FastHenry: A Multipole-Accelerated 3-D Inductance Extraction Program

M. Kamon M. J. Tsuk J. White

Mass. Institute of Tech. Digital Equipment Corp. Mass. Institute of Tech.
Cambridge, MA 02139 Tewksbury, MA 01876 Cambridge, MA 02139

Abstract

In [1], it was shown that an equation formulation based on mesh analysis can be combined with a GMRES-style iterative matrix solution technique to make a reasonably fast 3-D frequency dependent inductance and resistance extraction algorithm. Unfortunately, both the computation time and memory required for that approach grow faster than n^2, where n is the number of volume-filaments. In this paper, we show that it is possible to use multipole-acceleration to reduce both required memory and computation time to nearly order n. Results from examples are given to demonstrate that the multipole acceleration can reduce required computation time and memory by more than an order of magnitude for realistic packaging problems.

1 Introduction

In high performance VLSI integrated circuits and integrated circuit packaging, there are many cases where accurate estimates of the coupling inductances of complicated three-dimensional structures are important for determining final circuit speeds or functionality. The most obvious examples are the pin-connect structures used in advanced packaging. For the past decade, volume-element techniques have been used to compute self and coupling inductances of complex three dimensional geometries, but the techniques were intended for geometries which could be represented with at most a few hundred volume filaments. However, the complex structures currently used in integrated circuit packaging can require up to ten thousand filaments to be accurately analyzed. Existing programs become extraordinarily computationally expensive for such large problems, and new algorithms whose computational cost grows more slowly with problem size must be developed.

In [1], it was shown that an equation formulation based on mesh analysis can be combined with a GMRES-based iterative matrix solution technique to make a reasonably fast 3-D frequency dependent inductance and resistance extraction algorithm. Unfortunately, both the computation time and memory required for the approach described in [1] grow faster than n^2, where n is the number of volume-filaments. In order to analyze very complicated structures, an algorithm whose complexity grows more slowly with problem size must be derived. In this paper, we show that it is possible to use multipole-acceleration to reduce both required memory and computation time to nearly order n. We start in the next section by reviewing the mesh formulation and iterative matrix solution approach from [1]. In Section 3, we show that the matrix-vector product required in the iterative algorithm can be reinterpreted as a sequence of electrostatic potential evaluations, and briefly describe how the multipole algorithm can be used to accelerate those potential evaluations. Results from our implementation in FASTHENRY are given in Section 4, and they show that for an example package, multipole-acceleration reduces memory and computation time by more than an order of magnitude. Finally, we give our conclusions and acknowledgements.

2 The Mesh-based Formulation

One approach to computing the frequency dependent inductance and resistance matrix associated with the terminal behavior of a collection of conductors involves first approximating each conductor with a set of piecewise-straight conducting sections. The volume of each straight section is then discretized into a collection of parallel thin filaments through which current is assumed to flow uniformly. The interconnection of these current filaments can be represented with a planar graph, where the n nodes in the graph are asso-

30th ACM/IEEE Design Automation Conference®

ciated with connection points between conductor segments, and the b branches in the graph represent the current filaments into which each conductor segment is discretized (See Fig. 1).

To derive a system of equations from which the resistance and inductance matrix, Z_r, can be deduced, we start by assuming the applied currents and voltages are sinusoidal, and that the system is in sinusoidal steady-state. Following the partial inductance approach in [2, 3], the branch current phasors can be related to branch voltage phasors (hereafter, phasors will be assumed and not restated) by

$$ZI_b = V_b, \tag{1}$$

where V_b, $I_b \in C^b$, b is the number of branches (number of current filaments), and $Z \in C^{b\times b}$ is the complex impedance matrix given by

$$Z = R + j\omega L, \tag{2}$$

where ω is excitation frequency. The entries of the diagonal matrix $R \in \Re^{b\times b}$ represent the dc resistance of each current filament, and $L \in \Re^{b\times b}$ is the dense matrix of partial inductances [4]. Specifically,

$$L_{i,j} = \frac{\mu_0}{4\pi a_i a_j} \int_{fil\ i} \int_{fil\ j} \frac{l_i(X_i)\cdot l_j(X_j)}{|X_i - X_j|} d^3x_i d^3x_j, \tag{3}$$

where $X_i, X_j \in \Re^3$ are the positions in filament i and j respectively, $l_i, l_j \in \Re^3$ are the unit vectors in the direction of current flow in filaments i and j, a_i, a_j are the cross sectional areas of filaments i and j, and $fil\ i, fil\ j$ represent the volumes of filaments i and j.

