Physical Communication 78 (2026) 103212

Contents lists available at ScienceDirect

Physical Communication

journal homepage: www.elsevier.com/locate/phycom

Full length article

Joint optimization of deployment and topology for irregular RIS in wireless communication systems

Xinying Guo [a,*], Rui Yuan [a], Jiankang Zhang [b], Yi Song [a], Sheng Chen [c]

[a] Key Laboratory of Grain Information Processing and Control (Henan University of Technology), Ministry of Education, and College of Information Science and Engineering, Henan University of Technology, Zhengzhou, 450001, China
[b] Department of Computing & Informatics, Bournemouth University, BH12 5BB, UK
[c] School of Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, UK

ARTICLE INFO

Keywords:
Irregular reconfigurable intelligent surface
Deployment optimization
Topology design
Block alternating optimization

ABSTRACT

This paper considers an irregular reconfigurable intelligent surface (IRIS)-aided multi-user wireless communication system. A novel block alternating optimization (BAO) algorithm framework is proposed to maximize the system's weighted sum-rate (WSR) by jointly optimizing base station (BS) active beamforming, IRIS deployment, IRIS passive beamforming and IRIS topology. This challenging non-convex optimization problem is decomposed into two blocks based on the continuous and discrete characteristics of the optimization variables. In the first block, BS active beamforming and IRIS deployment are optimized using convex optimization and successive convex approximation algorithms, respectively. In the second block, IRIS passive beamforming and topology are addressed through a neighbor extraction cross-entropy algorithm and a proposed genetic algorithm-tabu search (GA-TS) algorithm, respectively. These two blocks are alternately optimized. Simulation results reveal three key findings: 1) compared to the traditional alternating optimization framework, the proposed BAO framework achieves rapid convergence while significantly enhancing the system's achievable WSR; 2) optimizing the IRIS deployment is critical for mitigating the "multiplicative fading effect", exhibiting a distinct double-peak characteristic; and 3) the proposed GA-TS algorithm closely tracks the near-optimal performance of the exhaustive search scheme, reducing computational complexity even with a limited number of RIS elements.

1. Introduction

The air interface transmission conditions between the base station (BS) and the user equipment (UE) have long been considered uncontrollable [1]. However, recent advancements in meta-materials have introduced a game-changing solution: the reconfigurable intelligent surface (RIS) [2,3]. By enabling the adjustment of numerous RIS elements' reflection coefficients, this technology allows for the manipulation of the wireless propagation environment [4]. As a result, RIS is now widely considered as one of the most revolutionary and promising candidate technologies for 6G networks [5,6].

In recent years, extensive research has been conducted on the system performance and optimization algorithms of RIS-assisted wireless communication systems [7–10]. Among these, the deployment of RIS has been identified as a crucial factor in mitigating the "multiplicative fading effect" in such systems. Typically, a RIS is deployed between the BS and the UE, and its position significantly impacts the quality of the reflected wireless channel and the overall system performance [11–13]. Using the criterion of maximizing the received signal-to-noise ratio (SNR) at the UE, the authors in [11] concluded that the received SNR of the UE is maximized when RIS is deployed near the BS or the UE through numerical analysis. The work [12] found that the optimal position of the RIS depends on the relative positions among the BS, the RIS and the UE. The study [13] proposed a novel RIS deployment strategy aimed at maximizing the received signal power. However, the aforementioned studies [11–13] mainly focused on single-user systems, representing an oversimplification that fails to account for the complex inter-user interference in practical multi-user scenarios. To address this issue, the authors in [14] investigated an RIS-assisted multi-user wireless communication system by integrating antenna grouping with movable RIS deployment. Nevertheless, these deployment strategies are exclusively tailored for regular RIS architectures with uniform element arrangements, which lack the spatial degrees of freedom required to further manipulate the propagation environment.

On the other hand, for RIS-assisted systems, increasing the number of reflecting elements improves system capacity but also elevates

* Corresponding author.
E-mail addresses: guoxinying@haut.edu.cn (X. Guo), yuanrui@stu.haut.edu.cn (R. Yuan), jzhang3@bournemouth.ac.uk (J. Zhang), songyi@haut.edu.cn (Y. Song), sqc@ecs.soton.ac.uk (S. Chen).

https://doi.org/10.1016/j.phycom.2026.103212
Received 20 April 2026; Received in revised form 22 May 2026; Accepted 8 June 2026
Available online 9 June 2026

channel estimation overhead and beamforming complexity [15]. To address this challenge, the authors in [16] proposed the idea of using the irregular RIS (IRIS) to maximize system capacity with limited reflecting elements. In [17], IRIS was further introduced for maximizing the energy efficiency of multiple-input single-output wireless systems. Simulation results show that IRIS can significantly improve system energy efficiency. Furthermore, the potential of IRIS to enhance wireless communication security within dense urban environments was investigated in [18]. However, the aforementioned contributions [16–18] primarily focus on IRIS topology design under fixed deployment assumptions, overlooking the potential performance gains achievable through IRIS deployment optimization in multi-user environments. Therefore, a new framework for jointly optimizing deployment and topology is urgently needed for IRIS-aided multi-user downlink communication systems.

1.1. Contributions

To bridge this gap, we propose a joint optimization framework for deployment and topology design in IRIS-aided multi-user wireless communication systems. The main contributions are summarized as follows:

- To the best of our knowledge, this work is the first to investigate the joint optimization of deployment and topology for IRIS-aided multi-user wireless communication systems. We aim to maximize the system weighted sum-rate (WSR) by jointly optimizing the BS active beamforming, IRIS deployment, IRIS passive beamforming and IRIS topology. In contrast to existing literature that primarily considers fixed IRIS deployments [16–18], our approach integrates IRIS deployment with IRIS configuration to significantly enhance overall system capacity.
- To address the formulated non-convex problem, a novel block alternating optimization (BAO) framework is proposed, which integrates successive convex approximation (SCA), neighbor extraction cross-entropy (NECE) and the proposed genetic algorithm-tabu search (GA-TS) methods. Since the formulated problem involves the joint optimization of multiple variables, the conventional alternating optimization (AO) framework, which updates only a single variable per iteration, often suffers from oscillatory update behavior, search direction perturbations and redundant adjustments. To circumvent these drawbacks, the proposed BAO framework replaces the single variable update pattern with a strategic grouping of highly coupled variables into continuous and discrete blocks. Specifically, the original problem is decomposed into two blocks: 1) the first block is dedicated to the highly coupled continuous variables, where BS active beamforming and IRIS deployment are optimized using convex programming and the SCA technique, respectively; and 2) the second block focuses on highly coupled discrete configurations, addressing IRIS passive beamforming and topology through the NECE and the proposed GA-TS algorithms, respectively.
- Numerical results demonstrate the effectiveness and superiority of the proposed scheme. Specifically, the BAO framework converges significantly faster than the conventional AO framework while achieving higher WSR performance. Furthermore, the optimized IRIS deployment exhibits a distinct double-peak characteristic, which is essential for mitigating the "multiplicative fading effect." Moreover, compared with regular RIS architectures, the IRIS architecture provides additional spatial degrees of freedom, thereby overcoming inherent capacity limitations. Finally, the proposed GA-TS algorithm tightly tracks the near-optimal performance of the exhaustive search (ES) baseline while maintaining significantly lower computational complexity, validating the effectiveness of the proposed scheme.

1.2. Organization and notations

The remainder of this paper is organized as follows. Section 2 introduces the system model and formulates the optimization problem aimed at maximizing the system WSR. Section 3 develops the proposed algorithm to solve the resulting non-convex problem. Simulation results are provided in Section 4 to demonstrate the effectiveness of the proposed algorithm. Finally, Section 5 concludes this paper.

Notations: Scalars, vectors and matrices are represented by lower-case, boldface lower-case and boldface upper-case letters, respectively. $\mathbb{C}^{N\times M}$ denotes the set of all $N \times M$ complex-valued matrices. $\mathbb{E}[\mathbf{A}]$, $\mathbf{A}^{\mathrm{T}}$ and $\mathbf{A}^{\mathrm{H}}$ represent the statistical expectation operators, transpose and Hermitian transpose of matrix $\mathbf{A}$, respectively. diag($\mathbf{a}$) represents a diagonal matrix with the entries of vector $\mathbf{a}$ on its main diagonal.

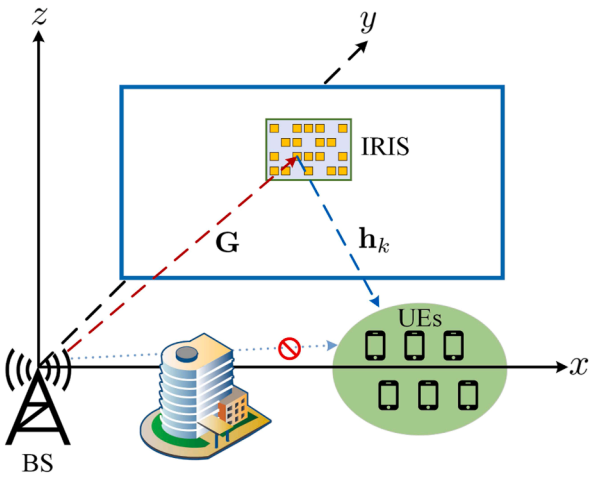

Fig. 1. The IRIS-aided multi-user communication system.

