SLAC-PUB-13268
June 2008

Computational science research in support of petascale electromagnetic modeling

L Lee[1], V Akcelik[1], L Ge[1], S Chen[1], G Schussman[1], A Candel[1], Z Li[1], L Xiao[1], A Kabel[1], R Uplenchwar[1], C Ng[1], and K Ko[1]

[1]Stanford Linear Accelerator Center, 2575 Sand Hill Road, Menlo Park, CA 94025

E-mail: liequan@stanford.edu

SciDAC CET/Institute Collaborators: E Ng, X Li, C Yang (LBNL), L Dianchin (LLNL), K Devine, E Boman (SNL), B Osting, D Keyes (Columbia), X Luo, M Shepard (RPI), R Barrett, S Hodson, R Kendall (ORNL), W Gropp (UIUC), Z Bai, K Ma(UCDavis)

Abstract. Computational science research components were vital parts of the SciDAC-1 accelerator project and are continuing to play a critical role in newly-funded SciDAC-2 accelerator project, the Community Petascale Project for Accelerator Science and Simulation (ComPASS). Recent advances and achievements in the area of computational science research in support of petascale electromagnetic modeling for accelerator design analysis are presented, which include shape determination of superconducting RF cavities, mesh-based multi-level preconditioner in solving highly-indefinite linear systems, moving window using h- or p- refinement for time-domain short-range wakefield calculations, and improved scalable application I/O.

1. Introduction

Particle accelerators, such as Spallation Neutron Source (SNS), Linac Coherent Light Source (LCLS), Rare Isotope Accelerator, Large Hadron Collider, and proposed International Linear Collider, are pivotal experimental facilities for discoveries in physical sciences. Electromagnetic modeling of those existing or proposed billion-dollar class accelerator facilities often requires petascale computing and drives the computational science research in the related areas. Computational science research components were vital parts of the SciDAC-1 accelerator project [1, 2, 3, 4] and continue to play a critical role in the SciDAC-2 accelerator project, the Community Petascale Project for Accelerator Science and Simulation (ComPASS). Recent advances and achievements in the area of computational science research in support of petascale electromagnetic modeling for accelerator design and optimization are presented.

2. Shape Determination of Superconducting RF Cavities

Shape deviations of the real cavity from the design may significantly impact cavity response. A shape determination tool has been developed at Stanford Linear Accelerator Center (SLAC) to infer for the unknown shape deviations using measurable quantities such as cavity frequencies, field distributions and external quality factor values. A nonlinear least square optimization problem constrained by a complex Maxwell eigenvalue problem is solved in the

Contributed to Scientific Discovery through Advanced Computing Program (SciDAC) 2008 Conference, 7/13/2008-7/17/2008, Seattle, WA, USA

Work supported in part by US Department of Energy contract DE-AC02-76SF00515

shape determination process.

$$\underset{\mathbf{e}_j,k_j,\mathbf{d}}{\text{minimize}} \quad \mathcal{J}(\mathbf{e}_j,\mathbf{k}_j,\mathbf{d}) \quad \text{subject to} \quad \mathbf{K}\mathbf{e}_j + ik_j\mathbf{W}\mathbf{e}_j - k_j^2\mathbf{M}\mathbf{e}_j = 0 \tag{1}$$

$$\mathbf{e}_j^H\mathbf{M}\mathbf{e}_j = 1 \quad and \quad \Re(\mathbf{e}_j)^T\mathbf{M}\Im(\mathbf{e}_j) = 0 \tag{2}$$

where $(\mathbf{e}_j, k_j)$ is the jth eigenpair and $\mathbf{d}$ are the shape deviation parameters. $\Re$ denotes the real part and $\Im$ the imaginary part. More discussion about the complex quadratic eigenvalue problem Eq(1) can be found in [1].

