Multiobjective Bayesian optimization for the shape design of rf cavity in particle accelerators

Yanhong Wang, Yungai Tang,* Cong-Feng Wu, and Guangyao Feng†

National Synchrotron Radiation Laboratory, University of Science and Technology of China, Hefei, Anhui, China

(Received 2 September 2025; accepted 11 February 2026; published 9 March 2026)

The design of radio frequency (rf) cavities in particle accelerators constitutes a multiobjective optimization problem (MOOP) characterized by strict equality constraints and computationally expensive evaluations. Here, a multiobjective Bayesian optimization (MOBO) framework is applied to rf cavity design for the first time to efficiently identify high-quality Pareto-optimal solutions under limited simulation budgets. To maintain high sample efficiency of the MOBO framework in problems with extremely strict equality constraints, two constraint-handling strategies are investigated: a penalty-based scalar acquisition function (sAF) strategy and a novel two-stage acquisition strategy (tAS). The tAS divides the sampling process into two stages. In the first stage, a vectorized acquisition function (vAF) is combined with a multiobjective evolutionary algorithm (MOEA) to efficiently search for candidate points in the full high-dimensional objective space. In the second stage, a sAF is employed to select the most promising point from candidate points for actual evaluation. Results on a test problem and the design of a practical 499.65 MHz rf cavity show that tAS outperforms penalty-based sAF in handling equality constraints within the MOBO framework. Furthermore, by using the MOBO framework with tAS, high-quality Pareto-optimal solutions are obtained within 1000 simulations, demonstrating a significant advantage of efficiency over conventional rf cavity optimization algorithms. Additionally, tAS allows for the flexible adjustment of MOEA, vAF, and sAF configurations, thereby providing flexible adaptation to other MOOPs in particle accelerators.

DOI: 10.1103/mrr8-z48f

I. INTRODUCTION

Radio frequency (rf) cavities have played an important role in particle accelerators since the advent of resonant acceleration techniques in the 1920s [1], with a wide range of applications in x-ray free-electron lasers [2], synchrotron radiation sources [3], neutron source [4], Accelerator driven systems [5], and so on. The accelerator systems set high demands on the performance of rf cavities to ensure stable and efficient acceleration of the beam. Therefore, it is important to optimize the cavity's geometric structure to simultaneously optimize multiple performance objectives, which are typically competing and subject to strict equality constraints. For example, the accelerating frequency f_{FM} needs to be strictly equal to a given target frequency, the shunt impedance usually has to be maximized, and the interaction of higher order modes (HOMs) with the beam minimized [6]. These objectives are evaluated using computationally complex electromagnetic (EM) simulation tools such as HFSS [7] and CST [8]. Consequently, rf cavity optimization is a multiobjective optimization problem (MOOP) with strict equality constraints and computationally expensive evaluations.

Traditional design approaches rely on manually tuning several geometric parameters and iteratively scanning the decision space through trial-and-error until a "good enough" geometry structure is obtained. This process is inefficient and dependent on expert experience. Therefore, investigating advanced multiobjective optimization algorithms for rf cavity design is of great importance. Such methods can improve both optimization efficiency and solution quality. There are three requirements for the algorithm for rf cavity optimization: (i) effectively balancing trade-offs among competing objectives; (ii) developing constraint handling method to identify the narrow feasible region which is formed by points that satisfy the constraints, (iii) reducing the number of EM simulations to conserve computational resources. However, identifying the narrow feasible region usually requires numerous EM simulations, making the design of an efficient optimization method challenging.

*Contact author: tygai@ustc.edu.cn.
†Contact author: fenggy@ustc.edu.cn.

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

Evolutionary algorithms (EAs) [9,10] have been extensively investigated for MOOPs. Various constraint-handling methods have been developed to address constrained optimization problems in EAs. These methods include penalty function methods (such as static [11], dynamic [12], death, and adaptive [13,14] penalties), feasibility-based rules [15], stochastic ranking [16], the ε-constrained method [17], multiobjective approaches [18], hybrid strategies [19], two- or multistage mechanisms [20,21], and cooperative frameworks [22]. The nondominated sorting genetic algorithm II (NSGA-II) [23], a classic multiobjective evolutionary algorithm (MOEA) [24], has been widely applied to rf cavity optimization [6,25–31]. Typically, tens of thousands of EM simulations are required when using NSGA-II. Methods for handling equality constraints of f_{FM} include penalty and tuning. Details of the rf cavity optimization algorithms are summarized in Table I. Although NSGA-II demonstrates good solution quality, it typically requires a large population size and a great number of generations to locate the feasible region and to converge to the ideal Pareto front. Each individual in the population must be evaluated by complex EM simulations, which increase the computational cost and time consumption.

To improve optimization efficiency, machine-learning-enhanced NSGA-II methods have been applied in the accelerator field [33–36]. The most common strategy is to employ machine-learning surrogate models as replacements for computationally expensive high-fidelity physics simulations, thereby significantly reducing the overall optimization cost [33,35,36]. Reference [35] employs a machine-learning surrogate model as a full replacement for high-fidelity physics simulation during the optimization process. Reference [36] proposes a neural-network-based multiobjective genetic algorithm. Reference [33] proposes the dynamically used neural-network-based multiobjective genetic algorithm (DNMOGA).

Several studies [32,33,37–39] have shown that surrogate models can also be effectively applied to rf cavity optimization and DNMOGA [33] achieves remarkable progress. In DNMOGA, a highly accurate neural network needs to be pretrained on a set of EM simulation samples and then dynamically integrated into the NSGA-II algorithm. The accuracy of the surrogate model has a considerable influence on the performance of the DNMOGA. However, training a highly accurate neural network typically requires a large number of EM simulation samples (for example, 1000 samples are required during the pretraining phase in Ref. [33]). The total number of EM simulations and the constraint-handling methods involved in this approach are also summarized in Table I. Compared to the conventional NSGA-II algorithm, the DNMOGA significantly reduces the number of required EM simulations.

Further reducing the number of EM simulations is necessary to save computational resources and enhance optimization efficiency in rf cavity optimization. Bayesian optimization (BO) [40–44] has shown good performance in diverse fields due to its exceptional sample efficiency, including materials discovery [45], hyperparameter tuning [46], and accelerator physics [47–52]. By constructing probabilistic surrogate models and designing acquisition strategy (typically formulated as an acquisition function), BO efficiently converges to the optimal solution with just a few number of sequential evaluations. For example, BO demonstrates better performance than the approaches based on machine-learning surrogate model in online accelerator tuning [35,50]. A recent review article about BO algorithms in accelerator physics is available in Ref. [51]. Although BO has been widely applied to accelerator tuning and rf-related problems [50,53–55], to the best of our knowledge, no prior work has applied BO to the shape design of rf cavity, which is characterized by a strict equality constraint. The motivation of this research is to study the application of multiobjective Bayesian optimization (MOBO) [56] to the shape design of rf cavity to further improve optimization efficiency.

However, common acquisition functions in MOBO, such as expected improvement (EI) and upper confidence bound (UCB), are mainly designed to maximize or minimize objective functions and lack the capability to handle strict equality constraints. To address inequality constraints, constraint-aware acquisition functions have been applied to guide sampling toward the feasible region during the

TABLE I. Summary of rf cavity optimization algorithms.

	Algorithm	EM simulations	Frequency constraint
rf gun for Korea-4GSR [28]	NSGA-II	120 000	`pymoo.algorithms.moo.nsga2.NSGA2`
rf cavity for Korea-4GSR [28]	NSGA-II	60 000	`pymoo.algorithms.moo.nsga2.NSGA2`
ALS-U [30]	NSGA-II	36 000	...
SLS-2 [6]	NSGA-II	20 000	A nondeath-penalty approach
FCC-ee [26]	MOEA	6000	Tuning by varying a parameter
Double spoke cavity [32]	DNMOGA	4000	Dynamical penalty, Feasibility-based rules
Spherical-shaped cavity [33]	DNMOGA	3000	Dynamical penalty, Feasibility-based rules, Tuning by varying a parameter

optimization process [50]. Equality constraints can be transformed into inequality constraints by allowing a very small tolerance value, but the effectiveness of these methods in optimization problems with extremely narrow feasible region has not been validated. Therefore, applying MOBO to rf cavity optimization directly is challenging, and it is important to develop efficient equality constraint-handling strategies suitable for MOBO.

In this paper, we investigate the application of MOBO to the shape design of rf cavity, with a particular focus on equality constraint-handling methods. A novel approach that uses a two-stage acquisition strategy to handle strict equality constraints within the MOBO framework is presented. The paper is organized as follows. Section II reviews the basic MOBO framework. Section III explores several constraint-handling strategies in MOBO and introduces a novel two-stage acquisition strategy (tAS). Section IV applies MOBO to the optimization of a 499.65 MHz normal-conducting rf cavity and evaluates the performance of MOBO with tAS through a series of experiments. Section V concludes the paper.

II. MULTIOBJECTIVE BAYESIAN OPTIMIZATION

A. Single-objective Bayesian optimization

A maximization problem with a single- objective function can be formulated as follows:

$$\begin{aligned} \max \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x} \in \Omega \end{aligned}. \tag{1}$$

Here, $\mathbf{x}$ denotes an n-dimensional decision variable in the decision space Ω, and $f(\mathbf{x})$ is the objective function to be maximized. If the original problem is a minimization problem, it can be equivalently converted to a maximization problem by negating the objective function.

The BO framework generally consists of two core components: a probabilistic surrogate model and an acquisition function. The workflow of the BO framework is presented in Algorithm 1.

Probabilistic surrogate models are employed as computationally efficient alternatives to computationally expensive blackbox objective functions. These models can provide not only predictions of the objective functions but also uncertainty estimates of their predictions. A commonly used probabilistic surrogate model in BO is Gaussian process (GP) model, which can capture the similarity between data points using kernel functions, enabling accurate predictions even with relatively few training samples. It models the objective function as a stochastic process, i.e., $f(\mathbf{x}) \sim \mathcal{GP}[\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')]$, where $\mu(\mathbf{x})$ and $k(\mathbf{x}, \mathbf{x}')$ denote the mean function and the covariance (or kernel) function, respectively. Given a set of observations $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$, the prior GP can be updated via Bayes' theorem to obtain the posterior distribution $f(\mathbf{x})|\mathcal{D} \sim \mathcal{GP}[\mu_{\mathbf{post}}(\mathbf{x}), k_{\mathbf{post}}(\mathbf{x}, \mathbf{x}')]$, where $\mu_{\mathbf{post}}(\mathbf{x})$ and $k_{\mathbf{post}}(\mathbf{x}, \mathbf{x}')$ are the posterior mean and kernel functions, respectively. This posterior enables both point-wise prediction of the objective function and estimation of predictive uncertainty across the input space.

