OPTIMIZATION

Evolutionary multi-objective optimization for RIS-aided MU-MISO communication systems

Mengke Li[1] · **Bai Yan**[1,2,3] · **Jin Zhang**[1,3]

Abstract

We study a multi-user multiple-input single-output downlink system aided by a reconfigurable intelligent surface (RIS). Users' sum rate and transmit power are two important performance indicators in such systems. However, most existing works only optimize one of them, resulting in severe performance degradation of the other. Motivated by this, in this paper, we formulate a multi-objective optimization problem to maximize the sum rate of users and minimize the transmit power simultaneously. According to our early work on fitness landscape analysis of sum rate maximization problems, the proposed problem is inferred to be multi-modal. To solve this non-convex and multi-modal problem, we propose a novel multi-objective evolutionary hybrid beamforming (MEHB) framework to find different trade-off solutions between the two conflicting objectives. In particular, we employ different kinds of multi-objective evolutionary algorithms and multi-modal multi-objective evolutionary algorithms as the baseline of MEHB framework, so as to design the passive beamforming. And the active beamforming at the base station is optimized by the classical zero-forcing method. The simulation results have verified the effectiveness of the dominance-based evolutionary algorithms in handling hybrid beamforming problems.

Keywords Reconfigurable intelligent surface (RIS) · Hybrid beamforming · multi-objective evolutionary algorithms · Multi-objective multi-modal optimization

1 Introduction

The fifth-generation (5 G) mobile communication standard promises to provide enhanced mobile broadband, massive connectivity and ultra-low latency through various technological advances (Wu and Zhang 2019a). In harsh wireless propagation environments, however, guaranteed quality of service requirements may not be fully achieved with the existing technologies. For example, the communication quality may be affected by the shadowing caused by fixed and moving obstacles, e.g., trees, and buildings. And the network energy consumption and hardware cost still remain vital issues.

To solve the above issues, reconfigurable intelligent surface (RIS), has emerged as a promising solution to achieve a smart and programmable wireless propagation environment for 5 G systems (Wu and Zhang 2019c). Specifically, RIS is a two-dimensional artificial structure, consisting of low-cost, passive, reconfigurable reflecting elements. The *passive beamforming* of RIS, i.e., each RIS element can induce a phase shift to the incident signal (Chen et al 2016), makes it possible to change the wireless environment intelligently. By densely deploying RISs in the wireless network and smartly coordinating their phase shifts, the signal propagation/wireless channels between transmitters and receivers can be flexibly reconfigured to achieve desired realizations. With proper passive beamforming, the RIS-assisted communications system can realize expected properties, such as extending signal coverage, improving energy efficiency, miti-

✉ Jin Zhang
zhangj4@sustech.edu.cn

Mengke Li
12032456@mail.sustech.edu.cn

Bai Yan
yanb@sustech.edu.cn

[1] Computer Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, Guangdong, China

[2] Research Institute of Trustworthy Autonomous Systems, Southern University of Science and Technology, Shenzhen 518055, Guangdong, China

[3] Guangdong Provincial Key Laboratory of Brain-inspired Intelligent Computation, Southern University of Science and Technology, Shenzhen 518055, Guangdong, China

gating interference, enhancing system security, and so on Wu et al (2021). Thus, how to select the optimal passive beamforming is very important.

With this in mind, several works explore the potential of RIS in multi-user (MU) communication scenarios by elaborately selecting the beamforming. In Guo et al (2020); Yu et al (2019); Wu and Zhang (2019b), the *active beamforming* (beamforming by active antennas) at the base station (BS) and the passive beamforming at the RIS are jointly optimized. Different optimization objectives are achieved, such as maximizing the user rate (Guo et al 2020), maximizing the spectral efficiency (Yu et al 2019), or minimizing the transmit power (Wu and Zhang 2019b). These works consider the continuous passive beamforming settings for the RIS, and gradient-based iterative algorithms can achieve good system performance. However, given the hardware limitations of real-world implementations, systems only support discrete passive beamforming. To achieve discrete passive beamforming, several methods are proposed, such as the quantization approach (Chen et al 2019), exhaustive search (Wu and Zhang 2019a), sequential algorithm (Di et al 2020), and the brand-and-bound (BB)-based method (Di et al 2020). Among these methods, exhaustive search and BB are able to locate the optimal solutions but with very high computational complexity; the quantization approach and sequential algorithm run fast but degrade the performance significantly. A more efficient method for passive beamforming design with lower complexity is required.

In these works (Guo et al 2020; Yu et al 2019; Wu and Zhang 2019b; Chen et al 2019; Wu and Zhang 2019a; Di et al 2020, 2016), only a single objective is considered, either maximizing the sum rate or minimizing transmit power. However, multiple objectives or features of real-world problems often need to be considered at the same time (Bakkouri and Afdel 2020, 2022). Integrating multiple features dramatically improves the performance of the systems, while only optimizing one of them would result in severe performance degradation in the other objective. Thus, it is better to optimize them simultaneously. A related close work is Khalili et al (2021), where two objectives, i.e., the data sum rate and the total harvested energy, are maximized simultaneously. The problem is transformed into a single-objective optimization problem by assigning a balancing parameter. However, the performance is very sensitive to this parameter, which is difficult to determine. Moreover, the proposed solver does not have the decision-making flexibility due to incapable of obtaining different trade-off solutions.

Inspired by the above analysis, we are motivated to jointly design the hybrid beamforming to simultaneously maximize the sum rate and minimize transmit power, in a downlink RIS-aided multi-user multiple-input single-out (MU-MISO) system. The *hybrid beamforming* includes the active beamforming at BS and passive beamforming at RIS

to be optimized. Here, we consider a practical case, i.e., the passive beamforming takes discrete values. Different from Khalili et al (2021), we formulate this hybrid beamforming problem as a multi-objective optimization problem, with the sum rate and transmit power being two objectives. This problem is a mix-integer programming problem, which is NP-hard and non-convex with highly coupled variables. To design a solver, two challenges should be considered:

1. Existing methods (Chen et al 2019; Wu and Zhang 2019a; Di et al 2020, 2016) ignore the high dependence between the passive beamforming but decouple the passive beamforming problem and estimate them separately, degrading the performance. Moreover, the methods cannot balance the high computational complexity and performance.
2. The sum rate maximization problem shows a multi-modal landscape, which has many scattered, uncorrelated local peaks (Yan et al 2021). These local peaks are great obstacles for hindering the solver to converge to the optima, especially in the cases with a large-scale RIS. Since one of our optimization objectives is maximizing the sum rate, the proposed multi-objective problem should be also multi-modal. The multi-modal property causes great difficulty to find the global optima.

To solve the proposed multi-objective hybrid beamforming problem, we propose a novel multi-objective evolutionary hybrid beamforming (MEHB) framework to find all Pareto-optimal solutions between the two conflicting objectives. In the MEHB framework, we employ different kinds of multi-objective evolutionary algorithms (MOEAs) and multi-modal multi-objective evolutionary algorithms (MMOEAs) for passive beamforming design, while the corresponding sub-optimal active beamforming is obtained by the classical zero-forcing method (Di et al 2020).

The contributions of our work are summarized as follows:

1. We propose a multi-objective optimization model. This model maximizes the sum rate and minimizes the transmit power simultaneously with a conflicting relationship. No balancing parameters are required to be set.
2. We propose a novel multi-objective evolutionary hybrid beamforming framework to analyze and solve the proposed model, which helps understand the problem better and design hybrid beamforming. In this framework, three MOEAs and two MMOEAs are employed to optimize the passive beamforming, while the corresponding sub-optimal active beamforming is obtained by the classical zero-forcing method. This framework is capable of obtaining different trade-off solutions in one run and providing selection flexibility for decision-makers.
3. We conduct numerical experiments and the simulation results validate that the effectiveness of the proposed

algorithm. In general, we conclude that to solve the passive beamforming, population-based algorithms are more effective with a large-scale RIS. Remarkably, the dominance-based evolutionary algorithms with/without multi-modal strategies outperform other MOEAs and MMOEAs under different numbers of RIS quantization bits, different numbers of users and signal–noise ratios.

The remainder of the paper is organized as follows. In Section II, we briefly review the background of the hybrid beamforming approaches and multi-objective evolutionary optimization. The system model is introduced in Section III. Section IV presents the novel multi-objective formulation and the multi-objective evolutionary hybrid beamforming framework. Section V gives the simulation results and discussions. Finally, section VI concludes the paper.

2 Background

2.1 Related works on hybrid beamforming

Most existing works regard RIS as a reflection-type device in point-to-point communication or multi-user (MU) systems. In these works, several hybrid beamforming methods were proposed to achieve different communication goals, for example, maximizing the sum rate (Guo et al 2020), maximizing the network's spectral or energy efficiency Yu et al (2019), maximizing the secrecy capacity for physical layer security, minimizing the transmit power (Wu and Zhang 2019b) or maximizing the harvested power (Nadeem et al 2019).

Hybrid beamforming design is very challenging due to the high-dimension property of the passive beamforming as well as the deep coupling between the active beamforming at BS and the passive beamforming at RIS. To solve hybrid beamforming problems efficiently, the original problem is generally decoupled into two sub-problems: active beamforming sub-problem and passive beamforming sub-problem. Then, the two sub-problems are solved alternatively by alternating optimization (AO) based methods. To solve the active and passive beamforming sub-problems, there are a number of existing methods as shown below.

2.1.1 Active beamforming design

Given that the passive beamforming at RIS is fixed, the sub-problem to optimize active beamforming is reduced to a traditional beamforming problem in wireless communication systems. For such beamforming problems, several methods such as the maximum-ratio transmission, the minimum mean square error criterion, the semi-definite relaxation (SDR) technique and the zero-forcing method are usually used. Among the methods, the maximum-ratio transmission method is suitable for the single-user case, and the others are usually used to cope with the multi-user interference. The optimization objectives of current RIS beamforming designs are divided into two categories, namely the sum rate maximization problems and the transmit power minimization problems. For the sum rate maximization problems, the minimum mean-square error method is a good option (Guo et al 2020). For the transmit power minimization problems, the SDR technique is a commonly-used method, but the computation complexity is extremely high for large-scale RISs (Wu and Zhang 2019b). Comparably, the zero-forcing method is an efficient method with low complexity (Di et al 2020)(Huang et al 2018).

