

# Detecting Resonance of Radio-Frequency Cavities Using Fast Direct Integral Equation Solvers and Augmented Bayesian Optimization

Yang Liu, *Member, IEEE*, Tianhuan Luo, Aman Rani, Hengrui Luo, and Xiaoye Sherry Li

arXiv:2305.05918v2 [physics.acc-ph] 30 Aug 2023

**Abstract**—This paper presents a computationally efficient framework for identifying resonance modes of 3D radio-frequency (RF) cavities with damping waveguide ports. The proposed framework relies on surface integral equation (IE) formulations to convert the task of resonance detection to the task of finding resonance frequencies at which the lowest few eigenvalues of the system matrix is close to zero. For the linear eigenvalue problem with a fixed frequency, we propose leveraging fast direct solvers to efficiently invert the system matrix; for the frequency search problem, we develop a hybrid optimization algorithm that combines Bayesian optimization with down-hill simplex optimization. The proposed IE-based resonance detection framework (IERD) has been applied to detection of high-order resonance modes (HOMs) of realistic accelerator RF cavities to demonstrate its efficiency and accuracy.

**Index Terms**—RF cavity, eigen solver, fast direct solver, Bayesian optimization, high-order resonance mode, down-hill simplex algorithm.

## I. INTRODUCTION

HIGH-quality radio-frequency (RF) cavities are critical components for high energy particle accelerators developed at national facilities and laboratories [1]–[3]. RF cavities are usually excited by a time-harmonic source field operated at the working resonance frequency to support a particular electromagnetic field pattern (called the fundamental resonance mode), which provides stable acceleration of the particle bunches passing through the cavity. However, the transient bunch, which can be treated as e.g., a Gaussian wavelet packet in time, can also excite undesired high-order resonance modes (HOMs) at higher resonance frequencies which can cause deterioration of beam quality and severe damage to the device. These HOMs are often suppressed by adding damping waveguide ports to the cavity, which couple the HOMs from the cavity to the external absorbing loads. Recall that the quality (Q) factor of a cavity is defined as the ratio of stored energy in the cavity and the dissipated power per cycle by the cavity wall losses and/or damping ports. While the fundamental mode operates with high Q factors, the

HOMs requires sufficiently low Q factors so that they cannot be established in practice.

Given a fixed design with cavity geometry, material, port position and port shapes, the fundamental mode and HOMs and their resonance frequencies need to be accurately and efficiently modeled by numerical tools. The most commonly-used numerical method for realistic 3D cavity modeling is the finite element method (FEM) [4]–[6], which formulates the problem as an eigenvalue problem whose eigenvalues and eigenvectors are respectively the resonance frequencies and resonant electrical fields in the cavity volume (or called eigenmodes). In this paper, we use the terms “eigenmode” and “resonance mode” interchangeably. However, FEM faces the following computational challenges: (1) The FEM method requires spatial discretization of the entire cavity volume with dense meshes and high-order basis functions to resolve the resonance modes, which lead to large numbers of unknowns and high CPU/memory usage. (2) When waveguide boundary conditions are used for the damping ports or the dissipative power loss is considered, the FEM method leads to a nonlinear eigenvalue problem, and linearization techniques such as Newton’s method [7], [8], contour integral methods [9], [10], and rational Krylov methods [11]. However, these methods suffer from either lack of convergence and/or high computational costs. Moreover, the FEM method tends to become less accurate when the resonance frequency lies near or below the lowest cutoff frequencies of the ports. (3) Although the Q factors can be computed using the FEM method, they cannot provide information about accuracy of the computed resonance frequencies regarding uncertainty in design parameters and numerical models.

This paper considers the surface integral equation (IE) method as an alternative to the FEM method to address the abovementioned computational challenges. The IE method formulates the problem as searching for the frequencies at which the system matrix’s smallest few eigenvalues are close to zero. In contrast to FEM, the IE method only requires spatial discretization of the cavity boundary leading to much smaller numbers of unknowns. In addition, at each trial frequency, a linear dense eigenvalue problem needs to be solved and traditional linear eigen solvers can be directly applied. Last but not least, the IE method naturally provides bandwidth information about the computed resonance frequencies by investigating a sequence of trial frequencies and the associated eigenvalues. That being said, the IE methods have been primarily considered only for traveling wave modeling

Manuscript received XX XX, XXXX. (*Corresponding author: Yang Liu.*)

Y. Liu, H. Luo, and X. S. Li are with Applied Mathematics and Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. (e-mail: liuyangzhan@lbl.gov, hrluo@lbl.gov, XSLi@lbl.gov).

T. Luo is with Accelerator Technology and Applied Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. (e-mail: TLuo@lbl.gov).

A. Rani is with the department of Mathematics and Statistics, Texas Tech University, Lubbock, TX, 79409, USA. (e-mail: aman.rani@ttu.edu).

[12]–[15] and characteristic mode analysis (CMA) [16]–[18], and the use of IE methods for resonance modeling remains largely unexplored [19], [20]. This is due to the following two computational difficulties: (1) Unlike traveling wave modeling and CMA, the resonance problem leads to intrinsically large-scale and ill-conditioned linear systems at resonance frequencies of the HOMs. Although fast iterative algorithms such as fast multipole methods [21] have been applied to large-scale CMA [22], they become computationally infeasible for resonance problems. This is due to the ill-conditioned IE matrix near the resonance frequencies which, even with good preconditioners, lead to large numbers of matrix-vector multiplications. (2) The identification of the resonance frequencies is typically performed with grid search or frequency sweep, which becomes computationally very expensive with fine frequency steps [19]. Optimization algorithms that aim at minimizing the eigenvalues corresponding to each mode can be considered instead of grid search, but the objective function w.r.t frequency exhibits sharp minimum for high-Q modes and flattened minimum for low-Q ones, and existing optimization tools either require too many samples to capture all resonance modes or suffer from convergence issues for these complex objective functions. Therefore IE methods that can both robustly solve the eigenvalue problems with given trial frequencies and efficiently propose promising trial frequencies are called for.

The proposed IE-based resonance detection framework (IERD) framework addresses the abovementioned two challenges by leveraging fast direct IE solvers and developing hybrid optimization algorithms. First, we rely on the shift-and-invert implicitly restarted Arnoldi method (IRAM) [23] for solving the eigenvalue problem for the smallest few eigenvalues, which requires supplying the inverse of the dense IE matrix. The matrix inverse is computed with fast direct solvers such as hierarchically off-diagonal low-rank (HODLR) matrices [24] and hierarchically off-diagonal butterfly (HODBF) [25], among the more general class of hierarchical matrix algorithms [26]–[30]. For working modes and the first few HOMs, the cavity remains electrically small and low-rank-based direct solvers, e.g., HODLR, are very efficient; for HOMs with higher frequencies, the cavity becomes electrically large. Therefore, the equation system show certain patterns for which butterfly-based direct solvers, e.g., HODBF becomes computational more efficient. In fact, butterfly-based direct solvers have been successfully applied to surface IE [25], [31], volume IE [32], and differential equation [33]-based large-scale electromagnetic simulations. The use of HODLR and HODBF direct solvers can significantly improve the robustness and efficiency for solving the eigenvalue problem with fixed frequencies. Next, we propose a hybrid frequency search algorithm that combines the global search algorithm such as Gaussian process (GP)-based Bayesian optimization (BO) [34] and local search algorithm such as the down-hill simplex method (also known as the Nelder–Mead method) [35]. Our proposed algorithm first uses surrogate-based Bayesian optimization to locate the resonance regions in the frequency band of interest and then uses down-hill simplex algorithm to refine the optimal frequency within region. Both steps require only smaller numbers of trial frequencies compared with brute-

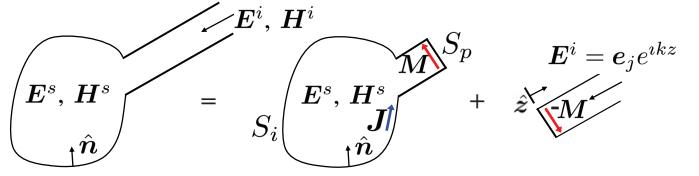


Fig. 1. Illustration of a RF cavity example with a waveguide port, decomposed as an interior problem and an exterior problem. It is assumed that the incident field  $\mathbf{E}^i = \mathbf{e}_j e^{ikz}$  has a single waveguide mode.

force frequency sweep or grid-search scheme.

The rest of this paper is organized as follows. Section II formulates the resonance detection problem as an IE-based eigenvalue problem with trial frequencies. The fast direct solvers for accelerating the eigenvalue problem for each frequency is summarized in Section III and the proposed hybrid Bayesian optimization and down-hill simplex algorithm is detailed in Section IV. Section V presents a few canonical and real-life numerical examples to demonstrate the efficiency, accuracy and robustness of the proposed IERD framework.

## II. FORMULATION

Consider the example of a RF cavity with a semi-infinite waveguide port in Fig. 1. In practice, the port serves as a damping port with a finite length and absorbing materials at the end. For illustration purposes, we initially consider a single port in this section. Later, in Section V, we demonstrate the applicability of IERD to cases with multiple ports by providing numerical results. The cavity is filled with material of permittivity  $\varepsilon_0$ , permeability  $\mu_0$ , and intrinsic impedance  $\eta_0 = \sqrt{\mu_0/\varepsilon_0}$ . To modify the cavity geometry and material properties for a specific problem, adjust the values of  $\varepsilon_0$ ,  $\mu_0$ , and  $\eta_0$  accordingly. Additionally, you can update the cavity and waveguide port geometries by changing their dimensions and shapes in the mathematical formulation. In order to derive a formulation for the resonance modes, it is assumed that the cavity is excited by an incident field  $\{\mathbf{E}^i(\mathbf{r}), \mathbf{H}^i(\mathbf{r})\}$  of frequency  $f$  (or wavenumber  $k$ ) through the port. The incident field excites surface equivalent source that produces a scattered field  $\{\mathbf{E}^s(\mathbf{r}), \mathbf{H}^s(\mathbf{r})\}$  in the cavity.