Now assume that sources attached to the conductor system's terminals generate explicit branches in the graph representing the discretized problem. Kirchoff's voltage law, which implies that the sum of branch voltages around each mesh (a mesh is any loop of branches in the graph which does not enclose any other branches) in the network is represented by

$$MV_b = V_s \tag{4}$$

where V_b is the vector of voltages across each branch, $V_s \in \Re^m$ is the mostly zero vector of source branch voltages, and $M \in \Re^{m\times b}$ is the mesh matrix. Here, $m = b - s + c$, where s is the number of conductor sections and c is the number of conductors. The M matrix has the property that most of its columns ($m-c$ of them) have no more than two nonzero entries, a plus one and minus one.

The mesh currents, that is the currents around each mesh loop, satisfy

$$M^t I_m = I_b, \tag{5}$$

Algorithm 1 (GMRES Alg. for $Ax = b$)

```
guess x^0
for k = 0,1,... until converged {
  Compute the error, r^k = b - Ax^k
  Find x^{k+1} to minimize r^{k+1}
    based on x^i and r^i, i = 0,...,k
}
```

where the superscript t denotes matrix transpose, and $I_m \in \Re^m$ is the vector of mesh currents. Note that one of the entries in the mesh current vector will be identically equal to the source branch current, shown as I_{m_source} in Fig. 1. Combining (5) with (4) and (1) yields

$$MZM^t I_m = V_s. \tag{6}$$

The complex admittance matrix which describes the terminal behavior of the conductor system, denoted $Y_r = Z_r^{-1}$, can by derived from (6) by noting that

$$\tilde{I}_s = Y_r \tilde{V}_s, \tag{7}$$

where $\tilde{I}_s$ and $\tilde{V}_s$ are the vectors of terminal source currents and voltages. Therefore, to compute the i^{th} column of Y_r, solve (6) with a V_s whose only nonzero entry corresponds to $\tilde{I}_{s_i}$, and then extract the entries of I_m associated with the source branches.

The standard approach to solving the complex linear system in (6) is Gaussian elimination, but the cost is m^3 operations. For this reason, inductance extraction of packages requiring more than a few thousand filaments is considered computationally intractable. To improve the situation, consider using a conjugate-residual style iterative method like GMRES [5]. Such methods have the general form given in Algorithm 1.

The GMRES algorithm can be applied to solving (6); however, the cost of each iteration of the GMRES algorithm is at least order m^2 operations. This follows from the fact that evaluating $\mathbf{r}^k$ implies computing a matrix-vector product, where in this case the matrix is MZM^t and is dense. Note also that forming MZM^t explicitly requires order m^2 storage.

3 The Multipole Approach

It is possible to approximately compute $MZM^t I_m^k$ in order b operations using a hierarchical multipole