2.1.2 Passive beamforming design

Keeping the active beamforming at BS fixed, existing works consider continuous or discrete passive beamforming. The passive beamforming is parameterized by phase shifts, hence the passive beamforming design is equivalent to optimizing the phase shifts. When the phase shifts take continuous values, that is, each phase shift can be independently adjusted between 0 and 2π. In this case, the optimal phase shifts can be estimated by phase alignment method (Ma et al 2021) or SDR technique (Wu and Zhang 2019b). The phase alignment method converges fast but can only work in single-user situation. The SDR technique is suitable for a wider range of scenarios, but the time complexity is higher. However, the continuous phase shifts are difficult to realize in practice because of the high cost of hardware implementations.

Such discrete settings of phase shifts at RIS lead to an integer programming problem, which complicates the passive beamforming design compared with continuous settings. To achieve discrete phase shifts, several methods are proposed, such as exhaustive search (Wu and Zhang 2019a), the quantization approach (Chen et al 2019), sequential algorithms (Di et al 2020), and the brand-and-bound (BB) based method (Di et al 2020). Exhaustive search incurs prohibitively high complexity with a large number of RIS elements, thus seriously limiting its applications. In the quantization approach, the optimal continuous phase shifts are obtained at first, then they are quantized to their nearest values in the corresponding discrete sets. Despite the low computational complexity, this approach suffers from great quantization errors, especially when fewer quantization bits are used. In the sequential algorithm, each phase shift is optimized in sequence while keeping the others fixed. The main drawback is that the performance is easy to be trapped into local optima due to variable decoupling. The computational complexity of the BB method is exponential over the number of RIS elements; hence, this method is incapable of handling large-scale RISs.

In general, due to the nature of such problems, it is difficult to achieve good results while maintaining low time complexity. Such problems become especially complicated when the numbers of RIS elements and quantization bits increase. Therefore, a more efficient method with a good balance between convergence and complexity is required for passive beamforming design.

2.2 Multi-objective evolutionary optimization

A multi-objective optimization problem (MOP) with constraints can be formulated as follows:

$$\min_{\mathbf{x}} \quad \mathbf{F}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), ..., f_M(\mathbf{x}))$$

$$\text{subject to} \quad g_i(\mathbf{x}) \leq 0, i \in \{1, 2, ..., J\}, \quad \text{and} \quad \mathbf{x} \in \mathbb{R}^D \tag{1}$$

where $\mathbf{x}$ and f_m are the decision variables and the m-th objective, $\mathbb{R}^D$ is the feasible space, g_i denote the i-th constraint, and M, D and J are the numbers of objectives, decision variables and constraints, respectively (Mohseni-Bonab et al 2015).

Compared with single-objective optimization problems, MOPs are supposed to obtain different trade-off solutions simultaneously in a single run. To compare the qualities of two solutions of a MOP, the *Pareto dominance* relationship is employed (Zhang et al 2018). Solution $\mathbf{x}_1$ is better than another solution $\mathbf{x}_2$, i.e., $\mathbf{x}_1$ Pareto dominates $\mathbf{x}_2$ ($\mathbf{x}_1 \prec \mathbf{x}_2$), if the following inequations are satisfied (Zhang et al 2015; Cheng et al 2016; Li et al 2015):

$$\begin{cases} \forall i \in 1, 2, ..., M, f_i(\mathbf{x}_1) \leq f_i(\mathbf{x}_2) \\ \exists j \in 1, 2, ..., M, f_j(\mathbf{x}_1) < f_j(\mathbf{x}_2). \end{cases} \tag{2}$$

If $\mathbf{x}$ Pareto dominates all the other solutions in $\mathbb{R}^D$, $\mathbf{x}$ is said to a *Pareto-optimal solution*. All Pareto-optimal solutions, i.e., $\{\mathbf{x} \in \mathbb{R}^D | \nexists \mathbf{y} \in \mathbb{R}^D, \mathbf{y} \prec \mathbf{x}\}$, form the Pareto-optimal set (PS), and the corresponding performances of PS in the objective space, i.e., $\{F(\mathbf{x}) | \mathbf{x} \in PS\}$, form the Pareto-optimal front (PF). It is incapable to compare any two Pareto-optimal solutions if there is no additional information.

To solve a MOP, the beginning idea is to transform it into a single-objective optimization problem, by aggregating all objectives with a sensitive weighting parameter. These methods perform poorly on the Pareto-optimality since only a single optimal solution can be found in each simulation run. To obtain different trade-off solutions and provide flexibility in decision making, many multi-objective optimization evolutionary algorithms (MOEAs) have been proposed (Schaffer 1985). Collectively, there are about three kinds of MOEAs, namely the Pareto-based MOEAs, the decomposition-based MOEAs, and the performance indicator-based MOEAs (He

et al 2017). The main feature of the Pareto-based MOEAs is adopting the Pareto dominance relationship to the selection mechanism. A typical algorithm is the elitist non-dominated sorting genetic algorithm (NSGA-II) (Deb et al 2002). NSGA-II proposes the fast non-dominated sorting and the crowding distance. The former is employed to divide solutions into different layers, thus guiding the choice of solutions to guarantee convergence. The latter is used to compare solutions in the same layer and guarantee diversity. Based on NSGA-II, the reference vector guided evolutionary algorithm (RVEA) (Cheng et al 2016) propose an adaptive strategy to handle unnormalized objective functions and can handle high-dimensional MOPs. The decomposition-based MOEAs decompose a MOP into multiple sub-problems and employ the neighbor relationship to aggregate all sub-problems. The multi-objective evolutionary algorithm based on decomposition (MOEA/D) is proposed in Zhang and Li (2007). MOEA/D defines the neighbor relationship between sub-problems through the distance between the weight vectors. Information is shared among adjacent sub-problems through the evolution process. Apart from genetic algorithms, particle swarm optimization (PSO) can also be extended to handle MOPs. The multi-objective particle swarm optimization (MOPSO) (Coello and Lechuga 2002) is based on PSO and is applied to optimize multiple conflicting objectives.

In some problems such as space mission design (Schutze et al 2011), path-planning (Yue et al 2017) and traveling salesman problems, the MOEAs may fail to find any optimum when the problem size increases. This is because the MOPs contain multiple global and local peaks, i.e., the multimodal optimization problems (MMOPs). In order to handle the multi-modal property, a number of multi-modal MOEAs (MMOEAs) are proposed (Ray et al 2022). They attempt to locate as many as possible (almost) equivalent Pareto-optimal solutions, so as to catch all the global optima. The MMOEAs mainly include three categories (Lin et al 2020):

- *Pareto-Based MMOEAs:* Based on NSGA-II, two algorithms are developed for solving MMOPs. The authors propose Omni-optimizer (Deb and Tiwari 2005), presenting and employing the concept of crowding distance in decision space. Another MMOEAs is double-niched NSGA-II (DN-NSGA-II) (Liu et al 2018). DN-NSGA-II employs a niche sharing method to maintain diversity not only in the objective spaces, but also in decision spaces. The niche sharing method degrades the fitness of individuals in the same neighborhood (usually near a local peak) according to its scale. Thus, outside individuals are discouraged to occupy the same niche. Both methods are able to find multiple global PSs for the same PF.
- *Decomposition-Based MMOEAs:* There exist specially designed MMOEAs based on MOEA/D to cope with MMOPs. Among them, one representative algorithm is
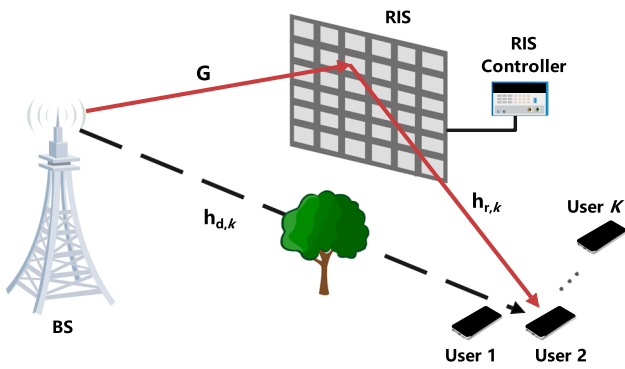
Fig. 1 RIS-aided MU-MISO system

MOEA/D-AD (Tanabe and Ishibuchi 2018). The most notable difference between MOEA/D and MOEA/D-AD is that MOEA/D-AD proposes the perpendicular distance. Based on this, each solution is assigned to a sub-problem with the closest weight vector and a relative neighborhood size L was set in its decision space. After that, each offspring only updates the original solutions associated with the same sub-problem from its L nearest individuals in the decision space. The strategies are crucial to preserving multiple PSs.

- *Others:* There are also some extensions from other MMOEAs to solve MMOPs, such as MOPSO using ring topology and special crowding distance (MO_Ring_PSO_SCD) Yue et al (2017). In MO_Ring_PSO_SCD, an index-based ring topology (Ring) is introduced to induce some stable niches, encouraging identifying a number of PSs. Note that the designed crowding distance (SCD) is adopted to measure the density of solutions.

3 System model

A RIS-aided downlink MU-MISO system is considered, as shown in Fig. 1. The system includes a BS equipped with N_t antennas, a RIS equipped with R passive elements, and K single-antenna users. There exist some obstacles between the BS and users, causing a very weak signal on the direct path (Guo et al 2020). In this case, deploying the RIS becomes particularly important to provide a new path. In this work, we neglect the direct path due to its severe signal degradation and only consider signals reflected by the RIS.