The problem can be treated as the superposition of an interior problem and an exterior problem. The interior problem consists of a closed cavity with surface  $S = S_i + S_p$ , where  $S_i$  corresponds to the open surface of the original open cavity, and the added boundary  $S_p$  corresponds to the surface of the port that encloses  $S_i$ . The scattered field  $\{\mathbf{E}^s(\mathbf{r}), \mathbf{H}^s(\mathbf{r})\}$  is produced by a surface electric current density  $\mathbf{J}(\mathbf{r})$  for  $\mathbf{r} \in S$  and a surface magnetic current density  $\mathbf{M}(\mathbf{r})$  for  $\mathbf{r} \in S_p$  (see Fig. 1 middle). The exterior problem consists of the semi-infinite wavguide with the incident electric field  $\mathbf{E}^i = \mathbf{e}_j e^{ikz}$ , where  $z$  is the local coordinate with  $z = 0$  at  $S_p$ , and  $\mathbf{e}_j$  is the normalized eigenvector of the waveguide (see Fig. 1 right).  $\mathbf{e}_j$  can be TM or TE modes, and can have an analytical expression for rectangular or circular waveguides, or can be numerically computed for irregular-shaped waveguides. The scattered field  $\{\mathbf{E}^s(\mathbf{r}), \mathbf{H}^s(\mathbf{r})\}$  can be reproduced by introducing an additional surface magnetic current density  $-\mathbf{M}$  at  $S_p$ .

Enforcing the boundary conditions for electric fields on  $S_i$  and  $S_p$  and magnetic fields on  $S_p$  leads to the following equations:

$$\hat{\mathbf{n}} \times \hat{\mathbf{n}} \times \mathcal{T}[\mathbf{J}](\mathbf{r}) - \hat{\mathbf{n}} \times \mathcal{K}\left[\frac{\mathbf{M}}{\eta_0}\right](\mathbf{r}) - \frac{Z_s}{\eta_0} \mathbf{J}(\mathbf{r}) = 0, \mathbf{r} \in S_i \quad (1)$$

$$\hat{\mathbf{n}} \times \hat{\mathbf{n}} \times \mathcal{T}[\mathbf{J}](\mathbf{r}) - \hat{\mathbf{n}} \times \mathcal{K}\left[\frac{\mathbf{M}}{\eta_0}\right](\mathbf{r}) + \hat{\mathbf{n}} \times \frac{\mathbf{M}(\mathbf{r})}{\eta_0} = 0, \mathbf{r} \in S_p \quad (2)$$

$$\begin{aligned} & \frac{1}{\eta_0} \hat{\mathbf{n}} \times \mathbf{J}(\mathbf{r}) + \frac{1}{2} \sum_i \frac{\hat{\mathbf{z}} \times \mathbf{e}_i(\mathbf{r})}{\eta_i \eta_0} \int_{S_p} d\mathbf{r}' (\hat{\mathbf{z}} \times \mathbf{e}_i(\mathbf{r}')) \cdot \mathbf{M}(\mathbf{r}') \\ &= \frac{\hat{\mathbf{z}} \times \mathbf{e}_j(\mathbf{r})}{\eta_j \eta_0}, \mathbf{r} \in S_p \end{aligned} \quad (3)$$

Here,  $\hat{\mathbf{n}}$  denotes the inward surface normal,  $\hat{\mathbf{z}}$  denotes the unit vector along the waveguide.  $Z_s = (1 + i) \sqrt{\frac{\pi f \mu_0}{\sigma_s}}$  is the surface impedance of the cavity wall material. Here  $\sigma_s$  is the electric conductivity of the material and we use  $5.80 \times 10^7$  S/m of Copper in this paper.  $\eta_i$  and  $\mathbf{e}_i(\mathbf{r})$ ,  $i = 1, 2, \dots$ , are the impedance and normalized tangential electric field of waveguide mode  $i$  with cutoff frequency  $f_{c,i} < f$ . Note that for regular-shaped wave guide ports such as circular or rectangular waveguides,  $\mathbf{e}_i(\mathbf{r})$  are analytically defined; otherwise they can be numerically computed and tabulated from e.g., FEM methods. In (5), the integral operators  $\mathcal{T}$  and  $\mathcal{K}$  are:

$$\mathcal{T}[\mathbf{X}](\mathbf{r}) = ik \int_S d\mathbf{r}' \mathbf{X}(\mathbf{r}') \cdot \left( \mathbf{I} - \frac{1}{k^2} \nabla \nabla' \cdot \right) \mathbf{g}(\mathbf{r}, \mathbf{r}') \quad (4)$$

$$\mathcal{K}[\mathbf{X}](\mathbf{r}) = \hat{\mathbf{n}} \times \int_S d\mathbf{r}' \mathbf{X}(\mathbf{r}') \times \nabla' \mathbf{g}(\mathbf{r}, \mathbf{r}') \quad (5)$$

To numerically detect the resonance modes,  $\mathbf{J}$  is discretized with  $N_J$  local basis functions  $\mathbf{f}_n(\mathbf{r})$ ,  $n = 1, \dots, N_J$  such as the Rao-Wilton-Glisson (RWG) functions [36] or the high-order ( $p$ th order) Graglia-Wilton-Peterson (GWP( $p$ )) basis functions [37] (RWG is essentially GWP(0)),  $\mathbf{M}/\eta_0$  is discretized with  $N_M$  mode basis functions  $\hat{\mathbf{z}} \times \mathbf{e}_n$ ,  $n = 1, \dots, N_M$  with  $N_M$  denotes the number of propagating modes of the port whose cutoff frequencies  $f_{c,n} < f$ .

$$\mathbf{J}(\mathbf{r}) = \sum_{n=1}^{N_J} I_n^J \mathbf{f}_n(\mathbf{r}) \quad (6)$$

$$\frac{\mathbf{M}(\mathbf{r})}{\eta_0} = \sum_{n=1}^{N_M} I_n^M \hat{\mathbf{z}} \times \mathbf{e}_n(\mathbf{r}) \quad (7)$$

Next we test (1) and (2) with  $\mathbf{f}_n$ , (3) with  $\hat{\mathbf{z}} \times \mathbf{e}_n$  and set the right-hand side (RHS) of (3) to 0. Note if  $f$  is a resonance frequency in the band of interest  $[f_{\min}, f_{\max}]$ , the system will require no excitation of the RHS. The following eigenvalue system is constructed:

$$\begin{bmatrix} \bar{\mathbf{T}} & \bar{\mathbf{K}} \\ \bar{\mathbf{C}} & \bar{\mathbf{D}} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{I}}^J \\ \bar{\mathbf{I}}^M \end{bmatrix} \approx \bar{\mathbf{0}}. \quad (8)$$

Note that “ $\approx$ ” instead of “ $=$ ” is used to account for numerical errors in surface discretization and matrix evaluation. In (8),

$\bar{\mathbf{I}}^J$  and  $\bar{\mathbf{I}}^M$  are collections of  $I_n^J$  and  $I_n^M$ , respectively. The  $(m, n)^{\text{th}}$  entry of the sub-matrices is:

$$T_{mn} = \langle \mathbf{f}_m(\mathbf{r}), \hat{\mathbf{n}} \times \hat{\mathbf{n}} \times \mathcal{T}[\mathbf{f}_n](\mathbf{r}) \rangle - \frac{Z_s}{\eta_0} \langle \mathbf{f}_m(\mathbf{r}), \delta^i(\mathbf{r}) \mathbf{f}_n(\mathbf{r}) \rangle \quad (9)$$

$$K_{m,n} = \langle \mathbf{f}_m(\mathbf{r}), -\hat{\mathbf{n}} \times \mathcal{K}[\hat{\mathbf{z}} \times \mathbf{e}_n](\mathbf{r}) \rangle + \langle \mathbf{f}_m(\mathbf{r}), \hat{\mathbf{z}} \times \mathbf{e}_n(\mathbf{r}) \rangle \quad (10)$$

$$C_{m,n} = \langle \hat{\mathbf{z}} \times \mathbf{e}_m(\mathbf{r}), \frac{1}{\eta_0} \hat{\mathbf{n}} \times \mathbf{f}_n(\mathbf{r}) \rangle \quad (11)$$

$$D_{m,n} = \frac{\delta_{m,n}}{2\eta_m}. \quad (12)$$

Here,  $\delta^i(\mathbf{r}) = 1$  if  $\mathbf{r} \in S_i$  and 0 elsewhere,  $\delta_{m,n}$  is the Kronecker delta function, and  $\langle \cdot, \cdot \rangle$  denotes standard inner product. The system in (8) can be written as  $\bar{\mathbf{Z}}(f)\bar{\mathbf{I}} \approx \bar{\mathbf{0}}$  with  $\bar{\mathbf{I}} = [\bar{\mathbf{I}}^J; \bar{\mathbf{I}}^M]$  and  $\bar{\mathbf{Z}} = [\bar{\mathbf{T}}, \bar{\mathbf{K}}, \bar{\mathbf{C}}, \bar{\mathbf{D}}]$ . Therefore, the problem of resonance detection reduces to finding the trial frequency  $f$  at which the smallest few eigenvalues of  $\bar{\mathbf{Z}}$  are close to 0. The physical meaning of such a formulation is that at resonance, the system can sustain an approximately non-decaying surface electrical current density without any external excitation (i.e., setting the RHS of the linear system to  $\bar{\mathbf{0}}$ ).

