf{W}_k^H \mathbf{P}_k \mathbf{W}_k\right) - 2\Re\left\{\mathrm{Tr}\left(\mathbf{W}_k^H \mathbf{Q}_k\right)\right\}, \quad (39)
$$

where $\mathbf{P}_k = \rho\mathbf{I}_M + |\tau_k|^2 \tilde{\mathbf{h}}_k(\tilde{\mathbf{r}})\tilde{\mathbf{h}}_k(\tilde{\mathbf{r}})^H$, $\mathbf{Q}_k = \mathbf{F}_k + \rho\mathbf{W} + \mathbf{Z}_k$, $\mathbf{F}_k = [\mathbf{0}_{M\times k-1}, \sqrt{1 + \iota_k}\tau_k\tilde{\mathbf{h}}_k(\tilde{\mathbf{r}}), \mathbf{0}_{M\times K-k}]$. Then, the optimal solution to the subproblem (39) is derived in the closed forms as $\mathbf{W}_k^\star = (\mathbf{P}_k)^{-1}\mathbf{Q}_k$.

For the case $\boldsymbol{\Theta}_k \in \mathcal{X}_1$, we assume that $\boldsymbol{\psi}_k = \mathrm{diag}(\boldsymbol{\Theta}_k)$, $\forall k \in \mathcal{K}$, the subproblem for $\boldsymbol{\psi}_k$ can be expressed as [42]

$$
\boldsymbol{\psi}_k = \arg\min_{\boldsymbol{\psi}_k} \boldsymbol{\psi}_k^H \mathbf{D}_k \boldsymbol{\psi}_k - 2\Re\left\{\mathbf{d}_k^H \boldsymbol{\psi}_k\right\}, \quad (40)
$$

where

$$
\mathbf{D}_k = |\tau_k|^2 \mathrm{diag}(\mathbf{h}_k)\mathbf{G}(\tilde{\mathbf{r}})^* \mathbf{W}_k^* \mathbf{W}_k^T \mathbf{G}(\tilde{\mathbf{r}})^T \mathrm{diag}(\mathbf{h}_k^H) + \rho\mathbf{I}_M \quad (41)
$$

and

$$
\mathbf{d}_k = \rho\boldsymbol{\psi} + \mathrm{diag}(\mathbf{U}_k) + \sqrt{1 + \iota_k}\tau_k^* \mathrm{diag}(\mathbf{h}_k)\mathbf{G}(\tilde{\mathbf{r}})^* \mathbf{w}_{k,k}^*. \quad (42)
$$

The suboptimal solution of (40) is $\boldsymbol{\psi}_k^\star = (\mathbf{D}_k)^{-1}\mathbf{d}_k$. Next, we extend the solution (19) to address problem (40) when $\boldsymbol{\Theta}_k \in \mathcal{X}_2$. The subproblem for $\boldsymbol{\Theta}_k$ is

$$
\boldsymbol{\Theta}_k = \arg\min_{\boldsymbol{\Theta}_k} -2\sqrt{1 + \iota_k}\Re\left\{\tau_k^*\mathbf{h}_k^H \boldsymbol{\Theta}_k \mathbf{G}(\tilde{\mathbf{r}})\mathbf{w}_{k,k}\right\}
$$
$$
+ |\tau_k|^2 \left(\sum_{i\in\mathcal{K}} |\mathbf{h}_k^H \boldsymbol{\Theta}_k \mathbf{G}(\tilde{\mathbf{r}})\mathbf{w}_{i,k}|^2 + \sigma_k^2\right) \quad (43)
$$
$$
- 2\Re\left\{\mathbf{U}_k^H(\boldsymbol{\Theta} - \boldsymbol{\Theta}_k)\right\} + \rho \|\boldsymbol{\Theta} - \boldsymbol{\Theta}_k\|_2^2.
$$

By setting the derivative of (43) to $\mathbf{0}$, we obtain the stationary point

$$\mathbf{\Theta}_k = \mathrm{mat}\left(\left(|\tau_k|^2\left(\mathbf{G}\left(\tilde{\mathbf{r}}\right)\mathbf{w}_{i,k}\mathbf{w}_{i,k}^H\mathbf{G}\left(\tilde{\mathbf{r}}\right)^H\right)^T\otimes\sum_{i\in\mathcal{K}}\mathbf{h}_k\mathbf{h}_k^H+\rho\mathbf{I}\right)^{-1}\right.$$
$$\left.\mathrm{vec}\left(\sqrt{1+\iota_k}\cdot\tau_k^*\mathbf{h}_k\mathbf{w}_{k,k}^H\mathbf{G}\left(\tilde{\mathbf{r}}\right)^H-\mathbf{U}_k+\rho\mathbf{\Theta}\right)\right). \tag{44}$$

Then, during every ADMM time step, the dual variables $\mathbf{Z}_k$, $\mathbf{U}_k$ are updated as

$$\mathbf{Z}_k = \mathbf{Z}_k - \rho(\mathbf{W}_k - \mathbf{W}) \tag{45}$$

and

$$\mathbf{U}_k = \mathbf{U}_k - \rho(\mathbf{\Theta}_k - \mathbf{\Theta}). \tag{46}$$

When other variables are fixed, the subproblem related to $\boldsymbol{\iota}$ and $\boldsymbol{\tau}$ is characterized as an unconstrained convex optimization problem. Then the optimal solutions is obtained by setting $\partial f_k\left(\mathscr{A}_k\right)/\partial\iota_k = 0$ and $\partial f_k\left(\mathscr{A}_k\right)/\partial\tau_k = 0$, i.e.,

$$\iota_k = \frac{4}{\left(\sqrt{4+\left|\Re\left\{\mathbf{w}_{k,k}^H\tilde{\mathbf{h}}_k(\tilde{\mathbf{r}})\tau_k\right\}\right|^2}-\Re\left\{\mathbf{w}_{k,k}^H\tilde{\mathbf{h}}_k(\tilde{\mathbf{r}})\tau_k\right\}\right)^2}-1, \tag{47}$$

$$\tau_k = \frac{\sqrt{1+\iota_k}\,\tilde{\mathbf{h}}_k(\tilde{\mathbf{r}})^H\mathbf{w}_{k,k}}{\sum_{i\in\mathcal{K}}\left|\tilde{\mathbf{h}}_k(\tilde{\mathbf{r}})^H\mathbf{w}_{i,k}\right|^2+\sigma_k^2}. \tag{48}$$

Using the first-order derivative and the Cauchy-Schwarz inequality, we can derive the globally optimal solutions for $\mathbf{W}$ and $\mathbf{\Theta}$ in closed form, i.e.,

$$\mathbf{W} = \frac{\sum_{k\in\mathcal{K}}(\rho\mathbf{W}_k - \mathbf{Z}_k)}{\rho K}I_{\mathbf{W}}, \tag{49}$$

where $I_{\mathbf{W}}$ represents the normalization factor, i.e., $I_{\mathbf{W}} = \frac{\sqrt{P_{\max}}}{\left\|\sum_{k\in\mathcal{K}}(\rho\mathbf{W}_k-\mathbf{Z}_k)/\rho K\right\|_F}$ if $\left\|\sum_{k\in\mathcal{K}}(\rho\mathbf{W}_k-\mathbf{Z}_k)/\rho K\right\|_F > P_{\max}$, and $I_{\mathbf{W}} = 1$ otherwise. From (49), the weight vector of each user is aggregated at the BS to form its precoding vector. Similarly, the phase matrix $\mathbf{\Theta}$ can be updated as

$$\widehat{\mathbf{\Theta}} = \frac{\sum_{k\in\mathcal{K}}\left(\rho\mathbf{\Theta}_k - \mathbf{U}_k\right)}{\rho K}. \tag{50}$$

Then, we perform a projection operation to satisfy the BD-RIS constraints as according to Section III-D, i.e., $\mathbf{\Theta} = \mathrm{Proj}_{\mathrm{symuni}}(\widehat{\mathbf{\Theta}})$.

B. Proposed AS-ADMM Algorithm

As illustrated in Fig. 3(a), the synchronous execution process described previously reveals that the master node and faster worker nodes frequently experience significant idle periods. As a result, the available parallel computational resources are often underutilized. Based on the concept of asynchronous signal processing, the BS operates independently without waiting for responses from all user nodes. Instead, as shown in Fig. 3(b), the BS updates immediately upon receiving input from a subset of users. This approach effectively eliminates idle time, improves robustness against failures, and significantly enhances computational efficiency. Consequently, the
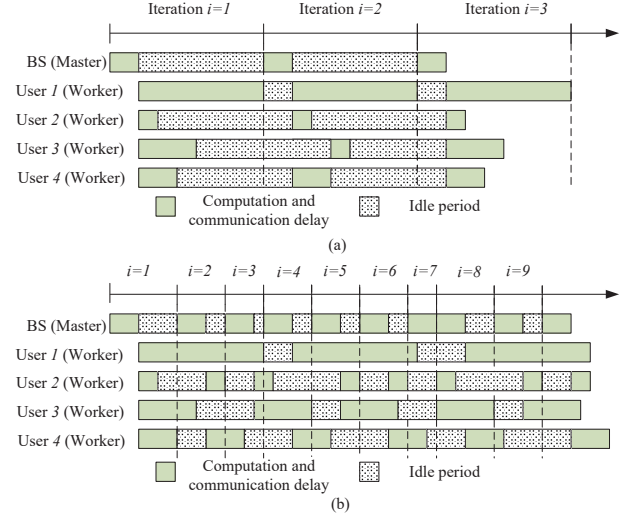

Fig. 3: Example of how (a) synchronous and (b) asynchronous ADMM algorithms affect the operations between BS and user nodes in the system.

asynchronous design removes the requirement for synchronization among worker nodes and eliminate the depending of the master node on the slowest worker, ensuring continuous and efficient resource utilization.