Denote the channel vector from BS to user k as $\mathbf{h}_{\mathrm{d},k} \in \mathbb{C}^{N_t \times 1}$, the channel vector from RIS to user k as $\mathbf{h}_{\mathrm{r},k} \in \mathbb{C}^{R \times 1}$, and the channel vector from BS to RIS as $\mathbf{G} \in \mathbb{C}^{R \times N_t}$, where $k = 1, 2, ..., K$. Owing to the severe signal degradation of the direct path from BS to users, we ignore the channel $\mathbf{h}_{\mathrm{d},k}$ for brevity. Existing systems only support discrete passive beamforming of the RIS due to hardware limitations. Hence, we consider a RIS of b-bit controllable, that is, 2^b possi-

ble configuration modes. Then the passive beamforming is expressed as a diagonal matrix

$$\Theta = \mathrm{diag}(\beta_1 e^{j\theta_1}, ..., \beta_R e^{j\theta_R}), \tag{3}$$

where $\boldsymbol{\beta} = [\beta_1, ..., \beta_r, ..., \beta_R]$ $(0 \leq \beta_r \leq 1)$ indicates the reflection efficiency of RIS. For simplicity, we set all β_r to 1. $\boldsymbol{\theta} = [\theta_1, ..., \theta_r, ..., \theta_R]$ denotes the phase shift vector with

$$\theta_r = \frac{m_r \pi}{2^{b-1}}, m_r \in \left\{0, 1, ..., 2^b - 1\right\}, \tag{4}$$

where $r = 1, 2, ..., R$.

We assume the channel state information of all channels is perfectly known at the BS and RIS and the mobility of the users is very limited, which is the same as Guo et al (2020). The transmitted signal at the BS is

$$\mathbf{x} = \sum_{k=1}^{K} \mathbf{w}_k s_k, \tag{5}$$

where $\mathbf{w}_k \in \mathbb{C}^{N_t \times 1}$ represents the active beamforming vector at the BS. s_k denotes the transmit data to user k, an independent random variable with zero mean and unit variance.

Then, the signal received by user k can be expressed as

$$y_k = \mathbf{h}_{\mathrm{r},k}^{\mathrm{H}} \Theta^{\mathrm{H}} \mathbf{G} \mathbf{x} + n_k, \tag{6}$$

where $(\cdot)^{\mathrm{H}}$ is the conjugate transpose, n_k denotes the white Gaussian noise at user k with zero mean and power σ^2.

The user k treats all the signals from the other users as interference. Hence, the rate at user k is defined as:

$$\begin{aligned} S_k &= \log_2(1 + \gamma_k) \\ &= \log_2(1 + \frac{|\mathbf{h}_{\mathrm{r},k}^{\mathrm{H}} \Theta^{\mathrm{H}} \mathbf{G} \mathbf{w}_k|^2}{\sum_{i=1, i \neq k}^{K} |\mathbf{h}_{\mathrm{r},k}^{\mathrm{H}} \Theta^{\mathrm{H}} \mathbf{G} \mathbf{w}_i|^2 + \sigma^2}), \end{aligned} \tag{7}$$

where γ_k indicates the signal–noise ratio (SNR) of user k, $|(\mathbf{h}_{\mathrm{r},k}^{\mathrm{H}} \Theta^{\mathrm{H}} \mathbf{G}) \mathbf{w}_k|^2$ is the received signal of the k-th user. $\sum_{i=1, i \neq k}^{K} |(\mathbf{h}_{\mathrm{r},k}^{\mathrm{H}} \Theta^{\mathrm{H}} \mathbf{G}) \mathbf{w}_i|^2$ is the sum of interference from other users.

4 Proposed multi-objective formulation and multi-objective evolutionary hybrid beamforming framework

4.1 Multi-objective formulation for hybrid beamforming

For the downlink RIS-aided MU-MISO system under consideration, we focus on two objectives, i.e., the sum rate of all users and the transmit power at BS, defined as

$$f_1(\mathbf{W}, \boldsymbol{\Theta}) = \sum_{k=1}^{K} S_k,$$

$$f_2(\mathbf{W}, \boldsymbol{\Theta}) = \sum_{k=1}^{K} \|\mathbf{w}_k\|^2 . \tag{8}$$

Both of the objectives are crucial. However, either maximizing the sum rate or minimizing transmit power would result in severe performance degradation in the other objective.

Therefore, we aim at the hybrid beamforming design to maximize the sum rate of all users and minimize the transmit power at the BS simultaneously. It is expected to achieve different trade-off solutions between the conflicting sum rate and transmit power. As the decision maker, the BS or RIS would have the flexibility to decide whether it needs to optimize one of the objectives or to strike a good balance between them. To implement this, we formulate a multi-objective optimization model

$$\min \mathbf{F}(\mathbf{W}, \boldsymbol{\Theta}) = \min \left(-f_1(\mathbf{W}, \boldsymbol{\Theta}), f_2(\mathbf{W}, \boldsymbol{\Theta}) \right),$$

$$\text{subject to (3), (4), } \sum_{k=1}^{K} \|\mathbf{w}_k\|^2 \leq P_T, \tag{9}$$

where the active beamforming matrix $\mathbf{W}$ at the BS and passive beamforming matrix $\boldsymbol{\Theta}$ at the RIS are decision variables, where $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_K] \in \mathbb{C}^{N_t \times K}$. $-f_1(\mathbf{W}, \boldsymbol{\Theta})$ and $f_2(\mathbf{W}, \boldsymbol{\Theta})$, represent the minus sum rate and the transmit power, respectively. $\mathbf{w}_k^2 \leq P_T$ means that the transmit power is constrained below P_T.

4.2 Multi-objective evolutionary hybrid beamforming framework

The problem (9) is a mix-integer programming problem, which is NP-hard and non-convex with highly coupled variables. Many approaches can be adopted to deal with the MOP problem, such as weighted sum maximization (Schaffer 1985), ϵ-constraint (Cheng et al 2016) and MOEAs (Zhou et al 2011). The weighted sum maximization introduces a weighting parameter to combine all optimization objectives together. Consequently, the performance highly depends on the weighting parameter. Moreover, the method only outputs a single solution in each run. The ϵ-constraint method reserves one objective and transforms the others into constraints. It can characterize the entire Pareto boundary of the target region and is more effective than the weighted sum maximization methods (Miettinen 2012). MOEAs deal with the MOP directly so that a series of Pareto solutions can be obtained. In other words, MOEAs are more likely to provide multiple options for the decision maker, without assigning weighting parameters. Therefore, MOEAs are more advan-

tageous in handling our proposed MOP (9) in the RIS-aided MU-MISO system.

It is found that the sum rate maximization problem shows a multi-modal landscape, i.e., it has many scattered, uncorrelated local peaks (Yan et al 2021). These local peaks are great obstacles for hindering the solver to converge to the optima, especially in the cases with a large-scale RIS (Yan et al 2021). Since one objective of the proposed multi-objective formulation (9) is maximizing the sum rate, the proposed multi-objective problem is inferred to be multi-modal, too. The multi-modal property causes great difficulty to find the global optima. Fortunately, MMOEAs show great advantages in handling multi-modal multi-objective optimization problems (Ray et al 2022). They attempt to locate as many as possible equivalent Pareto-optimal solutions, so as to catch all the global optima.

Therefore, based on the superiority of MOEAs and MMOEAs, we propose a multi-objective evolutionary hybrid beamforming (MEHB) framework to solve the proposed MOP (9). In contrast to Guo et al (2020); Yu et al (2019); Wu and Zhang (2019b); Chen et al (2019); Wu and Zhang (2019a); Di et al (2020, 2016), the proposed framework fundamentally characterizes their trade-off, even though the proposed MOP is probably multi-modal. Remarkably, the MEHB framework is suitable for dealing with discrete passive beamforming. It solves discrete variables directly without quantization and decoupling, and is thus able to locate optimal solutions and guarantee convergence with acceptable computational complexity. Furthermore, different from Khalili et al (2021), no balancing parameters are needed and a series of non-dominated solutions are obtained simultaneously. Specifically, the MEHB needs to accomplish two tasks, that is, the active beamforming and passive beamforming design. The main idea is to first optimize the discrete passive beamforming of the RIS by some MOEAs or MMOEAs. Then, with each obtained passive beamforming, the multi-objective problem (9) is reduced to the active beamforming sub-problem

$$\max f_1(\mathbf{W}),$$

$$\text{subject to } \sum_{k=1}^{K} \|\mathbf{w}_k\|^2 \leq P_T, \tag{10}$$

To simplify this sub-problem, we preset the transmit power randomly with P_T as being the upper bound. This sub-problem (10) is reduced to a well-known digital beamforming problem with preset transmit power P_r:

$$\max f_1(\mathbf{W}),$$

$$\text{subject to } \sum_{k=1}^{K} \|\mathbf{w}_k\|^2 = P_r, \tag{11}$$

where $P_r \leq P_T$. For different passive beamforming, different P_rs will be used, so as to maintain a good diversity of solutions. It is well-known knowledge that the sub-problem (11) can be well solved by the classical zero-forcing method (Di et al 2020). After that, the solutions are evaluated according to equation (9) and selected based on the specific selection criterion of the employed MOEA or MMOEA. These steps are repeated until the stopping criterion is satisfied.

Based on the above idea, the workflow of the proposed MEHB framework is shown in Algorithm 1. The population is represented by $\mathbf{S} = \{(\mathbf{W}_1, \mathbf{\Theta}_1), ..., (\mathbf{W}_i, \mathbf{\Theta}_i), ..., (\mathbf{W}_N, \mathbf{\Theta}_N)\}$, where N is the population size, $(\mathbf{W}_i, \mathbf{\Theta}_i)$ is the i-th solution with $\mathbf{W}$ and $\mathbf{\Theta}$ being the active beamforming at the BS and the passive beamforming at the RIS, respectively. The MEHB framework starts with initialization. A number of N passive beamformings $\{\mathbf{\Theta}_i^0\}_{i=1}^N$ are randomly generated. For each $\mathbf{\Theta}_i$, the corresponding active beamforming matrix $\mathbf{W}_i$ is estimated by the function "zero-forcing" (will be detailed next). $\{(\mathbf{\Theta}_i^0, \mathbf{W}_i^0)\}_{i=1}^N$ compose the initial population $\mathbf{S}^0$. The fitness $\mathbf{F}^0$ of the population $\mathbf{S}^0$ is calculated by the proposed bi-objective model (9). In each generation, the following steps are executed:

1. Offspring generation (Lines 5–6 of Algorithm 1). If not stated, we employ uniform crossover (Semenkin and Semenkina 2012) and random mutation (Kramer 2017) operators to produce offspring's passive beamforming $\mathbf{\Theta}_{o,i}^t$ (Line 5 of Algorithm 1). The corresponding active beamforming matrix $\mathbf{W}_{o,i}^t$ is obtained by the function "zero-forcing." $\mathbf{W}_{o,i}^t$ and $\mathbf{\Theta}_{o,i}^t$ constitute a new population $\mathbf{S}_o^t$. We calculate the fitness $\mathbf{F}^t$ of $\mathbf{S}_o^t$ according to the proposed bi-objective model (9).
2. Elite solutions selection (Lines 8–12 of Algorithm 1). Considering solutions' performance, use the fitness to select elite solutions into the next population $\mathbf{S}^{t+1}$. If less than N non-dominated solutions are found, we select N solutions with the highest fitness. Otherwise, we disregard the dominated solutions prior to applying selection. Thus, the first N non-dominated solutions with the highest fitness are retained, forming the Pareto front.