The proposed IERD framework aims at the accurately and efficiently finding all resonance modes in the frequency band of interest. IERD relies on a fast direct IE solver-based linear eigensolver to compute these eigenvalues for each trial frequency, as described in Section III, and an efficient optimization algorithm to search for the resonance frequencies of multiple modes, as detailed in Section IV.

### III. FAST DIRECT METHOD-BASED EIGENSOLVER

For a fixed frequency  $f$ , the proposed IERD framework utilizes the shift-and-invert IRAM method from ARPACK [23] with zero shift, which requires the input of the inverse operator  $\bar{\mathbf{Z}}^{-1}$ . IRAM is a fast iterative algorithm that can be used to quickly compute the first few eigenvalues of interest. When the number of required eigenvalues is small, IRAM typically converges in a small number of iterations. ARPACK returns the smallest few eigenvalues and their eigenvectors among which some eigenvectors do not correspond to a physical resonance, but rather artifact from numerical discretization errors. These eigenvectors are characterized as sparse vectors with few large elements and small ones elsewhere, representing non-physical and “local” resonance near shape geometry regions. The sparsity of the eigenvectors can be measured by  $\|\bar{\mathbf{I}}\|_1$  given that  $\|\bar{\mathbf{I}}\|_\infty = 1$ . Hence only those eigenvectors  $\bar{\mathbf{I}}$  with  $\|\bar{\mathbf{I}}\|_1/\|\bar{\mathbf{I}}\|_\infty > \epsilon_{\text{norm}}$  for a prescribed threshold  $\epsilon_{\text{norm}}$  (e.g.,  $\epsilon_{\text{norm}} = 1000$ ) are considered to be physically meaningful resonance vectors. In this paper, we use “eigenmodes” to denote these physically meaningful resonance vectors. Note that those eigenvectors with  $\|\bar{\mathbf{I}}\|_1/\|\bar{\mathbf{I}}\|_\infty < \epsilon_{\text{norm}}$  cannot be called eigenmodes as they are due to numerical modeling errors. One frequency  $f$  may yield no eigenmode or multiple eigenmodes. Ideally assuming no material loss, power damping, numerical error and degenerate eigenmodes, each eigenmode has only one resonance frequency (i.e.,  $\bar{\mathbf{Z}}$  is singular at the resonance frequency and the Q factor becomes infinity). In practice,

however, a cavity can operate at near-resonance regime (i.e., slightly perturbed resonance frequency) with the same eigenmode. Therefore, one trial frequency  $f$  can detect multiple eigenmodes when it's in the near-resonance regime of multiple eigenmodes.

HODLR and HODBF are two types of rank-structured fast direct solvers for fast computing  $\bar{\mathbf{Z}}^{-1}$  implemented in the software ButterflyPACK [38]. These solvers first reorder the matrix  $\bar{\mathbf{Z}}$  and partition the matrix into  $O(\log N)$  levels. At each level off-diagonal blocks representing non-self interactions are represented as low-rank or butterfly formats with a prescribed compression tolerance. Next,  $\bar{\mathbf{Z}}^{-1}$  is computed also with compressed representations using deterministic or randomized operations [39]. Once computed, a function  $\bar{\mathbf{y}} = \bar{\mathbf{Z}}^{-1}\bar{\mathbf{x}}$  with arbitrary input vector  $\bar{\mathbf{x}}$  can be supplied to ARPACK. Note that  $\bar{\mathbf{Z}}^{-1}$  is approximately computed and preconditioned Krylov solvers may be needed for this function. However, preconditioned iterative solvers can significantly slow down the ARPACK computation due to significantly more numbers of matrix-vector multiplication required by a sequence of Arnoldi factorization from ARPACK and instead we compute  $\bar{\mathbf{Z}}^{-1}$  with small compression tolerance and directly pass it to ARPACK without iterative solvers. It's worth mentioning that to properly reorder the matrix  $\bar{\mathbf{Z}}$ , we associate each unknown with a Cartesian coordinate and use e.g., KD tree clustering algorithms or cobblestone-like schemes [40]. For each flat/curvilinear triangle, there are  $(p+3)(p+1)$  associated GWP( $p$ ) bases  $f_n(\mathbf{r})$ . We use the edge centers as the coordinates for the first  $3(p+1)$  bases and triangle center for the next  $p(p+1)$  bases. For each basis function  $\hat{\mathbf{z}} \times \mathbf{e}_n(\mathbf{r})$ , we use the port center as the coordinate. This list of  $N_J + N_M$  coordinates is then passed to the clustering algorithm to compute the matrix reordering.

Note that in the context of 3D surface IE, HODLR displays asymptotically unfavorable computational complexity for HOM modeling, as it employs both weak admissibility and low-rank compression (formats not intended for high-frequency Green's functions). However, HODLR remains highly efficient for the first few HOMs of a cavity when considering moderate system sizes (e.g.,  $N < 20,000$ ), and as such, we employ HODLR as the default direct solver in IERD. For HOMs of very high frequencies, we can switch to HODBF whose CPU and memory complexities are asymptotically superior. See Section V-A for numerical results.

#### IV. AUGMENTED BAYESIAN OPTIMIZATION FOR SEARCHING RESONANCE FREQUENCIES

The section describes the proposed efficient frequency search algorithm (Algorithm 1) that combines the global convergent behavior of BO and local refinement behavior of the down-hill simplex algorithm.

Assume that within the frequency band of interest  $[f_{\min}, f_{\max}]$  there are a total of  $N_{\text{mod}}$  modes, and the modes indexed by  $m$  are sorted by their resonance frequency  $f^m$  in ascending orders. For each mode  $m$ , there is at most one eigenpair  $(e^m(f), \bar{\mathbf{I}})$  returned by ARPACK at any frequency  $f$ . If no such eigenpair for mode  $m$  exists at  $f$ , we simply let

$e^m(f) = 1$ . Such a function  $e^m(f)$  has only one global yet very sharp minimum at  $f^m$ , particularly for resonance modes without efficient power damping (i.e., with high Q factor). Efficient optimization algorithms need to efficiently locate the resonance regions with  $e^m(f) < 1$  and shrink to the global minimum  $f^m$  for all resonance mode  $m$ .

Let  $\{(f_i^m, e_i^m)\}$  be the samples of  $e^m(f)$ . In BO [34], [41], we assume a GP model  $y^m(f)$  for  $e^m(f)$  such that  $y^m(f) \sim GP(\mu, \Sigma)$  with mean  $\mu = 0$  and covariance

$$\Sigma(f, f') = \sigma \exp\left(\frac{(f - f')^2}{l}\right) + d\delta(f, f') \quad (13)$$

where  $\sigma, l, d$  are hyperparameters. From the samples  $\{(f_i^m, e_i^m)\}$ , these hyperparameters can be learned by optimizing the log-likelihood function of the GP model, known as the process of fitting a GP model to data. Once fitted, the GP surrogate can be used to quickly predict the mean and variance (or standard deviation) of  $e^m(f)$  at any unknown frequency  $f$ . The mean and variance information can be used to define an acquisition function such as the Expected Improvement (EI), which is used to propose new trial frequency  $f^*$ . IERD will then evaluate  $f^*$  with ARPACK to append the new sample  $(f^*, e^m(f^*))$  to  $\{(f_i^m, e_i^m)\}$ , such as the Fun\_Eval function in Algorithm 1. Note that everytime called, Fun\_Eval will add a new sample to all detected modes so far, depending on the eigen pairs returned by ARPACK. More specifically, the returned eigenvectors and those from existing modes can be grouped with the mode tracking technique [16] by computing their inner products. Two eigenvectors  $\bar{\mathbf{I}}$  and  $\bar{\mathbf{I}}^m$  are the same eigenmode only when  $\bar{\mathbf{I}}^* \cdot \bar{\mathbf{I}}^m > \epsilon_{\text{dot}}$  for a prescribed threshold  $\epsilon_{\text{dot}}$  ( $\epsilon_{\text{dot}} = 0.7 - 0.9$ ). Moreover, the recorded eigenvectors for existing modes  $\bar{\mathbf{I}}^m$  are replaced by  $\bar{\mathbf{I}}$  when  $\bar{\mathbf{I}}^* \cdot \bar{\mathbf{I}}^m > \epsilon_{\text{dot}}$  and the new eigenvalue  $e < \min_i e_i^m$  (see Line 35 of Algorithm 1). Note that for each trial frequency  $f$ , the Fun\_Eval function will generate a sample  $(f_i^m, e_i^m)$  for all modes, but only those with  $e_i^m$  returned by ARPACK (i.e.,  $e_i^m < 1$ ) denote a detected resonance.

The above procedure is repeatedly executed until the total number of affordable frequency trials per mode has reached a prescribed number  $n_{\text{BO}}$  (See Line 9 of Algorithm 1). The advantage of BO for IERD is that BO requires typically a much smaller number of trial frequencies compared with naive frequency sweep algorithms. That said, the objective function  $e^m(f)$  for each mode exhibits a sharp global minimum and the smooth and stationary kernel in (13) is not suited to detect the true non-smooth minimum [42], [43].