Inspired by this, we introduce an AS-ADMM algorithm to facilitate more efficient collaboration between the BS and users. As illustrated in Fig. 3, the BS updates the consensus variables in $\mathscr{B}$, while user k updates its local variables $\{\mathscr{A}_k,\mathbf{Z}_k,\mathbf{U}_k\}$. To support an asynchronous distributed implementation, the BS and user k maintain independent clocks, t_c and t_{d_k}, respectively. Without loss of generality, these clocks start at zero and increment by 1 after each update of $\mathscr{B}$ and $\{\mathscr{A}_k,\mathbf{Z}_k,\mathbf{U}_k\}$.

At clock t_c, the BS updates $\mathscr{B}$ upon receiving the latest $\{\mathscr{A}_k,\mathbf{Z}_k,\mathbf{U}_k\}$ from at least $D_{\min}$ users, denoted by the set $\mathcal{D}_{t_c}\subset\mathcal{K}$ with $\mathrm{size}(\mathcal{D}_{t_c})\geq D_{\min}$. The BS then broadcasts the updated $\{\mathbf{W},\mathbf{\Theta},\tilde{\mathbf{r}}\}$ to the users in $\mathcal{D}_{t_c}$ for local updates. To ensure the freshness of these updates across all users, a bounded asynchronous delay $\tau_{\max}$ is assumed. This implies that the latest $\{\mathscr{A}_k,\mathbf{Z}_k,\mathbf{U}_k\}$ at each user k is at most $\tau_{\max}$ clock cycles behind the BS's clock, i.e.,

$$k\in\mathcal{D}_{t_c}\cup\mathcal{D}_{t_c-1}\cup\cdots\cup\mathcal{D}_{t_c-\tau_{\max}+1}. \tag{51}$$

At the BS's clock t_c+1, the updates are computed via

$$\left\{\mathbf{W}_k^{(t_c+1)},\mathbf{\Theta}_k^{(t_c+1)},\mathbf{Z}_k^{(t_c+1)},\mathbf{U}_k^{(t_c+1)}\right\}$$
$$=\begin{cases}\left\{\widehat{\mathbf{W}}_k,\widehat{\mathbf{\Theta}}_k,\widehat{\mathbf{Z}}_k,\widehat{\mathbf{U}}_k\right\} & k\in\mathcal{D}_{t_c+1}\\\left\{\mathbf{W}_k^{(t_c)},\mathbf{\Theta}_k^{(t_c)},\mathbf{Z}_k^{(t_c)},\mathbf{U}_k^{(t_c)}\right\} & k\notin\mathcal{D}_{t_c+1}\end{cases},\forall k\in\mathcal{K} \tag{52}$$

$$\mathbf{W}^{(t_c+1)} = \frac{\sum_{k\in\mathcal{K}}(\rho\mathbf{W}_k^{(t_c+1)}-\mathbf{Z}_k^{(t_c+1)})}{\rho K}I_{\mathbf{W}}, \tag{53}$$

$$\mathbf{\Theta}^{(t_c+1)} = \frac{\sum_{k\in\mathcal{K}}(\rho\mathbf{\Theta}_k^{(t_c+1)}-\mathbf{U}_k^{(t_c+1)})}{\rho K}, \tag{54}$$

Algorithm 2 Asynchronous Distributed ADMM for Tackling Problem (35)

1: **Algorithm of the Master:**
2: **Input:** Initial $\left\{\mathbf{W}^{(0)}, \mathbf{\Theta}^{(0)}, \rho^{(0)}\right\}$, set clock $t_c = 0$, and initialize delays for all K users as $\tau_1 = \cdots = \tau_K = 0$.
3: **repeat**
4:     Wait until receiving the up-to-date $\{\widehat{\mathbf{W}}_k, \widehat{\mathbf{\Theta}}_k, \widehat{\mathbf{Z}}_k, \widehat{\mathbf{U}}_k\}$ from each user $k \in \mathcal{D}_{t_c+1}$, where both $\mathrm{size}(\mathcal{D}_{t_c}) \geq \mathcal{D}$ and $\tau_k \leq \tau_{\max}, \forall k \in \mathcal{K}$ are required.
5:     Compute $\left\{\mathbf{W}_k^{(t_c+1)}, \mathbf{\Theta}_k^{(t_c+1)}, \mathbf{Z}_k^{(t_c+1)}, \mathbf{U}_k^{(t_c+1)}\right\}_{k \in \mathcal{K}}$ based on (52).
6:     Compute $\left\{\mathbf{W}^{(t_c+1)}, \mathbf{\Theta}^{(t_c+1)}\right\}$ based on (53) and (54).
7:     Set $\tau_k = \tau_k + 1$ for any user $k \notin \mathcal{D}_{t_c+1}$.
8:     Broadcast $\left\{\mathbf{W}^{(t_c+1)}, \mathbf{\Theta}^{(t_c+1)}, \rho^{(t_c+1)}\right\}$ to each user $k \in \mathcal{D}_{t_c+1}$.
9:     set $t_c \leftarrow t_c + 1$.
10: **until** a predefined stopping criterion is satisfied.

1: **Algorithm of the k-th Worker:**
2: **Input:** Initial variables $\{\mathbf{W}_k^{(0)}, \mathbf{\Theta}_k^{(0)}, \tau_k^{(0)}, \iota_k^{(0)}, \mathbf{Z}_k^{(0)}, \mathbf{U}_k^{(0)}\}$, clock $t_{d_k} = 0$.
3: **repeat**
4:     Wait until receiving the up-to-date $\{\widehat{\mathbf{W}}, \widehat{\mathbf{\Theta}}\}$ from the BS.
5:     Compute $\{\mathbf{W}_k^{(t_{d_k}+1)}, \mathbf{\Theta}_k^{(t_{d_k}+1)}\}$ according to (55).
6:     Compute $\{\mathbf{Z}_k^{(t_{d_k}+1)}, \mathbf{U}_k^{(t_{d_k}+1)}\}$ according to (56) and (57).
7:     Send $\{\mathbf{W}_k^{(t_{d_k}+1)}, \mathbf{\Theta}_k^{(t_{d_k}+1)}, \mathbf{Z}_k^{(t_{d_k}+1)}, \mathbf{U}_k^{(t_{d_k}+1)}\}$ to the BS.
8:     Set $t_{d_k} \leftarrow t_{d_k} + 1$.
9: **until** a predefined stopping criterion is satisfied.

---

where $\{\widehat{\mathbf{W}}_k, \widehat{\mathbf{\Theta}}_k, \widehat{\mathbf{Z}}_k, \widehat{\mathbf{U}}_k\}$ in (52) represents the latest information received from D_{t_c+1} user. At the local clock time $t_{d_k}+1$, the user k executes the AS-ADMM updates as follows:

$$\left\{\mathbf{W}_k^{(t_{d_k}+1)}, \mathbf{\Theta}_k^{(t_{d_k}+1)}, \iota_k^{(t_{d_k}+1)}, \boldsymbol{\tau}_k^{(t_{d_k}+1)}\right\}$$
$$= \arg\max_{\mathscr{A}_k} f_k(\mathscr{A}_k) + 2\Re\left\{\mathrm{Tr}\left(\mathbf{W}_k^H \mathbf{Z}_k^{(t_{d_k})} + \mathbf{\Theta}_k^H \mathbf{U}_k^{(t_{d_k})}\right)\right\}$$
$$- \rho\left\|\mathbf{W}_k - \widehat{\mathbf{W}}\right\|_F^2 - \rho\left\|\mathbf{\Theta}_k - \widehat{\mathbf{\Theta}}\right\|_F^2, \tag{55}$$

$$\mathbf{Z}_k^{(t_{d_k}+1)} = \mathbf{Z}_k^{(t_{d_k})} - \rho(\mathbf{W}_k^{(t_{d_k}+1)} - \widehat{\mathbf{W}}), \tag{56}$$

$$\mathbf{U}_k^{(t_{d_k}+1)} = \mathbf{U}_k^{(t_{d_k})} - \rho(\mathbf{\Theta}_k^{(t_{d_k}+1)} - \widehat{\mathbf{\Theta}}). \tag{57}$$

The closed-form solution for the variable update can be found in Section IV-A. The overall AS-ADMM distributed beamforming algorithm is presented in Algorithm 2.

Remark 3. *While the consensus problem* (35) *has a higher dimension than the original problem* (9) *due to the increase of K, the proposed distributed AS-ADMM algorithm may require more total iterations than the centralized approach. However, solving problem* (35) *offers a key advantage: it enables each user to perform computations locally, leveraging its own information. This eliminates the substantial communication*

overhead between the users and the BS, highlighting the AS-ADMM algorithm's scalability and flexibility.