2. IRIS-Aided multi-user system

2.1. System model

The IRIS-assisted multi-user wireless communication system investigated is illustrated in Fig. 1. The system includes a BS, which is equipped with M antennas, and an IRIS, which is equipped with N reflective elements. The BS and IRIS work together to serve K single-antenna UEs, which are denoted by the set $k \in \mathcal{K} = \{1, \ldots, K\}$. The IRIS extension surface has $N_s = N_x \times N_y$ grid points in which N reflective elements are unevenly arranged, where $N_s > N$. The coordinates of BS and IRIS are given by $\mathbf{I}_{\mathrm{BS}} = (0,\ 0,\ 0)$ and $\mathbf{q} = (x_{\mathrm{R}},\ y_{\mathrm{R}},\ z_{\mathrm{R}})$, respectively. [1] The UEs are uniformly and randomly distributed in a circle with center $(x_0,\ 0,\ 0)$ and diameter c_0. The coordinate of the k-th UE is $\mathbf{I}_{\mathrm{UE},k} = (x_{\mathrm{UE},k},\ y_{\mathrm{UE},k},\ 0)$.

All channels in this paper experience quasi-static flat fading, similar to [21], and the direct link connecting the BS to the UEs is completely blocked by obstacles. The baseband equivalent channels linking the BS to the IRIS and linking the IRIS to the k-th UE can be expressed respectively as

$$\mathbf{G} = \sqrt{\rho d_{\mathrm{BR}}^{-\alpha_{\mathrm{BR}}}}\tilde{\mathbf{G}}, \tag{1}$$

$$\mathbf{h}_k = \sqrt{\rho d_{\mathrm{RU},k}^{-\alpha_{\mathrm{RU}}}}\tilde{\mathbf{h}}_k. \tag{2}$$

Here ρ denotes the channel power gain at unit distance, while α_{BR} and α_{RU} represent the corresponding path loss exponents. Furthermore, $d_{\mathrm{BR}} = \|\mathbf{q} - \mathbf{I}_{\mathrm{BS}}\|$ denotes the distance from the BS to the IRIS and $d_{\mathrm{RU},k} = \|\mathbf{q} - \mathbf{I}_{\mathrm{UE},k}\|$ is the distance from the IRIS to the k-th UE. Additionally, the

[1] Considering that practical RIS deployments are commonly integrated into building facades [19], similarly to [20], we solely optimize its horizontal deployment x_{R} along a corridor, while keeping other coordinates constant.

small-scale fading $\tilde{\mathbf{G}}$ and $\tilde{\mathbf{h}}_k$ are given respectively by

$$\tilde{\mathbf{G}}=\sqrt{\frac{\beta_{\mathrm{BR}}}{1+\beta_{\mathrm{BR}}}}\mathbf{G}_{\mathrm{BR}}^{\mathrm{LoS}}+\sqrt{\frac{1}{1+\beta_{\mathrm{BR}}}}\mathbf{G}_{\mathrm{BR}}^{\mathrm{NLoS}}, \tag{3}$$

$$\tilde{\mathbf{h}}_k=\sqrt{\frac{\beta_{\mathrm{RU}}}{1+\beta_{\mathrm{RU}}}}\mathbf{h}_k^{\mathrm{LoS}}+\sqrt{\frac{1}{1+\beta_{\mathrm{RU}}}}\mathbf{h}_k^{\mathrm{NLoS}}, \tag{4}$$

where β_{BR} and β_{RU} are the Rician factors and the non-line of sight (NLoS) channel matrix and vector, $\mathbf{G}_{\mathrm{BR}}^{\mathrm{NLoS}}$ and $\mathbf{h}_k^{\mathrm{NLoS}}$, obey the Rayleigh distribution, while the line-of-sight (LoS) components are expressed as $\mathbf{G}_{\mathrm{BR}}^{\mathrm{LoS}}=\mathbf{a}_{\mathrm{RIS}}(\hat{\varphi}_{\mathrm{R}},\tilde{\varphi}_{\mathrm{R}})\mathbf{a}_{\mathrm{BS}}^{\mathrm{H}}(\varphi_{\mathrm{B}})$ and $\mathbf{h}_k^{\mathrm{LoS}}=\mathbf{a}_{\mathrm{RIS}}(\hat{\phi}_k,\tilde{\phi}_k)$. The BS transmit array response vector, the IRIS receive array response vector and the IRIS transmit array response vector are given respectively by

$$\mathbf{a}_{\mathrm{BS}}(\varphi_{\mathrm{B}})=\left[1,e^{-j\frac{2\pi d}{\lambda_0}\cos(\varphi_{\mathrm{B}})},\ldots,e^{-j\frac{2\pi d}{\lambda_0}(M-1)\cos(\varphi_{\mathrm{B}})}\right]^{\mathrm{H}}, \tag{5}$$

$$\mathbf{a}_{\mathrm{RIS}}(\hat{\varphi}_{\mathrm{R}},\tilde{\varphi}_{\mathrm{R}})=\Big[1,\mathrm{e}^{-j\frac{2\pi d}{\lambda_0}(\sin(\hat{\varphi}_{\mathrm{R}})\sin(\tilde{\varphi}_{\mathrm{R}})+\cos(\hat{\varphi}_{\mathrm{R}})\sin(\tilde{\varphi}_{\mathrm{R}}))},\ldots,\mathrm{e}^{-j\frac{2\pi d}{\lambda_0}((N_x-1)\sin(\hat{\varphi}_{\mathrm{R}})\sin(\tilde{\varphi}_{\mathrm{R}})+(N_y-1)\cos(\hat{\varphi}_{\mathrm{R}})\sin(\tilde{\varphi}_{\mathrm{R}}))}\Big]^{\mathrm{H}}, \tag{6}$$

$$\mathbf{a}_{\mathrm{RIS}}(\hat{\phi}_k,\tilde{\phi}_k)=\Big[1,\mathrm{e}^{-j\frac{2\pi d}{\lambda_0}(\sin(\hat{\phi}_k)\sin(\tilde{\phi}_k)+\cos(\hat{\phi}_k)\sin(\tilde{\phi}_k))},\ldots,\mathrm{e}^{-j\frac{2\pi d}{\lambda_0}((N_x-1)\sin(\hat{\phi}_k)\sin(\tilde{\phi}_k)+(N_y-1)\cos(\hat{\phi}_k)\sin(\tilde{\phi}_k))}\Big]^{\mathrm{H}}, \tag{7}$$

where d is the spacing between adjacent antennas and λ_0 is the carrier wavelength, and φ_{B} is the BS's angle of departure, while $\hat{\varphi}_{\mathrm{R}}$ and $\tilde{\varphi}_{\mathrm{R}}$ are the azimuth and elevation angles of arrival from the BS to the RIS, and $\hat{\phi}_k$ and $\tilde{\phi}_k$ are the azimuth and elevation angles of departure from the RIS to the k-th UE.

The phase shift matrix is commonly used to represent the ability of an individual element on the IRIS to adjust the phase of an electromagnetic wave, which is given by

$$\mathbf{\Theta}=\mathrm{diag}\left(\beta_1 e^{j\theta_1},\cdots,\beta_n e^{j\theta_n},\cdots,\beta_{N_s}e^{j\theta_{N_s}}\right), \tag{8}$$

where the n-th IRIS element's the reflection amplitude and phase shift are denoted as $\beta_n=1$ and $\theta_n\in[0,2\pi]$, for $\forall n\in\mathcal{N}=\{1,2,\cdots,N_s\}$. In addition, the phase shift of any IRIS reflection element can only take a finite number of discrete values due to the practical hardware implementation of RIS [22,23]. Let the number of quantized bits of finite discrete phase shifts be b and set $L=2^b$. Then, the discrete phase shift values for each reflection element are defined by the set

$$\mathcal{F}=\{0,\Delta\theta,\cdots,(L-1)\Delta\theta\}, \tag{9}$$

in which $\Delta\theta=\frac{2\pi}{L}$. Define $\mathbf{z}=[z_1,z_2,\cdots,z_{N_s}]^{\mathrm{T}}$, where $z_n\in\{1,0\}$ is the binary indicator determining whether there is a reflection element deployed at the n-th grid point or not. Then the topology matrix of the IRIS can be expressed as

$$\mathbf{Z}=\mathrm{diag}(\mathbf{z}). \tag{10}$$

Let $\mathbf{w}_k\in\mathbb{C}^{M\times 1}$ be the transmitted precoding vector for the transmission symbol s_k of the k-th UE. Then the BS's complex baseband transmit signal can be written as

$$\mathbf{x}=\sum_{k=1}^{K}\mathbf{w}_k s_k, \tag{11}$$

where the symbol vector $\mathbf{s}=[s_1,s_2,\ldots,s_K]^{\mathrm{H}}\in\mathbb{C}^{K\times 1}$ is assumed to satisfy $\mathbb{E}[\mathbf{s}\mathbf{s}^{\mathrm{H}}]=\mathbf{I}$. Accordingly, the received signals at the k-th UE can be expressed as

$$y_k=\mathbf{h}_k^{\mathrm{H}}\mathbf{Z}\mathbf{\Theta}\mathbf{G}\mathbf{x}+n_k, \tag{12}$$

where $n_k\sim\mathcal{CN}(0,\sigma^2)$ denotes the additive white Gaussian noises. Therefore, the signal to interference plus noise ratio at the k-th UE is calculated as

$$\gamma_k=\frac{\left|\mathbf{h}_k^{\mathrm{H}}\mathbf{Z}\mathbf{\Theta}\mathbf{G}\mathbf{w}_k\right|^2}{\sum_{i=1,i\neq k}^{K}\left|\mathbf{h}_k^{\mathrm{H}}\mathbf{Z}\mathbf{\Theta}\mathbf{G}\mathbf{w}_i\right|^2+\sigma^2}. \tag{13}$$