To improve the search quality, we augment BO with a down-hill simplex refinement inside each resonance region using BO samples  $(f_i^m, e_i^m)$  (see Line 19 of Algorithm 1). The down-hill simplex optimization is a very efficient derivative-free local algorithm requiring a small number of function samples in our application. However, this algorithm heavily rely on good initial guesses and search regions. Given that the objective functions  $e^m(f)$  has a pre-defined "V"-like shapes, we can leverage BO samples to define the lower and upper frequency bounds for the simplex search. As an illustration, Fig. 2 shows three possible configurations of the boundary (shown in red circles) and optimal (shown in red squares) BO samples and

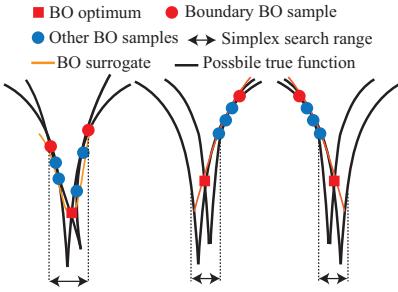


Fig. 2. Identification of the search range for down-hill simplex based on the BO samples  $\{(f_i^m, e_i^m)\}$ , particularly, the three possible relative locations of optimal ( $f_{\min}^m = \arg \min_i e_i^m$ ) and boundary BO samples ( $\min f_i^m$  and  $\max f_i^m$ ) shown in the left, middle and right.

potential locations of the global minimum. Note that standard GP surrogate model in BO (shown in orange curves) cannot represent the true objective function sufficient accuracy near a non-smooth optimum. More specifically, from the BO samples  $(f_i^m, e_i^m)$  of one mode let  $f_{\min}^m = \arg \min_i e_i^m$ . If  $f_{\min}^m$  appears in between  $\min f_i^m$  and  $\max f_i^m$ , the simplex search range is  $(f_{\min}^{SX}, f_{\max}^{SX}) = (\min f_i^m, \max f_i^m)$ ; otherwise if  $f_{\min}^m$  equals  $\min f_i^m$ , the simplex search range is  $(f_{\min}^{SX}, f_{\max}^{SX}) = ((1 - \alpha)f_{\min}^m, f_{n_r})$  where  $\alpha = 0.004$  is a small constant and  $f_{n_r}$  is the first trial frequency to the right of  $f_{\min}^m$ ; the case that  $f_{\min}^m$  equals  $\max f_i^m$  is treated in a similar fashion. Together with these search ranges, we can use their midpoints  $(f_{\min}^{SX} + f_{\max}^{SX})/2$  as the initial guess to down-hill simplex. See Line 19 of Algorithm 1 for more details.

In our implementation, we use two Python packages: the GPTune software [44] for the BO optimization and the SciPy software [45] for the down-hill simplex algorithm. For each trial frequency, the IE-based linear eigensolver (i.e., Butterfly-PACK and ARPACK) is invoked from the Python code with multiple MPIs ranks. Overall, the runtime for the proposed IERD framework is at most  $(n_{BO} + n_{SX})N_{mod}T_{eigen}$ , where  $T_{eigen}$  is the CPU time for each trial frequency  $f$ .

## V. NUMERICAL RESULTS

This section provides several numerical examples to demonstrate the efficiency, accuracy and robustness of the proposed IERD framework for modeling resonance modes in both canonical and realistic RF accelerator cavities. When the cavity involves waveguide ports with irregular-shaped cross sections, we use CST Studio Suite [46]'s 2D eigenmode solver to precompute the basis functions  $z \times e_j(r)$ . As the reference, we also show computational results using CST's 3D eigenmode solver [46] and the SLAC Omega3P eigenmode solver [5], [6], both of which are state-of-the-art FEM solvers for electromagnetic simulations. For Omega3P, we use the quadratic eigenproblem formulation as described in [6].

All numerical experiments are performed on the Cori Haswell system at NERSC. Cori Haswell is a Cray XC40 system and consists of 2388 dual-socket nodes with Intel Xeon E5-2698v3 processors running 16 cores per socket. The nodes are equipped with 128 GB of DDR4 memory and are connected through the Cray Aries interconnect.

---

### Algorithm 1 Hybrid BO and down-hill simplex optimization for frequency search

---

```

1: Input:
2:  $n_{BO}$ : number of trial frequencies per mode in BO
3:  $n_{SX}$ : number of trial frequencies per mode in down-hill simplex
4:  $n_{AK}$ : number of smallest eigenvalues for ARPACK
5:  $(f_{\min}, f_{\max})$ : lower and upper bounds for the search
6: Output:
7:  $N_{mod}$ : number of resonance modes (initialized to 0)
8:  $(f_i^m, e_i^m)$ : frequencies and the eigenvalues of each mode  $m \leq N_{mod}$  for all trials  $i$ .  $f^m = \arg \min_i e_i^m$  is the resonance frequency of mode  $m$ .
9: BO:
10: Generate  $n'_{BO} = n_{BO}/2$  random frequencies in  $[f_{\min}, f_{\max}]$ . For each frequency sample  $f$ , call Fun_Eval( $f, N_{mod}$ )
11:  $\triangleright$  The number of initial random samples  $n'_{BO} = n_{BO}/2$  is a heuristic choice.
12: for  $n = 1$  to  $n_{BO} - n'_{BO}$  do
13:   for each existing mode  $m \leq N_{mod}$  do
14:     Fit a GP model using data  $\{(f_i^m, e_i^m)\}$ 
15:     Use the EI function to propose the next trial frequency  $f$ 
16:     Fun_Eval( $f, N_{mod}$ )
17:   end for
18: end for

19: Downhill-simplex:
20: for each existing mode  $m \leq N_{mod}$  do
21:    $f_{\min}^m = \arg \min_i e_i^m$ 
22:   if  $\min f_i^m < f_{\min}^m < \max f_i^m$  then
23:      $(f_{\min}^{SX}, f_{\max}^{SX}) = (\min f_i^m, \max f_i^m)$ 
24:   else if  $\min f_i^m == f_{\min}^m$  then
25:      $(f_{\min}^{SX}, f_{\max}^{SX}) = ((1 - \alpha)f_{\min}^m, f_{n_r})$   $\triangleright f_{n_r}$  denotes the right neighbor of  $f_{\min}^m$ , and  $\alpha = 0.004$  is a small constant.
26:   else if  $\max f_i^m == f_{\min}^m$  then
27:      $(f_{\min}^{SX}, f_{\max}^{SX}) = (f_{n_l}, (1 + \alpha)f_{\min}^m)$   $\triangleright f_{n_l}$  denotes the left neighbor of  $f_{\min}^m$ 
28:   end if
29:   Call down-hill simplex algorithm with search range  $[f_{\min}^{SX}, f_{\max}^{SX}]$  for  $n_{SX}$  samples. Each sample calls Fun_Eval( $f, N_{mod}$ ).
30: end for

31: procedure FUN_EVAL( $f, N_{mod}$ )
32:   Call ARPACK with HODLR/HODBF for  $n_{AK}$  smallest eigenvalues. Ignore eigen pairs  $(e, \bar{I})$  with  $\|\bar{I}\|_1/\|\bar{I}\|_\infty < \epsilon_{norm}$ .
33:   for each existing mode  $m \leq N_{mod}$  with eigen vector  $\bar{I}^m$  do
34:     if  $\exists$  eigen pair  $(e, \bar{I})$  from ARPACK such that  $\bar{I}^* \cdot \bar{I}^m > \epsilon_{dot}$  then
35:       Let  $\bar{I}^m = \bar{I}$  if  $e < \min_i e_i^m$ 
36:       Remove  $(e, \bar{I})$  from ARPACK results and Append  $(f, e)$  to  $\{(f_i^m, e_i^m)\}$ 
37:     else
38:       Append  $(f, 1)$  to  $\{(f_i^m, e_i^m)\}$ 
39:     end if
40:   end for
41:   for each remaining eigen pair  $(e, \bar{I})$  from ARPACK do
42:      $N_{mod} = N_{mod} + 1$ 
43:     Let  $(f_1^{N_{mod}}, e_1^{N_{mod}}) = (f, e)$  and  $\bar{I}^{N_{mod}} = \bar{I}$ 
44:   end for
45: end procedure

```

---

#### A. Pillbox cavity with no port

In this subsection, we use a canonical pillbox cavity of radius 0.1 m and height 0.1 m to demonstrate the computational efficiency of the proposed IERD framework. First, we validate

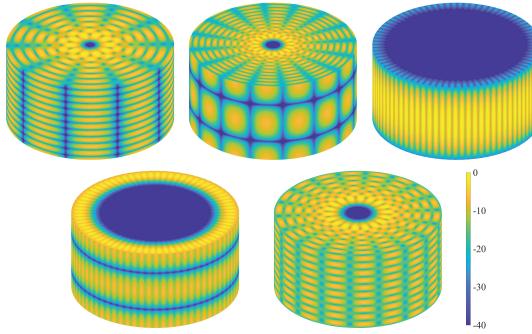


Fig. 3.  $|J|$  (in dB) for the five resonance modes of the pillbox cavity model detected at  $f=20.472$  GHz with the proposed solver.

the accuracy of IERD by setting  $f = 20.472$  GHz, one of the HOM frequencies without frequency search. This leads to a diameter of  $13.6 \lambda$  and we use a mesh of 213,792 triangle elements and  $N = N_J = 320,688$  RWG basis functions. When no port is considered, the system matrix  $\bar{\mathbf{Z}}$  reduces to  $\bar{\mathbf{T}}$  of (8). The inverse  $\bar{\mathbf{Z}}^{-1}$  is computed with the HODLR direct solver with compression tolerance  $\epsilon = 10^{-4}$  with 16 Cori nodes and supplied to ARPACK for the smallest 10 eigenvalues. The five detected eigenmodes  $\mathbf{J}$  corresponding to valid resonance are plotted in Fig. 3, which agree with the analytical solutions.