If the stopping criterion is satisfied, return the non-dominated solutions. Otherwise, repeat the above steps.

In the MEHB framework, for given $\mathbf{\Theta}$, we calculate the active beamforming at the BS via the "zero-forcing" function. Common choices for the active beamforming design at the BS include the zero-forcing method, SDR method and minimum mean square error criterion (Guo et al 2020). Here, we choose the classical zero-forcing method (Di et al 2020), since it has much lower computation complexity. Thus, given RIS passive beamforming $\mathbf{\Theta}$, the BS active beamforming is estimated by

---

Algorithm 1 The MEHB overall process

Require: Channel vectors $\mathbf{h}_{r,k}^H$, $\mathbf{G}$, population size N;
Ensure: BS active beamforming $\mathbf{W}$, RIS passive beamforming $\mathbf{\Theta}$;
1: $t = 0$;
2: Initialize solutions $\mathbf{S}^0$ randomly;
3: $\mathbf{F}^0 \leftarrow$ *Fitness_Calculation*$(\mathbf{S}^0)$;
4: **while** "*the stopping criterion is not met*" **do**
5:     Generate offspring $\mathbf{S}_o^t$ by crossover and mutation;
6:     $\mathbf{F}^t \leftarrow$ *Fitness_Calculation*$(\mathbf{S}_o^t)$;
7:     Find non-dominated solutions from $\mathbf{S}_o^t \cup \mathbf{S}^t$ to form a non-dominated solution set $\mathbf{Q}$;
8:     **if** the size of $\mathbf{Q}$ is smaller than N **then**
9:         Select N solutions with the highest fitness from $\mathbf{S}_o^t \cup \mathbf{S}^t$ and put into $\mathbf{S}^{t+1}$;
10:     **else**
11:         Select N solutions with the highest fitness from $\mathbf{Q}$ and put into $\mathbf{S}^{t+1}$;
12:     **end if**
13:     $t = t + 1$;
14: **end while**
15: $\mathbf{W}, \mathbf{\Theta} \leftarrow$ find elite solutions $\mathbf{Q}$ from $\mathbf{S}^t$.

$$\mathbf{W} = \mathbf{F}^H \left(\mathbf{FF}^H\right)^{-1} \mathbf{P}^{\frac{1}{2}} = \hat{\mathbf{W}} \mathbf{P}^{\frac{1}{2}}, \tag{12}$$

where $\mathbf{F} = [\mathbf{f}_1, ..., \mathbf{f}_k, ..., \mathbf{f}_K]$, $\mathbf{f}_k = \mathbf{h}_{r,k}^H \mathbf{\Theta}^H \mathbf{G}$ denote the reflection-channel matrix and $\hat{\mathbf{W}} = \mathbf{F}^H \left(\mathbf{FF}^H\right)^{-1}$. $\mathbf{P}$ is a diagonal matrix whose k-th element represents the power p_k for user k. The zero-forcing needs to satisfy two constraints: $|\mathbf{f}_k^H \mathbf{w}_k| = \sqrt{p_k}$ and $|\mathbf{f}_k^H \mathbf{w}_{k'}'| = 0, k \neq k'$. Problem (12) is cast as a power allocation problem (Tse and Viswanath 2005)

$$\max_{\{p_k \geq 0\}} \sum_{1 \leq k \leq K} \log_2 \left(1 + \frac{p_k}{\sigma^2}\right),$$
$$s.t. \ Tr\left(\mathbf{P}^{\frac{1}{2}} \hat{\mathbf{W}}^H \hat{\mathbf{W}} \mathbf{P}^{\frac{1}{2}}\right) \leq P_T, \tag{13}$$

where $Tr(\cdot)$ represents the trace of a matrix. By applying the water-filling algorithm, the optimal solution is obtained as $p_k = \frac{1}{v_k} \max\left\{\frac{1}{\delta} - v_k \sigma^2, 0\right\}$, where v_k denotes the k-th element of $\hat{\mathbf{W}}^H \hat{\mathbf{W}}$ and δ is a normalized factor satisfying $\sum_{1 \leq k \leq K} \max\left\{\frac{1}{\delta} - v_k \sigma^2, 0\right\} = P_T$.

5 Simulation results

In this paper, we have formulated a bi-objective optimization problem to simultaneously maximize the sum rate and minimize the transmit power. To solve this bi-objective optimization problem and investigate its properties, we conduct experiments by integrating three MOEAs and two MMOEAs into our proposed MEHB framework. We select these MOEAs/MMOEAs for passive beamforming design because they have employed various genetic search and selection operators as well as represented different kinds of multi-objective evolutionary algorithms.

Table 1 Settings of the numbers of decision variables and FEs for each test instances, respectively

R	16	36	64	100	150	200
FEs	4.00E+04	4.00E+04	5.00E+04	5.00E+04	8.00E+04	8.00E+04
b	1	2	3	4	5	6
FEs	5.00E+04	5.00E+04	5.00E+04	5.00E+04	5.00E+04	5.00E+04
K	1	2	3	4	5	6
FEs	4.00E+04	4.00E+04	5.00E+04	5.00E+04	6.00E+04	6.00E+04
SNR	−2	0	2	4	6	8
FEs	5.00E+04	5.00E+04	5.00E+04	5.00E+04	5.00E+04	5.00E+04

Three MOEAs: Three multi-objective evolutionary algorithms are employed, i.e., NSGA-II (Deb et al 2002), RVEA (Cheng et al 2016) and MOPSO (Coello and Lechuga 2002). NSGA-II is a popular fast sorting and elite multi-objective genetic algorithm. It reduces the complexity of the non-inferior sorting genetic algorithm and is a benchmark for the performance of other multi-objective optimization algorithms. RVEA is the reference vector guided evolutionary algorithm, in which an adaptive strategy is proposed to adjust the reference vectors to deal with objective functions that are not well normalized. MOPSO is based on PSO, which can be applied to multi-objective optimization. The algorithm saves the knowledge of good solutions for all particles, so the convergence speed is fast.

Two MMOEAs: For handling the multi-modal property (explained in Sect. 2.2), two MMOEAs are adopted for comparison, i.e., DN-NSGA-II (Liu et al 2018) and MO_Ring_PSO_SCD (Yue et al 2017). DN-NSGA-II is based on NSGA-II, considering the solution distance and employing a niche sharing method to maintain diversity not only in the objective spaces, but also in decision spaces. MO_Ring_PSO_SCD adopts the key idea of MOPSO, proposing an index-based ring topology and special crowding distance to maintain the diversity of solutions. By comparing the results of DN-NSGA-II and NSGA-II, MO_Ring_PSO_SCD and MOPSO, we can compare the effect of the presence or absence of multi-modal strategies on the results convincingly.

Two benchmark algorithms: Except for the above MOEAs and MMOEAs, we also consider the random method (Guo et al 2020) and the gradient descent method (Di et al 2020) for performance comparison. In the random method, the passive beamforming of the RIS is randomly designed from the feasible set. In the gradient descent approach, the continuous passive beamforming is firstly optimized by gradient information, then quantized to their nearest values in the corresponding discrete sets. Note that these two benchmark algorithms cannot deal with MOP directly; therefore, we obtain a series of solutions under different transmit power settings, forming the Pareto-front. Thus, it is possible to compare benchmark algorithms with the above population-based algorithms in terms of Pareto-front quality.

We construct a test suite for the bi-objective hybrid beamforming problem, which involves a scalable number of decision variables. This test suite would promote the analysis and understanding of the variable interactions and scalability of variables. The test suite is listed in Table 1. For each test problem, the numbers of decision variables include, i.e., number of RIS elements (R), number of RIS quantization bits (b), number of users (K) and SNR, and the maximum number of function evaluations (FEs).

5.1 Experimental settings

5.1.1 System settings

We consider a downlink RIS-aided MU-MISO system, as shown in Fig. 1. The BS and RIS are located at (0, 0) and (200m, 50m). There are four users ($K = 4$) randomly distributed in a circle with radius 30m centering at (200m, 0m). The BS is equipped with five antennas ($N_t = 5$), the maximum transmit power is 20dBm, and the transmission bandwidth is 200kHz. The path-loss for BS-RIS and RIS-user link is set as $27 + 20 \lg d$ (d is the distance of channels), and the noise power spectral density is -170 dBm/Hz.

5.1.2 Algorithm settings

The population size is set to 100 for all population-based algorithms. The uniform crossover (Semenkin and Semenkina 2012) and random mutation (Kramer 2017) are adopted for offspring generation in RVEA, NSGA-II and DN-NSGA-II. As for the parameters of reproduction, the probabilities of crossover and mutation are $p_c = 0.5$ and $p_m = 1/D$, respectively, where D denotes the scale of decision variables (Justesen 2009). And the distribution indices of crossover and mutation are set to $n_c = 20$ and $n_m = 20$, respectively. The standard PSO is adopted in MOPSO and MO_Ring_PSO_SCD (Coello and Lechuga 2002).

5.1.3 Termination condition

For all the test instances, the total number of FEs is different according to specific parameters of the experiment scenario, and detailed values are shown in Table 1. For each test question, each algorithm is run 20 times, and statistical results are obtained after significance tests. Wilcoxon's rank sum test at a significance level of 0.05 is performed on all algorithms (Zhou et al 2011).