2.2. System optimization formulation

The joint optimization problem of BS active beamforming, IRIS deployment, IRIS passive beamforming and IRIS topology is constructed with the objective of WSR maximization, which is formulated as

$$\mathcal{P}_1:\ \max_{\mathbf{W},\mathbf{q},\mathbf{\Theta},\mathbf{Z}} f_1(\mathbf{W},\mathbf{q},\mathbf{\Theta},\mathbf{Z})=\sum_{k=1}^{K}\omega_k\log(1+\gamma_k), \tag{14a}$$

$$\text{s.t.}\quad \mathrm{C}_1:\ P_{\mathrm{T}}\le P, \tag{14b}$$

$$\mathrm{C}_2:\ \theta_n\in\mathcal{F},\ \forall n\in\mathcal{N}, \tag{14c}$$

$$\mathrm{C}_3:\ z_n\in\{1,0\},\ \forall n\in\mathcal{N}, \tag{14d}$$

$$\mathrm{C}_4:\ \mathbf{1}^{\mathrm{T}}\mathbf{z}=N, \tag{14e}$$

$$\mathrm{C}_5:\ x_{\min}\le x_{\mathrm{R}}\le x_{\max}, \tag{14f}$$

where $\mathbf{W}=[\mathbf{w}_1,\mathbf{w}_2,\cdots,\mathbf{w}_K]\in\mathbb{C}^{M\times K}$ is the beamforming matrix, $P_{\mathrm{T}}=\sum_{k=1}^{K}\|\mathbf{w}_k\|^2$ is the total transmit power, ω_k and P denote the weight for the k-th UE and the maximum power threshold at the BS, respectively. Furthermore, C_1 restricts the total transmit power, C_2 represents the discrete phase shift constraint, C_3 and C_4 are the sparse constraints, and C_5 denotes the IRIS deployment region constraint.

3. Joint optimization algorithm

In order to effectively solve the challenging non-convex optimization problem $\mathcal{P}_1$, we propose a novel BAO algorithm framework to maximize the system WSR by jointly optimizing $\mathbf{W}$, $\mathbf{q}$, $\mathbf{\Theta}$ and $\mathbf{Z}$. The optimization problem $\mathcal{P}_2$ is divided into two blocks based on the continuous and discrete characteristics of the variables, as shown in Fig. 2.

Specifically, the original objective function f_1 is first converted into a tractable equivalent form. Then, the optimization task is decomposed into two distinct blocks, with the first block jointly optimizing the BS active beamforming $\mathbf{W}$ and IRIS deployment $\mathbf{q}$, the second block addressing the IRIS passive beamforming $\mathbf{\Theta}$ and topology $\mathbf{Z}$.

3.1. Transformation of the optimization problem

Owing to the high coupling of the optimization variables $\mathbf{W}$, $\mathbf{q}$, $\mathbf{\Theta}$ and $\mathbf{Z}$, along with the non-convexity of the constraints, the complex non-convex optimization problem $\mathcal{P}_1$ is difficult to solve. To tackle this challenge, the original problem is converted into a multiple-ratio fractional programming optimization task [24]. Specifically, by leveraging the quadratic transformation and Lagrangian dual transformation while introducing the non-negative auxiliary variable α_k, the objective function (14a) can be equivalently transformed into

$$f_2(\boldsymbol{\alpha},\mathbf{W},\mathbf{q},\mathbf{\Theta},\mathbf{Z})=\sum_{k=1}^{K}\left(\omega_k\log(1+\alpha_k)-\omega_k\alpha_k+\frac{\tilde{\alpha}_k\gamma_k}{1+\gamma_k}\right), \tag{15}$$

where $\boldsymbol{\alpha}=[\alpha_1,\alpha_2,\cdots,\alpha_K]^{\mathrm{T}}$, $\tilde{\alpha}_k=\omega_k(1+\alpha_k)$, and the optimal $\alpha_k=\gamma_k$ is achieved by setting $\frac{\partial f_2}{\partial\alpha_k}=0$. For a given $\boldsymbol{\alpha}$, the optimization problem $\mathcal{P}_1$ can be reformulated as

$$\mathcal{P}_2:\quad \max_{\mathbf{W},\mathbf{q},\mathbf{\Theta},\mathbf{Z}} f_3(\mathbf{W},\mathbf{q},\mathbf{\Theta},\mathbf{Z})=\sum_{k=1}^{K}\frac{\tilde{\alpha}_k\gamma_k}{1+\gamma_k}, \tag{16a}$$

$$\text{s.t.}\quad (14\text{b})-(14\text{f}). \tag{16b}$$

Despite the reformulation, obtaining the global optimum of $\mathcal{P}_2$ directly remains mathematically intractable due to the persistent non-convex constraints and the intricate coupling among the variables $\mathbf{W},\mathbf{q},\mathbf{\Theta}$ and $\mathbf{Z}$. To circumvent this obstacle, a BAO framework is developed to decouple the optimization variables into two tractable subproblems. The detailed procedures for solving these two subproblems are elaborated in the following subsections.

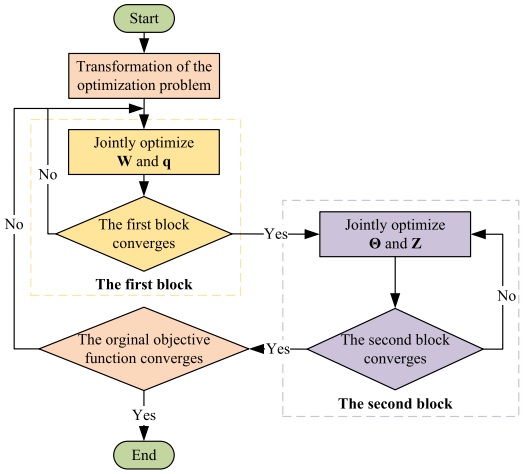

Fig. 2. The BAO algorithm framework.

3.2. Block 1: Joint optimization of BS active beamforming and IRIS deployment

Given the optimization variables $\mathbf{\Theta}$ and $\mathbf{Z}$, $\mathcal{P}_2$ can be transformed into

$$\mathcal{P}_{3-1}: \max_{\mathbf{W},\,\mathbf{q}} \sum_{k=1}^{K} \frac{\tilde{\alpha}_k \gamma_k}{1+\gamma_k}, \tag{17a}$$

$$\text{s.t.} \quad (14b), (14f). \tag{17b}$$

To solve the non-convex subproblem $\mathcal{P}_{3-1}$, an inner iterative procedure is performed within Block 1. Specifically, for a fixed IRIS deployment $\mathbf{q}$, the BS active beamforming $\mathbf{W}$ is first optimized. Subsequently, the deployment $\mathbf{q}$ is updated based on the previously optimized beamforming $\mathbf{W}$. This internal process iterates until the first block converges, as detailed in the following steps.

3.2.1. Optimization of BS active beamforming

Given $\mathbf{q}$, by utilizing the quadratic transformation algorithm while introducing the auxiliary variable $\boldsymbol{\tau} = [\tau_1, \tau_2, \cdots, \tau_K]^{\mathrm{T}}$, $\mathcal{P}_{3-1}$ is equivalently transformed into

$$\mathcal{P}_{3-2}: \max_{\boldsymbol{\tau},\mathbf{W}} f_4(\boldsymbol{\tau}, \mathbf{W}), \tag{18a}$$

$$\text{s.t.} \quad (14b), \tag{18b}$$

where $f_4(\boldsymbol{\tau}, \mathbf{W}) = \sum_{k=1}^{K} \left(2\sqrt{\tilde{\alpha}_k}(\tau_k^{\dagger}\mathbf{f}_k\mathbf{w}_k) - |\tau_k|^2\left(\sum_{i=1}^{K}|\mathbf{f}_k\mathbf{w}_i|^2 + \sigma^2\right)\right)$, $\mathbf{f}_k = \mathbf{h}_k^{\mathrm{H}}\mathbf{Z}\mathbf{\Theta}\mathbf{G}$. Given $\mathbf{W}$, the optimal τ_k can be obtained by setting $\frac{\partial f_4}{\partial \tau_k} = 0$, which can be expressed as

$$\tau_k^* = \frac{\sqrt{\tilde{\alpha}_k}\mathbf{f}_k\mathbf{w}_k}{\sum_{i=1}^{K}|\mathbf{f}_k\mathbf{w}_i|^2 + \sigma^2}. \tag{19}$$

Given τ_k^*, $\mathcal{P}_{3-2}$ is a convex optimization problem. By using the Lagrange multiplier method, the optimal $\mathbf{w}_k$ can be achieved as

$$\mathbf{w}_k^* = \sqrt{\tilde{\alpha}_k}\tau_k\left(\lambda\mathbf{I}_M + \sum_{i=1}^{K}|\tau_i|^2\mathbf{f}_i^{\mathrm{H}}\mathbf{f}_i\right)^{-1}\mathbf{f}_k^{\mathrm{H}}, \tag{20}$$

where λ denotes the Lagrange multiplier, which can be optimized via the bisection search method.