Next, different solver options are compared to demonstrate the efficiency of IERD. In addition to the HODLR direct solver with RWG basis functions, we also tested the HODBF direct solver with RWG basis functions, and the HODLR direct solver with GWP(1) basis functions and curvilinear elements. The system size  $N$  is kept similar across the three solvers and technical data is listed in Table I. Note that the use of GWP(1) basis on a coarser mesh can lead to significant reduction in CPU time. This is mainly due to that the time for compressing  $\bar{\mathbf{Z}}$  is dominated by that of computing entries of submatrices of  $\bar{\mathbf{Z}}$  required by the adaptive cross approximation (ACA) [47]. Given similar numbers of unknowns  $N$  (and similar ranks), GWP(1) requires a much less number of triangles than RWG. As the matrix is assembled on a triangle basis, the entry evaluation using GWP(1) becomes much faster. Also, the use of HODBF instead of HODLR can significantly reduce the memory usage for very high-order HOM modeling as HODBF can achieve significantly better compression performance than HODLR for high-frequency problems [32]. As the reference, we also used the Omega3P solver on 16 Cori nodes to detect resonance modes around 20.5 GHz with a tetrahedron mesh of similar mesh density. This leads to a FEM matrix of 17 million unknowns and the CPU time is similar to our IERD framework. That said, IERD with HODBF requires only 51.1 GB memory while Omega3P requires 1.56 TB memory.

#### B. Single-cell cavity with two rectangular waveguide ports

In this subsection, we demonstrate the efficacy of IERD for a realistic single-cell cavity with two rectangular waveguide ports, a simulation model for the third harmonic cavity in the Advance Light Source at Lawrence Berkeley

TABLE I  
COMPARISON OF THE HODLR AND HODBF-ENHANCED EIGEN SOLVER FOR THE PILLBOX CAVITY MODEL WITH  $f=20.472$  GHZ USING 16 CORI HASWELL NODES.

Algorithm	HODLR	HODBF	HODLR	Omega3P
Diameter	$13.6 \lambda$	$13.6 \lambda$	$13.6 \lambda$	$13.6 \lambda$
#of triangles	213,792	213,792	53,240	-
#of tetrahedrons	-	-	-	14,861,793
Max edge length	2.1 mm	2.1 mm	4 mm	2.7 mm
Basis	RWG	RWG	GWP(1)	Nedelec
#DoFs	320,688	320,688	266,200	17,185,459
Compression tolerance	$10^{-4}$	$10^{-4}$	$10^{-4}$	-
Assembly&compress	39 min	8 min	12 min	16.8 min
Max. rank	4113	380	4037	-
Inversion time	54 s	38 min	52 s	-
Total memory	250 GB	51.1 GB	200.8 GB	1.56 TB
Eigen solver time	13.4 s	49.2 s	11.9 s	7.4 min
#of detected modes	5	5	5	5

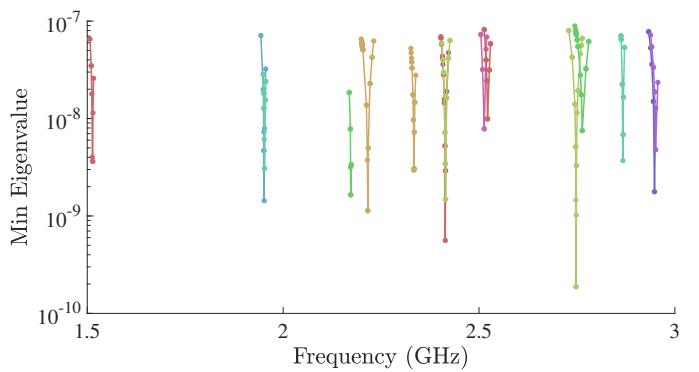


Fig. 4. Eigenvalues of the first 15 resonance modes in [1.5 GHz, 3 GHz] for the single-cell cavity computed by the proposed solver on 16 Cori Haswell nodes. The rectangular waveguides are treated as closed boundaries. Each color represents one detected resonance with trial frequencies (the dots) suggested by the proposed solver. The detected resonance frequency for each mode is the frequency where the eigenvalue attains the minimum. Note that the frequency samples  $f_i^m$  with  $e_i^m = 1$  are not plotted.

National Laboratory [48]. The two ports have an identical cross section and mode vectors for the waveguide are  $\mathbf{e}_j = \frac{m}{b} \cos \frac{n\pi x}{a} \sin \frac{m\pi y}{b} \hat{x} - \frac{n}{a} \sin \frac{n\pi x}{a} \cos \frac{m\pi y}{b} \hat{y}$  for  $TE_{nm}$  modes and  $\mathbf{e}_j = \frac{n}{a} \cos \frac{n\pi x}{a} \sin \frac{m\pi y}{b} \hat{x} + \frac{m}{b} \sin \frac{n\pi x}{a} \cos \frac{m\pi y}{b} \hat{y}$  for  $TM_{nm}$  modes, where  $x, y$  and  $\hat{x}, \hat{y}$  are local coordinates, and  $a = 70$  mm,  $b = 20$  mm in this model. The lowest cutoff frequency is  $f_c = 2.141$  GHz of the  $TE_{10}$  mode, and we search the resonance modes in [1.5 GHz, 3 GHz]. Therefore  $N_M = 0$  when  $f < f_c$ , and  $N_M = 2$  when  $f \geq f_c$  (i.e., each port supports at most one propagating mode). The model discretization leads to  $N_J = 25,455$  RWG basis functions and ARPACK computes the smallest 20 eigenvalues. Each trial frequency  $f$  requires about 35 s CPU time when  $n_{AK}=20$  and 5 GB memory using HODLR on 16 Cori nodes. First, we test the proposed frequency searching algorithm by setting  $N_M = 0$  regardless of the trial frequency (i.e., the waveguide ports are treated as closed boundaries). Fig. 4 plots the eigenvalues corresponding to the first 15 modes each represented by one color. Since closed boundaries are used, each mode exhibits a narrow bandwidth and a small minimum eigenvalue. Fig. 5 plots the eigenvalues

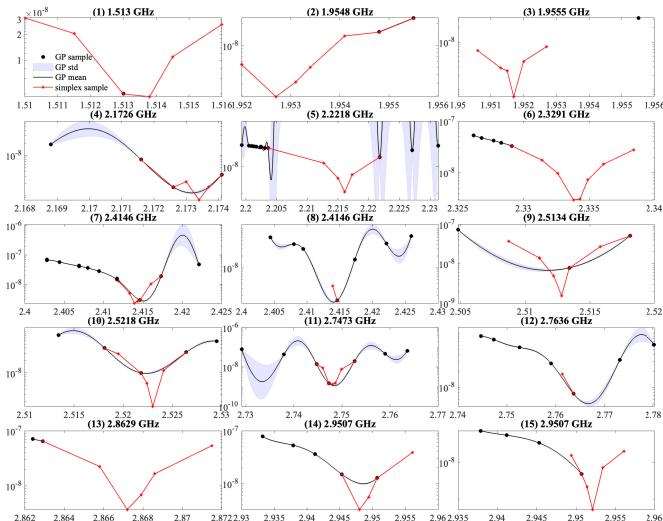


Fig. 5. Zoomed view of the 15 resonance modes in Fig. 4. Each subfigure shows one resonance mode numbered by its resonance frequency. The black dots, curves and blue shaded areas represent the samples, predicted mean and predicted standard deviation (std) by the Bayesian optimization, respectively. The red curves represent samples generated by the downhill simplex optimization. Note that the sensitivity or Q factor of the resonance is reflected by the bandwidth of the data.

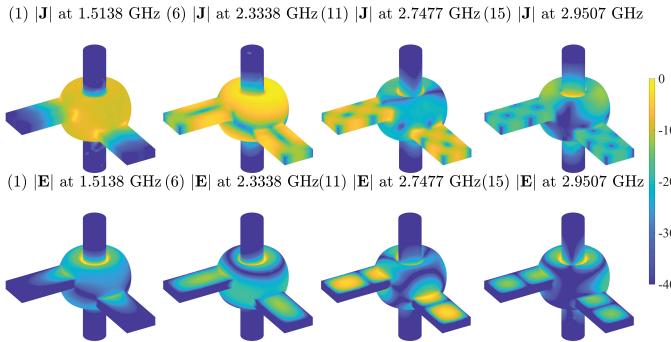


Fig. 6.  $|J|$  and surface  $|E|$  (in dB) for the 1<sup>st</sup>, 6<sup>th</sup>, 11<sup>th</sup>, and 15<sup>th</sup> resonance modes for the single-cell cavity with closed boundaries at the ports computed by the proposed solver. Note that the 1<sup>st</sup> mode represents the working mode.

for each mode separately. The black dots and red dots represent trial frequencies selected by the Bayesian optimization and down-hill simplex algorithm, respectively. Clearly, Bayesian optimization can locate the resonance region for all modes and down-hill simplex can effectively refine the resonance frequencies to find the sharp minimum. The detected resonance frequencies are labeled in the title of each subfigure. Fig. 6 shows the  $|J|$  and surface  $|E|$  for the 1<sup>st</sup>, 6<sup>th</sup>, 11<sup>th</sup>, and 15<sup>th</sup> modes where 1<sup>st</sup> mode is the working mode. Here  $E$  is simply computed by  $E = \frac{\nabla \cdot J}{i\omega_0}$ . Note that  $|E| = 0$  on the ports.