Algorithm 1. Bayesian optimization framework.

Input: Initial dataset size N_{init}, maximum number of iterations N_{iter}
Output: Best y recorded after optimization
Randomly sample an initial dataset $D_0 = \{X, y\}$;
for $t \leftarrow 0$ **to** N_{iter} **do**
  Construct a GP regression model with D_t;
  $x_t \leftarrow \arg\max_x \alpha(x; D_t)$;
  $y_t \leftarrow f(x_t)$;
  $D_{t+1} \leftarrow D_t \cup \{(x_t, y_t)\}$;
return best y

The acquisition function $\alpha(\mathbf{x})$ is defined based on the posterior distribution of the GP to select the most promising point for true evaluation. By maximizing the acquisition function, the most promising point is selected. The maximization of acquisition functions typically employs gradient-based optimization methods (e.g., L-BFGS, quasi-Newton methods) [57], discretization approaches [58], and EAs [40]. A well-designed acquisition function balances the exploration of unknown regions and the exploitation of known information to meet optimization needs. Two commonly used acquisition functions in BO are expected improvement (EI) and upper confidence bound (UCB).

EI quantifies the expected improvement over the best observed objective value $f_{\max}$ as follows:

$$\begin{aligned} \alpha(\mathbf{x}) &= E[\max(f(\mathbf{x}|\mathcal{D}) - f_{\max}, 0)] \\ &= (\hat{f}(\mathbf{x}) - f_{\max})\Phi\left(\frac{\hat{f}(\mathbf{x}) - f_{\max}}{\hat{s}(\mathbf{x})}\right) \\ &\quad + \hat{s}(\mathbf{x})\phi\left(\frac{\hat{f}(\mathbf{x}) - f_{\max}}{\hat{s}(\mathbf{x})}\right), \end{aligned} \tag{2}$$

where $\hat{f}(\mathbf{x})$ and $\hat{s}(\mathbf{x})$ denote the posterior predictive mean and standard deviation, respectively, and $\Phi(\cdot)$ and $\phi(\cdot)$ denote the cumulative distribution function and probability density function of the standard normal distribution.

UCB is defined as

$$\alpha(\mathbf{x}) = \hat{f}(\mathbf{x}) + \beta\hat{s}(\mathbf{x}), \tag{3}$$

where the exploration parameter β controls the trade-off between exploitation and exploration. A larger β encourages exploration, and a smaller β (even zero) favors exploitation.

B. Multiobjective Bayesian optimization

A maximization problem with m objectives can be formulated as

$$\begin{aligned} \max \quad & \mathbf{F}(\mathbf{x}) = [f_1(\mathbf{x}), \cdots, f_m(\mathbf{x})] \\ \text{s.t.} \quad & \mathbf{x} \in \Omega \end{aligned} \tag{4}$$

Here, $\mathbf{x}$ is an n-dimensional decision variable in the decision space Ω, and $\mathbf{F}$ maps $\mathbf{x}$ to m objective functions. If there exist minimization objectives, their corresponding objective functions $f_i(\mathbf{x})$ can be transformed into maximization problems by negating them.

In MOOPs, a solution consists of a solution vector, and since the objectives often compete with each other, no single solution can simultaneously optimize all objectives. As a result, the optimization result is typically a set of mutually nondominated solutions, called the Pareto-optimal set or Pareto-optimal solutions. The corresponding mapping of this set into the objective space is called the Pareto front.

To evaluate the quality of a Pareto front, the hypervolume (HV) is one of the most commonly used metrics. It measures the volume in the objective space bounded by the Pareto front and a reference point $\mathbf{z}$. A larger HV indicates a better Pareto front. Mathematically, it is defined as

$$HV(A) = \mathrm{Vol}\left(\bigcup_{\mathbf{x} \in A} [f_1(\mathbf{x}), z_1] \times \cdots \times [f_m(\mathbf{x}), z_m] \right) \tag{5}$$

where A denotes the nondominated solution set and $\mathbf{z} = (z_1, \cdots, z_m)$ is a reference point.

We now present the MOBO algorithm. For MOOPs with m objectives, each objective function is modeled as an independent GP model, and each GP model has its own kernel function. Since each solution in this case consists of a solution vector, a scalar acquisition function (sAF) α: $\mathcal{X} \to \mathbb{R}$ is constructed as a new indicator to proceed with multiobjective. Then, a single-objective optimizer can be employed to find the Pareto front by maximizing this indicator.

One approach to construct the scalar acquisition function is Pareto efficient global optimization, proposed by Knowles [59], using a weighted sum of the objective functions with random weights generated at each iteration. Another approach that scalarizes the acquisition function is hypervolume improvement (HVI), which can reframe the MOOP as a single-objective optimization by using a new objective that implicitly increases solution diversity. The HVI of a point $\boldsymbol{y}$ relative to the current Pareto front $\mathcal{P}$ is defined as

$$\mathrm{HVI}(\mathcal{P}, \boldsymbol{y}) = \mathrm{HV}(\mathcal{P} \cup \{\boldsymbol{y}\}) - \mathrm{HV}(\mathcal{P}). \tag{6}$$

Therefore, the EI and UCB can be extended to MOOPs. These extensions are defined as expected hypervolume improvement (EHVI) and upper confidence bound hypervolume improvement (UCB-HVI), respectively. The definitions are as follows:

$$\alpha_{\mathrm{EHVI}}(\mu, \boldsymbol{\sigma}, \mathcal{P}, \mathbf{r}) = \int_{\mathbb{R}^m} \mathcal{H}_I(\mathcal{P}, \mathbf{y}, \mathbf{r}) \cdot \xi_{\mu,\boldsymbol{\sigma}}(\mathbf{y}) d\mathbf{y}, \tag{7}$$

$$\alpha_{\text{UCB-HVI}}(\mu, \boldsymbol{\sigma}, \mathcal{P}, \mathbf{r}, \beta) = \mathcal{H}_I(\mathcal{P},\ \mu + \sqrt{\beta}\boldsymbol{\sigma}, \mathbf{r}), \tag{8}$$

where μ and $\boldsymbol{\sigma}$ are the posterior mean and standard deviation vectors from the GPs, $\mathcal{P}$ represents the current Pareto front approximation, $\mathbf{r}$ is the reference point, $\mathcal{H}_I(\cdot)$ denotes the hypervolume improvement function, $\xi_{\mu,\boldsymbol{\sigma}}(\mathbf{y})$ is the joint probability density function of the multivariate Gaussian distribution, and β is the exploration parameter.

III. STRATEGIES FOR HANDLING EQUALITY CONSTRAINTS IN MOBO

In practice, optimization problems usually comprise both objective functions (to be minimized or maximized) and constraint functions (including equality constraints and inequality constraints). The constrained MOOPs can be formulated as follows:

$$\begin{aligned} \max \quad & \mathbf{F}(\mathbf{x}) = [f_1(\mathbf{x}), \cdots, f_m(\mathbf{x})] \\ \text{s.t.} \quad & h_j(\mathbf{x}) = 0, \quad j = 1, \cdots, J \\ & g_k(\mathbf{x}) \le 0, \quad k = 1, \cdots, K \\ & \mathbf{x} \in \Omega \end{aligned} \tag{9}$$

where $h_j(\mathbf{x})$ denotes the jth equality constraint, and $g_k(\mathbf{x})$ denotes the kth inequality constraint. The nondominated solutions in the feasible region are called feasible Pareto-optimal set or feasible Pareto-optimal solutions. The corresponding mapping of this set into the objective space is called the feasible Pareto front.

Here, we investigate two acquisition strategies within the MOBO framework for addressing equality-constrained MOOPs. The strategies are as follows: (i) Penalty-based sAF strategy: A sAF with a penalty term is employed in this strategy to bias the sampling process toward the feasible region, which has proven effective in solving some inequality constraints with BO [60,61]. However, a sAF as a single indicator can only exploit limited information to guide the sampling process, which may hinder optimization efficiency under strict equality constraints and complex decision spaces. To address this potential limitation, we propose the following strategy. (ii) Two-stage acquisition strategy (tAS). A new acquisition strategy called two-stage acquisition strategy is employed. This strategy divides the sampling process into two stages. In the first stage, a vectorized acquisition function (vAF) is employed and a MOEA is used to search the feasible Pareto-optimal solutions defined by the vAF on GP models as candidate solutions. In the second stage, a sAF is employed to select the most promising solution in the candidate solutions, which is then used as the next point for actual evaluations.

This section provides a theoretical introduction of each strategy and presents experiments on a simple

equality-constrained MOOP with different levels of stringency to demonstrate the effectiveness of these strategies.

A. Penalty-based scalar acquisition function

The underlying principle of this strategy is to apply a penalty term $P_g(\mathbf{x})$ to the acquisition function, thereby guiding sampling points toward the feasible region by maximizing the penalized acquisition function:

$$\hat{\alpha}(\mathbf{x}) = \alpha(\mathbf{x}) P_g(\mathbf{x}). \tag{10}$$

One strategy is a truncated penalty-based sAF strategy [60]. This strategy defines a binary feasibility function:

$$P_g(\mathbf{x}) = \begin{cases} 1 & \mathbf{x} \in \mathcal{F}, \\ 0 & \text{otherwise}, \end{cases} \tag{11}$$

where $\mathcal{F}$ denotes the feasible region. This function ensures that the acquisition function is nonzero only in the feasible region.

The second strategy is a probabilistic penalty-based sAF strategy [61]. This strategy employs a soft constraint mechanism that allows sampling in both feasible and infeasible regions. In this case, $P_g(\mathbf{x})$ denotes the probability of satisfying the equality constraints:

$$P_g(\mathbf{x}) = \Pr[-\epsilon \le h(\mathbf{x}) \le \epsilon] = \int_{-\epsilon}^{\epsilon} p(h(\mathbf{x})|\mathcal{D}_h)\, \mathrm{d}h(\mathbf{x}) \tag{12}$$

In this function, $h(\mathbf{x})$ is the constraint function, ϵ denotes a very small tolerance value, and $p(h(\mathbf{x})|\mathcal{D}_h)$ denotes the predictive probability density of $h(\mathbf{x})$ under the current surrogate model. Under this function, points far from the feasible region receive low acquisition function values even if they offer considerable HVI due to their low probability of feasibility. Conversely, points near the feasible region tend to be favored, as their higher probability of feasibility compensates for their limited contribution to HVI.