3.2.2. Optimization of IRIS deployment

Given $\mathbf{W}$, $\mathcal{P}_{3-1}$ is nonconvex since $\mathbf{q}$ is coupled in both $d_{\mathrm{RU},k}$ and d_{BR}. To address this challenge, the first-order Taylor expansion method is employed to determine the optimal $\mathbf{q}$ using the SCA algorithm. The details are as follows.

Given $\mathbf{W}$, the objective function of $\mathcal{P}_{3-1}$ can be written as

$$f_5(\mathbf{q}) = \sum_{k=1}^{K} \frac{\tilde{\alpha}_k \mathbf{l}_k}{\mathbf{v}_k + \sigma^2 d_{\mathrm{RU},k}^{\alpha_{\mathrm{RU}}} d_{\mathrm{BR}}^{\alpha_{\mathrm{BR}}}}, \tag{21}$$

where $\mathbf{l}_k = \rho^2|\tilde{\mathbf{f}}_k\mathbf{w}_k|^2$, $\mathbf{v}_k = \rho^2\sum_{i=1}^{K}|\tilde{\mathbf{f}}_k\mathbf{w}_i|^2$, and $\tilde{\mathbf{f}}_k = \tilde{\mathbf{h}}_k\mathbf{Z}\mathbf{\Theta}\tilde{\mathbf{G}}$. Then $\mathcal{P}_{3-1}$ can be re-expressed as

$$\mathcal{P}_{4-1}: \max_{\mathbf{q}} \sum_{k=1}^{K} \frac{\tilde{\alpha}_k \mathbf{l}_k}{\mathbf{v}_k + \sigma^2 d_{\mathrm{RU},k}^{\alpha_{\mathrm{RU}}} d_{\mathrm{BR}}^{\alpha_{\mathrm{BR}}}}, \tag{22a}$$

$$\text{s.t.} \quad (14f). \tag{22b}$$

Introducing the slack variables u_k and μ, $\mathcal{P}_{4-1}$ can be transformed into

$$\mathcal{P}_{4-2}: \max_{\mathbf{q},u_k,\mu} f_6(\mathbf{q}, u_k, \mu) = \sum_{k=1}^{K} \frac{\tilde{\alpha}_k \mathbf{l}_k}{\mathbf{v}_k + \sigma^2 u_k^{\frac{\alpha_{\mathrm{RU}}}{2}} \mu^{\frac{\alpha_{\mathrm{BR}}}{2}}}, \tag{23a}$$

$$\text{s.t.} \quad u_k \geq d_{\mathrm{RU},k}^2, \tag{23b}$$

$$\mu \geq d_{\mathrm{BR}}^2, \tag{23c}$$

$$(14f). \tag{23d}$$

Using the SCA algorithm, the first-order Taylor expansion of f_6 at the fixed points $u_{k,0}$ and μ_0 is

$$f_6 \geq \sum_{k=1}^{K} A_{k,0} + B_{k,0}(u_k - u_{k,0}) + C_{k,0}(\mu - \mu_0), \tag{24}$$

where

$$A_{k,0} = \frac{\tilde{\alpha}_{k,0}\mathbf{l}_{k,0}}{\mathbf{v}_{k,0} + \sigma^2 u_{k,0}^{\frac{\alpha_{\mathrm{RU}}}{2}} \mu_0^{\frac{\alpha_{\mathrm{BR}}}{2}}}, \tag{25}$$

$$B_{k,0} = -\frac{\alpha_{\mathrm{RU}}\sigma^2\tilde{\alpha}_{k,0}\mathbf{l}_{k,0}u_{k,0}^{\frac{\alpha_{\mathrm{RU}}-2}{2}}\mu_0^{\frac{\alpha_{\mathrm{BR}}}{2}}}{2\left(\mathbf{v}_{k,0} + \sigma^2 u_{k,0}^{\frac{\alpha_{\mathrm{RU}}}{2}}\mu_0^{\frac{\alpha_{\mathrm{BR}}}{2}}\right)^2}, \tag{26}$$

$$C_{k,0} = -\frac{\alpha_{\mathrm{BR}}\sigma^2\tilde{\alpha}_{k,0}\mathbf{l}_{k,0}u_{k,0}^{\frac{\alpha_{\mathrm{RU}}}{2}}\mu_0^{\frac{\alpha_{\mathrm{BR}}-2}{2}}}{2\left(\mathbf{v}_{k,0} + \sigma^2 u_{k,0}^{\frac{\alpha_{\mathrm{RU}}}{2}}\mu_0^{\frac{\alpha_{\mathrm{BR}}}{2}}\right)^2}. \tag{27}$$

Based on the aforementioned derivations, $\mathcal{P}_{4-2}$ can be transformed into

$$\mathcal{P}_{4-3}: \max_{\mathbf{q},\mathbf{u},\mu} f_7(\mathbf{q}, \mathbf{u}, \mu) = \sum_{k=1}^{K} B_{k,0}u_k + C_{k,0}\mu, \tag{28a}$$

$$\text{s.t.} \quad (14f), (23b), (23c), \tag{28b}$$

where $\mathbf{u} = [u_1, u_2, \ldots, u_K]^{\mathrm{T}}$. Problem $\mathcal{P}_{4-3}$ is a convex problem, which can be solved by standard convex optimization tools such as CVX.

3.3. Block 2: Joint optimization of IRIS passive beamforming and IRIS topology

Given $\mathbf{W}$ and $\mathbf{q}$, $\mathcal{P}_2$ can be transformed into

$$\mathcal{P}_5: \max_{\mathbf{\Theta},\mathbf{Z}} f_8(\mathbf{\Theta}, \mathbf{Z}) = \sum_{k=1}^{K} \frac{\tilde{\alpha}_k \gamma_k}{1+\gamma_k}, \tag{29a}$$

$$\text{s.t.} \quad (14c), (14d), (14e). \tag{29b}$$

To solve the non-convex subproblem $\mathcal{P}_5$, an inner iterative procedure is performed within Block 2. Specifically, for a fixed IRIS topology $\mathbf{Z}$, the IRIS passive beamforming $\mathbf{\Theta}$ is first optimized using the NECE algorithm. Subsequently, the IRIS topology $\mathbf{Z}$ is updated based on the previously optimized passive beamforming $\mathbf{\Theta}$ by employing the proposed GA-TS algorithm. This internal process iterates until the second block converges, as detailed in the following steps.

3.3.1. Optimization of IRIS passive beamforming

For given $\mathbf{Z}$, $\mathcal{P}_5$ can be rewritten as

$$\mathcal{P}_{5-1}: \max_{\mathbf{\Theta},\mathbf{Z}} f_8(\mathbf{\Theta}) = \sum_{k=1}^{K} \frac{\tilde{\alpha}_k \gamma_k}{1+\gamma_k}, \tag{30a}$$

$$\text{s.t.} \quad (14c). \tag{30b}$$

It is observed that $\mathcal{P}_{5-1}$ is intractable owing to the finite discrete phase shift constraint. Moreover, traditional methods relying on continuous relaxation followed by simple quantization operations often lead to severe performance degradation. To address this challenge, we leverage the NECE algorithm [16] to solve $\mathcal{P}_{5-1}$. The specific procedures of the NECE algorithm are detailed as follows.

Let $\mathbf{P}=[\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{N_s}]$ denote the probability matrix of $\mathbf{\Theta}$, where $\mathbf{p}_n \in \mathbb{R}^{2^b \times 1}$ is the probability vector for θ_n satisfying $||\mathbf{p}_n||_1 = 1$. Moreover, each component $p_{n,l}$ of $\mathbf{p}_n$ represents the probability that θ_n takes a specific value from the feasible set $\mathcal{F}$, where $l \in \{1, 2, \cdots, 2^b\}$. Specifically, let $\mathcal{F}(l)$ be the l-th element of the set $\mathcal{F}$. To mathematically formulate this probabilistic selection process, we introduce an indicator function, which can be expressed as

$$\delta(t) = \begin{cases} 1, & t = 0, \\ 0, & t \neq 0. \end{cases} \tag{31}$$

Initially, the probability matrix is set to represent a uniform distribution over all feasible phase shifts, which is given by

$$\mathbf{P}^{(1)} = \frac{1}{2^b} \times \mathbf{1}_{2^b \times N_s}. \tag{32}$$

At the i-th iteration, we randomly generate C candidates $\{\mathbf{\Theta}^c\}_{c=1}^{C}$ based on the probability distribution function (PDF) $\Xi(\mathbf{\Theta}; \mathbf{P}^{(i)})$, which is formulated as

$$\Xi(\mathbf{\Theta}; \mathbf{P}^{(i)}) = \prod_{n=1}^{N_s} \left(\prod_{l=1}^{2^b} (p_{n,l}^{(i)})^{\delta(\theta_n - \mathcal{F}(l))} \right), \tag{33}$$

where $\mathbf{\Theta}^i$ denotes the generated i-th candidate. Once these candidates are obtained, the corresponding WSR $\{R(\mathbf{\Theta}^c)\}_{c=1}^{C}$ for each candidate is calculated and then sorted in descending order, which can be expressed as

$$R(\mathbf{\Theta}^{[1]}) \geq R(\mathbf{\Theta}^{[2]}) \geq \cdots \geq R(\mathbf{\Theta}^{[C]}). \tag{34}$$

Based on (34), the top C_{pr} candidates are selected as the primary elites, denoted by the set $\{\mathbf{\Theta}^{[1]}, \mathbf{\Theta}^{[2]}, \ldots, \mathbf{\Theta}^{[C_{\text{pr}}]}\}$.