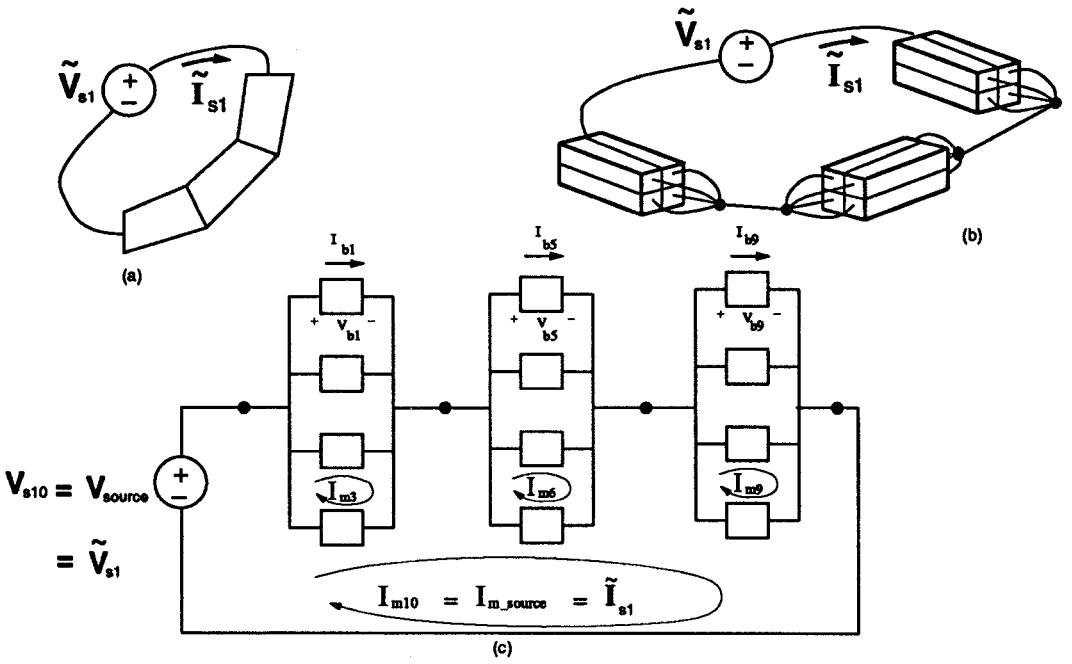

Figure 1: One conductor, (a) as piecewise-straight sections, (b) discretized into filaments, (c) modelled as a circuit.

algorithm [6]. Such algorithms also avoid explicitly forming MZM^t, and so reduce the memory required to order b. To show how a multipole algorithm can be applied to computing $MZM^tI_m^k$, consider expanding the matrix-vector product using (2),

$$MZM^tI_m^k = MRM^tI_m^k + j\omega MLM^tI_m^k. \quad (8)$$

The $MRM^tI_m^k$ term can be computed in order m operations because R is the diagonal matrix derived from the filament resistances, and M is the sparse mesh matrix with order m nonzero elements. Forming $MLM^tI_m^k$ is more expensive, requiring order m^2 operations as L is dense. By separating $LM^tI_m^k$, or equivalently LI_b, from the product, and by using the definition in (3), it is easily shown that LI_b can be written in terms of the vector potential, $A \in C^3$, which is given by

$$A(X) = \frac{\mu_0}{4\pi}\sum_j \frac{I_{b_j}}{a_j}\int_{filament_j}\frac{l_j(X_j)}{|X-X_j|}d^3x_j. \quad (9)$$

Specifically, the i^{th} element in the LI_b vector is the average over the filament cross section of the integral of the vector potential along filament i,

$$(LI_b)_i = \frac{1}{a_i}\int_{filament_i} A(X_i)\cdot l_i(X_i)d^3x_i. \quad (10)$$

This decomposition makes clear that LI_b can be evaluated by integrating the vector potential A over each filament. Also, from (9), each component of the vector potential can be identified precisely with a scalar electrostatic potential due to a collection of filament charges. That is, for $k \in \{1,2,3\}$, the k^{th} component of $A(X)$, denoted $\psi_k(X) \in C$, is a scalar potential given by

$$\psi_k(X) = \frac{\mu_0}{4\pi}\sum_j \frac{I_{b_j}}{a_j}\int_{filament_j}\frac{(l_j(X_j))_k}{|X-X_j|}d^3x_j, \quad (11)$$

and therefore $(I_{b_j}/a_j)\,l_j(X_j)_k$ can be interpreted as a charge density.

In summary, LI_b can be computed by combining the results of evaluating the electrostatic potential along b filaments due to b filament charges *for three separate sets of filament charges.* It is the evaluation of these electrostatic potentials which can be accelerated with the hierarchical multipole algorithm [6]. That is, the electrostatic potential due to b charges can be evaluated at b points in order b operations using the hierarchical multipole algorithm. This implies that by using the algorithm three times, LI_b can be computed in order b operations.

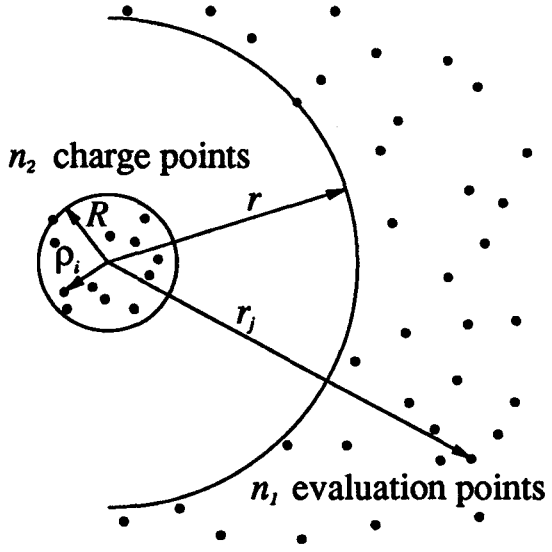

Figure 2: The evaluation point potentials are approximated with a multipole expansion.