5.1.4 Performance metrics

To evaluate the performance of MOEAs, there are mainly two evaluation criteria: convergence and diversity. For avoiding the possible limitations of a single performance indicator and reflecting the two evaluation criteria at the same time, we adopt two widely used performance indicators to measure the performance of all adopted algorithms. The two performance indicators are hypervolume (HV) indicator (While et al 2006) and inverted generational distance (IGD) indicator (Zhou et al 2006). HV is a quantitative indicator of the quality of a multi-objective trade-off surface, measuring the amount of objective space dominated by the obtained non-dominated front. IGD is used to calculate the average Euclidean distance between all solutions in the true Pareto front and the non-dominated solution obtained by the algorithm. A higher HV value and a lower IGD value indicate a higher solution quality. A reference point/set is needed to calculate the two indicators. This is usually a set of solutions that are evenly distributed in true PF. Since there are no true PFs of the test questions, we collect all the non-dominated solutions yielded over 20 runs into a set, termed P, as the reference points. Then, the reference point is set to $1.01 * \mathbf{n}$ ($\mathbf{n}$ is the nadir point of set P) for HV calculation, and the reference point set is set to $0.99 * P$ for IGD calculation (He et al 2020).

5.2 Simulation results

5.2.1 Convergence behavior analysis

We utilize the HV to observe the convergence behavior. Figure 2 shows a typical convergence behavior of NSGA-II under different numbers of RIS elements R with $b = 2$. We can observe that when R value takes 36, 64 and 100, it takes about 100, 200 and 300 iterations for NSGA-II to converge. The fewer RIS elements, the faster the algorithm converges. Moreover, the hypervolume value increases with more RIS elements. This implies that NSGA-II achieves much better performance on the two objectives with more RIS elements. Considering the parallel nature of evolutionary algorithms (Gong et al 2015), we suggest accelerating MOEAs by par-
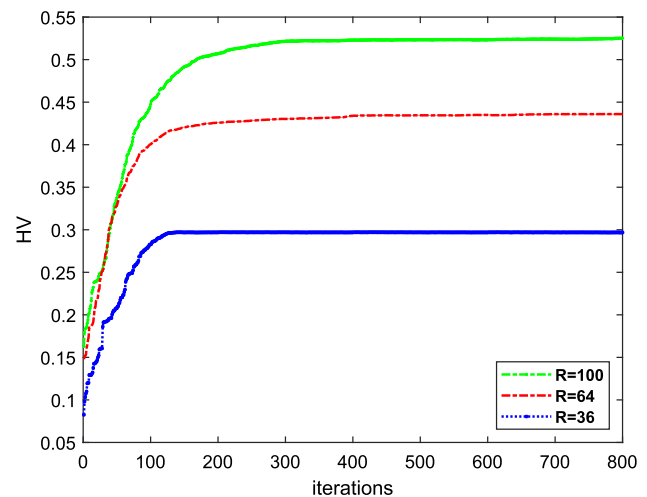

Fig. 2 The convergence behavior under different numbers of RIS elements R

allel implementation if there is a high demand for running time.

5.2.2 Comparison of different MEHB algorithms

The statistical results of the HV and IGD values achieved by the two benchmark algorithms and five MOEAs on four series of problems are summarized in Tables 2 and 3, respectively. It can be observed from these two tables that all population-based algorithms perform better than benchmark algorithms, i.e., random and gradient descent algorithms. The results confirm the potential of MOEAs to deal with hybrid beamforming problems. Moreover, DN-NSGA-II has achieved most of the best results, especially on problems with varying R and SNR. Meanwhile, NSGA-II and RVEA have achieved similar optimization results. Nevertheless, due to the complex variable interactions and correlations to different objectives in hybrid beamforming problems, MOPSO and MO_Ring_PSO_SCD have obtained worse results in comparison with other algorithms. To be specific, we discuss the simulation results based on four different test cases as follows.

Scalability over different numbers of RIS elements R: The first part of Tables 2 and 3 (Lines 2–7) investigates the scalability of the sum rate and transmit power trade-off over the number of RIS elements R with $b=2$, $K=4$ and SNR=2dB. Including the baseline methods (random and gradient descent algorithms), all methods obtain higher HV and IGD results with more RIS elements equipped. As the number of R increases, the gap between these population-based methods and the two benchmark algorithms gradually increases. RVEA, NSGA-II and DN-NSGA-II retain the highest HV and IGD with all Rs. The performance of MO_Ring_PSO_SCD is a little worse than the above three

Table 2 HV results obtained by seven algorithms on 24 test instances

Problem	Dim	Random	Gradient	RVEA	NSGA-II	DN-NSGA-II	MOPSO	MO_Ring_PSO_SCD
R	16	2.24E-01(1.51E-02)	3.56E-01(3.60E-02)	**4.22E-01(4.80E-03)**	4.06E-01(1.02E-02)	4.06E-01(7.90E-03)	3.48E-01(2.00E-04)	4.16E-01(2.80E-04)
	36	2.26E-01(6.20E-03)	4.42E-01(7.10E-03)	5.14E-01(6.90E-03)	**5.18E-01(5.30E-03)**	5.16E-01(6.40E-03)	4.46E-01(2.10E-03)	5.12E-01(1.80E-03)
	64	5.02E-01(1.21E-02)	6.24E-01(4.20E-03)	7.00E-01(3.30E-03)	7.06E-01(3.10E-03)	**7.14E-01(2.20E-03)**	6.02E-01(4.60E-03)	7.02E-01(1.70E-03)
	100	5.30E-01(4.40E-03)	7.96E-01(3.10E-03)	9.16E-01(6.40E-03)	**9.26E-01(3.70E-03)**	9.22E-01(6.20E-03)	8.04E-01(6.10E-03)	9.10E-01(3.70E-03)
	150	6.50E-01(1.09E-02)	9.52E-01(2.80E-03)	1.03E+00(3.80E-03)	1.04E+00(5.20E-03)	**1.04E+00(4.90E-03)**	9.96E-01(5.10E-03)	1.04E+00(6.30E-03)
	200	6.44E-01(2.16E-02)	1.05E+00(2.40E-03)	1.34E+00(6.50E-03)	1.35E+00(3.70E-03)	**1.36E+00(3.90E-03)**	1.08E+00(3.00E-03)	1.32E+00(7.00E-03)
b	1	2.42E-01(1.59E-02)	2.44E-01(9.80E-03)	**5.48E-01(8.80E-03)**	5.38E-01(7.40E-03)	5.38E-01(1.37E-02)	2.90E-01(8.00E-03)	5.28E-01(1.00E-02)
	2	2.40E-01(1.29E-02)	3.66E-01(5.50E-03)	6.06E-01(2.50E-03)	6.32E-01(7.90E-03)	**6.32E-01(5.90E-03)**	3.56E-01(5.50E-03)	6.06E-01(6.40E-03)
	3	2.28E-01(4.20E-03)	4.56E-01(1.30E-03)	6.38E-01(5.00E-03)	**6.44E-01(7.30E-03)**	6.10E-01(1.67E-02)	4.40E-01(1.00E-03)	6.14E-01(1.10E-02)
	4	2.30E-01(5.40E-03)	4.94E-01(7.40E-03)	6.30E-01(1.10E-02)	**6.48E-01(7.70E-03)**	6.04E-01(7.30E-03)	4.74E-01(7.10E-03)	6.14E-01(4.50E-03)
	5	2.38E-01(1.10E-02)	5.04E-01(3.50E-03)	6.38E-01(8.10E-03)	**6.42E-01(8.30E-03)**	5.94E-01(1.75E-02)	4.82E-01(4.20E-03)	6.20E-01(2.71E-02)
	6	2.34E-01(1.16E-02)	5.06E-01(4.70E-03)	6.38E-01(7.40E-03)	**6.44E-01(8.50E-03)**	5.90E-01(1.13E-02)	5.12E-01(7.10E-03)	6.24E-01(1.13E-02)
K	1	3.02E-01(2.50E-03)	3.86E-01(1.50E-03)	4.74E-01(6.00E-04)	4.76E-01(5.00E-04)	**4.76E-01(1.00E-04)**	4.00E-01(1.00E-03)	4.60E-01(9.00E-04)
	2	4.36E-01(3.90E-03)	4.60E-01(1.90E-03)	4.84E-01(1.09E-02)	**5.10E-01(7.70E-03)**	5.08E-01(3.90E-03)	4.42E-01(1.30E-03)	4.76E-01(1.11E-02)
	3	5.66E-01(1.21E-02)	6.36E-01(2.30E-02)	6.94E-01(2.00E-03)	**6.98E-01(1.60E-03)**	6.96E-01(2.90E-03)	6.22E-01(2.60E-02)	6.82E-01(6.00E-03)
	4	5.40E-01(9.40E-03)	7.26E-01(4.92E-02)	9.02E-01(3.00E-02)	**9.14E-01(4.60E-03)**	8.98E-01(5.20E-03)	7.74E-01(1.39E-02)	8.96E-01(1.00E-02)
	5	3.52E-01(3.21E-02)	7.38E-01(2.76E-02)	1.01E+00(4.30E-03)	1.02E+00(5.20E-03)	**1.04E+00(2.50E-03)**	8.26E-01(3.07E-02)	1.00E+00(4.09E-02)
	6	5.78E-02(1.37E-02)	7.94E-01(5.10E-02)	1.05E+00(1.49E-02)	**1.06E+00(9.20E-03)**	1.01E+00(2.15E-03)	8.68E-01(4.93E-02)	1.06E+00(1.03E-02)
SNR	−2	1.14E-02(9.00E-04)	1.59E-01(9.80E-03)	1.65E-01(6.40E-03)	**1.68E-01(5.40E-03)**	1.30E-01(6.10E-03)	1.42E-01(9.80E-03)	1.64E-01(4.10E-03)
	0	4.18E-02(2.30E-03)	2.10E-01(6.90E-03)	2.48E-01(6.20E-03)	**2.48E-01(5.40E-03)**	2.20E-01(1.78E-02)	2.02E-01(4.60E-03)	2.42E-01(5.90E-03)
	2	1.05E-01(7.10E-03)	4.02E-01(5.80E-03)	4.12E-01(9.00E-03)	4.20E-01(5.20E-03)	**4.22E-01(4.20E-03)**	4.02E-01(7.80E-03)	4.06E-01(6.00E-03)
	4	2.40E-01(9.10E-03)	6.20E-01(8.30E-03)	**6.46E-01(1.04E-02)**	6.26E-01(9.70E-03)	6.40E-01(4.10E-03)	6.24E-01(6.70E-03)	6.38E-01(1.12E-02)
	6	4.04E-01(1.18E-02)	7.98E-01(7.50E-03)	8.44E-01(7.10E-03)	8.54E-01(5.10E-03)	**8.62E-01(4.10E-03)**	8.18E-01(7.10E-03)	8.52E-01(7.20E-03)
	8	6.36E-01(1.33E-02)	9.78E-01(5.10E-03)	1.09E+00(4.70E-03)	1.09E+00(6.50E-03)	**1.10E+00(1.40E-03)**	9.98E-01(5.00E-03)	1.09E+00(5.80E-03)