Then, to broaden the search space, a neighbor extraction technique is developed, which operates by changing the phase shift associated with each diagonal entry in $\mathbf{\Theta}^{[1]}$. It is worth noting that only N diagonal elements in $\mathbf{\Theta}^{[1]}$ are effective, whose phase shifts are selected from the feasible set $\mathcal{F}$. Thus, we can acquire $N(2^b - 1)$ additional candidate solutions and subsequently compute their system WSRs.

Let $C_{\text{elite}} = C_{\text{pr}} + C_{\text{sup}}$ denote the updated total number of elites, incorporating the C_{sup} supplementary candidates. To quantify the contribution of each elite, we define a weighting factor η_c. For the c-th elite, this weight is evaluated by normalizing its individual WSR against the average WSR of the entire elite pool, which is formulated as

$$\eta_c = \frac{R(\mathbf{\Theta}^{[c]}) C_{\text{elite}}}{\sum_{c=1}^{C_{\text{elite}}} R(\mathbf{\Theta}^{[c]})}. \tag{35}$$

Consequently, candidates exhibiting superior WSR performance are assigned proportionally higher weights.

With the elite weights established, the probability matrix for the subsequent iteration, denoted as $\mathbf{P}^{(i+1)}$, is updated based on the probability transfer criterion. This process is mathematically expressed as

$$\mathbf{P}^{(i+1)} = \arg\max_{\mathbf{P}^{(i)}} \frac{1}{C_{\text{elite}}} \sum_{c=1}^{C_{\text{elite}}} \eta_c \ln \Xi\left(\mathbf{\Theta}^{[c]}; \mathbf{P}^{(i)}\right). \tag{36}$$

The refined matrix $\mathbf{P}^{(i+1)}$ is subsequently employed in the next iteration. This iterative process continues until the predefined maximum number of iterations I_{N} is reached, ultimately yielding the near-optimal passive beamforming solution. For clarity, the overall procedure of the NECE method is summarized in Algorithm 1.

Algorithm 1: NECE algorithm for IRIS passive beamforming.

Inputs: $\mathbf{Z}$, $\mathcal{F}$, the number of iterations I_{N}, the number of candidates C and the number of primary elites C_{pr}.

Outputs: $\hat{\mathbf{\Theta}}_{\text{opt}}$.

Initialize the probability matrix $\mathbf{P}^{(1)}$.
for $i = 1 : I_{\text{N}}$ **do**
  Randomly generate C candidates $\{\mathbf{\Theta}^c\}_{c=1}^{C}$ based on the PDF $\Xi(\mathbf{\Theta}; \mathbf{P}^{(i)})$.
  Calculate the WSR $\{R(\mathbf{\Theta}^c)\}_{c=1}^{C}$.
  Sort $\{R(\mathbf{\Theta}^c)\}_{c=1}^{C}$ in descending order.
  Select C_{pr} best candidates as primary elites.
  Obtain $N(2^b - 1)$ extra candidates through neighbor extraction based on $\mathbf{\Theta}^{[1]}$.
  Select C_{sup} candidates with better performance than $\mathbf{\Theta}^{[1]}$ as supplementary elites.
  Update $C_{\text{elite}} = C_{\text{pr}} + C_{\text{sup}}$ and calculate the weight η_c of each elite $\mathbf{\Theta}^{[c]}$, $c = 1, 2, \ldots, C_{\text{elite}}$.
  Update $\mathbf{P}^{(i+1)}$ based on $\{\mathbf{\Theta}^{[c]}\}_{c=1}^{C_{\text{elite}}}$ and $\{\eta_c\}_{c=1}^{C_{\text{elite}}}$.
end for

3.3.2. Optimization of IRIS topology

For given $\mathbf{\Theta}$, $\mathcal{P}_5$ can be recast as

$$\mathcal{P}_{6-1}: \max_{\mathbf{Z}} f_{12}(\mathbf{Z}) = \sum_{k=1}^{K} \frac{\tilde{\alpha}_k \gamma_k}{1+\gamma_k}, \tag{37a}$$

$$\text{s.t.} \quad (14d), (14e). \tag{37b}$$

Since the feasible set of the topology matrix $\mathbf{Z}$ for the IRIS is discrete and finite, the optimal topology $\mathbf{Z}_{\text{opt}}$ can be obtained via the ES method when the number of RIS elements is small. Specifically, all $\binom{N_s}{N}$ possible topology configurations are traversed. Subsequently, for each candidate topology, the phase shift of each IRIS element is continuously selected to generate the IRIS passive beamforming $\mathbf{\Theta}$ from the feasible set of 2^{bN} configurations. This approach can serve as a benchmark for other potential solutions due to its global optimality. However, when the number of RIS elements is large, the complexity of the ES method becomes computationally prohibitive.

To reduce the complexity, a low-complexity topology design based on the GA-TS algorithm is applied to optimize the topology matrix. While the GA excels in global search, it falls short in local search efficiency. Conversely, the TS algorithm is highly effective in local search but relies heavily on the initial solution [25]. Therefore, the GA-TS algorithm combines the advantages of both methods to boost the overall performance. The proposed GA-TS algorithm is detailed in Algorithm 2, as shown at the top of the next page. Its explanations are detailed as follows.

In step 1, an initial population of size O is generated by randomly arranging N ones and $N_s - N$ zeros. Given the population O, the fitness value f_8 of each individual is evaluated in Step 3. To preserve high-quality individuals for subsequent generations, the roulette-wheel selection method is adopted in step 4. The step 5 involves the generation of new offspring via single-point crossover. In step 6, the mutation is performed by randomly selecting two genes with values of 1 and 0, and

Algorithm 2: Proposed GA-TS algorithm for IRIS topology.

Inputs: Number of GA iterations I_{G}, individual selection rate $p_{\mathrm{c}1}$, crossover rate $p_{\mathrm{c}2}$, mutation rate $p_{\mathrm{c}3}$, TS ratio $p_{\mathrm{c}4}$, number of individuals in population O, number of TS iterations I_{T}, neighbor distance $\tilde{r}$, tabu list $\mathcal{H} = \emptyset$, neighborhood size Q, storage size H_{size}.

Output: IRIS topology $\hat{\mathbf{Z}}_{\mathrm{opt}}$.

1: Each individual represents a possible IRIS topology, and each individual consists of a randomized arrangement of N ones and $N_s - N$ zeros.
2: **for** $i = 1 : I_{\mathrm{G}}$ **do**
3: Compute fitness value f_8 of each individual.
4: Individuals are selected for reproduction with probability $p_{\mathrm{c}1}$ from i-th population based on their fitness values.
5: Perform crossover operation with probability $p_{\mathrm{c}2}$ to generate new offspring individuals.
6: Perform variation operation with probability $p_{\mathrm{c}3}$ to introduce population variation.
7: **end for**
8: Calculate fitness value of newly generated offspring and rank them.
9: Top $p_{\mathrm{c}4}O$ elite solutions $\{\tilde{\mathbf{Z}}_{ii}\}_{ii=1}^{p_{\mathrm{c}4}O}$ (with corresponding fitness values $\{\tilde{f}_{8,ii}\}_{ii=1}^{p_{\mathrm{c}4}O}$) are selected for TS.
10: **for** $ii = 1 : p_{\mathrm{c}4}O$ **do**
11: Take $\tilde{\mathbf{Z}}_{ii}$ as initial solution $\mathbf{Z}_1$ of TS and save $\mathbf{Z}_{\mathrm{opt}} = \tilde{\mathbf{Z}}_{ii}$ and $f_{8,\mathrm{opt}} = \tilde{f}_{8,ii}$.
12: **for** $j = 1 : I_{\mathrm{T}}$ **do**
13: Adjust $\tilde{r}$ based on value of j. Randomly exchange $\tilde{r}$ ones with $\tilde{r}$ zeros in diagonal elements of $\mathbf{Z}_j$ to create candidates of $\mathbf{Z}_j$.
14: Eliminate candidates that are already in $\mathcal{H}$ from candidates to obtain Q valid candidates $\{\mathbf{Z}_q\}_{q=1}^{Q}$.
15: Evaluate fitness value f_8 of each candidate $\mathbf{Z}_q$. Choose top candidate with highest f_8 as new topology $\mathbf{Z}_{j+1}$. Update sub-optimal solution $\mathbf{Z}_{\mathrm{opt}}$ and its corresponding fitness value $f_{8,\mathrm{opt}}$. Store $\mathbf{Z}_{ii} = \mathbf{Z}_{\mathrm{opt}}$ and $\tilde{f}_{8,ii} = f_{8,\mathrm{opt}}$.
16: **if** $|\mathcal{H}| < H_{\mathrm{size}}$ **then**
17: Add $\mathbf{Z}_j$ into tabu list $\mathcal{H}$.
18: **else**
19: Append $\mathbf{Z}_j$ to end of tabu list $\mathcal{H}$ and eliminate front elements.
20: **end if**
21: **end for**
22: Select individual with the highest fitness value from $\{\tilde{\mathrm{f}}_{8,ii}\}_{ii=1}^{p_{\mathrm{c}4}O}$ as sub-optimal solution $\hat{\mathbf{Z}}_{\mathrm{opt}}$.
23: **end for**

then flipping them to 0 and 1, respectively. The GA stage is terminated after the predefined maximum number of GA iterations I_{G} is reached.