The objective function $\mathcal{J}$ includes weighted least squares misfit of the computed and measured cavity response. A Gauss-Newton algorithm with a discrete adjoint method is used to solve the above nonlinear least square minimization problem. This problem is ill-posed, and a truncated singular value decomposition method is used to overcome the ill-posedness. Each nonlinear iteration of the algorithm requires solution of the forward problem (complex Maxwell eigenvalue problem), solution of adjoint problems (linear system of equations), and evaluation of the eigenvector and eigenvalue sensitivities. The algorithm is tested with examples and the solution of the shape deviation converges within a handful of nonlinear iterations.

The shape determination tool has been used to find the causes of high external quality factor in a high gradient prototype cryomodule at Jefferson Lab [5]. The algorithm predicted that the manufactured cavity was deformed significantly from the original shape, and its length was 6 to 8 mm shorter than that of the original design depending on the choice of inversion parameter set. The results agree well with those from the measurements.

3. Mesh-based Multilevel Preconditioner

In analyzing eigen-modes of accelerator cavities, the governing Maxwell's equations can be simplified to the following harmonic vector wave equation:

$$\nabla \times \left(\frac{1}{\mu}\nabla \times \overrightarrow{\mathbf{E}}\right) - \varepsilon k^2 \overrightarrow{\mathbf{E}} = 0 \tag{3}$$

With finite-element discretization using tangentially-continuous Nedelec basis functions [6], $\overrightarrow{\mathbf{E}} = \sum_i x_i \overrightarrow{\mathbf{N}}_i$, the above equation becomes an eigenvalue problem:

$$\mathbf{K}\mathbf{x} = k^2\mathbf{M}\mathbf{x} \tag{4}$$

where $\mathbf{K} = \int_\Omega \frac{1}{\mu}\left(\nabla \times \overrightarrow{\mathbf{N}}_i\right) \cdot \left(\nabla \times \overrightarrow{\mathbf{N}}_j\right) d\Omega$ and $\mathbf{M} = \int_\Omega \varepsilon \overrightarrow{\mathbf{N}}_i \cdot \overrightarrow{\mathbf{N}}_j \, d\Omega$. Note that matrix $\mathbf{M}$ is symmetric positive definite and matrix $\mathbf{K}$ is symmetric positive semi-definite with a large null space. Since the interior eigenvalues are of the interest in the accelerator cavity modeling, a shift-and-invert transform is often applied in the process of solving the above eigenvalue problem. This requires a solution of a highly indefinite linear system in every eigenvalue iteration, which is notoriously difficult to solve with iterative methods.

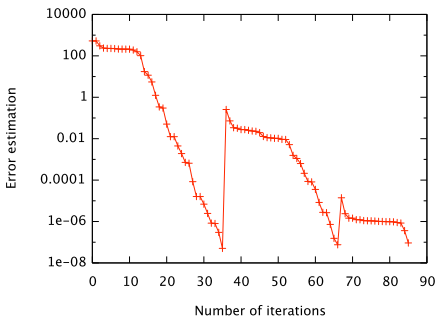

Figure 1. The convergence history using GMRES with multi-level preconditioner on solving a shifted linear system. Note that the error estimation of inner iterations of GMRES is also plotted.

$$(\mathbf{K} - \sigma\mathbf{M})\,\mathbf{x} = \mathbf{b} \tag{5}$$

where σ is a prescribed shift close to the eigenvalues of the interest.

A mesh-based multilevel preconditioner has recently been developed in solving Eq (5). Given the preconditioner system $\mathbf{Ax} = \mathbf{b}$, the vector $\mathbf{b}$, which is on a dense mesh, is first restricted onto a coarse mesh and becomes $\mathbf{b^c}$ (from now on, the variables with superscript c are denoted to those on the coarse mesh). Then $(\mathbf{K}^c - \sigma\mathbf{M}^c)\,\mathbf{x}^c = \mathbf{b}^c$ is solved with a sparse direct solver since its size can be very small. At the last step, the vector $\mathbf{x}^c$ is prolongated back onto the dense matrix and is regarded as the solution of the preconditioner system $\mathbf{x}$. For the restriction operation, a linear system is solved as follows