The best result in each row is highlighted

Table 3 IGD results obtained by seven algorithms on 24 test instances

Problem	Dim	Random	Gradient	RVEA	NSGA-II	DN-NSGA-II	MOPSO	MO_Ring_PSO_SCD
R	16	6.92E+01(6.49E-01)	6.68E+01(4.67E-01)	**6.54E+01(3.34E-01)**	6.66E+01(6.76E-01)	6.60E+01(4.99E-01)	6.60E+01(4.36E-02)	6.68E+01(3.23E-01)
	36	6.50E+01(3.88E-01)	6.26E+01(6.21E-01)	6.08E+01(3.23E-01)	6.04E+01(2.79E-01)	**5.98E+01(3.70E-01)**	6.26E+01(5.01E-01)	6.02E+01(2.18E-01)
	64	6.04E+01(3.38E-01)	5.56E+01(4.57E-01)	5.10E+01(2.07E-01)	5.16E+01(1.45E-01)	**5.00E+01(1.33E-01)**	5.32E+01(5.91E-01)	5.08E+01(3.16E-01)
	100	5.12E+01(5.77E-01)	4.64E+01(5.96E-01)	4.08E+01(2.97E-01)	4.02E+01(1.63E-01)	**4.00E+01(2.56E-01)**	4.40E+01(6.23E-01)	4.10E+01(2.04E-01)
	150	4.56E+01(6.25E-01)	4.02E+01(1.97E-01)	3.62E+01(1.72E-01)	3.50E+01(2.34E-01)	**3.46E+01(2.22E-01)**	3.98E+01(4.21E-01)	3.64E+01(1.89E-01)
	200	4.04E+01(7.60E-01)	2.90E+01(4.30E-01)	2.46E+01(2.30E-01)	2.22E+01(1.36E-01)	**2.20E+01(1.48E-01)**	2.46E+01(4.00E-01)	2.38E+01(2.03E-01)
b	1	6.04E+01(6.64E-01)	6.24E+01(2.56E-01)	**4.94E+01(3.84E-01)**	4.98E+01(3.14E-01)	4.94E+01(4.45E-01)	6.00E+01(3.06E-01)	4.96E+01(1.96E-01)
	2	5.92E+01(4.85E-01)	5.76E+01(4.45E-01)	4.68E+01(1.21E-01)	4.56E+01(3.36E-01)	**4.52E+01(1.69E-01)**	5.66E+01(4.19E-01)	4.58E+01(1.64E-01)
	3	6.08E+01(3.35E-01)	5.32E+01(2.56E-01)	5.56E+01(2.43E-01)	**4.50E+01(3.17E-01)**	4.60E+01(6.92E-01)	4.92E+01(1.99E-01)	4.46E+01(2.12E-01)
	4	6.14E+01(2.24E-01)	5.14E+01(3.78E-01)	4.58E+01(4.99E-01)	**4.48E+01(3.31E-01)**	4.60E+01(2.10E-01)	4.84E+01(4.36E-01)	4.54E+01(3.89E-01)
	5	6.10E+01(6.19E-01)	5.06E+01(2.43E-01)	4.54E+01(3.35E-01)	**4.52E+01(3.48E-01)**	4.64E+01(7.01E-01)	4.84E+01(5.12E-01)	4.56E+01(4.10E-01)
	6	6.08E+01(5.96E-01)	5.06E+01(2.14E-01)	4.54E+01(3.27E-01)	**4.50E+01(3.73E-01)**	4.70E+01(5.68E-01)	4.80E+01(5.24E-01)	4.54E+01(2.08E-01)
K	1	6.02E+01(3.08E-01)	5.78E+01(1.25E-01)	5.40E+01(6.27E-02)	5.36E+01(1.61E-02)	**5.32E+01(1.73E-02)**	5.40E+01(2.46E-01)	5.38E+01(3.46E-02)
	2	5.34E+01(3.17E-01)	5.32E+01(8.73E-02)	**5.08E+01(1.00E-00)**	5.92E+01(3.73E-01)	5.20E+01(1.22E-01)	5.36E+01(5.03E-02)	5.32E+01(9.15E-02)
	3	4.60E+01(3.23E-01)	4.64E+01(2.26E-01)	4.44E+01(8.84E-02)	4.38E+01(1.73E-02)	**4.34E+01(9.32E-02)**	4.54E+01(6.18E-01)	4.40E+01(5.48E-02)
	4	4.60E+01(5.35E-01)	4.22E+01(6.28E-01)	3.56E+01(1.33E-01)	**3.46E+01(3.17E-01)**	3.48E+01(1.56E-01)	3.80E+01(6.08E-01)	3.62E+01(1.01E-01)
	5	4.84E+01(1.35E-00)	4.12E+01(1.90E-01)	3.10E+01(1.47E-01)	3.00E+01(3.31E-01)	**2.94E+01(6.89E-02)**	3.54E+01(7.90E-01)	3.22E+01(1.45E-01)
	6	4.24E+01(6.58E-01)	3.86E+01(2.30E-01)	2.86E+01(4.65E-01)	**2.80E+01(3.48E-01)**	2.94E+01(7.25E-01)	3.08E+01(6.30E-01)	2.96E+01(3.08E-01)
SNR	−2	7.48E+01(2.06E-01)	6.82E+01(1.67E-01)	6.54E+01(3.17E-01)	**6.53E+01(2.74E-01)**	6.60E+01(3.76E-01)	6.66E+01(1.87E-01)	6.54E+01(2.15E-01)
	0	7.16E+01(3.87E-01)	6.26E+01(2.05E-01)	6.02E+01(3.34E-01)	**5.98E+01(3.56E-01)**	6.02E+01(8.73E-01)	6.16E+01(3.64E-01)	6.04E+01(2.65E-01)
	2	6.72E+01(7.45E-01)	5.72E+01(4.20E-01)	5.10E+01(3.88E-01)	5.06E+01(2.59E-01)	**5.02E+01(1.71E-01)**	5.94E+01(2.69E-01)	5.12E+01(1.89E-01)
	4	6.12E+01(7.22E-01)	4.90E+01(6.10E-01)	4.60E+01(5.03E-01)	4.58E+01(4.16E-01)	**4.50E+01(1.07E-01)**	4.68E+01(2.66E-01)	4.58E+01(4.10E-01)
	6	5.82E+01(4.35E-01)	4.56E+01(1.96E-01)	4.24E+01(2.93E-01)	4.16E+01(2.14E-01)	**4.10E+01(1.55E-01)**	4.58E+01(3.12E-01)	4.22E+01(2.34E-01)
	8	6.04E+01(2.73E-00)	4.12E+01(4.09E-01)	3.70E+01(1.87E-01)	3.62E+01(2.67E-01)	**3.56E+01(5.33E-02)**	4.00E+01(4.33E-01)	3.60E+01(2.07E-01)

The best result in each row is highlighted

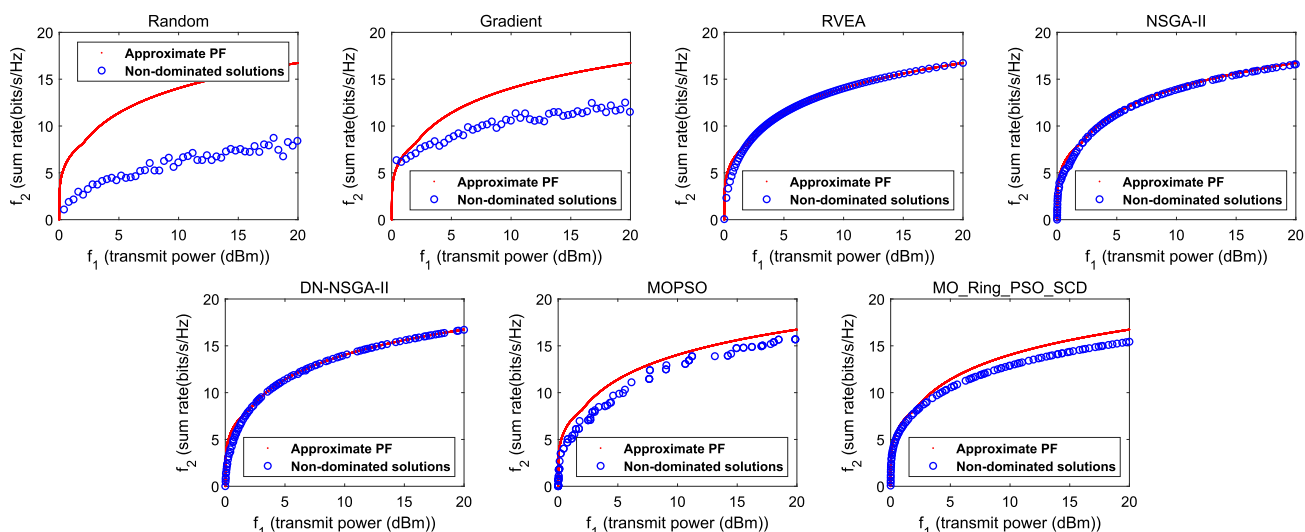**Fig. 3** Non-dominated solutions obtained by seven algorithms on $R=16$ ($b=2$, $K=4$ and SNR=2dB) in the run associated with the median fitness value

algorithms, but it achieves significant gain compared with MOPSO.