B. Two-stage acquisition strategy

Using a single scalar indicator to represent multiple objectives essentially maps a multidimensional objective space onto a one-dimensional scale. This dimensionality reduction results in a loss of information, which in turn curtails the capacity of algorithm to balance objective improvements with constraint satisfaction. Therefore, the core limitation of the truncated penalty-based sAF and probabilistic penalty-based sAF stems from their reliance on a single sAF. To address this issue, we propose a two-stage acquisition strategy to better handle equality constraints in MOBO. The workflow of the MOBO with tAS is presented in Algorithm 2. The core components of tAS are as follows.

Algorithm 2. Workflow of MOBO with tAS.

Input: Initial dataset size d_0, maximum number of iterations T
Output: Final feasible Pareto-optimal solutions
Randomly sample d_0 points to construct initial dataset $\mathcal{D}_0$;
Train $m + J + K$ GP models on $\mathcal{D}_0$ for each objective;
for $t \leftarrow 1$ **to** T **do**
  Construct the vAF $\boldsymbol{\alpha}(\mathbf{x})$ as in Eq. (13);
  Search candidate points $\mathcal{P}_t$ with MOEA;
  Select $\mathbf{x}_t \in \mathcal{P}_t$ with the largest sAF value;
  Actual evaluate $f(\mathbf{x}_t)$;
  Update dataset: $\mathcal{D}_{t+1} \leftarrow \mathcal{D}_t \cup \{[\mathbf{x}_t, f(\mathbf{x}_t)]\}$;
  Retrain all GP models with updated dataset $\mathcal{D}_{t+1}$;
return Final feasible Pareto-optimal solutions

Vectorized acquisition function: To achieve thorough search in the full high-dimensional objective space, we formulate the acquisition function in vector form.

Considering the optimization problem defined in Eq. (9), which comprises m maximization objectives, J equality constraints, and K inequality constraints, the vAF is defined as a constrained MOOP:

$$\begin{aligned} \max \quad & \boldsymbol{\alpha}(\mathbf{x}) = [\alpha_1(\mathbf{x}), \cdots, \alpha_m(\mathbf{x})] \\ & \alpha_i(\mathbf{x}) = \hat{f}_i(\mathbf{x}) + \beta_i \hat{s}_i(\mathbf{x}), \quad i = 1, \cdots, m \\ \text{s.t.} \quad & \hat{h}_j(\mathbf{x}) = 0, \quad j = 1, \cdots, J \\ & \hat{g}_k(\mathbf{x}) \le 0, \quad k = 1, \cdots, K \\ & \mathbf{x} \in \Omega \end{aligned} \tag{13}$$

where $\hat{f}_i(\mathbf{x})$ and $\hat{s}_i(\mathbf{x})$ represent the predictions of the mean and standard deviation of the ith objective, respectively; $\hat{h}_j(\mathbf{x})$ denotes the prediction of the jth equality constraint; $\hat{g}_k(\mathbf{x})$ denotes the prediction of the kth inequality constraint; and β_i controls the balance between exploration and exploitation for the ith objective. It is worth noting that we adopt the UCB to define the objective function to be maximized in this case. In fact, other types of acquisition functions such as EI can also be employed to accommodate the characteristics of different problems. This flexibility allows the framework to adapt to exploration–exploitation trade-offs and problem-specific requirements.

MOEA-based search: We employ a MOEA to search the feasible Pareto front of the vAF defined in Eq. (13), which is then treated as candidate points for actual evaluations. The selection of the MOEA is motivated by its proven effectiveness in handling complex MOOPs, particularly those involving strict constraint functions and complex decision spaces. This stage not only achieves thorough search in the full high-dimensional objective space but also integrates the advantages of the MOEA with the high sample efficiency of BO. After this stage, the search space for the next stage is effectively localized to the region surrounding the feasible Pareto front.

sAF-Based point selection: Upon obtaining the candidate points, the one with the largest sAF value is selected

for the next actual evaluation. The sAF can be designed in different forms depending on the specific needs of the problem. For most MOOPs, UCB-HVI, and EHVI are recommended.

It is worth noting that there are several staged acquisition approaches in the literature, including feasibility-first Bayesian optimization, staged constraint handling, and other two-phase optimization strategies [44,62–70]. While our method is also stage based, it differs from these works in several important aspects. First, our method is designed for problems with extremely narrow feasible regions, where feasibility-first methods may be challenged due to the absence of initial feasible points. Second, we employ a vectorized acquisition function, which allows us to evaluate candidates by jointly considering multiple criteria rather than optimizing a single metric.

C. Experiments on a simple test problem

Here, we assess the effectiveness of the strategies discussed above in handling equality-constrained MOOPs through experimental analysis on a simple test problem with different levels of stringency. The test problem is defined as follows:

$$\begin{aligned} \min \quad & \mathbf{F}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x})] \\ & f_1(\mathbf{x}) = \|\mathbf{x} - \mathbf{1}\| \\ & f_2(\mathbf{x}) = \|\mathbf{x} + \mathbf{1}\| \\ \text{s.t.} \quad & h(\mathbf{x}) = |f_1(\mathbf{x}) - f_2(\mathbf{x})| = 0 \\ & -2 < x_n < 2, \quad n = 1, 2. \end{aligned} \tag{14}$$

The tolerance value in this problem is set to 0.1 and 0.01, respectively.

The experiment process and setup are detailed as follows. First, we extract five sample points using latin hypercube sampling (LHS) [71] and train the GP models. Next, in the penalty-based sAF strategy, we maximize the acquisition function using the particle swarm optimization (PSO) algorithm [72]. In the tAS, NSGA-II is adopted as the MOEA. Equality constraints are handled using a penalty approach in NSGA-II. The sAF is defined by Eq. (8). The population size and number of iterations for both PSO and NSGA-II are set to 50. As a comparison, an unconstrained MOBO is employed as the baseline, where the equality constraint is treated as an additional objective $f_3 = |h(\mathbf{x}) - 0|$ to be minimized. The results after 20 optimization iterations using MOBO with these strategies are illustrated in Fig. 1.

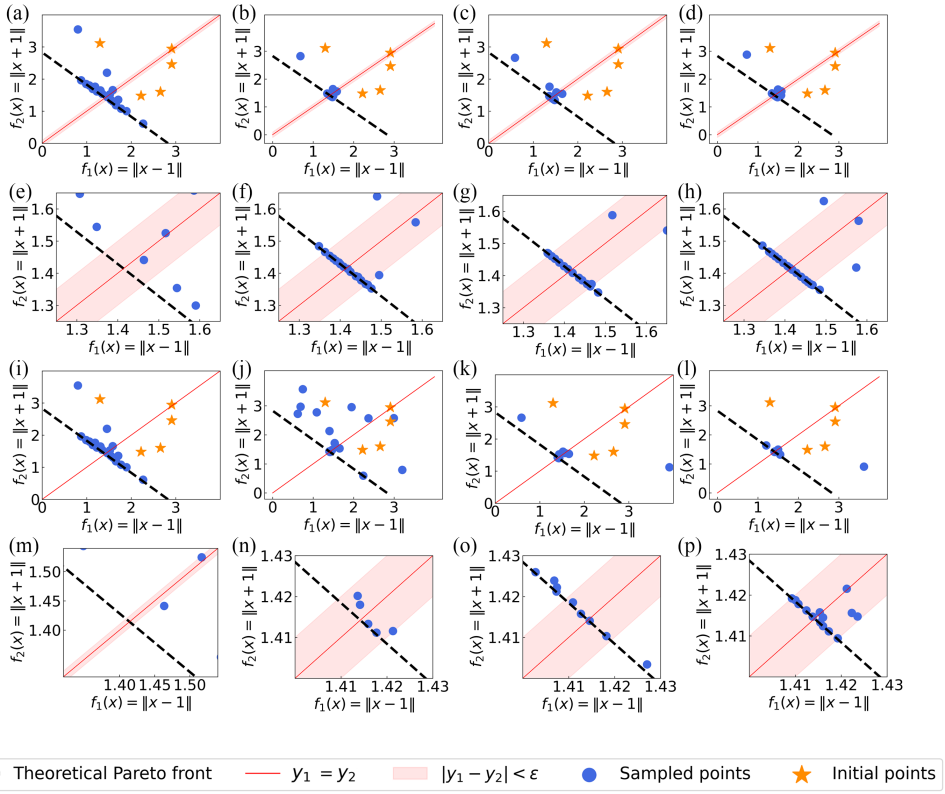

FIG. 1. Optimization results after 20 samples on the problem defined in Eq. (14). (a) Unconstrained MOBO with $\epsilon = 0.1$. (b) Truncated penalty-based sAF strategy with $\epsilon = 0.1$. (c) Probabilistic penalty-based sAF strategy with $\epsilon = 0.1$. (d) tAS with $\epsilon = 0.1$. (e) Magnified view of (a). (f) Magnified view of (b). (g) Magnified view of (c). (h) Magnified view of (d). (i) Unconstrained MOBO with $\epsilon = 0.01$. (j) Truncated penalty-based sAF strategy with $\epsilon = 0.01$. (k) Probabilistic penalty-based sAF strategy with $\epsilon = 0.01$. (l) tAS with $\epsilon = 0.01$. (m) Magnified view of (i). (n) Magnified view of (j). (o) Magnified view of (k). (p) Magnified view of (l).

Figures 1(a)–1(d) illustrate the sampling results when the tolerance value is set to 0.1. The regions near the feasible Pareto front are magnified in Figs. 1(e)–1(h). Compared to unconstrained MOBO, MOBO with all three strategies successfully identifies feasible Pareto-optimal solutions and effectively confines sampling to the feasible region.

Figures 1(i)–1(l) illustrate the sampling results when the tolerance value is set to 0.01. The equality constraint in this problem is stricter than when the tolerance is set to 0.1. We magnify the regions near the feasible Pareto front in Figs. 1(m)–1(p). Among the three strategies, only tAS effectively reduces sampling in the infeasible region and successfully identifies a large number of feasible Pareto-optimal solutions.

In conclusion, our experiment results show the imperative of incorporating a constraint-handling method into the BO framework for effectively solving strictly constrained MOOPs. All of the three strategies perform well under general situations. Compared to penalty-based sAF strategies, tAS is more efficient when the feasible region is extremely narrow.

IV. APPLICATION TO RF CAVITY OPTIMIZATION

A 499.65 MHz normal-conducting spherical- shaped (SS) cavity, which has previously been adopted to evaluate the performance of the DNMOGA algorithm [33], was employed in this study to assess the effectiveness of MOBO. The geometric structure of the SS cavity is illustrated in Fig. 2. With the radius (R_t) and length (L_t) of the beam tube fixed, 15 geometric parameters were selected as decision variables.