Following the GA stage, the top $p_{\mathrm{c}4}O$ elite individuals are selected and passed to the TS stage in steps 8–9. In step 13, neighborhood solutions are defined as newly generated topology matrices, where $\tilde{r}$ denotes the neighborhood distance that is adaptively adjusted according to the iteration index j. Specifically, a relatively large $\tilde{r}$ is adopted in the early iterations to expand the search region, while a smaller $\tilde{r}$ is utilized in later iterations for fine local refinement. In step 14, neighbor solutions contained in the tabu list are removed from the candidate set to avoid redundant searches, ensuring that Q admissible candidates are retained. The candidate topology with the maximum WSR is then selected as the new topology $\mathbf{Z}_{j+1}$, and the suboptimal solution $\mathbf{Z}_{\mathrm{opt}}$, along with its corresponding fitness value $f_{8,\mathrm{opt}}$, is updated accordingly. For each selected elite solution, the TS procedure is performed until the predefined maximum number of TS iterations I_{T} is reached. Finally, after the tabu list is updated in steps 16–20, and all the selected $p_{\mathrm{c}4}O$ elite solutions have been processed by the TS procedure, the individual with the highest fitness value among those processed by the TS procedure is selected as the final suboptimal topology matrix $\hat{\mathbf{Z}}_{\mathrm{opt}}$ in Step 22.

3.4. Convergence analysis

The original non-convex problem is decomposed into two block-wise subproblems, which are alternately optimized until the prescribed stopping criterion is satisfied.

For the continuous-variable block, with fixed IRIS passive beamforming $\boldsymbol{\Theta}$ and IRIS topology $\mathbf{Z}$, the BS active beamforming $\mathbf{W}$ and IRIS deployment $\mathbf{q}$ are updated successively. The BS active beamforming subproblem (18) can be optimally solved after the quadratic transformation, while the IRIS deployment subproblem (28) is handled by the SCA method. Within Block 1, given a feasible initial point, the algorithm maintains feasibility across iterations, and each update of the continuous-variable block does not decrease the system WSR because each subproblem is a maximization problem and is solved via a monotonic optimization procedure.

For the discrete-variable block, with fixed $\mathbf{W}$ and $\mathbf{q}$, the IRIS passive beamforming $\boldsymbol{\Theta}$ and topology $\mathbf{Z}$ are optimized by the NECE and GA-TS algorithms, respectively. Different from the continuous-variable block, the discrete-variable block involves finite feasible sets, since the phase shifts are selected from a finite quantization set and the topology vector is constrained by binary variables with a fixed number of active reflecting elements. Moreover, an elitist preservation mechanism is adopted in both the NECE and GA-TS algorithms, where the best-so-far solution is retained during the iterative search. Therefore, the discrete-variable update also yields a non-decreasing best-so-far WSR sequence.

Based on the above analysis, the WSR after the tth outer iteration of the BAO framework satisfies

$$R(\mathbf{W}^{t+1}, \mathbf{q}^{t+1}, \boldsymbol{\Theta}^{t+1}, \mathbf{Z}^{t+1}) \geq R(\mathbf{W}^{t+1}, \mathbf{q}^{t+1}, \boldsymbol{\Theta}^{t}, \mathbf{Z}^{t}), \tag{38a}$$

$$\geq R(\mathbf{W}^{t}, \mathbf{q}^{t}, \boldsymbol{\Theta}^{t}, \mathbf{Z}^{t}). \tag{38b}$$

Consequently, we can obtain that the system objective function is monotonically non-decreasing. Furthermore, since the transmit power is constrained, the deployment region is bounded and the feasible sets for both phase shifts and topology are finite, the convergence of the overall BAO framework is guaranteed.

3.5. Complexity analysis

The computational complexity of the proposed algorithm primarily arises from solving the subproblems within the two distinct blocks of the BAO framework.

Specifically, the first block involves the joint optimization of BS active beamforming and IRIS deployment. The optimization of BS active beamforming is accomplished by the conic relaxation approach, with a complexity of approximately $\mathcal{O}\big(L_1(M^3 + KM^2)\big)$, where L_1 is the number of iterations. The optimization of IRIS deployment is achieved by the SCA method. The variables and constraints involved determine the complexity of this process, which is given by $\mathcal{O}\big(L_2(K+2)^{3.5}\big)$, with L_2 denoting the number of iterations. Consequently, the total complexity of the first block is expressed as $\mathcal{O}\big(I_0(L_1(M^3 + KM^2) + L_2(K+2)^{3.5})\big)$, with I_0 denoting the number of inner alternating iterations.

The second block addresses the joint optimization of IRIS passive beamforming and topology. The complexity of the passive beamform-

Table 1
Interpretation of different schemes in the simulations.

Optimization variables \ Schemes	Proposed scheme (IRIS)	Scheme 1 (IRIS)	Scheme 2 (IRIS)	Scheme 3 (IRIS)	Scheme 4 (Regular RIS)
W	Convex optimization	Convex optimization	Convex optimization	Convex optimization	Convex optimization
q	SCA	SCA	SCA	random	SCA
Θ	NECE	NECE	NECE	NECE	NECE
Z	GA-TS	TS	ES	GA-TS	—

ing design is $\mathcal{O}\big(I_{\mathrm{N}}\big(C+N(2^b-1)\big)\big)$, and the IRIS topology optimization entails a complexity of $\mathcal{O}\big(I_{\mathrm{G}}O+p_{\mathrm{c4}}OI_{\mathrm{T}}Q\big)$. Here, I_{N}, I_{G} and I_{T} signify the iterations required for the convergence of the NECE, GA and TS algorithms, respectively. Thus, the total complexity of the second block is $\mathcal{O}\big(I_{00}\big(I_{\mathrm{N}}\big(C+N(2^b-1)\big)\big)\big(I_{\mathrm{G}}O+p_{\mathrm{c4}}OI_{\mathrm{T}}Q\big)\big)$, where I_{00} denotes the number of inter-block alternating iterations. This complexity is significantly lower than that of the traditional ES method, which entails a prohibitive complexity of $\mathcal{O}\Big(\binom{N_s}{N}2^{bN}\Big)$.

4. Simulation study

This section provides comprehensive simulation results to evaluate the effectiveness of our proposed algorithm. To verify its superiority, we conduct a comparative analysis against four benchmark schemes:

- *Scheme 1:* In this scheme, the IRIS topology is optimized leveraging the TS algorithm to evaluate the contribution of the GA stage utilized in the proposed scheme. This baseline aims to reveal the TS algorithm's heavy reliance on initialization and its comparatively limited global search capability.
- *Scheme 2:* This benchmark employs the ES method for IRIS topology optimization while keeping all other variables optimized identically to the proposed scheme. By traversing all possible topology configurations to find the global optimum, this scheme provides a performance upper bound for the IRIS setup.
- *Scheme 3:* In this baseline, the deployment position of the IRIS is generated randomly to serve as a comparative benchmark. Consequently, this scheme is specifically designed to quantify the performance gains yielded by strategic physical placement.
- *Scheme 4:* This benchmark utilizes a regular RIS to facilitate auxiliary communication links for the UEs. As a result, topology optimization is not involved in this configuration, with all other variables optimized exactly as in the proposed scheme.

For clarity, the explanations of different schemes in the simulations are summarized in Table 1.

4.1. Simulation setup

The parameters of the simulation system are set as follows. The BS is located at $(0,0,0)$ m. The coordinate parameters of IRIS are set as $y_{\mathrm{R}}=30$ m and $z_{\mathrm{R}}=20$ m, while its deployment along the x-axis is constrained within $0\le x_{\mathrm{R}}\le 100$ m. The UEs are uniformly and randomly distributed in a circular area centered at $(x_0,0,0)$ with $x_0=100$ m and the diameter is $c_0=10$ m. The numbers of BS antennas, UEs, RIS elements, and IRIS grid points are $M=4$, $K=4$, $N=8$, and $N_s=16$, respectively. The number of quantization bits $b=1$[26], the noise power $\sigma^2=-100$ dBm, the BS transmit power $P_{\mathrm{T}}=20$ dBm, the UE weight ω_k is set to be 1, $\forall k\in\mathcal{K}$. The parameters of the NECE algorithm are set to $I_{\mathrm{N}}=15$, $C=200$, $C_{\mathrm{pr}}=40$. The iteration parameters of the proposed GA-TS algorithm are set to $O=50$, $p_{\mathrm{c1}}=0.8$, $p_{\mathrm{c2}}=0.3$, $p_{\mathrm{c3}}=0.04$, $p_{\mathrm{c4}}=0.1$, $Q=15$, $I_{\mathrm{G}}=15$ and $I_{\mathrm{T}}=20$. The neighbor distance $\tilde{r}$ changes dynamically, starting with $\tilde{r}=3$ and reducing $\tilde{r}$ to 2 once the iteration arrives at the midpoint of the maximum number of iterations.

