

# PILOT: A FAST ALGORITHM FOR ENHANCED 3D PARASITIC EXTRACTION EFFICIENCY

Dipanjan Gope and Vikram Jandhyala

Department of Electrical Engineering, Box-352500, University of Washington, Seattle, WA-98195, Telephone: 206-543-2186 Fax: 206-543-3842, {dip, jandhyala}@ee.washington.edu

## ABSTRACT

*Integral equation methodologies applied to extract parasitics at the board, package, and on-chip levels involve solving a dense system of equations. In this paper, we present an improved oct-tree based multilevel QR compression technique for fast iterative solution. The regular tree and interaction structure of the fast multipole method, and the QR compression scheme applied to interaction sub-matrices as in IES<sup>3</sup> are combined to achieve superior time and memory efficiency. As is demonstrated by numerical simulation results presented herein, the new algorithm is found to be faster than both existing QR based methods and FastCap.*

## I. Introduction

Quasi-static parasitic extraction is an important problem in digital circuits and in mixed signal IC analysis. Due to the complexity of on-chip design structures, numerical techniques that utilize field solutions for parasitic extraction are preferred when high accuracy is necessitated. Among the existing approaches, a surface-based integral-equation methodology based on the Method of Moments (MoM) is ideally suited to address the problem. It leads to a well-conditioned system with reduced size compared to volumetric approaches, but the system of equations generated is inherently dense, thereby creating a time and memory bottleneck. Several fast iterative techniques have been developed to efficiently store and solve a MoM system. All these methods (e.g. QR-based methods, fast multipole methods (FMMs), FFT-based techniques etc.) rely on algorithms to accelerate matrix vector products and therefore expedite the iterative solution significantly to linear time and memory complexity.

The multilevel QR-based fast iterative solver (IES<sup>3</sup>) [1] is particularly attractive for circuit problems. The approach is based on low-rank decomposition of MoM sub-matrices by Singular Value Decomposition (SVD) or the Modified Gram Schmidt (MGS) method [2]. It is independent of the Green's function kernel, and can be applied directly to multi-layered dielectrics without increasing the size of the problem unlike the other methods. Even in terms of free-space capacitance extraction IES<sup>3</sup> has been demonstrated as being more efficient in terms of memory and solve time. The method however, suffers from a higher setup time cost. For problems with large number of nets, that require solving for many right-hand-sides (RHS's), the higher setup cost is more than offset by faster matrix vector multiplies. However for well-conditioned systems with fewer nets and consequently fewer matrix vector products, the time efficiency of IES<sup>3</sup> could be inferior.

In this paper, we present a Predetermined Interaction List Oct Tree (PILOT) QR algorithm, that greatly reduces the setup time while maintaining the memory and solve time efficiency of the Rank-Map based Binary Tree QR (RMBT-QR), which is based on the same principles as IES<sup>3</sup>. The PILOT-QR algorithm exploits the properties of a multilevel oct-tree implementation (common to FMM approaches), to create a predetermined set of interaction lists, thereby reducing the setup time considerably. In short the regular geometry structure of FMM and the compression efficiency of QR are combined together to yield an improved capacitance extraction algorithm. Though not discussed in this paper, the compression scheme is amenable to full-wave multi-layered dielectric kernel solutions and is stable for low frequency analysis without any modifications unlike some full-wave FMM techniques.

## II. Integral Equation

Capacitance problems formulated using the MoM are solved via Poisson's Equation  $\nabla^2\phi(\mathbf{r}) = -\rho(\mathbf{r})/\epsilon$  relating potential  $\phi$  and charge-density  $\rho$ . The discretization of the integral form of this equation results in a matrix system of the form  $\bar{\mathbf{Z}}\mathbf{I} = \mathbf{V}$  where the  $N \times N$  MoM matrix  $\bar{\mathbf{Z}}$  is a dense Green's function matrix,  $\mathbf{I}$  represent the unknown coefficients of known charge density basis functions, and  $\mathbf{V}$  is the known potential excitation. Each element of the MoM matrix denotes the interaction between a testing and a basis function and is written as follows:

$$\bar{Z}_{(j,i)} = \int_{S_j} ds t_j(\mathbf{r}) \int_{S_i} ds' g(\mathbf{r}, \mathbf{r}') f_i(\mathbf{r}') \quad (1)$$

where  $t_j$  is the testing function defined over  $S_j$ ,  $f_i$  is the basis function defined over  $S_i$  and  $g(\mathbf{r}, \mathbf{r}')$  is the relevant Green's function. In the electrostatic case for  $P$  disconnected conductors, each column of the required  $P \times P$  capacitance matrix is obtained by enforcing a voltage of 1V on the excited conductor, 0V on all other conductors, solving the above system, and integrating the charge density over each conductor. The  $N \times N$  system of equations is therefore solved  $P$  times to obtain the capacitance matrix.