V. CONVERGENCE AND COMPLEXITY ANALYSIS

The convergence conditions for the AS-ADMM algorithm, established in [32], can be adapted to our framework as follows:

1) The asynchronous updates from each user exhibit a bounded delay; specifically, the communication delays from each worker are upper-bounded.
2) For any $\{\iota_k, \tau_k\}$, the objective functions are twice differentiable with respect to $\mathbf{W}_k$ and $\mathbf{\Theta}_k$, and their gradients are L-Lipschitz continuous.
3) The optimal value of the sum rate maximization problem is finite.

Under these conditions, the sequence $\{\{\mathscr{A}_k\}_{k \in \mathcal{K}}, \mathscr{B}\}$ produced by the AS-ADMM framework converges to the set of stationary points of problem (35).

To demonstrate the convergence properties of the AS-ADMM algorithm, we consider the single-connected BD-RIS. In terms of assumption 1), a maximum asynchronous delay $\tau_{\max}$ is assumed for each user. To demonstrate assumption 2), we firstly know that the $f_k(\mathscr{A}_k)$ is a quadratic function and the gradients of $f_k(\mathscr{A}_k)$ are respectively given by

$$\nabla_{\boldsymbol{\psi}_k} f(\mathscr{A}_k) = \sqrt{1+\iota_k} \tau_k^H \widehat{\mathbf{G}}_k \mathbf{W}_k \mathbf{1}_k - |\tau_k|^2 \widehat{\mathbf{G}}_k \mathbf{W}_k^* \mathbf{W}_k^T \mathbf{h}_{\boldsymbol{\psi}_k}^*, \tag{58}$$

$$\nabla_{\mathbf{W}_k} f(\mathscr{A}_k) = \mathbf{F}_{\boldsymbol{\psi}_k} - |\tau_k|^2 \mathbf{h}_{\boldsymbol{\psi}_k} \mathbf{h}_{\boldsymbol{\psi}_k}^H \mathbf{W}_k, \tag{59}$$

where

$$\begin{aligned} &\widehat{\mathbf{G}}_k = \mathrm{diag}(\mathbf{h}_k) \mathbf{G}^*, \ \mathbf{1}_k = [\mathbf{0}_{1 \times k-1}, 1, \mathbf{0}_{1 \times K-k}]^T, \\ &\mathbf{h}_{\boldsymbol{\psi}_k} = \widehat{\mathbf{G}}_k^T \boldsymbol{\psi}_k^*, \\ &\mathbf{F}_{\boldsymbol{\psi}_k} = \left[\mathbf{0}_{N \times k-1}, \sqrt{1+\iota_k} \tau_k \mathbf{h}_{\boldsymbol{\psi}_k}, \mathbf{0}_{N \times K-k}\right]. \end{aligned} \tag{60}$$

If $f_k(\mathscr{A}_k)$ satisfies Lipschitz continuous gradient, there must exist a constant L such that

$$\left(\|\nabla_{\boldsymbol{\psi}_k} f(\mathscr{A}_k^1) - \nabla_{\boldsymbol{\psi}_k} f(\mathscr{A}_k^2)\|_2^2 + \|\nabla_{\mathbf{W}_k} f(\mathscr{A}_k^1) - \nabla_{\mathbf{W}_k} f(\mathscr{A}_k^2)\|_F^2\right)^{\frac{1}{2}}$$
$$\leq L\left(\|\boldsymbol{\psi}_k^1 - \boldsymbol{\psi}_k^2\|_2^2 + \|\mathbf{W}_k^1 - \mathbf{W}_k^2\|_F^2\right)^{\frac{1}{2}}, \tag{61}$$

where $\mathscr{A}_k^i = \{\mathbf{W}_k^i, \boldsymbol{\psi}_k^i, \iota_k, \tau_k\}, i \in \{1, 2\}$. We prove (61) with inequalities for the gradient $\nabla_{\boldsymbol{\psi}_k} f(\mathscr{A}_k)$ as (62), where equality 1) is obtained through auxiliary terms. Inequalities 2) and 3) follow from the matrix norm bounds $\|\mathbf{a}+\mathbf{b}\|_2 \leq \|\mathbf{a}\|_2 + \|\mathbf{b}\|_2$ and $\|\mathbf{ABC}\|_F \leq \lambda_{\max}(\mathbf{A})\|\mathbf{B}\|_F \lambda_{\max}(\mathbf{C})$, respectively. Additionally, for (63) we have $\lambda_{\max}(\mathbf{AB}) \leq \lambda_{\max}(\mathbf{A})\lambda_{\max}(\mathbf{B})$. For the gradient $\nabla_{\mathbf{W}_k} f(\boldsymbol{\psi}_k, \mathbf{W}_k)$, we derive

$$\|\nabla_{\mathbf{W}_k} f(\mathscr{A}_k^1) - \nabla_{\mathbf{W}_k} f(\mathscr{A}_k^2)\|_F$$
$$= \left\|\mathbf{F}_{\boldsymbol{\psi}_k^2} - \mathbf{F}_{\boldsymbol{\psi}_k^1} - |\tau_k|^2 \mathbf{h}_{\boldsymbol{\psi}_k^2} \mathbf{h}_{\boldsymbol{\psi}_k^2}^H \mathbf{W}_k^2 + |\tau_k|^2 \mathbf{h}_{\boldsymbol{\psi}_k^1} \mathbf{h}_{\boldsymbol{\psi}_k^1}^H \mathbf{W}_k^1\right\|_F$$
$$\leq \ell_3 \|\mathbf{W}_k^2 - \mathbf{W}_k^1\|_F + \ell_4 \|\boldsymbol{\psi}_k^2 - \boldsymbol{\psi}_k^1\|_2, \tag{64}$$

where the constants are defined as $\ell_3 = \sqrt{1+\iota_k}|\tau_k|\lambda_{\max}(\widehat{\mathbf{G}}_k) + |\tau_k|^2 \lambda_{\max}(\widehat{\mathbf{G}}_k^* \mathbf{W}_k^2)\lambda_{\max}(\mathbf{h}_{\boldsymbol{\psi}_k^1}) + |\tau_k|^2 \lambda_{\max}(\widehat{\mathbf{G}}_k)\lambda_{\max}(\mathbf{h}_{\boldsymbol{\psi}_k^2}^H \mathbf{W}_k^2)$, $\ell_4 = |\tau_k|^2(\sqrt{N}\lambda_{\max}(\widehat{\mathbf{G}}_k))^2$.

$$\left\|\nabla_{\boldsymbol{\psi}_k} f(\mathscr{A}_k^1) - \nabla_{\boldsymbol{\psi}_k} f(\mathscr{A}_k^2)\right\|_2$$

$$= \left\|\sqrt{1+\iota_k}|\tau_k|^2 \widehat{\mathbf{G}}_k(\mathbf{W}_k^1 - \mathbf{W}_k^2)\mathbf{1}_k + |\tau_k|^2 \widehat{\mathbf{G}}_k(\mathbf{W}_k^1)^*(\mathbf{W}_k^1)^T \mathbf{h}_{\boldsymbol{\psi}_k^1}^* + |\tau_k|^2 \widehat{\mathbf{G}}_k(\mathbf{W}_k^2)^*(\mathbf{W}_k^2)^T \mathbf{h}_{\boldsymbol{\psi}_k^2}^*\right\|_2$$

$$\overset{1)}{=} \left\|\sqrt{1+\iota_k}|\tau_k|^2 \widehat{\mathbf{G}}_k(\mathbf{W}_k^1 - \mathbf{W}_k^2)\mathbf{1}_k + |\tau_k|^2 \widehat{\mathbf{G}}_k(\mathbf{W}_k^1)^*(\mathbf{W}_k^2 - \mathbf{W}_k^1)^T \mathbf{h}_{\boldsymbol{\psi}_k^2}^* + |\tau_k|^2 \widehat{\mathbf{G}}_k(\mathbf{W}_k^2 - \mathbf{W}_k^1)^*(\mathbf{W}_k^2)^T \mathbf{h}_{\boldsymbol{\psi}_k^2}^*\right.$$
$$\left. + |\tau_k|^2 \widehat{\mathbf{G}}_k(\mathbf{W}_k^1)^*(\mathbf{W}_k^1)^T \widehat{\mathbf{G}}_k(\boldsymbol{\psi}_k^2 - \boldsymbol{\psi}_k^1)\right\|_2$$

$$\overset{2)}{\leq} \sqrt{1+\iota_k}|\tau_k|^2 \left\|\widehat{\mathbf{G}}_k(\mathbf{W}_k^1 - \mathbf{W}_k^2)\mathbf{1}_k\right\|_2 + |\tau_k|^2 \left\|\widehat{\mathbf{G}}_k(\mathbf{W}_k^1)^*(\mathbf{W}_k^2 - \mathbf{W}_k^1)^T \mathbf{h}_{\boldsymbol{\psi}_k^2}^*\right\|_2$$
$$+ |\tau_k|^2 \left\|\widehat{\mathbf{G}}_k(\mathbf{W}_k^2 - \mathbf{W}_k^1)^*(\mathbf{W}_k^2)^T \mathbf{h}_{\boldsymbol{\psi}_k^2}^*\right\|_2 + |\tau_k|^2 \left\|\widehat{\mathbf{G}}_k(\mathbf{W}_k^1)^*(\mathbf{W}_k^1)^T \widehat{\mathbf{G}}_k(\boldsymbol{\psi}_k^2 - \boldsymbol{\psi}_k^1)\right\|_2$$

