An Efficient and Parallel Electromagnetic Solver for Complex Interconnects in Layered Media

Damian Marek, Shashwat Sharma, Piero Triverio
Edward S. Rogers Sr. Dept. of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada
damian.marek@mail.utoronto.ca, shash.sharma@mail.utoronto.ca, piero.triverio@utoronto.ca

***Abstract*—A novel parallel solver based on the adaptive integral method (AIM) is proposed for the electromagnetic analysis of electrical interconnects in layered media. We show that graph partitioning techniques can be used to optimally distribute, across thousands of processes, the computations related to both matrix filling and system solution. The proposed workload distribution strategy is compared to existing techniques through a scalability study on a large realistic interposer model in layered media.**

***Index Terms*—surface integral equation method, adaptive integral method, parallel algorithms, skin effect modeling**

I. Introduction

Compared to volumetric formulations, surface integral equation (SIE) methods offer an efficient approach for the electromagnetic analysis of interconnects in layered media [1]–[8]. Although efficient, SIE methods are always partnered with a suitable acceleration scheme to increase their scalability. In layered media, the most popular choice is the adaptive integral method (AIM), which accelerates matrix-vector products through fast Fourier transforms (FFTs) [3], [9]. Even with acceleration, existing electromagnetic solvers can hardly handle realistic portions of the intricate interconnect network present in typical integrated circuits, especially those featuring 3D integration. This scenario calls for the development of efficient parallel solvers suitable for large-scale electromagnetic analyses.

To efficiently parallelize an electromagnetic solver one must: i) distribute the workload evenly across processors, and ii) minimize communication between them. In [6], [7], a 3D block decomposition is proposed to efficiently distribute AIM computations. This approach works well for integral equation methods using volumetric meshes, since mesh elements cover the simulation domain in a fairly uniform fashion. However, parallelization efficiency tends to be suboptimal for sparsely-populated problems or for "hollow" meshes, such as those that naturally arise with SIE formulations [10].

In this work, we propose an efficient parallelization strategy for SIE formulations involving layered media, generalizing our previous work in free space [11]. The new method leverages graph partitioning to distribute AIM operations among thousands of computing cores in a balanced fashion that minimizes inter-node communication. A numerical test on a realistic interposer structure, performed with up to 1,600 cores, demonstrates better performance than previously-published solutions.

978-1-7281-6161-7/20/$31.00 © 2020 IEEE

II. Formulation

We consider a structure made by lossy conductors and embedded in a medium stratified along the z axis. The goal is to solve Maxwell equations in order to determine the scattering (S) parameters between some given ports. To model the induced electric current density in conducting objects, we use the augmented electric field integral equation (AEFIE) [5] along with an approximate surface impedance boundary condition (SIBC) [12]. After discretization with Rao-Wilton-Glisson (RWG) [13] and pulse basis functions, the AEFIE can be written as:

$$\begin{bmatrix} jk_0\mathbf{Z}_A + \eta_0^{-1}\mathbf{Z}_s & -\mathbf{D}^{\mathrm{T}}\mathbf{Z}_\Phi\mathbf{B} \\ \mathbf{FD} & \mathbf{C} \end{bmatrix} \begin{bmatrix} \mathbf{J}_s \\ c_0\boldsymbol{\rho}_r \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{I}_s \end{bmatrix}. \quad (1)$$

In (1), $\mathbf{Z}_A$ and $\mathbf{Z}_\Phi$ are the discretized vector and scalar potential parts of the single-layer potential operator [14], which involve the Green's function of a layered medium [1]. The SIBC is enforced by $\mathbf{Z}_s$, while $\mathbf{C}$ provides coupling to a Thevenin-equivalent circuit excitation [2], and $\mathbf{I}_s$ is the excitation current. The remaining matrices are identical to [5].

In the AIM, (1) is solved iteratively, and the required matrix-vector products involving $\mathbf{Z}_A$ and $\mathbf{Z}_\Phi$ are accelerated with FFTs. Both $\mathbf{Z}_A$ and $\mathbf{Z}_\Phi$ are expressed as

$$\mathbf{Z} \approx \mathbf{Z}_{NR} + \mathbf{WHP}, \quad (2)$$

where $\mathbf{Z}_{NR}$ models electromagnetic interactions at a short distance, in the so-called near region. Far-region interactions are described by the second term, where $\mathbf{P}$ projects sources onto a regular grid (the AIM grid), $\mathbf{H}$ computes the resulting potentials via FFTs, which are finally interpolated by matrix $\mathbf{W}$ onto the original mesh elements [3].