According to Ref. [33], the objective of the optimization is to tune these 15 geometric parameters to (i) maintain f_{FM} of the cavity within 499.65 MHz $\pm$ 0.05 MHz; (ii) maximize the R/Q of the fundamental mode (R/Q_{FM}) and the corresponding shunt impedance (R_a); (iii) increase the frequency separation between the first HOM and the

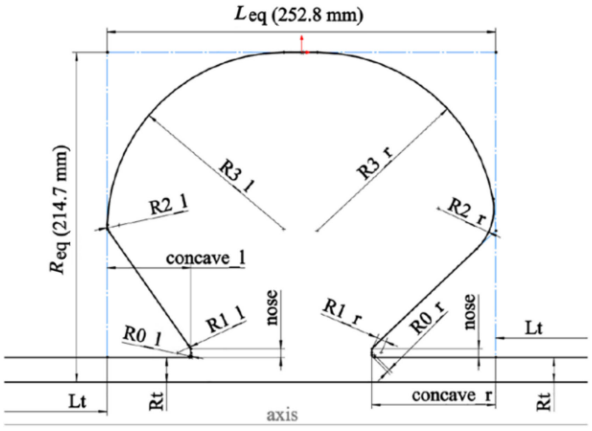

FIG. 2. Geometric structure of the 499.65 MHz normal-conducting SS cavity. Reproduced from Ref. [33].

TABLE II. Optimization objectives.

Objective type	Description
Constraint	$f_{\mathrm{FM}} = 499.65 \pm 0.05$ MHz
Maximization	Maximize R/Q_{FM} Maximize $f_{\mathrm{HOM}} - f_{\mathrm{FM}}$ Maximize R_a
Minimization	Minimize R/Q_{HOM}

fundamental mode ($f_{\mathrm{HOM}} - f_{\mathrm{FM}}$); and (iv) minimize the R/Q of the first HOM (R/Q_{HOM}). The specific optimization objectives are summarized in Table II.

The complete workflow is illustrated in Fig. 3(a). First, an initial dataset D is constructed by randomly generating d sets of samples in the decision space via LHS and the corresponding objective values are obtained by EM simulations. The dataset D is then used as the training set for building individual GP surrogate models for each objective. Second, an acquisition strategy is used to select one or more points P. Third, the points P are evaluated by EM simulations, and their frequency is tuned according to empirical formulas to generate new design points. We adopt the same tuning method as used in Ref. [33], assuming a linear physical relationship between R_{eq} and f_{FM}, and empirically setting the coefficient to 2.55 MHz/mm. It is worth noting that the empirical formula used here is specific to the 500 MHz SS cavity and cannot be directly applied to other cavities. For other rf cavities, an appropriate empirical formula suited to the corresponding cavity should be used. These new points are then also evaluated by EM simulations. Finally, all points evaluated by EM simulations are incorporated into dataset D to update the surrogate models. This process is repeated until a termination condition is met.

In this workflow, the different acquisition strategies differ at the "Select points P by AS" step [pink in Fig. 3(a)]. For the penalty-based sAF strategy, points P are selected by directly maximizing the sAF. For the tAS, it searches directly in the full high-dimensional objective space [as illustrated by the red dashed box in the upper part of Fig. 3(b)] during its first stage, thereby avoiding the inherent information loss in scalar compression. With the integration of a MOEA that enhances both the efficiency and effectiveness of the search within this complex space, the algorithm guides the optimization toward regions near the feasible Pareto front, thus effectively narrowing the sampling space to a refined, near-optimal region [as illustrated by the red dashed box in the lower part of Fig. 3(b)]. In the second stage, a sAF is employed to select the point with the highest improvement potential within this refined space as the actual evaluation point.

In conclusion, the general algorithmic framework of the tAS divides the sampling process into two stages. In the first stage, an advanced MOEA is used to search the

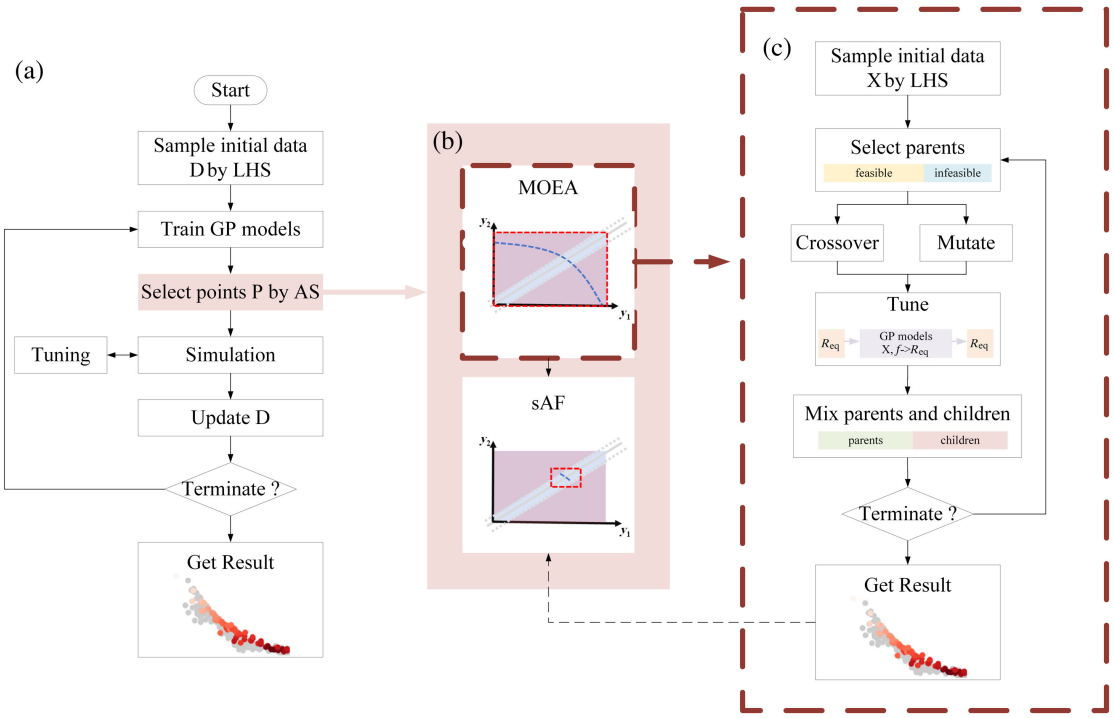

FIG. 3. Overview of the complete workflow: (a) The MOBO framework, (b) procedure of the tAS in MOBO, and (c) procedure of the MOEA in tAS.

feasible Pareto-optimal solutions defined by the vAF on GP models as candidate solutions. In the second stage, a sAF is employed to select the most promising solution. The detailed configuration of the MOEA, vAF, and sAF should be selected according to the specific problem, which provides substantial flexibility to our method (detailed discussion can be found in Sec. IV B). However, improper configuration of the MOEA, vAF, or sAF may adversely affect the algorithm's performance.

In this study, the optimization algorithm was implemented in MATLAB, and the selected sample points were evaluated using CST. The mesh is set as 20 cells per wavelength. All experiments were conducted on an AMD Ryzen 7 1700 Eight-Core Processor (3.00 GHz).

A. Comparison of equality constraint strategies in MOBO for rf cavity design

The strategies investigated above are applied to the design of the 499.65 MHz SS cavity to evaluate their effectiveness in rf cavity optimization.

In our experiments, all geometric parameters were normalized and treated as independent decision variables during the optimization process. A total of 100 initial samples were generated. An anisotropic radial basis function (RBF) kernel was adopted to construct the GP surrogate models. The models were trained using MATLAB's `fitrgp()` function. The GP surrogate models were updated after every ten new samples were collected. This iterative process was repeated until a total of 500 EM simulations were completed.

In the penalty-based sAF strategy, PSO is used to maximize the acquisition function. In the tAS, NSGA-II is used to optimize the vAF. The crossover and mutation rates are set to 1.0 and 0.4, respectively. To satisfy the equality constraint of f_{FM}, penalty mechanism [73] and a sorting strategy based on relative feasibility degree [15] are employed and the sorting preference is given to feasible solutions (feasibility-first strategy). For both PSO and NSGA-II, the population size and the number of iterations are set to 100. UCB-HVI offers faster computation than EHVI by avoiding costly posterior integration, particularly in high-dimensional objective spaces [50]. Therefore, we adopt UCB-HVI for all hypervolume-related computations in our experiments.

After 500 EM simulations, the feasible Pareto front obtained by each strategy is illustrated in Fig. 4. It can be seen that the feasible Pareto front obtained by the truncated penalty-based sAF strategy exhibits poor performance in both quantity and convergence. In contrast, the feasible Pareto front obtained by the tAS demonstrates superior performance.

B. Configuration of tAS

As shown in the previous subsection, MOBO with tAS achieved the best optimization results. The detailed settings of the tAS, including the design of the MOEA, the vAF, and

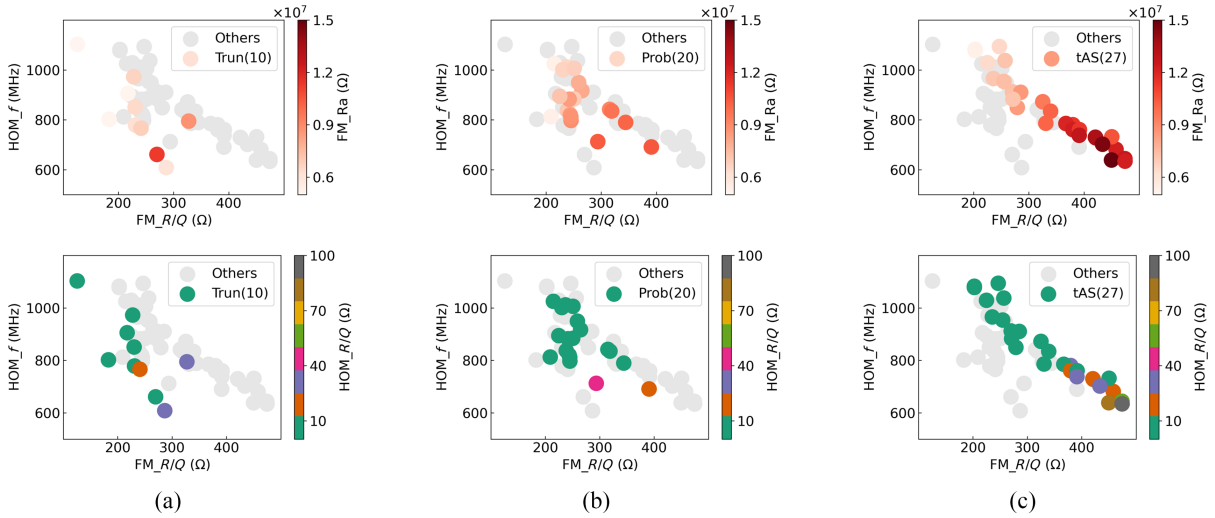

FIG. 4. Feasible Pareto fronts obtained by different strategies after 500 EM simulations. (a) Truncated penalty-based sAF strategy, (b) probabilistic penalty-based sAF strategy, and (c) tAS.

the sAF components, may affect the optimization results. Here, we explore how to adapt the tAS to better meet the optimization requirements.