$$\overset{3)}{\leq} L_{\boldsymbol{\psi}}^1 \left\|\mathbf{W}_k^1 - \mathbf{W}_k^2\right\|_F + L_{\boldsymbol{\psi}}^2 \left\|\boldsymbol{\psi}_k^2 - \boldsymbol{\psi}_k^1\right\|_2 \leq \ell_1 \left\|\mathbf{W}_k^1 - \mathbf{W}_k^2\right\|_F + \ell_2 \left\|\boldsymbol{\psi}_k^2 - \boldsymbol{\psi}_k^1\right\|_2. \tag{62}$$

$$L_{\boldsymbol{\psi}}^1 = \sqrt{1+\psi_k}\,|\tau_k|\,\lambda_{\max}\left(\widehat{\mathbf{R}}_k\right) + |\tau_k|^2\,\lambda_{\max}\left(\widehat{\mathbf{G}}_k\left(\mathbf{W}_k^1\right)^*\right)\lambda_{\max}\left(\mathbf{h}_{\boldsymbol{\psi}_k^*}^*\right) + |\tau_k|^2\,\lambda_{\max}\left(\widehat{\mathbf{G}}_k\right)\lambda_{\max}\left(\left(\mathbf{W}_k^2\right)^T \mathbf{h}_{\boldsymbol{\psi}_k^2}^*\right)$$

$$\leq \sqrt{1+\psi_k}\,|\tau_k|\,\lambda_{\max}\left(\widehat{\mathbf{G}}_k\right) + 2\,|\tau_k|^2\,\sqrt{P_{\max}}\lambda_{\max}\left(\widehat{\mathbf{G}}_k\right)\left(\|\mathbf{h}_k\|_2 + \sqrt{N}\lambda_{\max}\left(\widehat{\mathbf{G}}_k\right)\right) = \ell_1. \tag{63}$$

$$L_{\boldsymbol{\psi}}^2 = |\tau_k|^2\,\lambda_{\max}\left(\widehat{\mathbf{G}}_k\left(\mathbf{W}_k^1\right)^*\left(\mathbf{W}_k^1\right)^T \widehat{\mathbf{G}}_k^H\right) \leq |\tau_k|^2\,P_{\max}\lambda_{\max}^2\left(\widehat{\mathbf{G}}_k\right) = \ell_2.$$

The derivation of (64) mirrors that of (62), with details omitted for brevity. Substituting (62) and (64) into the left-hand side of (61), we obtain:

$$\|\nabla_{\boldsymbol{\psi}_k} f(\mathscr{A}_k^1) - \nabla_{\boldsymbol{\psi}_k} f(\mathscr{A}_k^2)\|_2^2 + \|\nabla_{\mathbf{w}_k} f(\mathscr{A}_k^1) - \nabla_{\mathbf{w}_k} f(\mathscr{A}_k^2)\|_F^2$$
$$\leq (\ell_2^2 + \ell_4^2)\|\boldsymbol{\psi}_k^1 - \boldsymbol{\psi}_k^2\|_2^2 + (\ell_1^2 + \ell_3^2)\|\mathbf{W}_k^1 - \mathbf{W}_k^2\|_F^2$$
$$+ 2(\ell_1\ell_2 + \ell_3\ell_4)\|\boldsymbol{\psi}_k^1 - \boldsymbol{\psi}_k^2\|_2\|\boldsymbol{\psi}_k^1 - \boldsymbol{\psi}_k^2\|_2$$
$$\leq L^2\left(\|\boldsymbol{\psi}_k^1 - \boldsymbol{\psi}_k^2\|_2^2 + \|\mathbf{W}_k^1 - \mathbf{W}_k^2\|_F^2\right), \tag{65}$$

where

$$L = \max\left\{\sqrt{\ell_2^2 + \ell_4^2 + \ell_1\ell_2 + \ell_3\ell_4},\right.$$
$$\left.\sqrt{\ell_1^2 + \ell_3^2 + \ell_1\ell_2 + \ell_3\ell_4}\right\}. \tag{66}$$

Therefore, building on the above analysis, problem (35) is proved to satisfy the three assumptions required for the AS-ADMM convergence. The computational complexity of the proposed AS-ADMM algorithm is evaluated against centralized schemes. In fact, the overall complexity of AS-ADMM is primarily determined by the variable updates at each worker. For instance, analyzing the per-iteration complexity of worker k in problem (35), the closed-form updates for $\mathbf{W}_k$ and $\boldsymbol{\Theta}_k$ incur complexities of $\mathcal{O}(M^3 + M^2K + M^2)$ and $\mathcal{O}(N^3 + 2NM + NM^2 + (K+1)N^2)$, respectively. In contrast, the per-iteration complexity of the AO-based centralized scheme, i.e., Penalty-BCD algorithm, is $\mathcal{O}(M^3 + M^2K + M^2 + N^3 + 2KNM + KNM^2 + 2KN^2)$. Therefore, the centralized scheme exhibits significantly higher complexity compared to the AS-ADMM algorithm.

VI. SIMULATION RESULTS

This section presents numerical results demonstrating the performance of our proposed AS-ADMM framework. The system operates at a central carrier frequency of 3 GHz. The BS is equipped with $M = 4$ antennas for serving $K = 4$ users. The BD-RIS consists of $N = 16$ elements. The channel
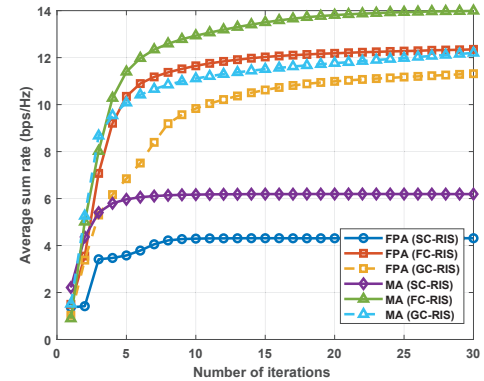

Fig. 4: Average sum rate versus number of iterations of centralized beamforming algorithms with $M = 4$, $K = 4$ and SNR $= 10$ dB.

between BD-RIS and the UEs is characterized by small-scale Rayleigh fading and large-scale distance-dependent path loss. The path loss model is expressed as $\zeta(d) = \zeta_0 d^{-\gamma}$, where $\zeta_0 = -30$ dB represents the reference path loss at a distance of 1 meter, d is the link distance, and γ is the path loss exponent, which is set to 2.2. Azimuth $\phi_{n,l}$ and elevation $\theta_{n,l}$ angles drawn from $[-1, 1]$, $\forall n, l$. The transmit region for the MAs, i.e., $\mathcal{C}$, is set as an $A \times A$ square area, the minimum and maximum coordinates bounded by $\mathbf{x}_{\min} = [-A/2, -A/2]^T$ and $\mathbf{x}_{\max} = [A/2, A/2]^T$, respectively [13], [35]. To avoid coupling effects, MAs require a minimum spacing of $D = \lambda/2$. The initial penalty factor are set as $\rho = 10$ and $\varrho = 10$, respectively. Also, we define the received signal-to-noise ratio (SNR) as SNR $= P_{\max}/\sigma_k^2$ and $D_{\min}$ is set as 2.

A. Centralized Algorithm Performance

We first evaluate the convergence performance of the proposed centralized beamforming algorithm. For clarity, we denote the fully, group, and single-connected architectures of
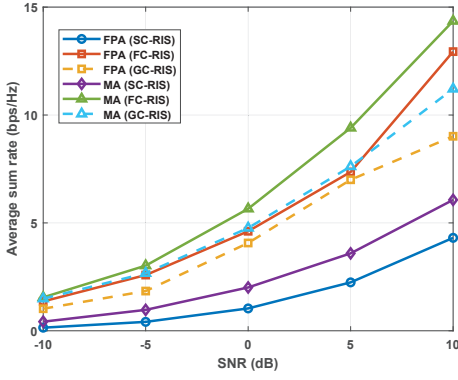
Fig. 5: Average sum rate versus SNR of centralized beam-forming algorithms with $M = 4$ and $K = 4$.

Fig. 6: Channel gain of BS-RIS link against number of iterations.