### III. Existing Multilevel QR algorithm

The IES<sup>3</sup> fast iterative solver, based on the QR method reduces the cost of performing the matrix vector product  $\bar{\mathbf{Z}}\mathbf{I}$  to  $O(N \log N)$  from quadratic time. It exploits smoothness of the Green's function to decompose the numerically rank-deficient far-field sub-matrices of the MoM using QR decomposition; a sub-matrix  $\bar{\mathbf{A}}$  of the MoM matrix  $\bar{\mathbf{Z}}$  can be decomposed as  $\bar{\mathbf{A}}_{m \times n} = \bar{\mathbf{Q}}_{m \times r} \bar{\mathbf{R}}_{r \times n}$  where  $\bar{\mathbf{R}}$  is upper triangular,  $\bar{\mathbf{Q}}$  is unitary i.e.  $\bar{\mathbf{Q}}^T \bar{\mathbf{Q}} = \bar{\mathbf{I}}$  and  $r \ll (m, n)$ . At the same time it is possible to construct the compressed representation without forming the entire sub-matrix from sampled rows and columns, thereby reducing the setup time to  $O(N \log N)$ .

In RMBT-QR, which is based on the same principles as IES<sup>3</sup>, the algorithm has 3 main steps:

a) *Geometry subdivisions into cells*: Binary decompositions with density balancing and tight bounds technically known as tight-bound k-d trees [3] are employed, similar to that used in IES<sup>3</sup>. As shown in figure 1 each cell is recursively decomposed along the largest side into 2 cells with equal number of entities.

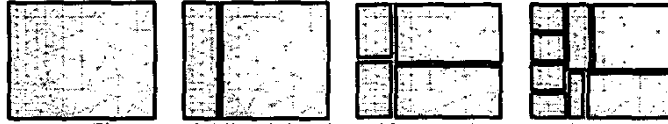


Figure 1: Multilevel tight bound k-d tree decomposition

It is interesting at this point to note that a tight-bound k-d tree may yield cells of any shape and size separated from each other by arbitrary distances depending on the geometry under consideration.

b) *Rank-Map predicted QR formations*: A rank-map is a statically-determined look up table that identifies low-ranked MoM sub-matrices on which QR formation from samples should be attempted. Large and low-ranked sub-matrices are the chief candidates to be identified by the rank-map so as to ensure maximum compression with minimum setup time. Each entry of the table outlines the expected rank of a cell-to-cell interaction, which is a function of various parameters pertaining to the source and observer cells:

$$r_E = f(l_s, b_s, h_s, l_o, b_o, h_o, d, g) \quad (2)$$

where  $r_E$  is the expected rank,  $l, b, h$  stands for the length, breadth and height of a cell, the subscripts  $s$  and  $o$  indicate source and observer cells respectively,  $d$  is the distance between the centers of the source and observer cells and  $g$  denotes the type of Green's function used e.g. free-space Green's function. The construction of this table is an expensive process considering all possible pairs of source and observer cells.

c) *Fine-tuning through splits and merges*: The rank-map only predicts the starting block structure for a MoM matrix. The rank estimation is often inaccurate and may result in underestimation of rank or missing larger low-rank blocks. These problems are addressed by splits and merges respectively [1].

The setup cost of the algorithm is largely controlled by the accuracy of rank map predictions. An accurate and exhaustive rank map would preclude the necessity for merges and unnecessary splits, and the optimum matrix block structure would be achieved without any backtracking or fine-tuning. However, a foolproof rank-map is difficult if not impossible to construct owing to the fact that the algorithm can lead to cells with any shape and size. This leads to a high constant being associated with the setup time cost of the algorithm.

### IV. New Multilevel QR Algorithm

The proposed PILOT QR algorithm develops a predetermined multilevel matrix structure for the geometry under consideration, which guarantees maximum compression. The algorithm has 3 main steps:

a) *Oct-tree spatial decomposition in 3D*: Each cube is recursively decomposed by loosely bounded, spatially balanced splits along orthants leading to 8 child cubes in 3D. The cell data structure is in the form of an oct-tree, identical to that in multilevel FMMs [4].