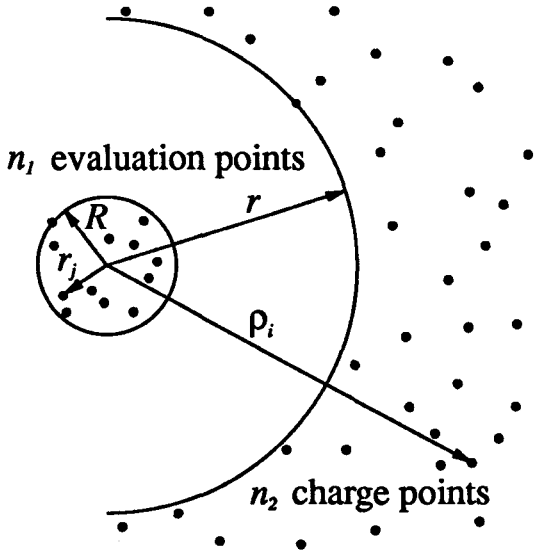

Figure 3: The evaluation point potentials are approximated with a local expansion.

3.1 The Hierarchical Multipole Algorithm

A complete description of the fast multipole algorithm is quite lengthy, and can be found in [6], or in the context of 3-D capacitance extracation, in [7, 8]. To see roughly what the algorithm exploits to achieve its efficiency consider the two configurations given in Figs. 2 and 3, depicted in 2-D for simplicity. In either figure, the obvious approach to determining the electrostatic potential at the n_1 evaluation points from the n_2 point-charges involves $n_1 * n_2$ operations; at each of the n_1 evaluation points one simply sums the contribution to the potential from n_2 charges.

An accurate approximation for the potentials for the case of Fig. 2 can be computed in many fewer operations using *multipole expansions*, which exploit the fact that $r >> R$ (defined in Fig. 2). That is, the details of the distribution of the charges in the inner circle of radius R in Fig. 2 do not strongly affect the potentials at the evaluation points outside the outer circle of radius r. It is also possible to compute an accurate approximation for the potentials at the evaluation points in the inner circle of Fig. 3 in many fewer than $n_1 * n_2$ operations using *local expansions*, which again exploit the fact that $r >> R$ (as in Fig. 3). In this second case, what can be ignored is the details of the evaluation point distribution.

This brief description of the hierarchical multipole algorithm is only intended to make clear that the algorithm's efficiency stems from coalescing charges and evaluation points using multipole and local expansions. A few points about the algorithm's application to computing LI_b should be considered however. When filaments are very near each other, a multipole expansion representation would lead to excessive error, so the interaction is evaluated directly using (3). Direct evaluations are also used for small groups of distant filaments when the computation required to build the multipole and local expansions exceeds the direct evaluation cost, thus making the algorithm adaptive. Therefore, when the hierarchical multipole algorithm is used to compute LI_b, the evaluation is typically a combination of three sets of multipole and local expansion evaluations, along with a single set of nearby-filament direct evaluations.

3.2 Accelerating Iteration Convergence

In general, the GMRES iterative method applied to solving (6) can be significantly accelerated by *preconditioning* if there is an easily computed good approximation to the inverse of MZM^t. We denote the approximation to $(MZM^t)^{-1}$ by P, in which case preconditioning the GMRES algorithm is equivalent to using GMRES to solve

$$(MZM^t)Px = V_s \tag{12}$$

for the unknown vector x. The mesh currents are then computed with $I_m = Px$. Clearly, if P is precisely $(MZM^t)^{-1}$, then (12) is trivial to solve, but then P will be very expensive to compute.