Figure 3 shows the convergence and diversity of solutions on all seven algorithms with $R=16$, $b=2$, $K=4$ and SNR=2dB. The red dots indicate the convergence approximate PF, and the blue dots are used to show the distribution of obtained non-dominated solutions of different algorithms. We can see that the sum rate grows rapidly with lower transmit power (less than 5dBm) and increases slowly as the transmit power continues to increase in all algorithms, including the two benchmark algorithms. This is because under this setting, the number of RIS elements is rather small, and even the benchmark algorithms can work well. In terms of Pareto-front quality, RVEA, NSGA-II and DN-NSGA-II perform best, followed by MOPSO and MO_Ring_PSO_SCD, and the two benchmark algorithms worst. RVEA, NSGA-II and DN-NSGA-II perform slightly better than MOPSO and MO_Ring_PSO_SCD, since the PSO-based algorithms only find a small number of feasible solutions clustered in several regions, and their poor performance could be attributed to their poor capabilities in diversity maintenance. All five population-based algorithms outperform the other two benchmark algorithms. This is because their global search ability can maintain good diversity and encourage solutions to communicate with each other, thereby finding high-quality solutions.

Figure 4 depicts the trade-off with $R=200$, $b=2$, $K=4$ and SNR=2dB. By contrast, $R=200$ is difficult as it involves 200 elements of RIS, which is more large-scale and has complex variable interactions. It can be observed that the performance of all seven algorithms is almost the same ranking as with $R=16$ in terms of Pareto-front quality. RVEA, NSGA-II and DN-NSGA-II obtain the best converged and diverse solu-

tions again, followed by MOPSO and MO_Ring_PSO_SCD, and the two benchmark algorithms worst. In addition, we observe some differences with $R=200$ since the increase of RIS elements brings more challenges to this problem. In the first echelon, RVEA performs slightly worse than NSGA-based algorithms since it cannot maintain the diversity well. The performance gap between the first echelon and the second echelon has further widened; that is, the convergence and diversity of MOPSO and MO_Ring_PSO_SCD further decrease with $R=200$. And MO_Ring_PSO_SCD can only get a few uneven solutions. As for the two benchmark algorithms, the sum rate obtained is less than half of that of all population-based algorithms under the same transmit power. In this case, there is little difference between the two benchmark algorithms, which means that the gradient descent method fails to find good solutions with $R=200$. As our early work (Yan et al 2021), the multi-modal property becomes more obvious when the number of RIS elements grows, resulting in a more difficult problem. In such a difficult case, the performance of random and gradient descent algorithms is very limited.

Performance versus the number of quantization bits b for RIS: The second part of Table 2 and Table 3 (Lines 8–13) shows the performance of the sum rate and transmit power trade-off versus the number of RIS quantization bits b with $R=64$, $K=4$ and SNR=2dB. Except for the random method, all the other six methods obtain higher HV and IGD results with more RIS quantization bits. For the random method, even if b increases, the random passive beamforming still leads to poor results. RVEA, NSGA-II and DN-NSGA-II achieve the highest HV and IGD under all numbers of quantization bits. As b increases from 1 to 2, the HV obtained by gradient descent and population-based algorithms grow
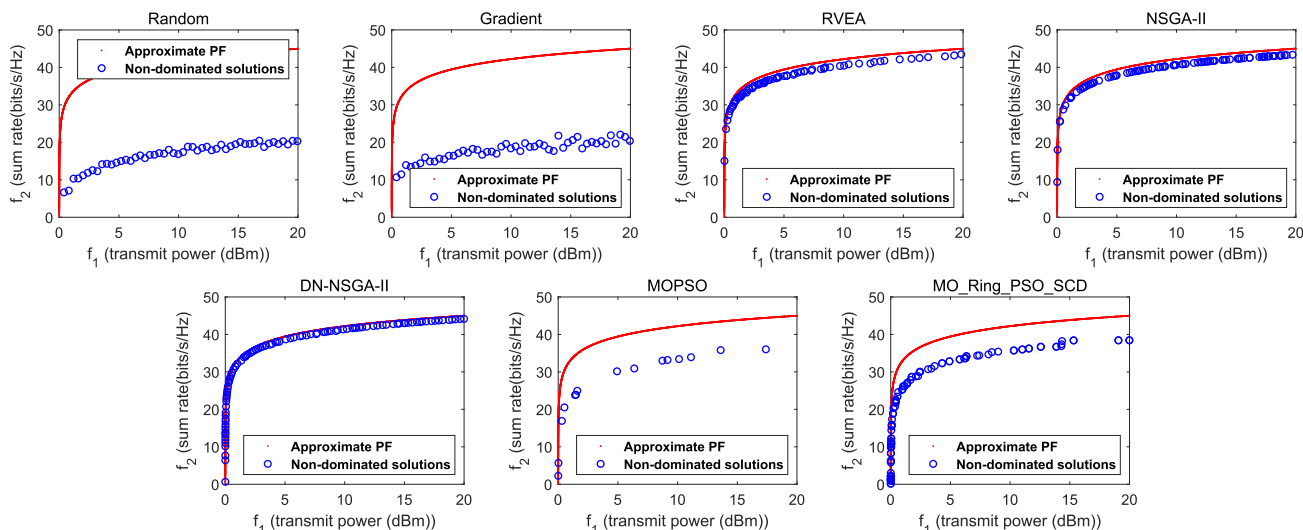
Fig. 4 Non-dominated solutions obtained by seven algorithms on $R=200$ ($b=2$, $K=4$ and SNR=2dB) in the run associated with the median fitness value

dramatically, since the RIS provides more accurate passive beamforming. This trend stops when b exceeds 2; that is, the gap of Pareto-front quality under b = 3,4,5 is very small. Hence, in practical implementation, b = 2 is enough to obtain satisfying solutions.

Scalability over different numbers of users K: The third part of Table 2 and Table 3 (Lines 14-19) depicts the scalability of the sum rate and transmit power trade-off over the number of users K with $R=64$, $b=2$ and SNR=2dB. As can be observed, when K increases, all methods perform better on both the transmit power and sum rate except the random method. This is due to the fact that both the gradient descent and population-based methods are able to exploit multi-user diversity. However, the random method fails to mitigate multi-user interference, which quickly leads to uncertainty of the performance.

Performance versus SNR: The fourth part of Table 2 and Table 3 (Lines 20–25) displays the performance of the sum rate and transmit power trade-off versus the SNR with $R=64$, $b=2$ and $K=4$. Including the baseline methods, all methods obtain higher HV and IGD results with higher SNR. It can be inferred that the sum rate increases with SNR since more power resources are allocated by the BS. RVEA, NSGA-II and DN-NSGA-II outperform all comparison methods, demonstrating the efficacy of population-based algorithms in hybrid beamforming design.

5.3 Computational complexity analysis

Below we analyze the computational complexity of the population-based algorithms in the case of the bi-objective optimization:

1. *RVEA:* The REVA consists of four important procedures, namely objective value translation, population partition, angle-penalized distance calculation, and elitism selection. The computational complexity of these four procedures is $O(2N)$, $O(2N^2)$, $O(N^2)$ and $O(2N^2)$, respectively (N is the population size). The dominant operator determines that the computational complexity of RVEA is $O(N^2)$.

2. *NSGA-II:* The NSGA-II consists of three important procedures, namely objective non-dominated sorting, crowding-distance assignment and sorting on the crowded-comparison operator. The computational complexity of the procedures is $O\left(2(2N)^2\right)$, $O(2(2N)\log(2N))$ and $O(2N\log(2N))$, respectively. Therefore, the total computational complexity of NSGA-II is $O(N^2)$, which is governed by the non-dominated sorting part of the algorithm (Deb et al 2002).

3. *DN-NSGA-II:* Compared with NSGA-II, the proposed double-niched strategy (Liu et al 2018) does not affect the whole computational complexity of the algorithm and thus is $O(N^2)$.

4. *MOPSO:* The MOPSO consists of three important procedures, namely objective function computation, local search, crowding distance computation and non-dominated comparison in the population and external archive. P, B and S are the numbers of particles, neighbors and size of the external archive, respectively. The computational complexity of the operators is $O(2P)$, $O(B)$, $O(2S\log S)$ and $O\left(2P^2\right)$. Collectively, the whole computational complexity of MOPSO within one generation is $O\left(P^2\right)$ (Coello and Lechuga 2002).

5. *MO_Ring_PSO_SCD:* The computational complexity of MO_Ring_PSO_SCD is same as that of MOPSO, i.e., $O\left(P^2\right)$ (Yue et al 2017).

It can be seen from the above analysis that the computational complexity of the three MOEAs and the two MMOEAs is roughly equivalent and acceptable, which mainly depends on the size of the population and the size of particles. Therefore, to avoid the problem of slow convergence that may be caused by particle size, algorithms that do not require grouping, namely RVEA, NSGA-II and DN-NSGA-II, can be preferentially selected, especially when the problem size is small.

5.4 Discussions

In summary, the three MOEAs and two MMOEAs exhibit good performance over benchmark algorithms in dealing with the challenges in our proposed hybrid beamforming. It indicates the promising potential of evolutionary multi-objective optimization algorithms in solving hybrid beamforming problems. We notice many interesting phenomena, according to which we discussed as follows:

1. Comparing the results of the seven algorithms, MMOEAs perform best, followed by MOEAs, then the gradient descent method, and the random method worst. MMOEAs outperform MOEAs since they incorporate strategies to handle multi-modal property, thus jumping out of the local optima and approaching the global optima. The performances of MOEAs are better than that of the gradient descent method. This is because the MOEAs are population-based and possess global search ability, encouraging communication between solutions. Therefore, the obtained solutions perform better in terms of convergence and diversity. The gradient descent method outperforms randomly choosing the passive beamforming since the search ability becomes stronger after considering the gradient information.
2. Comparing the results of three MOEAs (RVEA, NSGA-II and MOPSO) and two MMOEAs (DN-NSGA-II and MO_Ring_PSO_SCD), RVEA, NSGA-II and DN-NSGA-II perform best in most cases, followed by MO_Ring_PSO_SCD, and MOPSO worst. RVEA, NSGA-II and DN-NSGA-II outperform PSO-based and decomposition-based algorithms. This is due to the PSO-based and decomposition-based algorithms only finding a small number of feasible solutions clustered in several regions, and their poor performance could be attributed to their poor capabilities in diversity maintenance.
3. No matter how the parameters of the system change, the performances of NSGA-II and DN-NSGA-II are relatively satisfactory. If the algorithm precision is pursued,

it is recommended to use NSGA-II. When the system is not sensitive to the computational complexity, it is recommended to use the DN-NSGA-II algorithm to obtain better solutions in terms of convergence and diversity.