b) *Basic multilevel interaction list*: The basic multilevel interaction list of FMM is used as the starting block in the process of building the optimal multilevel compressed matrix structure. In multilevel FMM every cube  $c_i$  has a nearest neighbor list and an interaction list. The nearest neighbor list is defined as:  $K_{c_i} = \{c_j | c_j \text{ is in the same level as } c_i \text{ and has at least 1 common vertex with } c_i\}$  and the interaction list is denoted by:  $I_{c_i} = \{c_j | P_{c_j} \in K_{c_i}; c_j \notin K_{c_i}\}$ , where  $P_{c_i}$  denotes parent of  $c_i$ . Direct Green's function computations are used only at the lowest level for interactions between cube  $c_i$  and  $K_{c_i}$ . Multipole expansions are used to construct  $T(c_j, c_i) \forall j | c_j \in I_{c_i}$ , where  $T(c_j, c_i)$  denotes the interaction between testing functions of  $c_j^l$  and basis functions of  $c_i^l$ . Since PILOT does not explicitly require cubical regions but simply deals with interaction matrices, there is scope for further compression by combining cubes in  $I_{c_i}$  in an *a priori* manner into a new interaction list called the merged interaction list (MIL).

c) *Merged interaction list*: It is observed that the multilevel FMM interaction lists of siblings share many cubes in common as illustrated in figure 2. (2D version shown here for ease of illustration).

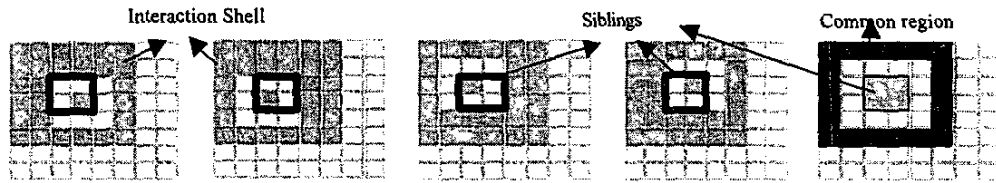


Figure 2: Common interaction shell of siblings in 2D

It is possible to group source and observer cubes of different interaction lists (merged interaction) in order to compress larger low-rank matrices and thereby gain in terms of overall compressibility. It must be noted that the entire common interaction list does not directly translate into a merged interaction because the rank of such an interaction sub-matrix will not be low. The common interaction list is carefully decomposed into disjointed parts such that the overall compression is optimized. Each such disjointed part is an interaction between grouped source cubes and observer cubes and forms an entry of the MIL denoted by  $\mu$ , which can be expressed as a combination of multilevel FMM cube-to-cube interactions:  $m_k = \{T_p(c_j, c_i)\} \forall p | 1 \leq p \leq n_g$ , where  $n_g$  is the number of FMM interactions grouped. Higher compression is achieved since a larger matrix is compressed to low rank:

$$(m_{\mu_k} + n_{\mu_k})r_{\mu_k} < \sum_{i=1}^{n_g} (m_i + n_i)r_i \quad (3)$$

where  $m$ ,  $n$  and  $r$  denote the number of rows, number of columns and the rank of a sub-matrix. The subscript  $i$  denotes a regular multilevel FMM interaction list entry that is now a constituent of the MIL. Figure 3 demonstrates the decomposition of the common interaction list of figure 2 into merged interactions.



Figure 3: Merged interaction list entries corresponding to the common interaction list

The merged interactions shown in figure 3 are only a subset of all MIL entries. In PILOT there are in total 16 MIL entries in 2D and 40 in 3D. The same MIL pattern is valid for all sibling pairs across levels. The MIL can thus be visualized as an accurate replacement for rank-map used in RMBT-QR and leads to a predetermined matrix block structure. MoM sub-matrices pertaining to interactions of the MIL are compressed by forming QRs from samples. At the lowest level, dense blocks are retained for interactions of the smallest cube with its neighbors.

## V. Simulation Results

In this section, simulation results are presented to demonstrate the accuracy and time and memory efficiency of PILOT. For a comparative analysis, results obtained from RMBT-QR, which is based on IES<sup>3</sup> and FastCap, which is an open source code based on FMM are presented side-by-side. A QR decomposition tolerance of 1e-3 is used for both PILOT and RMBT-QR whereas for FastCap the adaptive algorithm with multipole order of 2 is employed. An

absolute residual of  $1e-3$  is used for the Krylov sub-space iterative solution. All tests were run on a processor with 4GB RAM and 1.6GHz CPU speed.