The efficient parallelization of matrix-vector products involving (2) is difficult, because two different workload decomposition strategies would ideally be required to: 1) distribute the multiplication by $\mathbf{Z}_{NR}$, whose entries are related to the original mesh, and 2) distribute the multiplications and FFTs arising from the $\mathbf{WHP}$ term, which instead involve the AIM grid. Furthermore, undesirable communication will occur in matrix-vector products involving $\mathbf{W}$ and $\mathbf{P}$ if a mesh element and the corresponding AIM grid points are assigned to different processes.

Existing strategies to parallelize AIM-based solvers focus on the efficiency of computations related to the AIM grid, like FFTs, which are typically quite expensive [6], [7], [10].

However, for complex multiscale layouts, near-region computations may also be very costly, and need to be carefully balanced across computing nodes to achieve high parallelization efficiency.

III. Proposed Method

We present a new parallelization strategy for AIM-based solvers which efficiently distributes both near-region and far-region computations across thousands of computing cores, while minimizing the communication required by projection and interpolation steps.

A. 2D Fast Fourier Transforms

In layered media, multiplications by matrix $\mathbf{H}$ in [3] are computed through many 2D FFTs. We parallelize these operations with the pencil decomposition method proposed in [7]. The pencils in [7] decompose the AIM grid into long thin cuboids aligned with the x axis. In this starting configuration, each process computes 1D FFTs along the x axis. Then the result is transposed in a communication step, so that the pencils become aligned to the y axis. Finally, each process computes the remaining 1D FFTs along the y axis.

B. Near-Region Operations

Near-region operations include the computation of the $\mathbf{Z}_{A,NR}$ and $\mathbf{Z}_{\Phi,NR}$ matrices, and the corresponding matrix-vector products. In layered media, the construction of these matrices becomes even more time consuming due to the complexity of the multilayer Green's function.

In [7], a 3D block decomposition method is proposed to distribute near-region computations across processes. In this method, the simulation domain is first partitioned along the x axis into P_x equally sized slabs. Then, each slab is split along the y axis into P_y equally sized pencils. Finally each pencil is split along the z axis into P_z equally sized blocks. The computations associated to each block will be later assigned to $P = P_x \times P_y \times P_z$ processes. Prior to assignment, the workload is balanced with the following procedure [10]:

1) The slab boundaries are shifted one at a time, by small increments, to balance the number of near-region interactions associated to each slab;
2) Then, in each slab, the pencil boundaries are shifted to balance the number of near-region interactions associated to each pencil;
3) Finally, in each pencil the block boundaries are shifted to balance the number of near-region interactions associated to each block.

This method is simple and effective for many cases. However, as will be shown in Section IV, this method may not optimally balance the workload and communication associated with near-region operations for complex and multiscale layouts.

We propose a more efficient solution by generalizing, to structures in layered media, the graph-based approach advocated in [11] for problems in free space. In the proposed approach, near-region computations are distributed in an optimal fashion with the following steps:

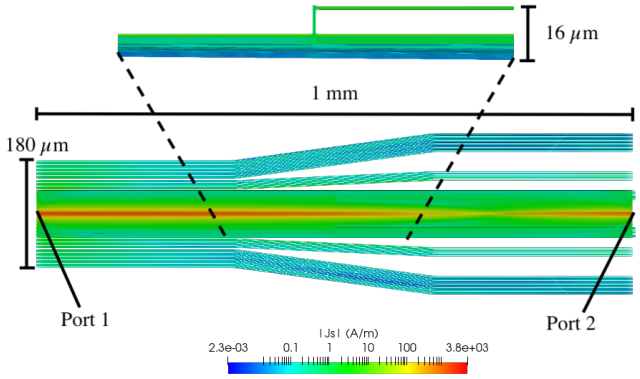

Figure 1. Geometry of the interposer-level interconnect considered in Sec. IV with current density computed at 50 GHz when port 1 is excited.