Next, we test the proposed frequency search algorithm with  $N_M = 2$  for  $f \geq f_c$ . Fig. 7 plots the eigenvalues corresponding to the first 12 modes each represented by one color. Note that the first three modes are below  $f_c$  and 9<sup>th</sup> mode is not damped efficiently via the wavguide ports. Therefore they exhibit narrow bandwidths and small minimum eigenvalues. All the other eight modes are damped efficiently and they

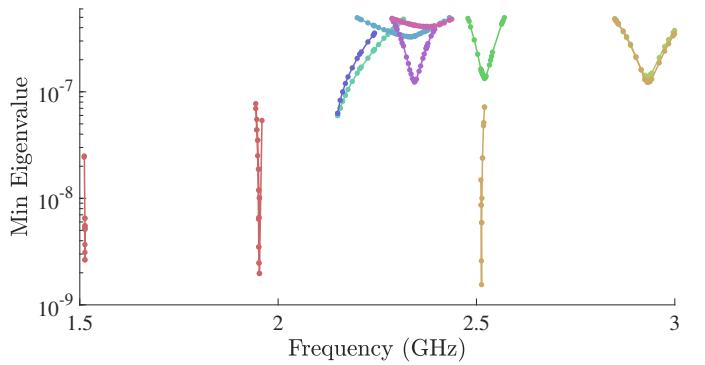


Fig. 7. Eigenvalues of the first 12 resonance modes in [1.5 GHz, 3 GHz] for the single-cell cavity with two rectangular waveguide ports computed by the proposed solver on 16 Cori Haswell nodes. Each color represents one detected resonance with trial frequencies (the dots) suggested by the proposed solver. The detected resonance frequency for each mode is the frequency where the eigenvalue attains the minimum. Note that most HOMs exhibit larger eigenvalues and higher bandwidth as the power is efficiently damped by the ports.

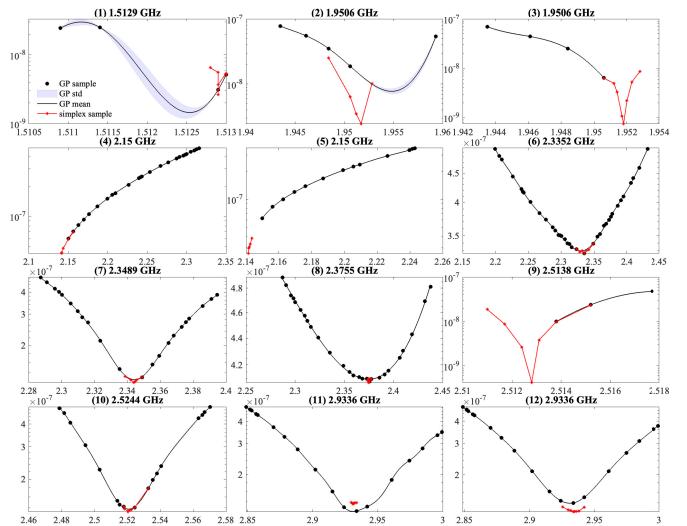


Fig. 8. Zoomed view of the 12 resonance modes in Fig. 7. Each subplot shows one resonance mode numbered by its resonance frequency. The black dots, curves and blue shaded areas represent the samples, predicted mean and predicted standard deviation (std) by the Bayesian optimization, respectively. The red curves represent samples generated by the downhill simplex optimization. Note that the sensitivity or Q factor of the resonance is reflected by the bandwidth of the data.

exhibit higher bandwidths and larger minimum eigenvalues. Fig. 8 plots the eigenvalues for each mode separately. The black dots and red dots represent trial frequencies selected by the Bayesian optimization and down-hill simplex algorithm, respectively. For the aforementioned four modes (subfigures (1)(2)(3)(9)) simplex refinement is critical for identifying more accurate resonance frequencies. That said, for the other modes whose power is damped efficiently, Bayesian optimization itself can yield sufficiently accurate solutions. Fig. 9 shows the  $|J|$  and surface  $|E|$  for the 4<sup>th</sup>, 7<sup>th</sup>, 8<sup>th</sup>, and 9<sup>th</sup> modes. Note that the 9<sup>th</sup> mode doesn't have sufficient power damping.

Furthermore, we measure the damping efficiency of each mode by computing the ratio  $P_w/P_p$ , where  $P_w = \frac{1}{2}\text{Re}\{<$

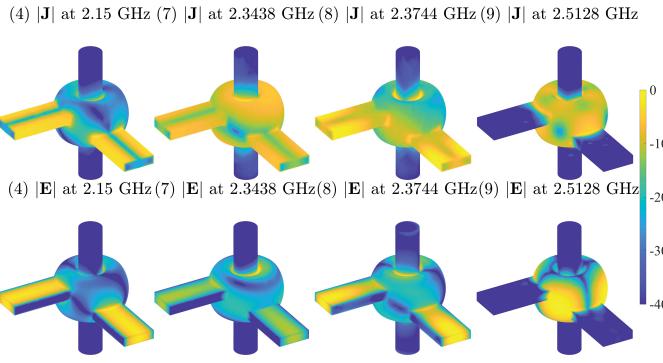


Fig. 9.  $|\mathbf{J}|$  and surface  $|\mathbf{E}|$  (in dB) for the 4<sup>th</sup>, 7<sup>th</sup>, 8<sup>th</sup>, and 9<sup>th</sup> HOMs for the single-cell cavity with open boundaries at the ports computed by the proposed solver. Note that the 4<sup>th</sup>, 7<sup>th</sup>, and 8<sup>th</sup> modes show efficient power damping.

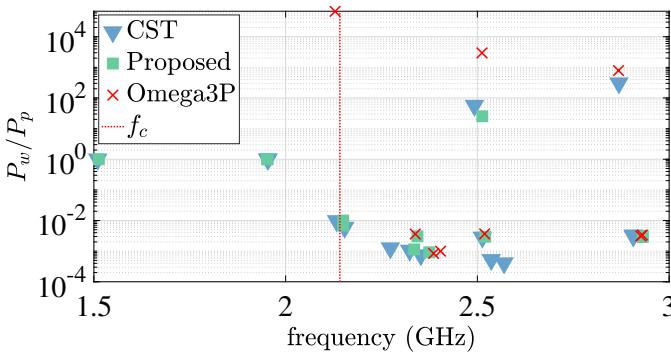


Fig. 10. The power ratios  $P_w/P_p$  of all resonance modes in [1.5 GHz, 3 GHz] detected by CST and the proposed solver, and for all resonance modes in [2.141 GHz, 3 GHz] detected by Omega3P, for the single-cell cavity with two rectangular waveguide ports. The red vertical line denotes the lowest cutoff frequency  $f_c = 2.141$  GHz of the ports, below which  $P_p = 0$  and we simply set  $P_w/P_p = 1$ .

$\mathbf{J}^*, Z_s \mathbf{J} >$  } is the power dissipated at the cavity wall, and  $P_p = \frac{1}{2} \text{Re}\{< \mathbf{J}^*, \mathbf{M} >\}$  is the power damped from the waveguide ports. The ratio is plotted in Fig. 10 for the 12 resonance modes. For modes detected below  $f_c$  we simply set  $P_w/P_p = 1$  for plotting purpose. As the reference, the ratios for the modes detected by CST and Omega3P are also plotted. It's worth mentioning that in this comparison study, for modes above  $f_c$ , CST doesn't yield accurate resonance frequencies. In contrast, Omega3P results agree better with IERD for  $f > f_c$ . That said, Omega3P cannot compute resonance modes accurately for  $f \leq f_c$  and  $f \approx f_c$  when waveguide port boundaries are used. For example, see the ratio for the mode near  $f = f_c$ .

#### C. Five-cell cavity with six H-shaped waveguide ports

In this subsection, the applicability of IERD to complicated cavity geometry with irregular-shaped ports are demonstrated using a five-cell cavity model with six H-shaped waveguide ports with identical cross sections, a simulation model for the Energy Recovery LINAC in Electron Ion Collider at Brookhaven National Laboratory [49].

The cutoff frequencies for the first four propagating modes of the ports are  $f_c = 0.776$  GHz, 2.132 GHz, 2.132 GHz, and

2.185 GHz. The model discretization leads to  $N_J = 41,223$  RWG basis functions and ARPACK computes the smallest 100 eigenvalues. Depending on the trial frequency  $f$ , one can have  $0 \leq N_M \leq 24$  (i.e., each port supports at most four propagating modes). Each trial frequency  $f$  requires about 85 s CPU time when  $n_{AK}=100$  and 8.5 GB memory using HODLR on 16 Cori nodes. We perform two frequency searches for [0.6 GHz, 1 GHz] and [2.1 GHz, 2.21 GHz], separately. IERD finds 22 and 35 resonances modes, respectively. Their eigenvalues are plotted in Fig. 11 where each color denotes one eigenmode and each dot represents one trial frequency. Again, high-Q modes exhibit narrow bandwidths and smaller minimum eigenvalues, and low-Q modes exhibit higher bandwidths and larger minimum eigenvalues. Fig. 12 shows the surface  $|\mathbf{E}|$  for the first 14 modes in [0.6 GHz, 1 GHz] and 7 selected modes in [2.1 GHz, 2.21 GHz] with their resonance frequencies shown in the subfigure titles. Note that the first 5 modes are below the lowest cutoff frequencies  $f_c=0.776$  GHz and the 5<sup>th</sup> mode at 0.65244 GHz is the working mode.

Finally, the power ratios  $P_w/P_p$  are plotted in Fig. 13. Note that CST cannot yet handle waveguide ports not aligned with X/Y/Z plane in Cartesian coordinates and we only compare with Omega3P. We added the eigenmodes detected by Omega3P in [0.776 GHz, 1 GHz] which allows one propagating mode per waveguide port. As can be seen from Fig. 13, Omega3P and IERD results match well, but Omega3P is missing a few resonance modes near  $f_c = 0.776$  GHz.