1. MOEA and vAF adjustment

We begin by adjusting the MOEA. First, we implement frequency tuning. Reference [33] performed tuning by modifying R_{eq} with an empirical formula in the mutation operator. The empirical formula may not be applicable to the tuning on the GP model. Therefore, we construct a surrogate model to directly learn the mapping between frequency and R_{eq}, which is updated at each training iteration. After each crossover operation and mutation operation, it is directly used to perform frequency tuning, as illustrated in the tune step of Fig. 3(c). Second, we retain infeasible solutions. Reference [74] demonstrates that retaining and utilizing a portion of infeasible solutions through archiving can facilitate the exploration of better solutions. Therefore, in this study, we incorporated a strategy that retains partially infeasible solutions, as illustrated in the Select parents step of Fig. 3(c). Third, we increase the number of iterations. Since retaining infeasible solutions may slow down the convergence of the MOEA, we set the maximum number of search iterations to 400.

Then, we adjust the vAF while retaining the above adjustments. For rf cavities of similar types, the fundamental mode typically achieves an R/Q value above 200 Ω, and the frequency separation between the fundamental mode and the first HOM is usually greater than 200 MHz. Therefore, we can further impose these additional inequality constraints in the vAF to limit the objective space.

It should be noted that the above settings for the vAF are based on expert-informed priors specific to the 500 MHz SS cavity. For other cavity geometries or frequency ranges, appropriate priors can generally be derived and should be adopted; otherwise, inappropriate priors may mislead the optimization process and degrade performance.

We retain the same MOEA configuration as in Fig. 4(c) and perform 1000 sampling iterations as the baseline, as shown in Fig. 5(a). The optimization results with the MOEA adjustment are shown in Fig. 5(b). The optimization results with the MOEA and vAF adjustment are shown in Fig. 5(c). All three methods use UCB-HVI as the sAF. It can be seen that all of these methods are capable of obtaining high-quality solutions. The solutions in Fig. 5(b) exhibit improved convergence and diversity over the baseline. Figure 5(c) exhibits the best convergence within the constrained space.

Since our goal is to validate the algorithm's search capability across the entire objective space, we use the configuration of tAS corresponding to Fig. 5(b) for subsequent experiments. It should be noted that the subsequent experiment, which does not include the additional inequality constraints in the vAF, still achieve competitive results. This indicates that the core approach is not highly sensitive to these particular priors.

2. sAF adjustment

We investigate three types of sAF here. The first sAF focuses on constraint satisfaction (Con-First). If the goal is to obtain as many feasible solutions as possible, the sAF can be defined as the probability of constraint satisfaction:

$$\begin{aligned}\alpha_{\text{Con-First}}(\mathbf{x}) &= \Pr\left[f_0 - \epsilon \le f_{\mathrm{FM}}(\mathbf{x}) \le f_0 + \epsilon\right] \\ &= \int_{f_0-\epsilon}^{f_0+\epsilon} p(f_{\mathrm{FM}}|\mathcal{D}_{f_{\mathrm{FM}}})\, df_{\mathrm{FM}}, \end{aligned} \tag{15}$$

in this expression, $f_0 = 499.65$ MHz, $\epsilon = 0.05$ MHz. The sAF $\alpha_{\text{Con-First}}(\mathbf{x})$ computes the probability that the predicted

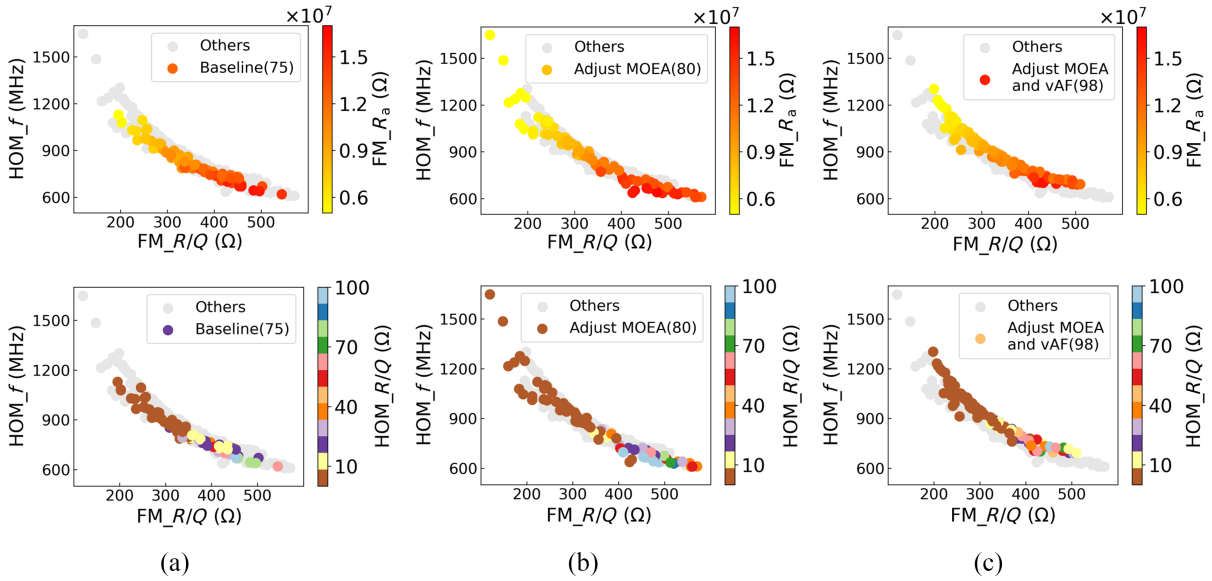

FIG. 5. Feasible Pareto fronts obtained by different configuration of tAS after 1000 samples. (a) Baseline, (b) adjust MOEA, and (c) adjust MOEA and vAF.

frequency $f_{\mathrm{FM}}(\mathbf{x})$ falls within the acceptable range $[f_0-\epsilon, f_0+\epsilon]$, based on the GP model trained on the current dataset $\mathcal{D}_{f_{\mathrm{FM}}}$. The second sAF focuses on objective optimization (HVI-First). The specific form of the sAF $\alpha_{\text{HVI-First}}(\mathbf{x})$ adopted here is UCB-HVI, which is the same as the setting in Fig. 5. The third sAF balances both aspects by adopting the probabilistic penalty-based sAF strategy (Con-HVI), which simultaneously considers constraint satisfaction and objective optimization:

$$\alpha_{\text{Con-HVI}}(\mathbf{x}) = \alpha_{\text{Con-First}}(\mathbf{x}) \times \alpha_{\text{HVI-First}}(\mathbf{x}). \tag{16}$$

The optimization results of these three methods are shown in Figs. 6(a)–6(c). Figure 6(a) explores the largest number of feasible Pareto-optimal solutions but suffers from poor convergence and diversity. Figure 6(b) achieves the best convergence and diversity but the number of feasible Pareto-optimal solutions is the least. Figure 6(c) yields a sufficient number of feasible Pareto-optimal solutions while covering most of the feasible region, with only a few sparse regions remaining unexplored.

The results in Figs. 5 and 6 demonstrate that the design of the MOEA, the vAF, and the sAF affects the optimization results. Specifically, carefully tailoring

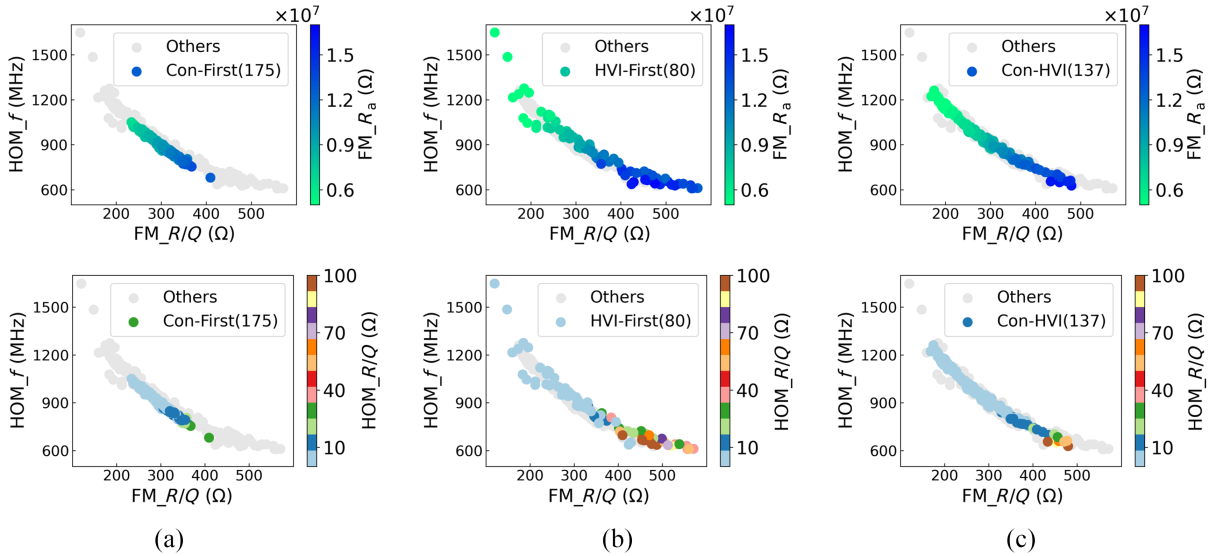

FIG. 6. Feasible Pareto fronts obtained by using different sAF in tAS after 1000 samples. (a) Con-First, (b) HVI-First, and (c) Con-HVI.

the MOEA according to the characteristics of the problem helps achieve convergence to higher-quality solutions. Incorporating practical constraints into the vAF can obtain better solutions in the region of interest. For problems with strict constraints, using the Con-First acquisition function $\alpha_{\text{Con-First}}(\mathbf{x})$ facilitates the discovery of more feasible solutions. In practical optimization, the choice of sAF can be tailored to the problem requirements and even adaptively adjusted during the optimization process.