To investigate the performance upper bound of the proposed BAO-based joint optimization framework, we assume the channel state information (CSI) of the UEs is available at the BS. The case of imperfect CSI is beyond this scope and is reserved for future studies. The channel power gain per unit distance $\rho=-30$ dB, the path loss exponents $\alpha_{\mathrm{BR}}=\alpha_{\mathrm{RU}}=2$, the Rician factors $\beta_{\mathrm{BR}}=\beta_{\mathrm{RU}}=1$. The iterative algorithm is considered to have converged and terminates when the fractional change in the WSR between two consecutive iterations falls below the convergence tolerance $\epsilon=10^{-4}$. Furthermore, all simulation results are averaged over 100 independent channel realizations.

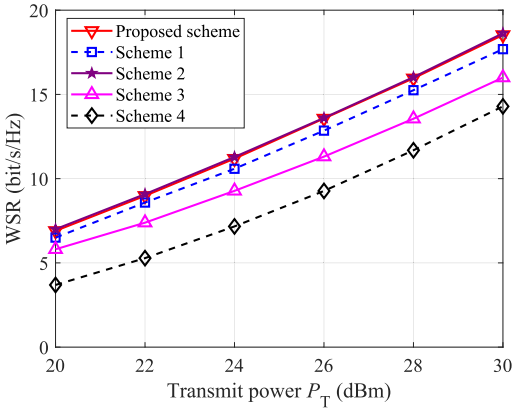

Fig. 3. WSR versus transmit power.

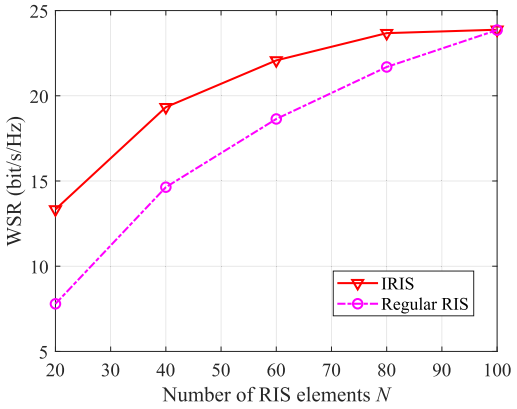

Fig. 4. WSR versus the number of RIS elements.

4.2. Performance analysis

Fig. 3 depicts the WSRs as the functions of the transmit power attained by various schemes. Three key phenomena can be observed. First, compared with Scheme 4, the proposed scheme offers significant performance gain, primarily because IRIS introduces additional spatial degrees of freedom and optimizes signal propagation paths, effectively mitigating the "multiplicative fading effect". Second, the performance of the proposed scheme is infinitely close to that of Scheme 2. However, as illustrated in Section 3.5, the proposed scheme's algorithmic complexity is significantly lower than that of Scheme 2, especially as N and N_s increase. Third, the proposed scheme outperforms Scheme 1 due to the GA-TS algorithm's incorporation of the GA to obtain a better initial

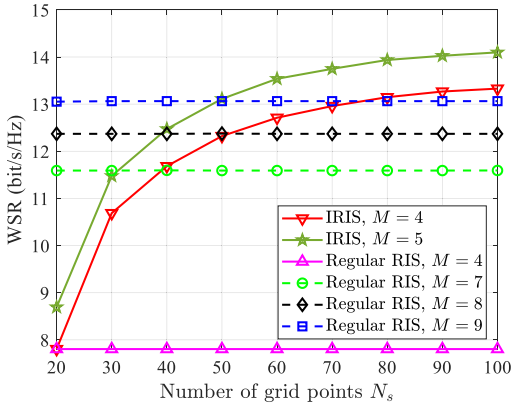

Fig. 5. WSR versus the number of IRIS grid points.

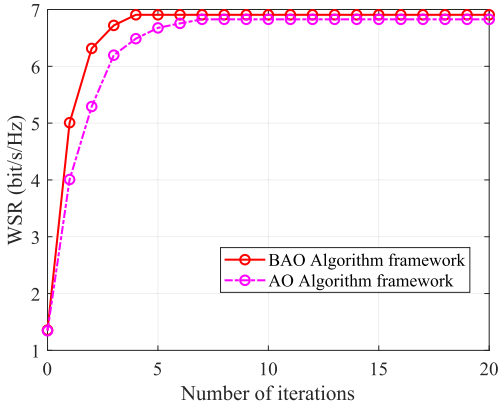

Fig. 7. Convergence behaviors of two algorithm frameworks.

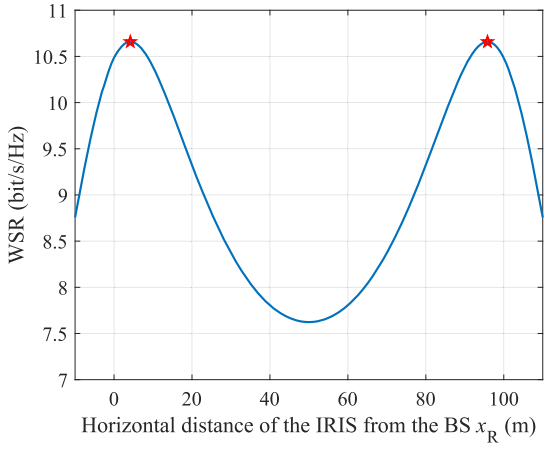

Fig. 6. WSR versus the IRIS horizontal distance from the BS.

solution, which enhances the TS algorithm's global search capability and ultimately achieves a better topology. These observations collectively validate the effectiveness of the proposed scheme, as well as the necessity of jointly optimizing the physical deployment and irregular topology.

The relationship between the WSR and the number of RIS elements is shown in Fig. 4, where $N_s = 100$. Compared to the regular RIS scheme, the IRIS scheme achieves higher WSR. However, as the irregularity ratio increases, the WSR performance gap between the IRIS scheme and the regular RIS scheme narrows. At very high irregularity ratios, the performance gain from increasing the RIS elements becomes negligible. Thus, it is crucial to balance the cost of the irregularity ratio against the performance improvement it provides.

As depicted in Fig. 5, the relationship between the WSR and the number of IRIS grid points is illustrated for $N = 20$. It can be observed that as the number of BS antennas M increases, the system WSRs for both the IRIS and regular RIS schemes exhibit a steady increase. This performance enhancement is primarily attributed to the additional spatial degrees of freedom introduced by the larger antenna array, which facilitate higher array gains and more precise active beamforming capabilities. Furthermore, taking the regular RIS scheme with a fixed surface size as a benchmark, it is evident that enlarging the surface area to modify the irregular ratio of RIS elements to grid points can significantly enhance the WSR. More remarkably, when the number of grid points is increased to $N_s = 40$, $N_s = 60$, $N_s = 80$, IRIS with $M = 4$ demonstrates superior performance compared to the regular RIS configurations with $M = 7$, $M = 8$, $M = 9$, respectively. A similar trend is observed for the IRIS with $M = 5$, which also surpasses regular RIS setups equipped with more antennas by moderately scaling up the number of grid points. This is particularly notable as these regular RIS setups involve a higher number of BS antennas. Therefore, the IRIS can improve the system performance without increasing the antennas and radio frequency (RF) chains at the BS.

Fig. 6 illustrates the system WSR versus the horizontal distance from the IRIS to the BS. It is observed that the curve exhibits a distinct double-peak characteristic, where the maximum WSR is achieved when the IRIS is deployed close to the BS or the UEs. This can be explained by the deployment-dependent path-loss product in (22). During the IRIS deployment optimization, the term $d_{\mathrm{RU},k}^{\alpha_{\mathrm{RU}}} d_{\mathrm{BR}}^{\alpha_{\mathrm{BR}}}$ plays a dominant role in determining the quality of the cascaded channel. When the IRIS is deployed close to the BS, the reduced BS-IRIS path loss enables more efficient illumination of the IRIS by the BS beamforming. When the IRIS is deployed close to the UEs, the strengthened IRIS-UE links facilitate more effective delivery of the reflected signals. Consequently, these two favorable deployment regions lead to the two WSR peaks. In contrast, when the IRIS is located around the middle of the deployment region, both propagation links become relatively long, resulting in a larger cascaded path-loss product and a more severe multiplicative fading effect. Thus, the WSR decreases and forms a valley around the middle region. This significant performance fluctuation reveals the critical importance of the IRIS deployment in mitigating the "multiplicative fading effect" and enhancing multi-user system capacity.

Fig. 7 depicts the convergence behaviors of the proposed BAO framework and the conventional AO algorithm frameworks. The simulation results reveal that the BAO framework exhibits a significantly faster convergence rate, reaching a stable state in approximately 5 iterations while achieving a superior WSR performance. This is attributed to its unique block-based iteration mechanism, validating the effectiveness of the proposed framework in addressing highly coupled and non-convex joint optimization problems.