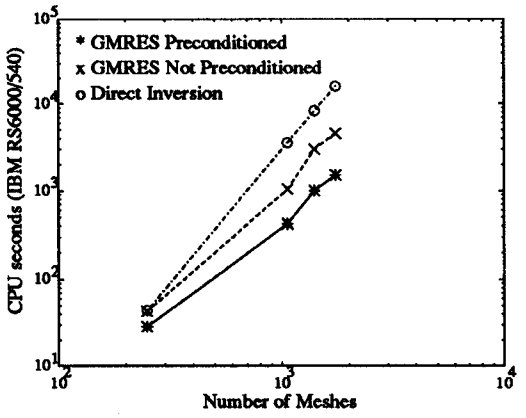

Figure 4: Comparison of the CPU time (IBM RS6000/540) to compute the reduced inductance matrix with and without a preconditioner.

A good approximation to $(MZM^t)^{-1}$ that is easily computed can be derived by exploiting the fact that the part of MZM^t corresponding to nearby, and therefore tightly coupled, meshes is computed explicitly. For each mesh i, the submatrix of MZM^t corresponding to all meshes near mesh i is inverted directly. Then, the row of the inverted submatrix associated with mesh i becomes the i^{th} row of P. That this preconditioner is effective is demonstrated in figure 4, which compares the CPU time required to compute the reduced inductance matrix with and without the preconditioner for the example in the next section.

4 Results

In this section we demonstrate the accuracy and computational efficiency of our multipole-accelerated version of FASTHENRY. We consider a typical industrial example, part of a 68-pin package, shown in Figure 5. Each pin consists of eight to ten conductor sections. For an accuracy comparison, we discretized each section into 2×2 filaments. This generated a problem with 1368 branches for which MZM^t is a 1061×1061 dense matrix. Note, using only four filaments per section is hardly sufficient to model the skin effect, though the coarse discretization does create a problem which is small enough to make possible an accuracy comparison between direct factorization, GMRES, and multipole-accelerated GMRES.

For the example package, the mutual inductance between pins 1 and 2 (labeled clockwise from the right) is much larger than the mutual inductance between

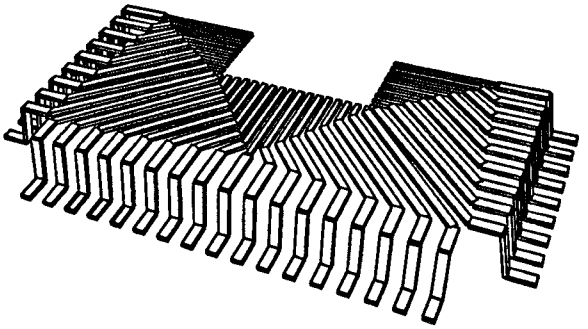

Figure 5: Half of a pin-connect structure. Thirty-five pins shown.

pin pair	direct	gmres	multipole
1 to 2	5.31870e+00	5.31867e+00	5.31403e+00
1 to 18	3.68292e-02	3.68223e-02	3.71027e-02

Table 1: Comparison of the accuracy of the computed inductance matrix entries between direct factorization, GMRES with explicit matrix-vector products, and the multipole-accelerated GMRES algorithm.

pins 1 and 18 which are perpendicular to each other except for their vertical sections. To show that the approximations used by the hierarchical multipole algorithm are sufficiently well-controlled to make it possible to accurately compute the small coupling inductances, consider the results in Table 1. The mutual inductance between pins 1 and 18 is more than two orders of magnitude smaller than the mutual inductance between pins 1 and 2, yet the solution computed using the multipole-accelerated algorithm is still within one percent of the solution computed using direct factorization.

To accurately model skin and proximity effects, each conductor section in the pin-connect structure should be divided into many more than 2×2 filaments. As the discretization is refined, the size of the problem will grow quickly, making the memory and CPU time advantage of the multipole-accelerated GMRES algorithm apparent (see figures 6 and 7). As the graphs clearly indicate, the cost of direct factorization grows like m^3, the cost of explicit GMRES grows as m^2, but the cost of multipole-accelerated GMRES grows only linearly with m. In addition, the memory requirement for multipole-accelerated GMRES algorithm grows linearly with m, but grows like m^2 for either explicit GMRES or direct factorization. In particular, for a 10,000 filament problem, the multipole