C. Sampling efficiency analysis in tAS

We investigate the impact of the number of EM simulations under the same settings corresponding to Fig. 6(b). Figures 7(a) and 7(b) visualize the projection of the HV of the feasible Pareto front obtained by MOBO after 500, 1000, 2000, and 3000 EM simulations. The results indicate that a large HV can be achieved with only 500 EM simulations. Further increasing the number of EM simulations to 2000 and 3000 yields only marginal improvements. This indicates that the optimization process essentially converges by 1000 EM simulations.

Scatter plots of the feasible Pareto front are subsequently shown in Figs. 7(c) and 7(d). The results show that the feasible Pareto front has not progressed noticeably apart from an increase in the number. In other words, high-quality solutions were already obtained with only 500 EM simulations. For instance, one representative solution achieved by 500 EM simulations achieves a fundamental mode frequency $f_{\text{FM}} = 499.68$ MHz, $R/Q = 357.22$ Ω, shunt impedance $R_a^{\text{FM}} = 11.32$ MΩ, and the frequency of first HOM $f_{\text{HOM}} = 822.09$ MHz with $R/Q = 5.39$ Ω.

However, although good convergence was achieved at 500 EM simulations, the number of solutions was limited. Increasing the number of EM simulations to 1000 improved the diversity and the number of the feasible Pareto-optimal solutions. Further increasing the number of EM simulations to 2000 and 3000 led to only minor improvements in convergence and diversity. This is consistent with the trend observed in Figs. 7(a) and 7(b). Therefore, we conclude that 1000 EM simulations are generally sufficient to obtain a well-converged and diverse feasible Pareto-optimal set.

D. Comparison of convergence

To assess the convergence speed of MOBO, comparisons were made with NSGA-II, multiple objective particle swarm optimization (MOPSO) [75], NSGA-II with the assisting method (NSGA-II-A), MOPSO with the assisting method (MOPSO-A), standard DNMOGA initialized with 1000 samples (DNMOGA1000), and DNMOGA initialized with 100 samples (DNMOGA100). The parameter settings are summarized in Table III. NSGA-II-A used the same parameter settings as NSGA-II, but its mutation operator was adjusted: individuals satisfying the frequency constraint were mutated using the polynomial mutation operator, while those violating the constraint were modified based on an empirical linear relationship between R_{eq} and f_{FM}. Similarly, MOPSO-A followed the same adjustment logic: particles that satisfied the frequency constraint followed the standard MOPSO update, while those that violated the constraint were guided by the same empirical relationship to enhance feasibility. It is worth noting that DNMOGA1000 used 1000 initial individuals as in Ref. [33], while DNMOGA100 and the other algorithms

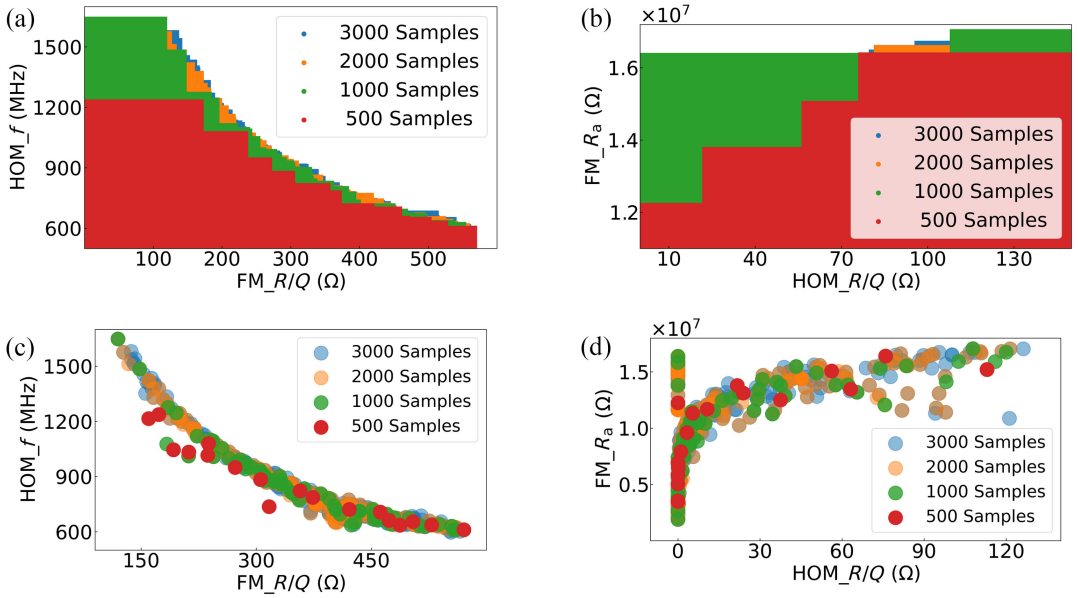

FIG. 7. Comparison of feasible Pareto fronts after 500, 1000, 2000, and 3000 EM evaluations. (a) Projected HV onto f_{HOM} and R/Q_{FM}. (b) Projected HV onto R/Q_{HOM} and $R_{a\text{FM}}$. (c) Scatter plots of f_{HOM} and R/Q_{FM}. (d) Scatter plots of R/Q_{HOM} and $R_{a\text{FM}}$.

TABLE III. Hyperparameter settings of baseline algorithms.

Algorithm	Parameter settings
NSGA-II	Population size: 50 Arithmetic crossover; 80% individuals participate per generation Polynomial mutation; 20% individuals participate per generation Mutation distribution index: 0.1 Deb's constraint-domination principle [76]
MOPSO	Population size: 50 Inertia weight $w = 0.6$ Learning factors: $c_1 = 1.7$, $c_2 = 1.7$ Gaussian mutation with probability $1/(\text{number of variables})$ Deb's constraint-domination principle applied in update and selection Archive size equals population size
DNMOGA	Population size: 50 Crossover and mutation: dynamically adjust Surrogate model: Transformer and ANN Penalty with dynamic coefficient

used the same 100 initial individuals. The HV of the feasible Pareto front was recorded every 100 EM simulations. Each algorithm was run for a total of 3000 EM simulations, and the results are presented in Figs. 8(a) and 8(b).

To assess the robustness of the algorithms, all algorithms were run 3 times using the same three sets of initial samples (each containing 100 samples). Each complete optimization process involves 1000 EM simulations. The results are presented in Figs. 8(c) and 8(d).

It can be observed that the MOBO algorithm achieves faster convergence than NSGA-II, NSGA-II-A, MOPSO, MOPSO-A, and DNMOGA, demonstrating a substantial improvement in convergence speed.

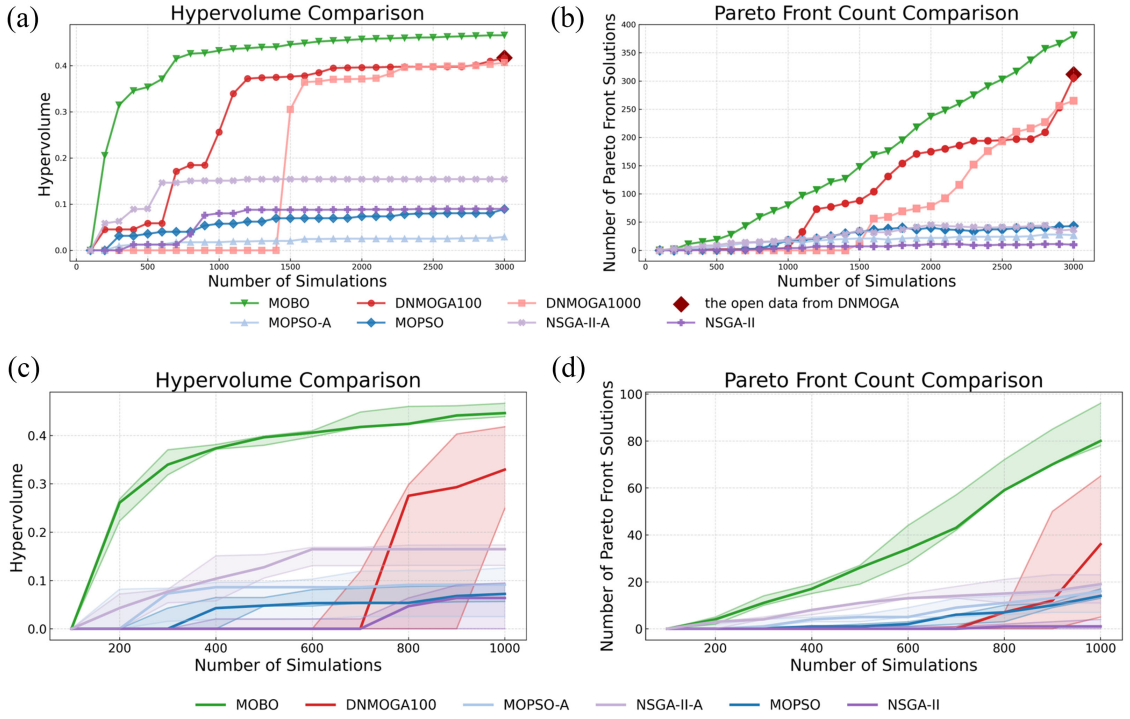

FIG. 8. Comparison of convergence speed for different algorithms. (a) Hypervolume comparison during 3000 EM simulations. (b) Comparison of the number of feasible Pareto-optimal solutions during 3000 EM simulations. (c) Hypervolume comparison during 1000 EM simulations. (d) Comparison of the number of feasible Pareto-optimal solutions during 1000 EM simulations. In (a) and (b), "the open data from DNMOGA" is obtained from Ref. [33]. In (c) and (d), the bold curves represent the median values from three independent runs, while the shaded areas indicate the corresponding upper and lower bounds.