5. Conclusions

In regular RIS-assisted multi-user wireless communication systems where the number of RIS elements is restricted, the system faces the problems of "multiplicative fading effect" and capacity constraints. In order to effectively improve the system achievable WSR, a BAO algorithm framework has been proposed by jointly optimizing BS active beamforming, IRIS deployment, IRIS passive beamforming and IRIS topology. For IRIS topology optimization, the GA-TS algorithm has been designed to address the strong dependence of the TS algorithm on the initial solution. Simulation results have demonstrated that the BAO algorithm framework achieves faster convergence than the AO framework, and our IRIS scheme significantly boosts the system achievable WSR. Furthermore, this work provides valuable design guidelines for practical communication systems. First, the IRIS should be strategically positioned in close proximity to either the BS or the user clusters to maximize the mitigation of the "multiplicative fading effect". Second, when

the number of BS antennas is limited, system designers are advised to increase the number of available grid points for the IRIS. This strategy effectively trades cost-efficient passive spatial degrees of freedom for expensive active RF chains, while guaranteeing the system performance.

CRediT authorship contribution statement

Xinying Guo: Writing – original draft, Supervision; **Rui Yuan:** Validation, Methodology; **Jiankang Zhang:** Conceptualization; **Yi Song:** Validation, Formal analysis; **Sheng Chen:** Writing – review & editing.

Data availability

Data will be made available on request.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Xinying Guo reports administrative support and article publishing charges were provided by Henan University of Technology. Xinying Guo reports a relationship with Henan University of Technology that includes: funding grants. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant 61901159, 62172141).

References

[1] Z. Lin, Z. Feng, K. Guo, A. Nauman, D. Niyato, J. Wang, AI-driven seamless and massive access in space-air-ground integrated networks, IEEE Wirel. Commun. 32 (3) (2025) 72–79. https://doi.org/10.1109/MWC.001.2400371

[2] J. Sang, Y. Yuan, W. Tang, Y. Li, X. Li, S. Jin, Q. Cheng, T.J. Cui, Coverage enhancement by deploying RIS in 5G commercial mobile networks: field trials, IEEE Wirel. Commun. 31 (1) (2024) 172–180. https://doi.org/10.1109/MWC.011.2200356

[3] X. Yue, X. Guo, X. Mu, J. Zhao, P. Yang, J. Mu, Z. Lu, Federated learning in active STARS-aided uplink networks, IEEE Trans. Veh. Technol. 75 (1) (2026) 1156–1170. https://doi.org/10.1109/TVT.2025.3590980

[4] Z. Lin, H. Niu, K. An, Y. Wang, G. Zheng, S. Chatzinotas, Y. Hu, Refracting RIS-aided hybrid satellite-terrestrial relay networks: joint beamforming design and optimization, IEEE Trans. Aerosp. Electron. Syst. 58 (4) (2022) 3717–3724. https://doi.org/10.1109/TAES.2022.3155711

[5] S. Lin, X. Li, H. Xu, S. Jin, Energy efficiency optimization in RIS-assisted communication systems with RIS power budget, IEEE Trans. Veh. Technol. 75 (2) (2026) 3434–3439. https://doi.org/10.1109/TVT.2025.3603852

[6] X. Zhu, Q. Wu, W. Chen, On the performance of RIS-aided spatial modulation for downlink transmission, IEEE Trans. Wirel. Commun. 23 (11) (2024) 16203–16217. https://doi.org/10.1109/TWC.2024.3438767

[7] J. Wang, Y. Han, J. Zhang, S. Jin, X. Li, C. Yuen, Deployment optimization of extremely large-scale RIS-Aided communication system, IEEE Trans. Commun. 73 (12) (2025) 15570–15582. https://doi.org/10.1109/TCOMM.2025.3606639

[8] Y. Cai, M.-M. Zhao, K. Xu, R. Zhang, Intelligent reflecting surface aided full-duplex communication: passive beamforming and deployment design, IEEE Trans. Wirel. Commun. 21 (1) (2022) 383–397. https://doi.org/10.1109/TWC.2021.3095939

[9] C. Chen, F. Wu, C. Zhang, D. Cheng, W. Fan, Y. Liu, Joint optimization of RIS-user association and beamforming design for multi-RIS-assisted multi-user systems, IEEE Wirel. Commun. Lett. 14 (11) (2025) 3575–3579. https://doi.org/10.1109/LWC.2025.3598573

[10] Z. Lin, H. Niu, Y. He, K. An, X. Zhong, Z. Chu, P. Xiao, Self-powered absorptive reconfigurable intelligent surfaces for securing satellite-terrestrial integrated networks, China Commun. 21 (9) (2024) 276–291. https://doi.org/10.23919/JCC.fa.2023-0437.202409

[11] Q. Wu, S. Zhang, B. Zheng, C. You, R. Zhang, Intelligent reflecting surface-aided wireless communications: a tutorial, IEEE Trans. Commun. 69 (5) (2021) 3313–3351. https://doi.org/10.1109/TCOMM.2021.3051897

[12] A.L. Moustakas, G.C. Alexandropoulos, M. Debbah, Reconfigurable intelligent surfaces and capacity optimization: a large system analysis, IEEE Trans. Wirel. Commun. 22 (12) (2023) 8736–8750. https://doi.org/10.1109/TWC.2023.3265420

[13] Y. Ren, R. Zhou, X. Teng, S. Meng, M. Zhou, W. Tang, X. Li, C. Li, S. Jin, On deployment position of RIS in wireless communication systems: analysis and experimental results, IEEE Wirel. Commun. Lett. 12 (10) (2023) 1756–1760. https://doi.org/10.1109/LWC.2023.3292125

[14] J. Tang, Z. Lv, J. Xiao, J. Wu, B. Shim, Weighted sum-rate maximization by joint antenna grouping and movable RIS deployment, IEEE Trans. Wirel. Commun. 25 (2026) 3863–3878. https://doi.org/10.1109/TWC.2025.3606657

[15] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, R. Zhang, Wireless communications through reconfigurable intelligent surfaces, IEEE Access 7 (2019) 116753–116773. https://doi.org/10.1109/ACCESS.2019.2935192

[16] R. Su, L. Dai, J. Tan, M. Hao, R. MacKenzie, Capacity enhancement for reconfigurable intelligent surface-aided wireless network: from regular array to irregular array, IEEE Trans. Veh. Technol. 72 (5) (2023) 6392–6403. https://doi.org/10.1109/TVT.2023.3236179

[17] H. Chen, C. He, G. Qian, X. Li, Energy-efficient wireless communication with irregular reconfigurable intelligent surfaces, IEEE Commun. Lett. 28 (12) (2024) 2754–2758. https://doi.org/10.1109/LCOMM.2024.3477948

[18] M. Wang, Y. Cui, Irregular RIS-assisted security for urban wireless communications, in: 2025 5th International Conference on Neural Networks, Information and Communication Engineering (NNICE), 2025, pp. 1035–1039. https://doi.org/10.1109/NNICE64954.2025.11064455

[19] C. Pan, G. Zhou, K. Zhi, S. Hong, T. Wu, Y. Pan, H. Ren, M.D. Renzo, A. Lee Swindlehurst, R. Zhang, A.Y. Zhang, An overview of signal processing techniques for RIS/IRS-Aided wireless systems, IEEE J. Sel. Top. Signal Process. 16 (5) (2022) 883–917. https://doi.org/10.1109/JSTSP.2022.3195671

[20] M. Fu, R. Zhang, Active and passive IRS jointly aided communication: deployment design and achievable rate, IEEE Wirel. Commun. Lett. 12 (2) (2023) 302–306. https://doi.org/10.1109/LWC.2022.3224457

[21] W. Lv, J. Bai, Q. Yan, H.M. Wang, RIS-assisted green secure communications: active RIS or passive RIS?, IEEE Wirel. Commun. Lett. 12 (2) (2023) 237–241. https://doi.org/10.1109/LWC.2022.3221609

[22] Q. Wu, R. Zhang, Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts, IEEE Trans. Commun. 68 (3) (2020) 1838–1851. https://doi.org/10.1109/TCOMM.2019.2958916

[23] J. Tan, S. Suo, H. Qin, Hybrid precoding codebook design in millimetre-wave massive MIMO systems with low-resolution phase shifters, IET Commun. 15 (15) (2021) 1982–1996. https://doi.org/10.1049/cmu2.12234

[24] K. Shen, W. Yu, Fractional programming for communication systems-part i: power control and beamforming, IEEE Trans. Signal Process. 66 (10) (2018) 2616–2630. https://doi.org/10.1109/TSP.2018.2812733

[25] F. Glover, Tabu search-part i, ORSA J. Comput. 1 (3) (1989) 190–206. https://doi.org/10.1287/ijoc.1.3.190

[26] H. Yang, F. Yang, S. Xu, Y. Mao, M. Li, X. Cao, J. Gao, A 1-bit 10×10 reconfigurable reflectarray antenna: design, optimization, and experiment, IEEE Trans. Antennas Propag. 64 (6) (2016) 2246–2254. https://doi.org/10.1109/TAP.2016.2550178