1) A dual graph of the mesh is constructed, where each triangle is associated with a graph node and graph edges connect adjacent triangles;
2) The number of triangles falling in the near region of a given triangle are assigned as a nodal weight to the corresponding node in the graph;
3) Finally, parMETIS [15], a distributed graph partitioning library, is used to partition the graph into P subgraphs, where P is the desired number of processes. The partition is generated to approximately satisfy two optimality criteria. First, the sum of the nodal weights in each partition should be approximately the same. In this way, the operations required to pre-compute $\mathbf{Z}_{A,NR}$ and $\mathbf{Z}_{\Phi,NR}$ and multiply them against a given vector are evenly distributed among processes. Second, the number of edges cut by the partitioning process is minimized, to minimize inter-process communication.

C. Interpolation and Projection

After the AIM grid pencils have been assigned to processes, there is still freedom for each process to choose one of the mesh partitions generated in Section III-B. To minimize communication in the matrix-vector products involving $\mathbf{W}$ and $\mathbf{P}$, there should be maximal overlap between the mesh partition and the AIM grid pencil assigned to each process. This goal is achieved by the following algorithm:

1) Each process analyzes all mesh partitions and calculates $N_E^{(p)}$, the number of edges of mesh partition p that are within its AIM pencil;
2) Each process ranks all partitions in order of decreasing $N_E^{(p)}$;
3) Finally, starting form the process with lowest rank, each process chooses the partition with highest $N_E^{(p)}$ among those still available.

IV. Results

The scalability of the proposed method was tested by extracting the S parameters of a complex interconnect network from an interposer used for 3D integration (courtesy of Dr.

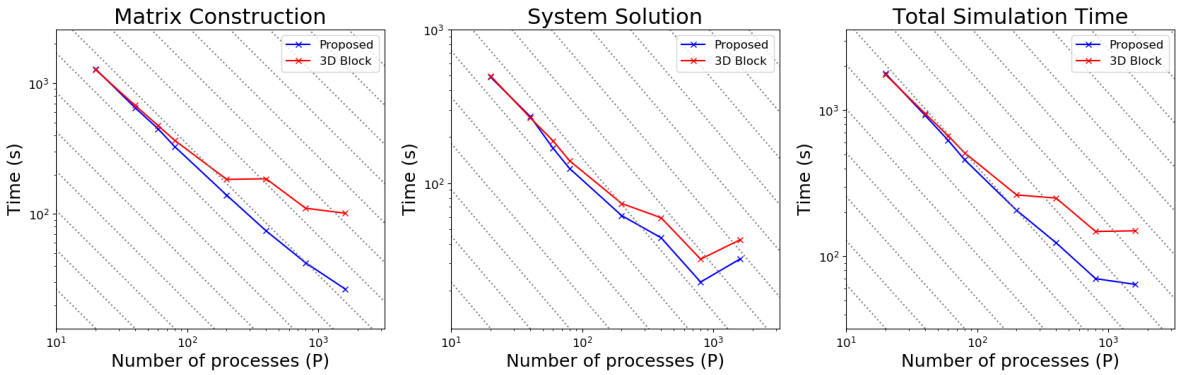

Figure 2. Scalability of the proposed solver and the "3D block" method in [10] for different steps of the solver. Dashed lines represent ideal efficiencies of parallelization and their relative spacing indicates a $2\times$ speedup.

Rubaiyat Islam, AMD). The interposer structure is illustrated in Fig. 1 together with the location of 2 of the 6 ports considered in this test. There are a total of 80 copper wires and one ground plane. The structure is embedded in a two-layer medium bounded above and below by air. The upper layer has $\epsilon_{r1} = 4$, $\sigma_1 = 0$ S/m, and thickness of $h_1 = 27.5$ μm. The lower layer has $\epsilon_{r2} = 11.9$, $\sigma_2 = 10$ S/m, and thickness of $h_2 = 47.5$ μm. The resulting mesh of the structure contained 357,000 triangles and 535,500 edges. The AIM grid dimensions were chosen to be $90 \times 400 \times 8$ $(N_x \times N_y \times N_z)$ and the near-region radius was set to 5 grid points.