$$(\mathbf{M}^c + \alpha\mathbf{K}^c)\mathbf{b}^c = \int_\Omega \left(\overrightarrow{\mathbf{F}}(\mathbf{b}) \cdot \overrightarrow{\mathbf{N}}_i^c + \alpha(\nabla \times \overrightarrow{\mathbf{F}}(\mathbf{b})) \cdot (\nabla \times \overrightarrow{\mathbf{N}}_i^c) \right) \, d\Omega \tag{6}$$

where α is a positive parameter to be specified and $\overrightarrow{\mathbf{F}}(\mathbf{b}) = \sum_{\mathbf{i}} \mathbf{b_i} \overrightarrow{\mathbf{N}}_{\mathbf{i}}$ on the dense mesh. Note that the matrix $\mathbf{M}^c + \alpha\mathbf{K}^c$ is symmetric positive definite when α is positive and the above linear system can be solved using conjugate gradient with incomplete Cholesky preconditioner. A similar linear system is solved for the prolongation operation. It is shown in [7] that the above operation balanced both the error of the interpolating field and that of the curl of the interpolating field, which is important in electromagnetic modeling. Therefore, it is more effective to use the $\mathbf{M}^c + \alpha\mathbf{K}^c$ than using the mass matrix $\mathbf{M}^c$ alone.

Figure 1 shows the convergence history of the GMRES with this mesh-based multilevel preconditioner in solving a shifted linear system. With 85 iterations, the converged solution with residual of less than 10^{-7} is achieved using this method. As a comparison, it took about 518 iterations to achieve the same converged solution using CG with SSOR preconditioner and it failed to converge using CG with incomplete factorization based preconditioner. Work on the parallel implementation of this multilevel preconditioner is in progress.

4. Moving Window with h- or p- Refinement

In calculating short-range wakefield inside an accelerator structure, only the small region in the vicinity of the particle beam is required in the simulation. A moving window technique, in which the domain of the simulation is limited to a small region of interest near the beam and moves with it, can greatly save the computational efforts. This technique has been widely used in the finite-difference simulation but not in the finite-element simulation with unstructured grids. In this section, a brief introduction of finite-element time-domain (FETD) method used in T3P is given. Moving window techniques with curvilinear tetrahedral meshes using either h-refinement or p-refinement are then presented.

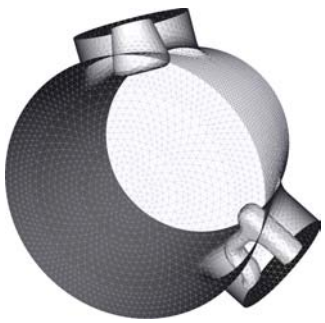

Figure 2. A couper region of an accelerating cavity.

In T3P, the following inhomogeneous vector wave equation is solved numerically for the electric field $\overrightarrow{\mathbf{E}}$:

$$\nabla \times \left(\frac{1}{\mu} \nabla \times \int_{-\infty}^{t} \overrightarrow{\mathbf{E}} \, \mathrm{d}\tau \right) + \varepsilon \frac{\partial^2}{\partial t^2} \int_{-\infty}^{t} \overrightarrow{\mathbf{E}} \, \mathrm{d}\tau = -\overrightarrow{\mathbf{J}}, \tag{7}$$

where $\overrightarrow{\mathbf{J}}$ is the electric current density, and ε and μ are the electric permittivity and magnetic permeability.