BD-RIS as FC, GC, and SC, respectively, corresponding to the solutions of problem (9) under constraints $\mathcal{X}_i$ ($i \in \{1, 2, 3\}$). Fig. 4 illustrates the convergence behavior of the centralized beamforming algorithm for the three BD-RIS architectures. The results confirm that MA-aided schemes achieve better performance than the FPA-aided ones. However, the MA-based approaches require additional number of iterations to converge due to the antenna position optimization process inherent in our proposed algorithm. On the other hand, compared to GC and SC architectures, the FC BD-RIS delivers performance gains of 17% and 57%, respectively. This is because the FC BD-RIS can more effectively exploit multiuser diversity, enabling superior signal reflection and broader service coverage. As a result, the FC BD-RIS offers significantly greater flexibility in application compared to "GC/SC" configurations. In Fig. 5, we evaluate the sum rate versus SNR for investigated algorithms. At low SNR, all configurations yield similar sum rates, but as SNR increases, MA (FC-RIS) outperforms others, achieving the highest sum rate with widening performance gaps. Indeed, with a higher transmit power budget, the proposed algorithm can leverage greater flexibility in power allocation during transmission, thus facilitating more effective passive beamforming by the RIS, especially in the FC structure, which has a higher DoF in tuning its elements.

Fig. 6 illustrates the evolution of the average BS-RIS channel gain during antenna position optimization. As the number of iterations increases, the channel power gain steadily improves, further demonstrating how MAs enhance the communication environment by mitigating small-scale fading from multipath effects. Fig. 7 illustrates the evolution of the average equivalent user channel gain $\|\tilde{\mathbf{h}}_k(\tilde{\mathbf{r}})\|^2$ and cross-correlation coefficient Φ, where the latter is computed as $\Phi = \sum_{1 \leq k \neq q \leq K} \frac{|\tilde{\mathbf{h}}_k(\tilde{\mathbf{r}})^H \tilde{\mathbf{h}}_q(\tilde{\mathbf{r}})|}{\|\tilde{\mathbf{h}}_k(\tilde{\mathbf{r}})\|_2 \|\tilde{\mathbf{h}}_q(\tilde{\mathbf{r}})\|_2}$. The joint optimization process reveals the channel gain increases marginally 71% while user similarity drops significantly 41.8%, demonstrating that the performance improvements arise primarily from multiuser interference mitigation.

To demonstrate the benefits of MAs, we compare them with three baseline schemes:

Fixed position antenna (FPA): The BS employs a uniform planar array (UPA) with M antennas uniformly spaced at $\lambda/2$
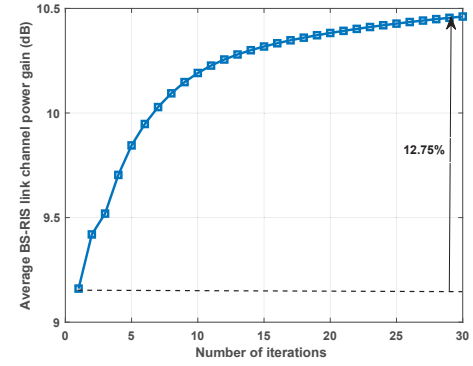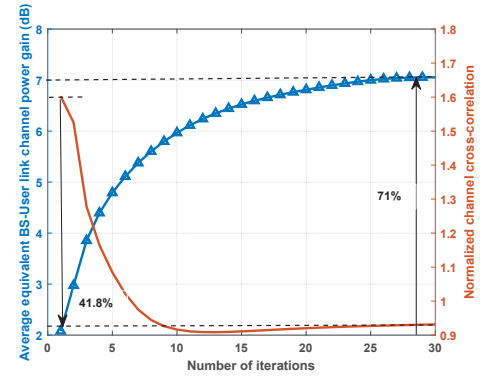

Fig. 7: Average equivalent BS-user link channel power gain and normalized crosscorrelation against number of iterations.

intervals.

Exhaustive antenna selection (EAS) [43]: The BS adopts a larger UPA comprising $2M$ antennas ($\lambda/2$-spaced), selecting the optimal M-antenna subset via exhaustive search.

Particle swarm optimization (PSO) [44]: The positions of the MAs in transmitter are updated by the PSO with a penalty function.

In Figs. 8 and 9, our proposed algorithm consistently outperforms all baselines. Specifically, compared with FPA and AS methods, the proposed optimization framework and other MA-based algorithms can flexibly optimize antenna position in a continuous spatial region, thus enabling a higher spatial DoF for significantly enhancing the achievable rate. Indeed, this performance improvement arises from two main factors. First, MAs dynamically position themselves to harness constructive multipath fading, thereby enhancing channel power gain (as demonstrated in Fig. 6). Second, and more importantly, MAs strategically determine favourable locations that increase spatial separation among users' channels, effectively reducing inter-user interference. Moreover, given a fixed total number of antennas, the performance gain of the EAS scheme gradually saturates as the number of selectable antennas increases, while its performance gap with the MA scheme continues to increase. Compared with the PSO algorithm, our proposed algorithm demonstrates significantly superior performance. This is particularly relevant in practical scenarios, where zeroth-order optimization methods such as PSO often involve tuning
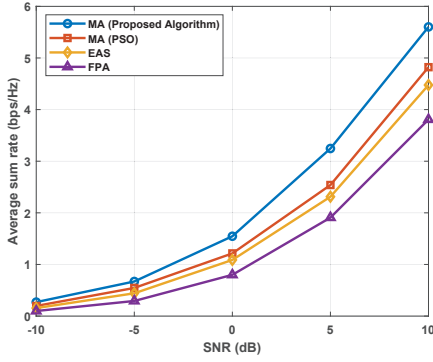
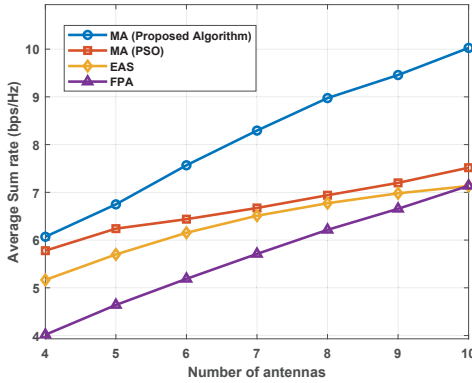Fig. 8: Average sum rate versus SNR of the proposed Penalty-BCD algorithm.

Fig. 10: Average sum rate versus SNR of distributed beam-forming algorithms with $M = 4$ and $K = 4$.

Fig. 9: Average sum rate versus the number of antennas of the proposed Penalty-BCD algorithm.

Fig. 11: Average sum rate versus SNR of distributed beam-forming algorithms with $M = 4$ and $K = 4$.

numerous hyperparameters, rendering it challenging to identify effective configurations.

B. Distributed Algorithm Performance

To demonstrate the effectiveness of the proposed AS-ADMM framework at different SNRs, we plot the sum rate as a function of SNR as shown in Figs. 10 and 11. Our results show that the AS-ADMM algorithm achieves performance comparable to S-ADMM in both FAS and MA scenarios. Moreover, Fig. 11 verifies that our algorithm is still applicable under three different RIS architectures. Additionally, the MA scheme achieves approximately 30% higher performance than the FPA scheme across all scenarios by fully exploiting spatial DoF. This gain stems from simultaneous improvements in three aspects: channel conditions, multiuser interference suppression, and antenna array geometry optimization.

Finally, we verify the computational overhead of the proposed algorithms. In the AS-ADMM framework, the effects of asynchronous clock settings can be quantified by the parameter $D_{\min}$, which is the minimum number of users the BS needs to wait for to return their update results before each new update. In Table I, it is clear that increasing $D_{\min}$ leads to longer competition time per iteration, with performance precisely approaching that of synchronized algorithm. Fig. 12 demonstrates that the distributed algorithm achieves approximately a 50.18% reduction in computational complexity compared to
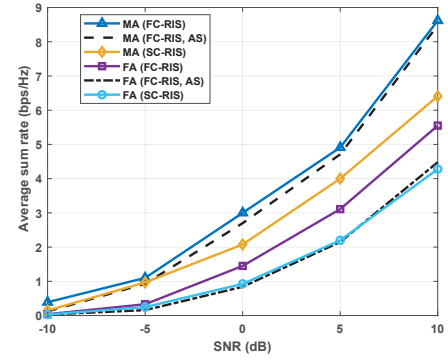
its centralized execution. This improvement stems from the distributed approach's ability to perform calculations locally at user devices, which inherently involves lower complexity than centralized processing at the BS.

In Table II, it can be seen that the performance of the proposed AS-ADMM algorithm is comparable to that of synchronization under several RIS architectures, but with lower computational and interaction overhead. As shown in Table III, for all considered values of K, the AS-ADMM algorithm converges faster than S-ADMM by relaxing strict synchronization requirements. This enables more frequent variable updates at both the BS and workers, resulting in significantly lower computational complexity compared to the synchronous scheme. We conclude from these simulation results that AS-ADMM distributed algorithm significantly outperforms S-ADMM distributed algorithm in the asynchronous network, even though the two have the comparable convergence behaviors.

VII. CONCLUSION

This paper studied downlink transmission of the BD-RIS-aided MA systems and proposed centralized algorithms to maximize the sum rate by optimizing the transmit beamformer, BD-RIS phase-shift, and antenna positions. Furthermore, to further reduce computational overhead, we proposed an asynchronous distributed optimization solution. Numerical results

TABLE I: Comparison of Normalized Running Time for Different $D_{\min}$.

$D_{\min}$	8	4	2
Normalized Running Time [s]	1	0.9186	0.9142
SE Performance [bps/Hz]	5.9143	4.6673	4.4641

TABLE II: Comparison of Normalized Running Time for Investigated Algorithms.