In the first example the capacitance matrix of the multi-net structure (figure 4) is simulated. The surface of the structure is meshed with 0.113 million patches. The absolute values of the first column of the matrix as obtained from the 3 algorithms are plotted in figure 4, where it can be seen that even very small entries relative to the large ones match very well. In the next example a 5-by-5 bus structure (figure 5) is considered. Each trace is  $1 \mu\text{m}$  in length and  $0.1 \mu\text{m} \times 0.1 \mu\text{m}$  in cross section. The distance between the layers of trace is  $0.3 \mu\text{m}$  and the separation between traces on the same layer is  $0.1 \mu\text{m}$ . The number of triangular patches is varied from 1000 to 0.7 million. The memory and time efficiency of PILOT is demonstrated first by comparison to regular direct and iterative solvers as in figure 5a and 5b and then to RMBT-QR and FastCap as in figure 5c and

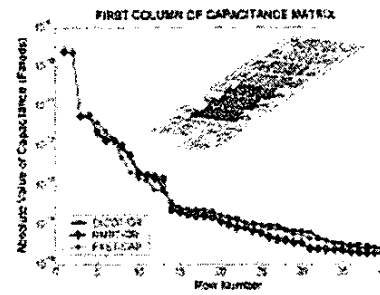


Figure 4: Absolute values of the first column of the capacitance matrix

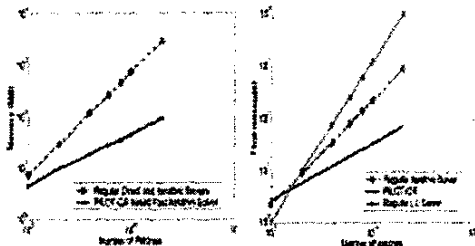


Figure 5a: Memory

Figure 5b: Time

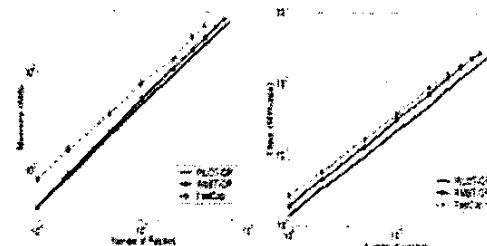
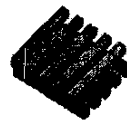


Figure 5c: Memory

Figure 5d: Time

The next example demonstrates the relative advantage of QR methods as compared to FMM based algorithms for higher number of nets and consequently larger number of matrix vector products. A package structure with 14 leads as illustrated in figure 6a is considered. The surface is meshed with 0.101 million patches and then solved for increasing number of right hand sides (1 to 14). The time requirements are plotted in figure 6b. The constant offset between the plots of PILOT and RMBT-QR is due to the superior one-time setup cost. The memory required for the process by PILOT is 441MB, by RMBT-QR is 445MB and FastCap 700MB. The largest problem solved using our method is an on-chip structure discretized with 0.913 million patches, similar to that in figure 4, with 30 times more traces in a  $10 \times 3$  array. The problem took 3.3Gb and 48 minutes to setup and 90 minutes to solve for 3 specific excited nets. Both the other methods required more than the 4 Gb available.

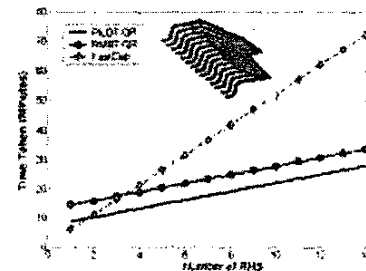


Figure 6) Setup and Solve Time

## VI. Conclusions

A new oct-tree based QR technique for fast parasitic extraction, PILOT, was presented. The best features of FMM and IES<sup>3</sup> are exploited and merged, along with the generation of new merged interaction lists in order to yield superior run times and reduced memory consumption in PILOT. In the battle of reducing constants in the era of mature linear complexity algorithms, PILOT can potentially emerge as an optimal paradigm for parasitic extraction. While this paper is related to parasitic extraction, the PILOT paradigm can also be applied to full-wave applications in multi-layered microelectronic environments. *This work was partially supported by US DARPA/MTO NeoCAD program under Grant N66001-01-1-8920 and NSF Career Award grant ECS-0093102.*

### References

- [1] S.Kapur and D.E.Long, "IES<sup>3</sup>: Efficient Electrostatic and Electromagnetic Solution" *IEEE Computer Sci. and Engg.* 5(4) pp. 60-67 Oct.-Dec 1998.
- [2] G.H.Golub and C.F.Van Loan Matrix Computations 2<sup>nd</sup> Ed. The Johns Hopkins Univ. Press Baltimore 1989.
- [3] R.J. Anderson, "Tree data-structures for N-body simulation", *SIAM J. Comp.*, Vol. 28, No. 6, pp. 1923-1940, 1999.
- [4] K. Nabors, S. Kim and J. White, "Fast capacitance extraction of general three-dimensional structures", *IEEE Trans. on Microwave Theory and Techniques*, vol. 40 issue 7 pp. 1496-1506, July 1992.