With finite-element spacial discretization [6], $\int_{-\infty}^{t} \overrightarrow{\mathbf{E}}(\mathbf{x},\tau)\,\mathrm{d}\tau = \sum_i e_i(t)\cdot\overrightarrow{\mathbf{N}}_i(\mathbf{x})$, and the implicit Newmark-beta scheme [8] for temporal discretization, the following linear system is solved for each time step:

$$\begin{aligned}\left(\mathbf{M}+\beta(\Delta t)^2\mathbf{K}\right)\mathbf{e}^{n+1} = &\left(2\mathbf{M}-(1-2\beta)(\Delta t)^2\mathbf{K}\right)\mathbf{e}^{n}-\left(\mathbf{M}+\beta(\Delta t)^2\mathbf{K}\right)\mathbf{e}^{n-1}\\ &-(\Delta t)^2\left(\beta\mathbf{f}^{n+1}+(1-2\beta)\mathbf{f}^{n}+\beta\mathbf{f}^{n-1}\right)\end{aligned} \tag{8}$$

where matrix $\mathbf{K}$ and $\mathbf{M}$ are the same as in Eq(4) and vector $\mathbf{f}=\int_\Omega \overrightarrow{\mathbf{N}}_i\cdot\overrightarrow{\mathbf{J}}\,d\Omega$ is the discretized representation of the current density. Note that the four discretized vectors $\mathbf{e}^n$, $\mathbf{e}^{n-1}$, $\mathbf{f}^n$, and $\mathbf{f}^{n-1}$, need to be transferred from the current window configuration to the next window configuration when the particle beam moves out of the window.

4.1. P-refinement

A window is defined with its front and back boundaries perpendicular to the velocity of the particle beam. Inside the window, the finite-element basis function order p of tetrahedral elements is set to be nonzero value while outside of the window p is zero. This effectively makes the number of degrees of freedom (DOF) to be zero outside of the window, therefore reducing the computational efforts. A padding zone is put between the front of the beam and the front boundary of the window so that the particle beam will stay in the window for a while. By changing the size of the padding zone, the window size is adjusted. When the particle beam moves out the padding zone, the window will move forward with a given distance. The vectors $\mathbf{e}^n$, $\mathbf{e}^{n-1}$, $\mathbf{f}^n$, and $\mathbf{f}^{n-1}$ as shown in Eq (8) are transferred element-wise according to the changes of the finite-element order p's. If p of a tetrahedral element increases, zeroes are filled for those additional coefficients. On the other hand, if p decreases, the corresponding coefficients are dropped.

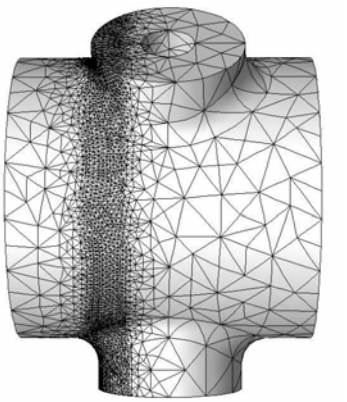

Figure 3. A picture of a mesh for the coupler region.

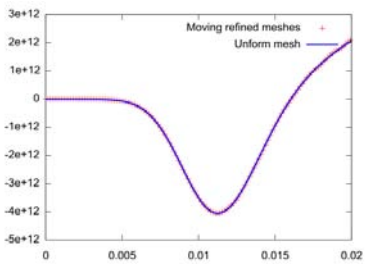

Figure 4. Wakefield comparison with moving h-refined meshes and with a uniform mesh.

The short-range wakefield of a coupler shown in Figure 2 is calculated with the moving window through p-refinement on a curvilinear tetrahedral mesh with 13 million elements. During the simulation, the window moved 5 times with each window having 2.37 million, 1.08 million, 1.02 million, 1.02 million, 1.50 million, and 1.78 million elements, respectively. The run-time of the simulation using the moving window technique is only *one-tenth* of that using all 13 million elements.