## VI. CONCLUSION

This paper presents IERD, an efficient Integral Equation (IE)-based resonance detection tool for modeling RF cavities with arbitrarily-shaped cross sections and damping waveguide ports supporting multiple propagating modes. IERD leverages fast direct solvers such as HODLR or HODBF from ButterlyPACK, the ARPACK eigensolver to identify resonance modes at specific trial frequencies, and a hybrid Bayesian optimization and downhill simplex algorithm to minimize the number of trial frequencies required. The tool also supports higher-order basis functions and curvilinear surface mesh representation.

It was shown with numerical experiments that IERD can detect HOMs effectively and accurately over wide frequency bands for realistic RF accelerator cavity modeling with significant lower memory usage compared to FEM-based formulations and tunable CPU requirement depending on the frequency band and allowed maximum numbers of trial frequencies.

IERD contributes to RF cavity modeling with a novel approach that combines IE methods, fast direct solvers, and advanced optimization algorithms. Its novel hybrid Bayesian optimization and downhill simplex algorithm efficiently locates resonance regions with minimal trial frequencies. Supporting arbitrarily-shaped cross sections, damping waveguide ports, higher-order basis functions, and curvilinear surface mesh, IERD offers a practical and efficient solution for addressing limitations of conventional techniques. Future work includes efficient schemes for post-computing the electric fields inside

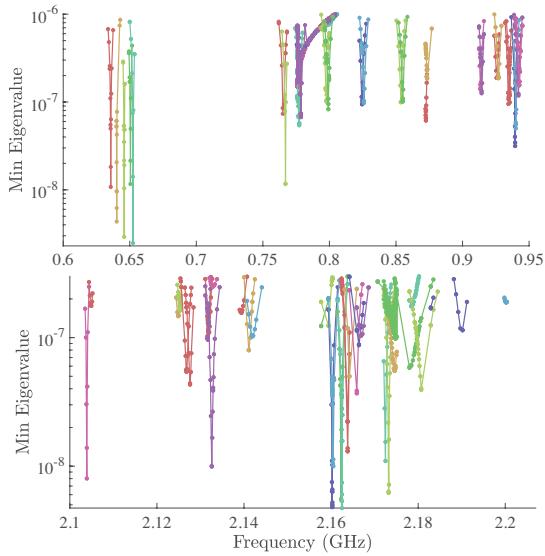


Fig. 11. Eigenvalues of the first 22 resonance modes in [0.6 GHz, 1 GHz] and the first 35 resonance modes in [2.1 GHz, 2.21 GHz] for the 5-cell cavity with 6 H-shaped waveguide ports computed by the proposed solver on 32 Cori Haswell nodes. Each color represents one detected resonance with trial frequencies (the dots) suggested by the proposed solver. The detected resonance frequency for each mode is the frequency where the eigenvalue attains the minimum. Note that most HOMs exhibit larger eigenvalues and higher bandwidth as power is efficiently damped by the ports.

the cavity from the surface current densities, and integrating IERD into the automatic cavity design process.

#### ACKNOWLEDGMENT

This research was supported in part by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration, and in part by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Scientific Discovery through Advanced Computing (SciDAC) program through the FAST-Math Institute under Contract No. DE-AC02-05CH11231 at Lawrence Berkeley National Laboratory. This work was supported in part by previous breakthroughs obtained through the Laboratory Directed Research and Development Program of Lawrence Berkeley National Laboratory under U.S. Department of Energy Contract No. DE-AC02-05CH11231. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231.

#### REFERENCES

- [1] J. Qiang, D. Bizzozero, L. Ge, C.-K. Ng, and L. Xiao, "Integration of cavity design and beam dynamics simulation using the parallel IMPACT and the ACE3P codes," SLAC National Accelerator Lab., Tech. Rep., 2019.
- [2] 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. A: Accel. Spectrom. Detect. Assoc. Equip.*, vol. 940, pp. 12–18, 2019.

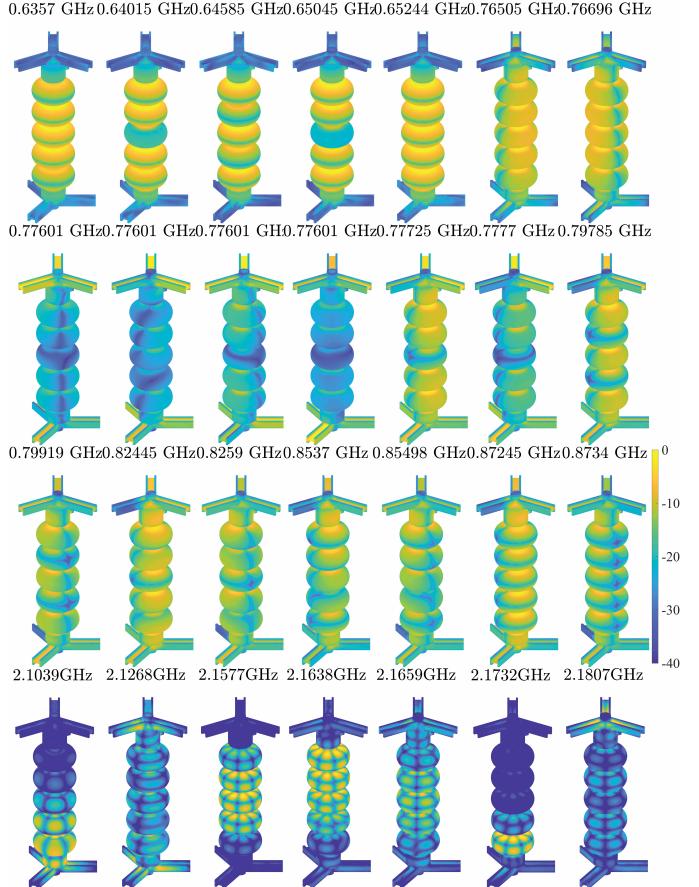


Fig. 12. surface  $|E|$  (in dB) for the first 14 resonance modes in [0.6 GHz, 1 GHz] and 7 selected resonance modes in [2.1 GHz, 2.21 GHz] for the 5-cell cavity with 6 H-shaped waveguide ports computed by the proposed solver. Note that the 5<sup>th</sup> mode is the working mode, and all modes above 0.6524 GHz are HOMs.

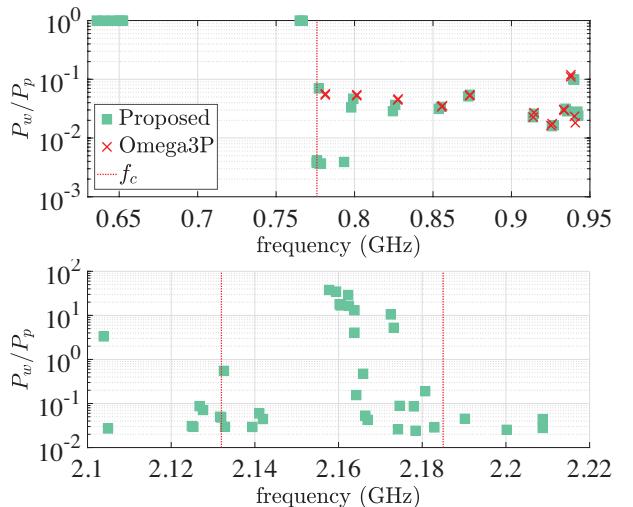


Fig. 13. The power ratios  $P_w/P_p$  for the 5-cell cavity with 6 H-shaped waveguide ports of all resonance modes in [0.6 GHz, 1 GHz] and [2.1 GHz, 2.21 GHz] detected by the proposed solver, and all resonance modes in [0.776 GHz, 1 GHz] detected by Omega3P. The red vertical lines denote the cutoff frequencies of the ports  $f_c = 0.776$  GHz, 2.132 GHz, 2.132 GHz, and 2.185 GHz. Below the lowest cutoff frequency 0.776 GHz,  $P_p = 0$  and we simply set  $P_w/P_p = 1$ .