Normalized Running Time	SC-RIS	GC-RIS	FC-RIS
S-ADMM	0.8286	0.9860	1.0000
AS-ADMM	0.7650	0.8320	0.8627

TABLE III: Comparison of Normalized Running Time for Investigated Algorithms.

Scheme	Normalized Running Time ($K = 4$)	Normalized Running Time ($K = 8$)	Normalized Running Time ($K = 16$)
FPA, S-ADMM	0.0176	0.0255	0.0853
FPA, AS-ADMM	0.0000	0.0085	0.0295
MA, S-ADMM	0.9986	0.8531	1.0000
MA, AS-ADMM	0.8905	0.8484	0.9629

demonstrated that MAs can effectively reduce user channel correlation, significantly mitigating multi-user interference, while antenna displacement helps harness multipath effects from channel fading and improve channel gain in BD-RIS systems. Moreover, our proposed asynchronous distributed algorithm achieves performance close to synchronous schemes at significantly lower complexity. Future work will explore robust beamforming design for BD-RIS-aided MA systems with channel uncertainty.

APPENDIX

A. Proof of Proposition 1

Applying the Lagrangian dual transform, we reformulate the objective function of (9) as

$$\sum_{k \in \mathcal{K}} \left(\log\left(1 + \iota_k\right) - \iota_k \right) + \sum_{k \in \mathcal{K}} \frac{\left(1 + \iota_k\right) \gamma_k}{1 + \gamma_k}, \quad (67)$$

where we set $g\left(\mathbf{W}, \tilde{\mathbf{r}}, \boldsymbol{\Theta}, \boldsymbol{\iota}\right) = \sum_{k \in \mathcal{K}} \frac{(1 + \iota_k) \gamma_k}{1 + \gamma_k}$. The optimal ι_k for maximizing (67) is given by $\iota_k^\star = \gamma_k$. Following the framework of [36], for a general complex quadratic form $\sum_{i \in \mathcal{K}} \mathbf{a}_i^H(\mathbf{x}) \mathbf{B}_i^{-1}(\mathbf{x}) \mathbf{a}_i(\mathbf{x})$, where $\mathbf{a}_i(\mathbf{x})$ is a complex vector and $\mathbf{B}_i(\mathbf{x})$ is a positive-definite matrix, the equivalent quadratic representation is $\sum_{i \in \mathcal{K}} \left(2\Re\{\mathbf{v}_i^H \mathbf{a}_i(\mathbf{x})\} - \mathbf{v}_i^H \mathbf{B}_i(\mathbf{x}) \mathbf{v}_i \right)$, where $\mathbf{v}_i$ is an auxiliary optimization variable. Applying this transformation, the $g\left(\mathbf{W}, \tilde{\mathbf{r}}, \boldsymbol{\Theta}, \boldsymbol{\iota}\right)$ can be expressed as the quadratic term

$$g\left(\mathbf{W}, \tilde{\mathbf{r}}, \boldsymbol{\Theta}, \boldsymbol{\iota}, \boldsymbol{\tau}\right) = \sum_{k \in \mathcal{K}} 2\sqrt{1 + \iota_k} \Re\{\tilde{\mathbf{h}}_k\left(\tilde{\mathbf{r}}\right)^H \mathbf{w}_k \tau_k^*\} \\ - \sum_{k \in \mathcal{K}} \left|\tau_k\right|^2 \left(\sum_{i \in \mathcal{K}} |\tilde{\mathbf{h}}_k(\tilde{\mathbf{r}})^H \mathbf{w}_i|^2 + \sigma_k^2\right). \quad (68)$$

The optimal τ_k for maximizing $g\left(\mathbf{W}, \tilde{\mathbf{r}}, \boldsymbol{\Theta}, \boldsymbol{\iota}, \boldsymbol{\tau}\right)$ is derived as $\tau_k^\star = \frac{\sqrt{1 + \iota_k} \mathbf{w}_k^H \left(\mathbf{G}^H \boldsymbol{\Theta}^H \mathbf{h}_k\right)}{\sum_{i \in \mathcal{K}} \left|\left(\mathbf{h}_k^H \boldsymbol{\Theta} \mathbf{G}\right) \mathbf{w}_i\right|^2 + \sigma_k^2}$. By integrating equations (67) and (68), we demonstrate that problem (9) is equivalent to problem (11), thus concluding the proof.

B. Proof of Proposition 2

The right-hand side of (23) can be expressed as

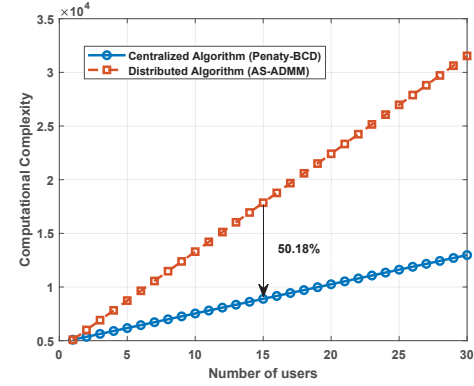$$\arg \min_{\mathbf{B} \in \mathcal{X}_2} \|\mathbf{A} - \mathbf{C} + \mathbf{C} - \mathbf{B}\|_F^2, \quad (69)$$

Fig. 12: Average sum rate versus SNR of distributed beamforming algorithms with $M = 4$ and $K = 4$.

where $\mathbf{C} = \text{Proj}_{\text{sym}}\left(\mathbf{A}\right) = \frac{1}{2}\left(\mathbf{A} + \mathbf{A}^T\right) = \mathbf{U}\mathbf{S}\mathbf{V}^H$. (69) can be further written as

$$\arg \min_{\mathbf{B} \in \mathcal{X}_2} \|\mathbf{C} - \mathbf{B}\|_F^2 + 2\Re\left\{\text{Tr}\left((\mathbf{A} - \mathbf{C})(\mathbf{C} - \mathbf{B})^H\right)\right\}. \quad (70)$$

Since the inner product of the skew-symmetric matrix $\mathbf{A} - \mathbf{C}$ and the symmetric matrix $\mathbf{C} - \mathbf{B}$ is zero, i.e., $\text{Tr}\left((\mathbf{A} - \mathbf{C})(\mathbf{C} - \mathbf{B})^H\right) = 0$, this reduces to

$$\arg \min_{\mathbf{B} \in \mathcal{X}_2} \|\mathbf{C} - \mathbf{B}\|_F^2, \quad (71)$$

where $\mathcal{X}_2$ is a subset of $\{\mathbf{B} \mid \mathbf{B}\mathbf{B}^H = \mathbf{I}\}$, and $\hat{\mathbf{U}}\mathbf{V}^H = \arg \min_{\mathbf{B}\mathbf{B}^H = \mathbf{I}} \|\mathbf{C} - \mathbf{B}\|_F^2$, which provides a lower bound for $\arg \min_{\mathbf{B} \in \mathcal{X}_2} \|\mathbf{C} - \mathbf{B}\|_F^2$. This bound is tight because $\hat{\mathbf{U}}\mathbf{V}^H \in \mathcal{X}_2$.

Substituting the SVD of $\mathbf{C}$ into $\mathbf{C} = \mathbf{C}^T$, we obtain:

$$\mathbf{U}\mathbf{S}\mathbf{V}^H = \mathbf{V}^*\mathbf{S}\mathbf{U}^T. \quad (72)$$

By left-multiplying (72) by $\mathbf{U}^H$ and right-multiplying by $\mathbf{U}^*$, we obtain $\mathbf{S}\mathbf{V}^H\mathbf{U}^* = \mathbf{U}^H\mathbf{V}^*\mathbf{S}$. Define $\mathbf{Y}_R = \mathbf{U}_R^H\mathbf{V}_R^*$. For its non-diagonal entries $y_{i,j}$, where $i \neq j$, $\forall i, j \in \{1, \ldots, R\}$, the condition $y_{i,j} = \frac{s_i}{s_j} y_{j,i}$ must hold. Similarly, multiplying (72) left by $\mathbf{V}^T$ and right by $\mathbf{V}$, respectively, gives $\mathbf{V}^T\mathbf{U}\mathbf{S} = \mathbf{S}\mathbf{U}^T\mathbf{V}$, leading to:

$$\begin{bmatrix} \mathbf{V}_R^T\mathbf{U}_R\mathbf{S}_R & \mathbf{0} \\ \mathbf{V}_{N-R}^T\mathbf{U}_R\mathbf{S}_R & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{S}_R\mathbf{U}_R^T\mathbf{V}_R & \mathbf{S}_R\mathbf{U}_R^T\mathbf{V}_{N-R} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \quad (73)$$

This implies $\mathbf{U}_R^H\mathbf{V}_{N-R}^* = \mathbf{0}$ and $y_{j,i}^* = \frac{s_i}{s_j} y_{i,j}^*$. Since $y_{i,j} = \frac{s_i}{s_j} y_{j,i}$, it follows that $y_{i,j} = 0$ for $i \neq j$, making $\mathbf{Y}_R$ a diagonal matrix. Given $\hat{\mathbf{U}} = [\mathbf{U}_R, \mathbf{V}_{N-R}]$, it can

This article has been accepted for publication in IEEE Transactions on Wireless Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TWC.2025.3613556

14

be verified that $\hat{\mathbf{U}}$ is unitary and $\hat{\mathbf{U}}^H\mathbf{V}^*$ is diagonal, i.e., $\hat{\mathbf{U}}^H\mathbf{V}^* = \mathbf{V}^H\hat{\mathbf{U}}^*$.