4.2. H-refinement

A series of windows with padding zones are defined in the vicinity of the moving particle beam. The corresponding series of meshes are generated by scientists from SciDAC ITAPS Center with a dense mesh region inside each window and a coarse mesh region outside the window. Figure 3 shows one such mesh for the coupler region shown in Figure 2. When the particle beam moves out of one window and into the next window, the vectors $\mathbf{e}^n$ and $\mathbf{e}^{n-1}$ are transferred onto the

next mesh using the projection as in Eq (6) while the vectors $\mathbf{f}^n$ and $\mathbf{f}^{n-1}$ are re-calculated with the new mesh. The wakefield monitors in a pillbox cavity with moving h-refined meshes and with a uniform mesh are plotted in Figure 4. It is shown that the results are in remarkable agreement, which indicates the validity of our h-refined moving window technique. Further work on moving window through combined h-refinement and p-refinement is in progress.

5. Scalable Application I/O

In this section, the improvement made on the parallel application I/O in the SLAC's simulation tool Omega3P and T3P is presented.

Due to the lack of the parallel I/O specification in the netcdf standard, a "rank 0 write all" approach for writing out the results in a parallel simulation had been used in the SLAC's code. This approach was non-scalable and the time for writing out data exceeded the time for the rest of the simulation when the number of executing processors is more than 4000. Working with scientists at ORNL and ANL, the application I/O has been overhauled with parallel-netcdf API [9]. With parallel-netcdf API the execution time for writing out data in the Omega3P simulations is reduced by two orders of magnitude.

Curvilinear tetrahedral elements are used in the SLAC's finite-element based application suite for high-fidelity modeling. To save disk space, only the mid-points of the curved edges are stored in the mesh file. During the simulation, all the mid-points of the curved edges are read in from the file and only those belonging to the local processor are kept in the memory. A more optimized read algorithm is developed by taking advantage of fast inter-processor communication. First the mid-points are partitioned with the number of processors. Each processor read in its partitioned portion of the mid-points in the second step. In the third step, an `MPI_Allgatherv()` is invoked to get all the mid-points for the selection of the mid-points in the local mesh. This scheme provides a scalable read algorithm and has a huge time-saving on NCCS Jaguar computer when number of processors in the simulation is large.

With the above-mentioned application I/O improvement and various communication pattern enhancement, SLAC's FETD code T3P has been efficiently used with a **half billion** degrees of freedom to simulate an 8-cavity cryomodule with the INCITE allocation[10].

Acknowledgments

This work is supported by the U.S. Department of Energy under contract number DE-AC02-76SF00515. The work used resources of the National Center for Computational Sciences at Oak Ridge National Laboratory and resources of the National Energy Research Scientific Computing Center, which are supported by the Office of Science of the Department of Energy under Contract DE-AC05-00OR22725 and DE-AC02-05CH11231, respectively.

References

[1] Lee L Q and et al 2005 *J. Phys.: Conf. Ser.* **16** 205–209

[2] Lee L Q, Akcelik V, Chen S, Ge L, Prudencio E, Li Z, Ng C, Xiao L and Ko K 2006 *Proc. of SciDAC 2006 Conference* (Denver, Colorado, USA)

[3] Lee L Q and et al 2007 *J. Phys.: Conf. Ser.* **78** 012040

[4] Akelik V, Ko K, Lee L Q, Li Z, Ng C K and Xiao L 2008 *Journal of Computational Physics* **227** 1722–1738

[5] Rimmer R and Li Z Private Communication

[6] Sun D K, Lee J F and Cendes Z 2001 *SIAM J. SCI. COMPUT* **23** 10531076

[7] Lee L Q, Candel A and Kabel A On projecting discretized electromagnetic fields in preparation

[8] Newmark N M 1959 *Journal of Engineering Mechanics Division* **85** 67–94

[9] Li J, Liao W K, Choudhary A, Ross R, Thakur R, Gropp W, Latham R, Siegel A, Gallagher B and Zingale M 2003 *Proc. of SC'03* (Phoenix, Arizona, USA)

[10] Lee L Q, Ng C, Ko K, Li Z and Kabel A 2008 *2008 INCITE Program*