The proposed method was compared to the 3D block decomposition of [10]. All simulations were run on the Scinet Niagara cluster where each node has 40 Intel Skylake cores running at 2.4 GHz and 202 GB of memory. In this scalability study, the number of processes, P, was varied from 20 to 1,600. The wall time required to precompute the matrices in (1) and solve the system for one frequency point are plotted in Fig. 2. For all P values, the CPU time required by the proposed method to determine an optimal workload distribution is negligible compared to the total simulation time (less than 2%).

The results show that, when the proposed method is adopted, a significant speedup can be realized for both the matrix construction step and the system solution step. For the matrix construction step, when $P = 1600$, the proposed method was $4\times$ faster than the 3D block decomposition method.

V. Conclusion

We presented a new strategy to efficiently parallelize large-scale electromagnetic simulations based on surface integral equations. A novel approach based on graph partitioning is shown to effectively distribute, across thousands of computing cores, the computation of electromagnetic interactions in both the near and the far range. When compared to other state-of-the-art strategies on a complex interconnect network from a 3D integrated circuit, the proposed method showed better scalability.

References

[1] K. A. Michalski and D. Zheng, "Electromagnetic Scattering and Radiation by Surfaces of Arbitrary Shape in Layered Media, Part I: Theory," *IEEE Trans. Antennas Propag.*, vol. 38, no. 3, pp. 335–344, Mar 1990.

[2] Y. Wang, D. Gope, V. Jandhyala, and C. J. R. Shi, "Generalized Kirchoff's current and voltage law formulation for coupled circuit-electromagnetic simulation with surface integral equations," *IEEE Trans. Microw. Theory Techn.*, vol. 52, no. 7, pp. 1673–1682, 2004.

[3] J. Phillips and J. White, "A precorrected-FFT method for electrostatic analysis of complicated 3-D structures," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 16, no. 10, pp. 1059–1072, Mar. 1997.

[4] Y. P. Chen, L. Jiang, Z. G. Qian, and W. C. Chew, "An augmented electric field integral equation for layered medium Green's function," *IEEE Trans. Antennas Propag.*, vol. 59, no. 3, pp. 960–968, 2011.

[5] Z. G. Qian and W. C. Chew, "Fast Full-Wave Surface Integral Equation Solver for Multiscale Structure Modeling," *IEEE Trans. Antennas Propag.*, vol. 57, no. 11, pp. 3594–3601, 2009.

[6] C. Liu, K. Aygün, H. Braunisch, V. I. Okhmatovski, and A. E. Yilmaz, "A parallel iterative layered-medium integral-equation solver for electromagnetic analysis of electronic packages," in *Proc. IEEE Conf. on Elect. Perf. of Electron. Packag. and Systems (EPEPS)*, 2017, pp. 1–3.

[7] C. Liu, K. Aygün, and A. E. Yılmaz, "A parallel FFT-accelerated layered-medium integral-equation solver for electronic packages," *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, pp. 1–17, Sep. 2019.

[8] S. Sharma and P. Triverio, "SLIM: A Well-Conditioned Single-Source Boundary Element Method for Modeling Lossy Conductors in Layered Media," submitted to IEEE Antennas Wireless Propag. Lett., 2020.

[9] E. Bleszynski, M. Bleszynski, and T. Jaroszewicz, "AIM: Adaptive integral method for solving large-scale electromagnetic scattering and radiation problems," *Radio Science*, vol. 31, no. 5, pp. 1225–1251, Sep. 1996.

[10] F. Wei and A. E. Yilmaz, "A More Scalable and Efficient Parallelization of the Adaptive Integral Method—Part I: Algorithm," *IEEE Trans. Antennas Propag.*, vol. 62, no. 2, pp. 714–726, Feb. 2014.

[11] D. Marek, S. Sharma, and P. Triverio, "An efficient parallelization strategy for the adaptive integral method based on graph partitioning," in *14th Euro. Conf. on Antennas and Propag. (EuCAP)*, 2020, pp. 1–5.

[12] S. Yuferev and N. Ida, *Surface Impedance Boundary Conditions*. CRC Press, 2009.

[13] 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, May. 1982.

[14] W. C. Gibson, *The Method of Moments in Electromagnetics*. Chapman and Hall/CRC, Jul. 2014.

[15] G. Karypis and V. Kumar, "A Parallel Algorithm for Multilevel Graph Partitioning and Sparse Matrix Ordering," *J. Parallel. Distrib. Comput.*, vol. 48, no. 1, pp. 71–95, Jan. 1998.