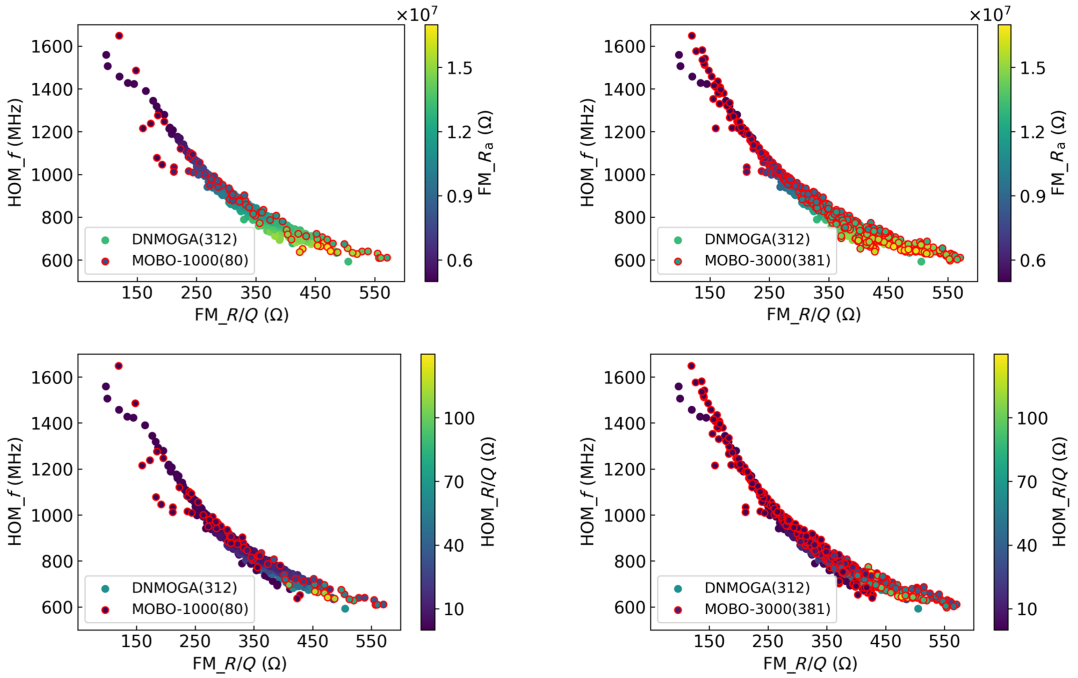

FIG. 9. Comparison of the feasible Pareto fronts obtained by MOBO (red circled markers) and DNMOGA (without red circled markers). (a) Results for MOBO with 1000 EM evaluations and DNMOGA with 3000 EM evaluations. (b) Results for MOBO with 3000 EM evaluations and DNMOGA with 3000 EM evaluations.

To evaluate the convergence quality of the feasible Pareto front, a comparison of MOBO algorithm with DNMOGA is presented in Fig. 9. As shown in Fig. 9(a), MOBO with 1000 EM simulations achieves a feasible Pareto front that is nearly comparable to that of DNMOGA with 3000 EM simulations. When the number of EM simulations in MOBO is increased to 3000, as shown in Fig. 9(b), the number of feasible solutions increases, especially in the high f_{HOM} region, and the overall Pareto front is comparable to that of DNMOGA in most regions.

V. CONCLUSION

This study demonstrates that the MOBO framework can be effectively applied to rf cavity optimization. Constraint handling in MOBO is a major challenge when dealing with complex MOOPs such as rf cavity optimization, which involve strict equality constraints and a complex decision space. To address this issue, we investigate several constraint-handling methods based on penalty and propose a novel approach called tAS. The tAS integrates a vAF with MOEA for effective search, followed by a sAF to select the most promising solution for actual evaluation. Experiments on both simple test problems and rf cavity optimization tasks show that tAS outperforms the penalty-based sAF strategy, particularly when the feasible region is extremely narrow.

The superior performance of MOBO with tAS stems from separating the sampling process into distinct search and selection stages. Leveraging the capability of MOEAs to handle complex MOOPs with strict constraints, it enables a more thorough exploration of the objective space and effectively locates the feasible Pareto front. Our experiments demonstrate that careful design of the MOEA, vAF, and sAF components enhances the quality of the results. Therefore, there remains great untapped potential in MOBO with tAS through further refinement of these components.

The MOBO with tAS demonstrates rapid convergence, which has been clearly validated through comparisons with the conventional rf cavity optimization algorithm NSGA-II. Moreover, experiments show that it can identify high-quality solutions with only 500 EM simulations. Increasing the number of EM simulations mainly improves the quantity rather than the quality of solutions, making this approach particularly suitable for situations with limited computational resources. The successful application of the MOBO framework with tAS to rf cavity optimization suggests its promising applicability to other challenging MOOPs in particle accelerators that are characterized by strict constraints and high computational costs.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Xuerui Hao for the helpful discussion. This work was supported by the Fundamental Research Funds for the Central Universities, Grant No. WK2310000118.

DATA AVAILABILITY

The data that support the findings of this article are openly available [77].

[1] G. Ising, Prinzip einer methode zur herstellung von kanalstrahlen hoher voltzahl, Ark. Mat. Astron. Fys. **18**, 1 (1924), https://cds.cern.ch/record/433984.

[2] N. Huang, H. Deng, B. Liu, D. Wang, and Z. Zhao, Features and futures of x-ray free-electron lasers, Innovation **2**, 100097 (2021).

[3] S. Shin, New era of synchrotron radiation: Fourth-generation storage ring, AAPPS Bull. **31**, 21 (2021).

[4] P. Zakalek, T. Gutberlet, and T. Brückel, Neutron sources for large scale user facilities: The potential of high current accelerator-driven neutron sources, Prog. Part. Nucl. Phys. **142**, 104163 (2025).

[5] P. K. Nema, Application of accelerators for nuclear systems: Accelerator driven system (ADS), Energy Procedia **7**, 597 (2011).

[6] M. Kranjčević, A. Adelmann, P. Arbenz, A. Citterio, and L. Stingelin, Multi-objective shape optimization of radio frequency cavities using an evolutionary algorithm, Nucl. Instrum. Methods Phys. Res., Sect. A **920**, 106 (2019).

[7] Ansys Inc., Ansys HFSS: 3D electromagnetic simulation software, Available at https://www.ansys.com/products/electronics/ansys-hfss.

[8] Dassault Systèmes, CST Studio Suite, Available at https://www.3ds.com/products-services/simulia/products/cst-studio-suite/.

[9] A. Slowik and H. Kwasnicka, Evolutionary algorithms and their applications to engineering problems, Neural Comput. Appl. **32**, 12363 (2020).

[10] A. Hofler, B. c. v. Terzić, M. Kramer, A. Zvezdin, V. Morozov, Y. Roblin, F. Lin, and C. Jarvis, Innovative applications of genetic algorithms to problems in accelerator physics, Phys. Rev. ST Accel. Beams **16**, 010101 (2013).

[11] A. Homaifar, C. X. Qi, and S. H. Lai, Constrained optimization via genetic algorithms, Simulation **62**, 242 (1994).

[12] J. A. Joines and C. R. Houck, On the use of non-stationary penalty functions to solve nonlinear constrained optimization problems with GA's, in *Proceedings of the first IEEE Conference on Evolutionary Computation. IEEE World Congress on Computational Intelligence* (IEEE, New York, 1994), pp. 579–584.

[13] J. Matias, A. Correia, P. Mestre, C. Serodio, P. Couto, C. Teixeira, and P. Melo-Pinto, Adaptive penalty and barrier function based on fuzzy logic, Expert Syst. Appl. **42**, 6777 (2015).

[14] C.-H. Lin, A rough penalty genetic algorithm for constrained optimization, Inf. Sci. (N.Y.) **241**, 119 (2013).

[15] C.-g. Cui, Y.-j. Li, and T.-j. Wu, A relative feasibility degree based approach for constrained optimization problems, J. Zhejiang Univ. Sci. C **11**, 249 (2010).

[16] T. P. Runarsson and X. Yao, Stochastic ranking for constrained evolutionary optimization, IEEE Trans. Evol. Comput. **4**, 284 (2000).

[17] T. Takahama and S. Sakai, Constrained optimization by the constrained differential evolution with gradient-based mutation and feasible elites, in *Proceedings of the IEEE Congress on Evolutionary Computation* (IEEE, New York, 2006), Vol. 1

[18] A. Isaacs, T. Ray, and W. Smith, Blessings of maintaining infeasible solutions for constrained multi-objective optimization problems, in *Proceedings of the 2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)* (IEEE, New York, 2008), pp. 2780–2787.

[19] R. Mallipeddi and P. N. Suganthan, Ensemble of constraint handling techniques, IEEE Trans. Evol. Comput. **14**, 561 (2010).

[20] Z. Fan, W. Li, X. Cai, H. Li, C. Wei, Q. Zhang, K. Deb, and E. Goodman, Push and pull search for solving constrained multi-objective optimization problems, Swarm Evol. Comput. **44**, 665 (2019).

[21] H. Ma, H. Wei, Y. Tian, R. Cheng, and X. Zhang, A multi-stage evolutionary algorithm for multi-objective optimization with complex constraints, Inf. Sci. (N.Y.) **560**, 68 (2021).

[22] J. Wang, G. Liang, and J. Zhang, Cooperative differential evolution framework for constrained multiobjective optimization, .IEEE Trans. Syst. Man Cybern. **49**, 2060 (2018).

[23] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Trans. Evol. Comput. **6**, 182 (2002).

[24] S. B. Selçuklu, Multi-objective genetic algorithms, in *Handbook of Formal Optimization* (Springer, New York, 2023), pp. 1–37.

[25] T. Luo, H. Feng, D. Filippetto, M. Johnson, A. Lambert, D. Li, C. Mitchell, F. Sannibale, J. Staples, S. Virostek *et al.*, RF design of APEX2 two-cell continuous-wave normal conducting photoelectron gun cavity based on multi-objective genetic algorithm, Nucl. Instrum. Methods Phys. Res., Sect. A **940**, 12 (2019).

[26] M. Kranjčević, S. Gorgi Zadeh, A. Adelmann, P. Arbenz, and U. Van Rienen, Constrained multiobjective shape optimization of superconducting rf cavities considering robustness against geometric perturbations, Phys. Rev. Accel. Beams **22**, 122001 (2019).

[27] M. Kranjčević. Multi-objective shape optimization of radio-frequency cavities, Ph.D. thesis, ETH Zurich, 2019.

[28] C. Kim, C. S. Park, E.-S. Kim, S. H. Shin, S. H. Park, C.-K. Min, and W. J. Byeon, Optimization study for rf cavity and beam dynamics simulation in the electron linear accelerator using the multi-objective genetic algorithm, J. Instrum. **19**, P01023

[29] H. Feng, T. Luo, D. Li, K. Baptiste, W. Huang, C. Tang, and Z. Liu, Optimization of rf cavities using MOGA for ALS-U, in *Proceedings of the 10th International Particle Accelerator Conference, IPAC-2019, Melbourne, Australia* (JACoW, Geneva, Switzerland, 2019), pp. 3007–3010.