Thus, the symmetric unitary projection of $\mathbf{Z}$ can be expressed as $\mathrm{Proj}_{\mathrm{symuni}}(\mathbf{Z}) = \hat{\mathbf{U}}\mathbf{V}^H = \mathbf{V}^*\hat{\mathbf{U}}^T = \mathrm{Proj}_{\mathrm{symuni}}(\mathbf{Z})^T$. Since $\mathrm{Proj}_{\mathrm{symuni}}(\mathbf{Z})\,\mathrm{Proj}_{\mathrm{symuni}}(\mathbf{Z})^H = \hat{\mathbf{U}}\mathbf{V}^H\mathbf{V}\hat{\mathbf{U}}^H = \mathbf{I}$, the proof is complete.

REFERENCES

[1] J. Zhang, Z. Liu, Y. Zhu, E. Shi, B. Xu, C. Yuen, D. Niyato, M. Debbah, S. Jin, B. Ai, and X. Shen, "Multi-agent reinforcement learning in wireless distributed networks for 6G," *arXiv:2502.05812*, 2025.

[2] B. Xu, J. Zhang, H. Du, Z. Wang, Y. Liu, D. Niyato, B. Ai, and K. B. Letaief, "Resource allocation for near-field communications: Fundamentals, tools, and outlooks," *IEEE Wireless Commun.*, vol. 31, no. 5, pp. 42–50, Jul. 2024.

[3] X. Shao, C. You, W. Ma, X. Chen, and R. Zhang, "Target sensing with intelligent reflecting surface: Architecture and performance," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 7, pp. 2070–2084, Mar. 2022.

[4] K. Zhao, Y. Mao, and Y. Shi, "Simultaneously transmitting and reflecting reconfigurable intelligent surfaces empowered cooperative rate splitting with user relaying," *Entropy*, vol. 26, no. 12, p. 1019, Nov. 2024.

[5] H. Li, S. Shen, and B. Clerckx, "Beyond diagonal reconfigurable intelligent surfaces: From transmitting and reflecting modes to single-, group-, and fully-connected architectures," *IEEE Trans. Wireless Commun.*, vol. 22, no. 4, pp. 2311–2324, Oct. 2023.

[6] T. Fang and Y. Mao, "A low-complexity beamforming design for beyond-diagonal RIS aided multi-user networks," *IEEE Commun. Lett.*, vol. 28, no. 1, pp. 203–207, Nov. 2024.

[7] O. Maraqa, M. H. Khoshafa, O. O. Oyerinde, and T. M. N. Ngatched, "Beyond diagonal RIS-aided wireless communications systems: State-of-the-art and future research directions," *arXiv:2503.08826*, 2025.

[8] K. Chen and Y. Mao, "Transmitter side beyond-diagonal RIS for mmwave integrated sensing and communications," in *Proc. IEEE 21st Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, 2024, pp. 951–955.

[9] C. Pan, G. Zhou, K. Zhi, S. Hong, T. Wu, Y. Pan, H. Ren, M. D. Renzo, A. Lee Swindlehurst, R. Zhang, and A. Y. Zhang, "An overview of signal processing techniques for RIS/IRS-aided wireless systems," *IEEE J. Sel. Top. Signal Process.*, vol. 16, no. 5, pp. 883–917, Aug. 2022.

[10] H. Li, S. Shen, M. Nerini, M. Di Renzo, and B. Clerckx, "Beyond diagonal reconfigurable intelligent surfaces with mutual coupling: Modeling and optimization," *IEEE Commun. Lett.*, vol. 28, no. 4, pp. 937–941, Feb. 2024.

[11] Z. Wu and B. Clerckx, "Optimization of beyond diagonal RIS: A universal framework applicable to arbitrary architectures," *arXiv:2412.15965*, 2024.

[12] X. Zhou, T. Fang, and Y. Mao, "Joint active and passive beamforming optimization for beyond diagonal RIS-aided multi-user communications," *IEEE Commun. Lett.*, vol. 29, no. 3, pp. 517–521, Jan. 2025.

[13] L. Zhu, W. Ma, and R. Zhang, "Modeling and performance analysis for movable antenna enabled wireless communications," *IEEE Trans. Wireless Commun.*, vol. 23, no. 6, pp. 6234–6250, Nov. 2024.

[14] K.-K. Wong and K.-F. Tong, "Fluid antenna multiple access," *IEEE Trans. Wireless Commun.*, vol. 21, no. 7, pp. 4801–4815, Dec. 2022.

[15] Y. Jin, Q. Lin, Y. Li, and Y.-C. Wu, "Handling distance constraint in movable antenna aided systems: A general optimization framework," in *Proc. Signal Processing Advances in Wireless Communications (SPAWC)*, 2024, pp. 396–400.

[16] Y. Jin, Q. Lin, Y. Li, H. Zhu, B. Cheng, Y.-C. Wu, and R. Zhang, "A general optimization framework for tackling distance constraints in movable antenna-aided systems," *arXiv:2503.02344*, 2025.

[17] Y. Wu, D. Xu, D. W. K. Ng, W. Gerstacker, and R. Schober, "Movable antenna-enhanced multiuser communication: Optimal discrete antenna positioning and beamforming," *arXiv:2308.02304*, 2023.

[18] H. Lu, Y. Zeng, S. Jin, and R. Zhang, "Group movable antenna with flexible sparsity: Joint array position and sparsity optimization," *IEEE Wirel. Commun. Lett.*, vol. 13, no. 12, pp. 3573–3577, Oct. 2024.

[19] R. Zhao, S. Hu, D. Mishra, and D. W. K. Ng, "Resource allocation for multi-waveguide pinching antenna-assisted broadcast networks," *arXiv:2507.03915*, 2025.

[20] S. Hu, R. Zhao, Y. Liao, D. W. K. Ng, and J. Yuan, "Sum-rate maximization for pinching antenna-assisted NOMA systems with multiple dielectric waveguides," *arXiv:2503.10060*, 2025.

[21] Y. Wu, D. Xu, D. W. K. Ng, W. Gerstacker, and R. Schober, "Globally optimal movable antenna-enabled multiuser communication: Discrete antenna positioning, power consumption, and imperfect CSI," *IEEE Trans. Commun.*, pp. 1–1, Jun. 2025.

[22] X. Shao, Q. Jiang, and R. Zhang, "6D movable antenna based on user distribution: Modeling and optimization," *IEEE Trans. Wireless Commun.*, vol. 24, no. 1, pp. 355–370, Jan. 2025.

[23] H. Lu, Z. Yu, Y. Zeng, S. Ma, S. Jin, and R. Zhang, "Wireless communication with flexible reflector: Joint placement and rotation optimization for coverage enhancement," *IEEE Trans. Wirel. Commun.*, pp. 1–1, May 2025.

[24] W. Ma, L. Zhu, and R. Zhang, "Compressed sensing based channel estimation for movable antenna communications," *IEEE Commun. Lett.*, vol. 27, no. 10, pp. 2747–2751, Aug. 2023.

[25] W. Kiat New, K.-K. Wong, H. Xu, F. Rostami Ghadi, R. Murch, and C.-B. Chae, "Channel estimation and reconstruction in fluid antenna system: Oversampling is essential," *IEEE Trans. Wireless Commun.*, vol. 24, no. 1, pp. 309–322, Nov. 2025.

[26] Y. Xu, E. G. Larsson, E. A. Jorswieck, X. Li, S. Jin, and T.-H. Chang, "Distributed signal processing for extremely large-scale antenna array systems: State-of-the-art and future directions," *IEEE Journal of Selected Topics in Signal Processing*, pp. 1–26, Feb. 2025.

[27] G. Tychogiorgos, A. Gkelias, and K. K. Leung, "A non-convex distributed optimization framework and its application to wireless Ad-hoc networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4286–4296, Aug. 2013.

[28] C. Shen, T.-H. Chang, K.-Y. Wang, Z. Qiu, and C.-Y. Chi, "Distributed robust multicell coordinated beamforming with imperfect csi: An ADMM approach," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2988–3003, Feb. 2012.

[29] Z. Wang, J. Zhang, E. Björnson, D. Niyato, and B. Ai, "Optimal bilinear equalizer for cell-free massive MIMO systems over correlated Rician channels," *IEEE Trans. Signal Process.*, to appear, 2025.

[30] F. Foukalas, R. Shakeri, and T. Khattab, "Distributed power allocation for multi-flow carrier aggregation in heterogeneous cognitive cellular networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2486–2498, Jan. 2018.

[31] M. Hong, "A distributed, asynchronous, and incremental algorithm for nonconvex optimization: An ADMM approach," *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 3, pp. 935–945, Jan. 2018.

[32] T.-H. Chang, M. Hong, W.-C. Liao, and X. Wang, "Asynchronous distributed ADMM for large-scale optimization-part I: Algorithm and convergence analysis," *IEEE Trans. Signal Process.*, vol. 64, no. 12, pp. 3118–3130, Mar. 2016.