- [3] R. Stanek, T. Arkan, J. Blowers, C. Ginsburg, A. Grassellino, C. Grimm, B. Hansen, E. Harms, B. Hartsell, J. Holzbauer *et al.*, “Overview of LCLS-II project status at Fermilab,” Fermi National Accelerator Lab., Tech. Rep., 2019.
- [4] J.-M. Jin, *The finite element method in electromagnetics*. John Wiley & Sons, 2015.
- [5] L. Xiao, L. Ge, Z. Li, and C.-K. Ng, “Advances in multiphysics modeling for parallel finite-element code suite ACE3P,” *IEEE J. Multiscale Multiphysics Comput. Tech.*, vol. 4, pp. 298–306, 2019.
- [6] L.-Q. Lee, Z. Li, C. Ng, and K. Ko, “Omega3P: A parallel finite-element eigenmode analysis code for accelerator cavities,” SLAC National Accelerator Lab., Tech. Rep., 2009.
- [7] B.-S. Liao, Z. Bai, L.-Q. Lee, and K. Ko, “Nonlinear Rayleigh-Ritz iterative method for solving large scale nonlinear eigenvalue problems,” *Taiwan. J. Math.*, vol. 14, no. 3A, pp. 869–883, 2010.
- [8] H. Voss, “An Arnoldi method for nonlinear eigenvalue problems,” *BIT Numer. Math.*, vol. 44, pp. 387–401, 2004.
- [9] W.-J. Beyn, “An integral method for solving nonlinear eigenvalue problems,” *Linear Algebra Appl.*, vol. 436, no. 10, pp. 3839–3863, 2012.
- [10] J. Asakura, T. Sakurai, H. Tadano, T. Ikegami, and K. Kimura, “A numerical method for nonlinear eigenvalue problems using contour integrals,” *SIAM Letters*, vol. 1, pp. 52–55, 2009.
- [11] R. Van Beeumen, O. Marques, E. G. Ng, C. Yang, Z. Bai, L. Ge, O. Kononenko, Z. Li, C.-K. Ng, and L. Xiao, “Computing resonant modes of accelerator cavities by solving nonlinear eigenvalue problems via rational approximation,” *J. Comput. Phys.*, vol. 374, pp. 1031–1043, 2018.
- [12] Q. Chang, Y. Wang, Y. Zhang, and X. Zhao, “Higher-order MoM combined waveport analysis of circular waveguide problems,” *IET Conference Proceedings*, 2015.
- [13] Y. Wang, X.-W. Zhao, Y. Zhang, S. Ting, T. K. Sarkar, and C. Liang, “Higher order MoM analysis of traveling-wave waveguide antennas with matched waveports,” *IEEE Trans. Antennas Propag.*, vol. 63, no. 8, pp. 3718–3721, 2015.
- [14] H. Auda and R. Harrington, “A moment solution for waveguide junction problems,” *IEEE Trans. Microwave Theory Tech.*, vol. 31, no. 7, pp. 515–520, 1983.
- [15] R. Bunger and F. Arndt, “Moment-method analysis of arbitrary 3-D metallic N-port waveguide structures,” *IEEE Trans. Microwave Theory Tech.*, vol. 48, no. 4, pp. 531–537, 2000.
- [16] Y. Chen and C.-F. Wang, *Characteristic modes: Theory and applications in antenna engineering*. John Wiley & Sons, 2015.
- [17] S. Huang, J. Pan, C.-F. Wang, Y. Luo, and D. Yang, “Unified implementation and cross-validation of the integral equation-based formulations for the characteristic modes of dielectric bodies,” *IEEE Access*, vol. 8, pp. 5655–5666, 2019.
- [18] R. Harrington and J. Mautz, “Theory of characteristic modes for conducting bodies,” *IEEE Trans. Antennas Propag.*, vol. 19, no. 5, pp. 622–628, 1971.
- [19] G. Angiulli, M. Versaci, and S. Calcagno, “Computation of the cutoff wavenumbers of metallic waveguides with symmetries by using a nonlinear eigenproblem formulation: A group theoretical approach,” *Mathematics*, vol. 8, no. 4, p. 489, 2020.
- [20] S. Huang, J. Pan, and Y. Luo, “Study on the relationships between eigenmodes, natural modes, and characteristic modes of perfectly electric conducting bodies,” *Int. J. Antennas Propag.*, vol. 2018, 2018.
- [21] W. C. Chew, E. Michielssen, J. Song, and J.-M. Jin, *Fast and efficient algorithms in computational electromagnetics*. Artech House, Inc., 2001.
- [22] Q. I. Dai, J. Wu, H. Gan, Q. S. Liu, W. C. Chew, and E. Wei, “Large-scale characteristic mode analysis with fast multipole algorithms,” *IEEE Trans. Antennas Propag.*, vol. 64, no. 7, pp. 2608–2616, 2016.
- [23] R. B. Lehoucq, D. C. Sorensen, and C. Yang, *ARPACK users’ guide: solution of large-scale eigenvalue problems with implicitly restarted Arnoldi methods*. SIAM, 1998.
- [24] S. Ambikasaran and E. Darve, “An O (N log N) fast direct solver for partial hierarchically semi-separable matrices,” *J. Sci. Comput.*, vol. 57, no. 3, pp. 477–501, 2013.
- [25] Y. Liu, H. Guo, and E. Michielssen, “An HSS matrix-inspired butterfly-based direct solver for analyzing scattering from two-dimensional objects,” *IEEE Antennas Wireless Propag. Lett.*, vol. 16, pp. 1179–1183, 2016.
- [26] E. Corona, P.-G. Martinsson, and D. Zorin, “An O(N) direct solver for integral equations on the plane,” *Appl. Comput. Harmon. Anal.*, vol. 38, no. 2, pp. 284 – 317, 2015.
- [27] L. Börm, Steffen Grasedyck and W. Hackbusch, “Introduction to hierarchical matrices with application,” *Eng. Anal. Bound. Elem.*, vol. 27, pp. 405–422, 2003.
- [28] R. Vandebril, M. Van Barel, G. Golub, and N. Mastronardi, “A bibliography on semiseparable matrices,” *Calcolo*, vol. 42, no. 3-4, pp. 249–270, 2005.
- [29] C. Gorman, G. Chávez, P. Ghysels, T. Mary, F.-H. Rouet, and X. S. Li, “Robust and accurate stopping criteria for adaptive randomized sampling in matrix-free hierarchically semiseparable construction,” *SIAM J. Sci. Comput.*, vol. 41, no. 5, pp. S61–S85, 2019.
- [30] M. Ma and D. Jiao, “Direct Solution of general  $\mathcal{H}^2$ -matrices with controlled accuracy and concurrent change of cluster bases for electromagnetic analysis,” *IEEE Trans. Microwave Theory Tech.*, vol. 67, no. 6, pp. 2114–2127, 2019.
- [31] H. Guo, Y. Liu, J. Hu, and E. Michielssen, “A butterfly-based direct integral-equation solver using hierarchical LU factorization for analyzing scattering from electrically large conducting objects,” *IEEE Trans. Antennas Propag.*, vol. 65, no. 9, pp. 4742–4750, 2017.
- [32] S. B. Sayed, Y. Liu, L. J. Gomez, and A. C. Yucel, “A butterfly-accelerated volume integral equation solver for broad permittivity and large-scale electromagnetic analysis,” *IEEE Trans. Antennas Propag.*, vol. 70, no. 5, pp. 3549–3559, 2022.
- [33] Y. Liu, P. Ghysels, L. Claus, and X. S. Li, “Sparse approximate multifrontal factorization with butterfly compression for high-frequency wave equations,” *SIAM J. Sci. Comput.*, no. 0, pp. S367–S391, 2021.
- [34] C. E. Rasmussen, C. K. Williams *et al.*, *Gaussian processes for machine learning*. Springer, 2006, vol. 1.
- [35] J. A. Nelder and R. Mead, “A simplex method for function minimization,” *Comput. J.*, vol. 7, no. 4, pp. 308–313, 1965.
- [36] S. Rao, D. Wilton, and A. Glisson, “Electromagnetic scattering by surfaces of arbitrary shape,” *IEEE Trans. Antennas Propag.*, vol. 30, no. 3, pp. 409–418, 1982.
- [37] R. D. Graglia, D. R. Wilton, and A. F. Peterson, “Higher order interpolatory vector bases for computational electromagnetics,” *IEEE Trans. Antennas Propag.*, vol. 45, no. 3, pp. 329–342, 1997.
- [38] Y. Liu and USDOE, “ButterflyPACK,” 11 2018. [Online]. Available: <https://www.osti.gov/biblio/1564244>
- [39] Y. Liu, X. Xing, H. Guo, E. Michielssen, P. Ghysels, and X. S. Li, “Butterfly factorization via randomized matrix-vector multiplications,” *SIAM J. Sci. Comput.*, vol. 43, no. 2, pp. A883–A907, 2021.
- [40] J. Shaeffer, “Direct Solve of Electrically Large Integral Equations for Problem Sizes to 1 M Unknowns,” *IEEE Trans. Antennas Propag.*, vol. 56, no. 8, pp. 2306–2313, 2008.
- [41] H. Luo, G. Nattino, and M. T. Pratola, “Sparse additive Gaussian process regression,” *J. Mach. Learn. Res.*, vol. 23, no. 61, pp. 1–34, 2022.
- [42] H. Luo, J. W. Demmel, Y. Cho, X. S. Li, and Y. Liu, “Non-smooth Bayesian optimization in tuning problems,” *arXiv:2109.07563*, 2021.
- [43] H. Luo, Y. Cho, J. W. Demmel, X. S. Li, and Y. Liu, “Hybrid models for mixed variables in Bayesian optimization,” *arXiv:2206.01409*, 2022.
- [44] Y. Liu, W. M. Sid-Lakhdar, O. Marques, X. Zhu, C. Meng, J. W. Demmel, and X. S. Li, “GPTune: Multitask learning for autotuning exascale applications,” in *Proceedings of the 26th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming*, ser. PPoPP ’21. Association for Computing Machinery, 2021, p. 234–246.
- [45] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, I. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python,” *Nat. Methods*, vol. 17, pp. 261–272, 2020.
- [46] E. S. Solvers and C. Compatibility, “CST studio suite,” *Vélizy-Villacoublay: Dassault Systèmes*, c2002–2022, 2020.
- [47] Y. Liu, W. Sid-Lakhdar, E. Rebrova, P. Ghysels, and X. S. Li, “A parallel hierarchical blocked adaptive cross approximation algorithm,” *Int. J. High Perform. Comput. Appl.*, vol. 34, no. 4, pp. 394–408, 2020.
- [48] J. Byrd, K. Baptiste, S. De Santis, S. Kosta, C. Lo, D. Plate, R. Rimmer, and M. Franks, “Design of a higher harmonic RF system for the advanced light source,” *Nucl. Instrum. Methods Phys. Res. A: Accel. Spectrom. Detect. Assoc. Equip.*, vol. 439, no. 1, pp. 15–25, 2000.
- [49] W. Xu, I. Ben-Zvi, H. Hahn, G. McIntyre, C. Pai, R. Porqueddu, K. Smith, J. Tuozzolo, J. Tuozzolo, and A. Zaltsman, “RF and mechanical design of 647 MHz 5-Cell BNL4 cavity for eRHIC ERL,” in *7th International Particle Accelerator Conference*, 2016, p. WEPMR041.