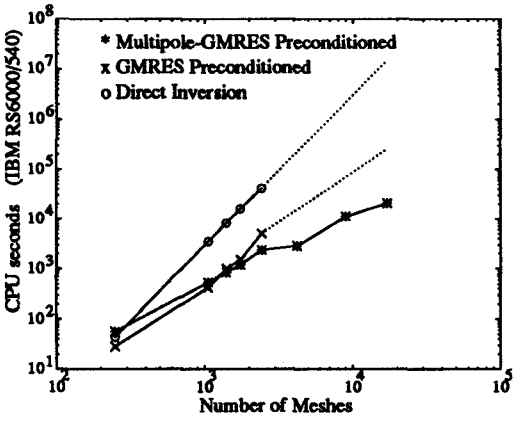

Figure 6: Comparison of the CPU time (IBM RS6000/540) to compute the reduced inductance matrix using direct factorization, GMRES, and GMRES with with multipole acceleration.

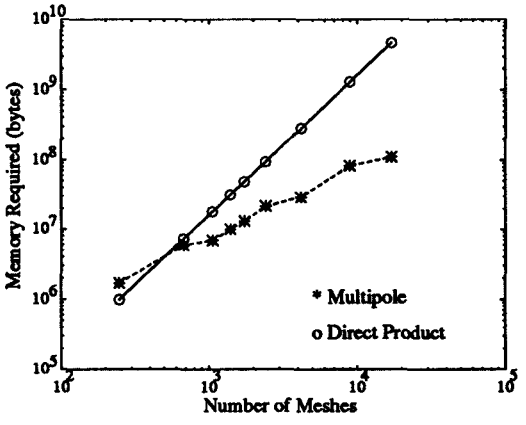

Figure 7: Comparison of the memory required using explicit matrix-vector products and using the multipole algorithm.

accelerated algorithm is two orders of magnitude faster than direct factorization, and uses an order of magnitude less time and memory than explicit GMRES.

5 Conclusions and Acknowledgements

In this paper, we show that 3-D inductance extraction can be substantially accelerated using the hierarchical multipole algorithm. Our multipole-accelerated inductance extraction program, FASTHENRY, was shown to be more than two orders of magnitude faster than direct factorization when used to extract the inductance matrix for a realistic packaging example. In addition, the multipole-accelerated algorithm uses an order of magnitude less time and memory than the explicit GMRES algorithm given in [1]. The authors would like to thank Keith Nabors, Songmin Kim, Dr. Sami Ali, and Joel Phillips for their help in understanding inductance. In addition, the authors would like to thank Dr. Albert Ruehli, Prof. Raj Mittra, and Dr. Colin Gordon for their helpful suggestions. This work was supported by the Defense Advanced Research Projects Agency contract N00014-91-J-1698, the National Science Foundation contract (MIP-8858764 A02), a National Science Foundation fellowship, and grants from I.B.M. and Digital Equipment Corporation.

References

[1] M. Kamon, M. Tsuk, C. Smithhisler, and J. White, "Efficient techniques for inductance extraction of complex 3-d geometries," in *Proceedings of the Int. Conf. on Comp. Aided Design*, November 1992.

[2] W. T. Weeks, L. L. Wu, M. F. McAllister, and A. Singh, "Resistive and inductive skin effect in rectangular conductors," *IBM Journal of Res. and Develop.*, vol. 23, pp. 652–660, November 1979.

[3] A. E. Ruehli, "Survey of computer-aided electrical analysis of integrated circuit interconnections," *IBM Journal of Research and Development*, vol. 23, pp. 626–639, November 1979.

[4] P. A. Brennan, N. Raver, and A. Ruehli, "Three dimensional inductance computations with partial element equivalent circuits," *IBM Journal of Res. and Develop.*, vol. 23, pp. 661–668, November 1979.

[5] Y. Saad and M. H. Schultz, "GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems," *SIAM Journal on Scientific and Statistical Computing*, vol. 7, pp. 856–869, July 1986.

[6] L. Greengard, *The Rapid Evaluation of Potential Fields in Particle Systems*. Cambridge, Massachusetts: M.I.T. Press, 1988.

[7] K. Nabors and J. White, "Fastcap: A multipole accelerated 3-D capacitance extraction program," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 10, pp. 1447–1459, November 1991.

[8] K. Nabors, S. Kim, and J. White, "Fast capacitance extraction of general three-dimensional structures," *IEEE Trans. on Microwave Theory and Techniques*, vol. 40, pp. 1496–1506, July 1992.