[33] 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, Jun. 2019.

[34] J. Zheng, J. Zhang, H. Du, D. Niyato, S. Sun, B. Ai, and K. B. Letaief, "Flexible-position MIMO for wireless communications: Fundamentals, challenges, and future directions," *IEEE Wireless Commun.*, pp. 1–9, Mar. 2024.

[35] S. Yang, W. Lyu, B. Ning, Z. Zhang, and C. Yuen, "Flexible precoding for multi-user movable antenna communications," *IEEE Wireless Commun. Lett.*, vol. 13, no. 5, pp. 1404–1408, Mar. 2024.

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

[37] S. Xie, S. Gong, H. Liu, C. Xing, J. An, and Y. Li, "Asynchronous distributed beamforming optimization framework for RIS-assisted wireless communications," *IEEE Trans. Signal Process.*, vol. 71, pp. 3083–3099, Aug. 2023.

[38] B. Xu, J. Zhang, Z. Chen, B. Cheng, Z. Liu, Y.-C. Wu, and B. Ai, "Channel estimation for Rydberg atomic receivers," *arXiv:2503.08985*, 2025.

[39] B. Xu, J. Zhang, Q. Lin, H. Xiao, Y.-C. Wu, and B. Ai, "Deep unfolding beamforming and power control designs for multi-port matching networks," *IEEE Trans. Wireless Commun.*, vol. 24, no. 2, pp. 1401–1414, Dec. 2025.

[40] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[41] A. I. Rikos, W. Jiang, T. Charalambous, and K. H. Johansson, "Asynchronous distributed optimization via ADMM with efficient communication," in *2023 62nd IEEE Conference on Decision and Control (CDC)*, 2023, pp. 7002–7008.

This article has been accepted for publication in IEEE Transactions on Wireless Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TWC.2025.3613556

15

[42] Z. Zhang and L. Dai, "A joint precoding framework for wideband reconfigurable intelligent surface-aided cell-free network," *IEEE Trans. Signal Process.*, vol. 69, pp. 4085–4101, Jun. 2021.

[43] S. Sanayei and A. Nosratinia, "Capacity maximizing algorithms for joint transmit-receive antenna selection," in *Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers, 2004.*, vol. 2, Nov. 2004, pp. 1773–1776 Vol.2.

[44] Y. Zuo, J. Guo, B. Sheng, C. Dai, F. Xiao, and S. Jin, "Fluid antenna for mobile edge computing," *IEEE Commun. Lett.*, vol. 28, no. 7, pp. 1728–1732, May 2024.

Zhe Wang received the Ph.D. degree from Beijing Jiaotong University, China, in 2025, and he was a visiting student at Nanyang Technological University, Singapore, from 2023 to 2024. He is currently a postdoctoral researcher at KTH Royal Institute of Technology, Sweden. His research interests include extremely large-scale MIMO, near-field communication, cell-free massive MIMO, and performance analysis of wireless systems. He received the Best Paper Award at the IEEE ICC 2023.

Bokai Xu received the B.S. degree in communication engineering from Beijing Jiaotong University, China, in 2024, where he is currently pursuing the Ph.D. degree. He has been a visiting student at the School of Electrical & Electronic Engineering, The University of Hong Kong, Hong Kong, since Nov. 2024. His research interests include trustworthy machine learning of wireless systems and Quantum MIMO Receiver.

Jiayi Zhang (Senior Member, IEEE) received the B.Sc. and Ph.D. degree of Communication Engineering from Beijing Jiaotong University, China in 2007 and 2014, respectively. Since 2016, he has been a Professor with School of Electronic and Information Engineering, Beijing Jiaotong University, China. From 2014 to 2016, he was a Postdoctoral Research Associate with the Department of Electronic Engineering, Tsinghua University, China. From 2014 to 2015, he was also a Humboldt Research Fellow in Institute for Digital Communications, Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Germany. His current research interests include cell-free massive MIMO, reconfigurable intelligent surface, XL-MIMO, near-field communications and applied mathematics.

Dr. Zhang received the Best Paper Awards at the IEEE ICC 2023, WCSP 2017 and IEEE APCC 2017, the URSI Young Scientist Award in 2020, and the IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award in 2020. He was the Lead Guest Editor of the special issue on "Multiple Antenna Technologies for Beyond 5G" of the IEEE Journal on Selected Areas in Communications, the Lead Guest Editor of the special issue on "Semantic Communications for the Metaverse" of the IEEE Wireless Communications and an Editor for IEEE Communications Letters from 2016-2021. He currently serves as an Associate Editor for IEEE Transactions on Communications and IEEE Transactions on Wireless Communications.

Bo Ai (Fellow, IEEE) received his Master degree and Ph. D. degree from Xidian University in China. He graduated from Tsinghua University with the honor of Excellent Postdoctoral Research Fellow at Tsinghua University in 2007. He was a visiting professor at EE Department, Stanford University in 2015. He is now working at Beijing Jiaotong University as a full professor and Ph. D. candidate advisor. He is the Deputy Director of State Key Lab of Rail Traffic Control and Safety, and the Deputy Director of International Joint Research Center. He is one of the main responsible people for Beijing "Urban rail operation control system" International Science and Technology Cooperation Base, and the backbone member of the Innovative Engineering Based jointly granted by Chinese Ministry of Education and the State Administration of Foreign Experts Affairs.

He has authored/co-authored 8 books and published over 300 academic research papers in his research area. He has hold 26 invention patents. He has been the research team leader for 26 national projects and has won some important scientific research prizes. Five papers have been the ESI highly-cited paper. He has been notified by Council of Canadian Academies (CCA) that, based on Scopus database, Prof. Bo Ai has been listed as one of the Top 1% authors in his field all over the world. Prof. Bo Ai has also been Feature Interviewed by IET ELECTRONICS LETTERS. His interests include the research and applications of channel measurement and channel modeling, dedicated mobile communications for rail traffic systems.

Prof. Bo Ai is a Fellow of the Institution of Engineering and Technology (IET Fellow), IEEE VTS Distinguished Lecturer. He is an IEEE VTS Beijing Chapter Vice Chair. IEEE BTS Xi'an Chapter Chair. He was as a Co-chair or a Session/Track Chair for many international conferences. He is an associate editor of IEEE ANTENNAS AND WIRELESS PROPAGATION LETTERS, IEEE TRANSACTIONS ON CONSUMER ELECTRONICS and an Editorial Committee Member of the Wireless Personal Communications journal. He is the Lead Guest Editor for Special Issues on IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE ANTENNAS AND PROPAGATIONS LETTERS, INTERNATIONAL JOURNAL ON ANTENNAS AND PROPAGATIONS. He has received many awards such as Distinguished Youth Foundation and Excellent Youth Foundation from National Natural Science Foundation of China, the Qiushi Outstanding Youth Award by Hong Kong Qiushi Foundation, the New Century Talents by the Chinese Ministry of Education, the Zhan Tianyou Railway Science and Technology Award by the Chinese Ministry of Railways, and the Science and Technology New Star by the Beijing Municipal Science and Technology Commission.

Ziheng Liu received the B.S. degree from the School of Information and Control Engineering, Qingdao University of Technology, Qingdao, China, in 2023. He is currently pursuing the Ph.D. degree with the School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, China. His research interests include massive MIMO systems, signal processing, and reinforcement learning.

Derrick Wing Kwan Ng (S'06-M'12-SM'17-F'21) received his bachelor's degree (with first-class Honors) and the Master of Philosophy degree in electronic engineering from The Hong Kong University of Science and Technology (HKUST), Hong Kong, in 2006 and 2008, respectively, and his Ph.D. degree from The University of British Columbia, Vancouver, BC, Canada, in November 2012. Following his Ph.D., he was a senior post-doctoral fellow at the Institute for Digital Communications, Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Germany. He is currently a Scientia Associate Professor with the University of New South Wales, Sydney, NSW, Australia. His research interests include global optimization, integrated sensing and communication (ISAC), physical layer security, IRS-assisted communication, UAV-assisted communication, wireless information and power transfer, and green (energy-efficient) wireless communications.

He has been recognized as a Highly Cited Researcher by Clarivate Analytics (Web of Science) since 2018. He was the recipient of the Australian Research Council (ARC) Discovery Early Career Researcher Award 2017, IEEE Communications Society Leonard G. Abraham Prize 2023, IEEE Communications Society Stephen O. Rice Prize 2022, Best Paper Awards at the WCSP 2020, 2021, IEEE TCGCC Best Journal Paper Award 2018, INISCOM 2018, IEEE International Conference on Communications (ICC) 2018, 2021, 2023, 2024, IEEE International Conference on Computing, Networking and Communications (ICNC) 2016, IEEE Wireless Communications and Networking Conference (WCNC) 2012, IEEE Global Telecommunication Conference (Globecom) 2011, 2021, 2023 and IEEE Third International Conference on Communications and Networking in China 2008. From January 2012 to December 2019, he served as an Editorial Assistant to the Editor-in-Chief of the IEEE Transactions on Communications. He is also the Editor of the IEEE Transactions on Communications and an Associate Editor-in-Chief for the IEEE Open Journal of the Communications Society.