[30] H. Feng, S. De Santis, K. Baptiste, W. Huang, C. Tang, and D. Li, Proposed design and optimization of a higher harmonic cavity for ALS-U, Rev. Sci. Instrum. **91** (2020).
[31] S. Smith, M. Southerby, S. Setiniyaz, R. Apsimon, and G. Burt, Multiobjective optimization and pareto front visualization techniques applied to normal conducting rf accelerating structures, Phys. Rev. Accel. Beams **25**, 062002 (2022).
[32] P. Wang, K. Ye, X. Hao, and J. Wang, Handling shape optimization of superconducting cavities with DNMOGA, Comput. Phys. Commun. **299**, 109136 (2024).
[33] P. Wang, K. Ye, X. Hao, and J. Wang, Combining multiobjective genetic algorithm and neural network dynamically for the complex optimization problems in physics, Sci. Rep. **13**, 880 (2023).
[34] Y. Li, W. Cheng, L. H. Yu, and R. Rainer, Genetic algorithm enhanced by machine learning in dynamic aperture optimization, Phys. Rev. Accel. Beams **21**, 054601 (2018).
[35] A. Edelen, N. Neveu, M. Frey, Y. Huber, C. Mayes, and A. Adelmann, Machine learning for orders of magnitude speedup in multiobjective optimization of particle accelerator systems, Phys. Rev. Accel. Beams **23**, 044601 (2020).
[36] J. Wan, P. Chu, and Y. Jiao, Neural network-based multiobjective optimization algorithm for nonlinear beam dynamics, Phys. Rev. Accel. Beams **23**, 081601 (2020).
[37] M. Mohamadian, H. Afarideh, and M. Ghergherehchi, Optimized feed-forward neural-network algorithm trained for cyclotron-cavity modeling, Chin. Phys. C **41**, 017003 (2017).
[38] A.-K. S. Hassan, H. L. Abdel-Malek, A. S. Mohamed, T. M. Abuelfadl, and A. E. Elqenawy, RF cavity design exploiting a new derivative-free trust region optimization approach, J. Advertising Res. **6**, 915 (2015).
[39] H. Gong, W. Fang, J. Tan, X. Huang, C. Wang, Y. Xu, and Z. Zhao, Design of a dual-mode transverse deflecting structure using neural network and multiobjective algorithms, Phys. Rev. Accel. Beams **27**, 042001 (2024).
[40] D. R. Jones, M. Schonlau, and W. J. Welch, Efficient global optimization of expensive black-box functions, J. Global Optim. **13**, 455 (1998).
[41] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, Taking the human out of the loop: A review of Bayesian optimization, Proc. IEEE **104**, 148 (2016).
[42] S. Greenhill, S. Rana, S. Gupta, P. Vellanki, and S. Venkatesh, Bayesian optimization for adaptive experimental design: A review, IEEE Access **8**, 13937 (2020).
[43] P. M. Pardalos, A. Zhigljavsky, J. Žilinskas *et al.*, *Advances in Stochastic and Deterministic Global Optimization* (Springer, New York, 2016), Vol. 107.
[44] J. Kirschner, M. Mutný, A. Krause, J. Coello de Portugal, N. Hiller, and J. Snuverink, Tuning particle accelerators with safety constraints using Bayesian optimization, Phys. Rev. Accel. Beams **25**, 062802 (2022).
[45] M. S. Priyadarshini, O. Romiluyi, Y. Wang, K. Miskin, C. Ganley, and P. Clancy, PAL 2.0: A physics-driven Bayesian optimization framework for material discovery, Mater. Horiz. **11**, 781 (2024).
[46] R. C. Soares, J. C. Silva, J. A. de Lucena Junior, A. C. Lima Filho, J. G. G. de Souza Ramos, and A. V. Brito, Integration of Bayesian optimization into hyperparameter tuning of the particle swarm optimization algorithm to enhance neural networks in bearing failure classification, Measurement **242**, 115829 (2025).
[47] A. Hanuka, X. Huang, J. Shtalenkova, D. Kennedy, A. Edelen, Z. Zhang, V. Lalchand, D. Ratner, and J. Duris, Physics model-informed Gaussian process for online optimization of particle accelerators, Phys. Rev. Accel. Beams **24**, 072802 (2021).
[48] F. Irshad, C. Eberle, F. M. Foerster, K. V. Grafenstein, F. Haberstroh, E. Travac, N. Weisse, S. Karsch, and A. Döpp, Pareto optimization and tuning of a laser wakefield accelerator, Phys. Rev. Lett. **133**, 085001 (2024).
[49] F. Irshad, S. Karsch, and A. Döpp, Multi-objective and multi-fidelity Bayesian optimization of laser-plasma acceleration, Phys. Rev. Res. **5**, 013063 (2023).
[50] R. Roussel, A. Hanuka, and A. Edelen, Multiobjective Bayesian optimization for online accelerator tuning, Phys. Rev. Accel. Beams **24**, 062801 (2021).
[51] R. Roussel *et al.*, Bayesian optimization algorithms for accelerator physics, Phys. Rev. Accel. Beams **27**, 084801 (2024).
[52] J. Kirschner, M. Mutný, A. Krause, J. Coello de Portugal, N. Hiller, and J. Snuverink, Tuning particle accelerators with safety constraints using Bayesian optimization, Phys. Rev. Accel. Beams **25**, 062802 (2022).
[53] C. Xu, T. Boltz, A. Mochihashi, A. Santamaria Garcia, M. Schuh, and A.-S. Müller, Bayesian optimization of the beam injection process into a storage ring, Phys. Rev. Accel. Beams **26**, 034601 (2023).
[54] J. Duris, D. Kennedy, A. Hanuka, J. Shtalenkova, A. Edelen, P. Baxevanis, A. Egger, T. Cope, M. McIntire, S. Ermon, and D. Ratner, Bayesian optimization of a free-electron laser, Phys. Rev. Lett. **124**, 124801 (2020).
[55] Y. Morita, T. Washio, and Y. Nakashima, Accelerator tuning method using autoencoder and Bayesian optimization, Nucl. Instrum. Methods Phys. Res., Sect. A **1057**, 168730 (2023).
[56] M. Emmerich, K. Yang, A. Deutz, H. Wang, and C. M. Fonseca, A multicriteria generalization of Bayesian global optimization, in *Advances in Stochastic and Deterministic Global Optimization* (Springer, New York, 2016), pp. 229–242.
[57] J. Wilson, F. Hutter, and M. Deisenroth, Maximizing acquisition functions for Bayesian optimization, in *Advances in Neural Information Processing Systems* (Curran Associates Inc., Red Hook, NY, 2018), Vol. 31.
[58] E. Brochu, V. M. Cora, and N. De Freitas, A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning, arXiv:1012.2599.
[59] J. Knowles, ParEGO: A hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems, IEEE Trans. Evol. Comput. **10**, 50 (2006).
[60] K. Yang, A. Deutz, Z. Yang, T. Back, and M. Emmerich, Truncated expected hypervolume improvement: Exact computation and application, in *Proceedings of the 2016*

IEEE Congress on Evolutionary Computation (CEC) (IEEE, New York, 2016), pp. 4350–4357.

[61] J. R. Gardner, M. J. Kusner, Z. E. Xu, K. Q. Weinberger, and J. P. Cunningham, Bayesian optimization with inequality constraints, in *Proceedings of the 31st International Conference on Machine Learning, Beijing, China* (JMLR.org, Beijing, China, 2014), pp. 937–945.

[62] J. Kirschner, M. Mutny, N. Hiller, R. Ischebeck, and A. Krause, Adaptive and safe Bayesian optimization in high dimensions via one-dimensional subspaces, in *Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA* (PMLR, Long Beach, CA, 2019), pp. 3429–3438.

[63] J. O. Lübsen, M. Schütte, S. Schulz, and A. Eichler, A safe Bayesian optimization algorithm for tuning the optical synchronization system at European XFEL, IFAC-PapersOnLine **56**, 3079 (2023), 22nd IFAC World Congress.

[64] Y. Sui, V. Zhuang, J. Burdick, and Y. Yue, Stagewise safe Bayesian optimization with Gaussian processes, in *Proceedings of the 35th International Conference on Machine Learning* (PMLR, Stockholm, 2018), pp. 4781–4789.

[65] J. Wu and P. Frazier, Practical two-step lookahead Bayesian optimization, in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., Vancouver, 2019).

[66] R. Lam and K. Willcox, Lookahead Bayesian optimization with inequality constraints, in *Advances in Neural Information Processing Systems*, edited by I. Guyon, U. V. Luxburg, S. Bengio, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates, Inc., Long Beach, CA, 2017), Vol. 30.

[67] Y. Zhang, X. Zhang, and P. Frazier, Constrained two-step look-ahead Bayesian optimization, in *Advances in Neural Information Processing Systems*, edited by M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan (Curran Associates, Inc., Red Hook, NY, 2021), Vol. 34, pp. 12563–12575.

[68] M. Imani and S. F. Ghoreishi, Two-stage Bayesian optimization for scalable inference in state-space models, IEEE Trans. Neural Networks Learn. Syst. **33**, 5138 (2022).

[69] D. Eriksson, M. Pearce, J. Gardner, R. D. Turner, and M. Poloczek, Scalable global optimization via local Bayesian optimization, in *Advances in Neural Information Processing Systems* (Curran Associates Inc., Red Hook, NY, 2019), Vol. 32, pp. 5497–5508.

[70] S. Liuzzo, N. Carmignani, L. Carver, L. Hoummi, I. Agapov, M. Boese, L. Malina, E. Musa, J. Keil, and B. Veglia, Optimisation of the touschek lifetime in synchrotron light sources using badger, in *Proceedings of the 19th International Conference on Accelerator and Large Experimental Physics Control Systems, ICALEPCS-2023, Cape Town, South Africa* (JACoW, Geneva, Switzerland, 2023).

[71] M. Stein, Large sample properties of simulations using Latin hypercube sampling, Technometrics **29**, 143 (1987).

[72] J. Kennedy and R. Eberhart, Particle swarm optimization, in *Proceedings of ICNN'95-International Conference on Neural Networks* (IEEE, New York, 1995), Vol. 4, pp. 1942–1948.

[73] A. Homaifar, C. X. Qi, and S. H. Lai, Constrained optimization via genetic algorithms, Simulation **62**, 242 (1994).

[74] Y. Wang, B.-C. Wang, H.-X. Li, and G. G. Yen, Incorporating objective function information into the feasibility rule for constrained evolutionary optimization, IEEE Trans. Syst. Man Cybern. **46**, 2938 (2015).

[75] X. Huang and J. Safranek, Nonlinear dynamics optimization with particle swarm and genetic algorithms for SPEAR3 emittance upgrade, Nucl. Instrum. Methods Phys. Res., Sect. A **757**, 48 (2014).

[76] K. Deb, An efficient constraint handling method for genetic algorithms, Comput. Methods Appl. Mech. Eng. **186**, 311 (2000).

[77] Y. Wang, Y. Tang, C.-F. Wu, and G. Feng, Data and code for Multiobjective Bayesian optimization for the shape design of RF cavity in particle accelerators, Zenodo (2026), 10.5281/zenodo.18829772.