6 Conclusion

In this work, we investigate the hybrid beamforming design in a RIS-aided downlink MU-MISO system. We formulate a bi-objective optimization problem, i.e., to maximize the sum rate and minimize the transmit power simultaneously. To solve this tricky problem, we propose a novel MEHB framework to strike multiple trade-offs between the two conflicting objectives. To be specific, active beamforming at the base station (BS) and passive beamforming at the RIS are jointly optimized in the framework. We adopt three multi-objective evolutionary algorithms and two multimodal multi-objective evolutionary algorithms for passive beamforming design, while the sub-optimal active beamforming is obtained by the zero-forcing method. According to our simulation results, we demonstrate the potential that NSGA-II and DN-NSGA-II handle the discrete passive beamforming well compared with state-of-the-art methods.

For future works, we intend to investigate the potential problems below to make up for the limitations of existing studies.

1. Considering the parallel nature of evolutionary algorithms, we suggest accelerating MOEAs by parallel implementation if there is a high demand for running time.
2. We will further handle the deployment of RIS and improve the proposed method with a given suitable RIS location.

Availability of data and materials The data underlying this study are available from the corresponding author upon reasonable request.

Declarations

Ethical approval This article does not contain any studies with human participants performed by any of the authors.

Conflict of interest The authors declare that they have no conflict of interest.

References

Bakkouri I, Afdel K (2020) Computer-aided diagnosis (cad) system based on multi-layer feature fusion network for skin lesion recognition in dermoscopy images. Multimedia Tools Appl 79(29):20,483-20,518

Bakkouri I, Afdel K (2022) Mlca2f: Multi-level context attentional feature fusion for covid-19 lesion segmentation from ct scans. Signal Image Video Process, pp 1–8

Chen HT, Taylor AJ, Yu N (2016) A review of metasurfaces: physics and applications. Rep Prog Phys 79(7):076,401

Chen J, Liang YC, Pei Y, et al (2019) Intelligent reflecting surface: a programmable wireless environment for physical layer security. IEEE Access 7:82599–82612

Cheng R, Jin Y, Olhofer M et al (2016) A reference vector guided evolutionary algorithm for many-objective optimization. IEEE Trans Evol Comput 20(5):773–791. https://doi.org/10.1109/TEVC.2016.2519378

Coello CC, Lechuga MS (2002) Mopso: A proposal for multiple objective particle swarm optimization. In: Proceedings of the 2002 congress on evolutionary computation. CEC'02 (Cat. No. 02TH8600), IEEE, pp 1051–1056

Deb K, Tiwari S (2005) Omni-optimizer: A procedure for single and multi-objective optimization. In: International conference on evolutionary multi-criterion optimization, Springer, pp 47–61

Deb K, Pratap A, Agarwal S et al (2002) A fast and elitist multiobjective genetic algorithm: Nsga-ii. IEEE Trans Evol Comput 6(2):182–197. https://doi.org/10.1109/4235.996017

Di B, Song L, Li Y (2016) Sub-channel assignment, power allocation, and user scheduling for non-orthogonal multiple access networks. IEEE Trans Wireless Commun 15(11):7686–7698

Di B, Zhang H, Song L et al (2020) Hybrid beamforming for reconfigurable intelligent surface based multi-user communications: Achievable rates with limited discrete phase shifts. IEEE J Select Areas Commun 38(8):1809–1822

Gong YJ, Chen WN, Zhan ZH et al (2015) Distributed evolutionary algorithms and their models: a survey of the state-of-the-art. Appl Soft Comput 34:286–300

Guo H, Liang YC, Chen J et al (2020) Weighted sum-rate maximization for reconfigurable intelligent surface aided wireless networks. IEEE Trans Wireless Commun 19(5):3064–3076

He C, Tian Y, Jin Y et al (2017) A radial space division based evolutionary algorithm for many-objective optimization. Appl Soft Comput 61:603–621

He C, Cheng R, Zhang C et al (2020) Evolutionary large-scale multiobjective optimization for ratio error estimation of voltage transformers. IEEE Trans Evol Comput 24(5):868–881

Huang C, Zappone A, Debbah M et al (2018) Achievable rate maximization by passive intelligent mirrors. 2018 IEEE Int Conf Acoust. Speech and Signal Processing (ICASSP), IEEE, pp 3714–3718

Justesen PD (2009) Multi-objective optimization using evolutionary algorithms. Department of Computer Science, University of Aarhus, Denmark, p 33

Khalili A, Zargari S, Wu Q et al (2021) Multi-objective resource allocation for IRS-Aided SWIPT. IEEE Trans Wireless Commun Lett 10(6):1324–1328. https://doi.org/10.1109/LWC.2021.3065844

p Kramer O (2017) Genetic algorithms. In: Genetic algorithm essentials. Springer, pp 11–19

Li K, Deb K, Zhang Q et al (2015) An evolutionary many-objective optimization algorithm based on dominance and decomposition. IEEE Trans Evol Comput 19(5):694–716. https://doi.org/10.1109/TEVC.2014.2373386

Lin Q, Lin W, Zhu Z et al (2020) Multimodal multiobjective evolutionary optimization with dual clustering in decision and objective spaces. IEEE Tran Evol Comput 25(1):130–144

Liu Y, Ishibuchi H, Nojima Y, et al (2018) A double-niched evolutionary algorithm and its behavior on polygon-based problems. In: International conference on parallel problem solving from nature, Springer, pp 262–273

Ma X, Guo S, Zhang H et al (2021) Joint beamforming and reflecting design in reconfigurable intelligent surface-aided multi-user communication systems. IEEE Trans Wireless Commun 20(5):3269–3283. https://doi.org/10.1109/TWC.2020.3048780

Miettinen K (2012) Nonlinear multiobjective optimization, vol 12. Springer Science & Business Media

Mohseni-Bonab SM, Rabiee A, Mohammadi-ivatloo B (2015) Voltage stability constrained multi-objective optimal reactive power dispatch under load and wind power uncertainties: A stochastic approach. Renew Energy 85:598–609. https://doi.org/10.1016/j.renene.2015.07.021

Nadeem QUA, Kammoun A, Chaaban A, et al (2019) Intelligent reflecting surface assisted wireless communication: modeling and channel estimation. arXiv preprint arXiv:1906.02360

Ray T, Mamun MM, Singh HK (2022) A simple evolutionary algorithm for multi-modal multi-objective optimization. arXiv preprint arXiv:2201.06718

Schaffer J (1985) Multiple objective optimization with vector evaluated genetic algorithms, pp 93–100

Schutze O, Vasile M, Coello CAC (2011) Computing the set of epsilon-efficient solutions in multiobjective space mission design. J Aerospace Comput Inf Commun 8(3):53–70

Semenkin E, Semenkina M (2012) Self-configuring genetic algorithm with modified uniform crossover operator. In: International conference in swarm intelligence, Springer, pp 414–421

Tanabe R, Ishibuchi H (2018) A decomposition-based evolutionary algorithm for multi-modal multi-objective optimization. In: International conference on parallel problem solving from nature, Springer, pp 249–261

Tse D, Viswanath P (2005) Fundamentals of wireless communication. Cambridge University Press, Cambridge

While L, Hingston P, Barone L et al (2006) A faster algorithm for calculating hypervolume. IEEE Trans Evol Comput 10(1):29–38

Wu Q, Zhang R (2019) Beamforming optimization for intelligent reflecting surface with discrete phase shifts. ICASSP 2019–2019 IEEE Int Conf Acoust. Speech and Signal Processing (ICASSP), IEEE, pp 7830–7833

Wu Q, Zhang R (2019) Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming. IEEE Trans Wireless Commun 18(11):5394–5409

Wu Q, Zhang R (2019) Towards smart and reconfigurable environment: intelligent reflecting surface aided wireless network. IEEE Commun Mag 58(1):106–112

Wu Q, Zhang S, Zheng B et al (2021) Intelligent reflecting surface-aided wireless communications: a tutorial. IEEE Trans Commun 69(5):3313–3351

Yan B, Zhao Q, Zhang J, et al (2021) Hybrid beamforming for RIS-aided communications: fitness landscape analysis and niching genetic algorithm. arXiv e-prints arXiv:2109.09054. [cs.NE]

Yu X, Xu D, Schober R (2019) MISO wireless communication systems via intelligent reflecting surfaces. In: 2019 IEEE/CIC international conference on communications in China (ICCC), IEEE, pp 735–740

Yue C, Qu B, Liang J (2017) A multiobjective particle swarm optimizer using ring topology for solving multimodal multiobjective problems. IEEE Trans Evol Comput 22(5):805–817

Zhang Q, Li H (2007) Moea/d: a multiobjective evolutionary algorithm based on decomposition. IEEE Trans Evol Comput 11(6):712–731. https://doi.org/10.1109/TEVC.2007.892759

Zhang X, Tian Y, Cheng R et al (2015) An efficient approach to non-dominated sorting for evolutionary multiobjective optimization. IEEE Trans Evol Comput 19(2):201–213. https://doi.org/10.1109/TEVC.2014.2308305

Zhang X, Zheng X, Cheng R et al (2018) A competitive mechanism based multi-objective particle swarm optimizer with fast convergence. Inf Sci 427:63–76. https://doi.org/10.1016/j.ins.2017.10.037 www.sciencedirect.com/science/article/pii/S0020025517310344

Zhou A, Jin Y, Zhang Q, et al (2006) Combining model-based and genetics-based offspring generation for multi-objective optimization using a convergence criterion. In: 2006 IEEE international conference on evolutionary computation, IEEE, pp 892–899

Zhou A, Qu BY, Li H et al (2011) Multiobjective evolutionary algorithms: a survey of the state of the art. Swarm Evol Comput 1(1):32–49