Cartesian Multipole Based Numerical Integration for 3D Capacitance Extraction

U. Geigenmüller and N.P. van der Meijs

Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands
e-mail: uggmllr@cas.et.tudelft.nl, nick@cas.et.tudelft.nl

Abstract

Application of the hierarchical Schur algorithm to the boundary element method for 3D capacitance extraction shifts the speed bottleneck from inversion of the influence matrix to its calculation. We show how the numerical integration required for the latter can be accelerated by an order of magnitude with the aid of a multipole expansion in Cartesian formulation. The scheme differs essentially from that of the FASTCAP *extractor.*

1 Introduction

In submicron integrated circuits, where the vertical size of the interconnects is typically not small compared to their minimum lateral size, a reliable determination of interconnect capacitances cannot be achieved with simple formulas based on area and perimeter of the conductors. Instead, a numerical solution of Poisson's equation in three dimensions is required. When the circuit is embedded in a stratified dielectric medium, this task can generally best be handled using the boundary element method (BEM) [1], since then only the conducting surfaces need to be discretized.

A BEM calculation consists of the following steps:

1. tile the conducting surfaces with boundary elements (BEs),
2. calculate for all pairs of BEs the *influence matrix* element $\mathcal{G}_{\mu\nu}$, *i.e.* the (weighted average of the) potential generated on BE μ due to a charge density on BE ν,
3. invert the influence matrix $\mathcal{G}$, and
4. determine the short-circuit capacitance matrix C as

$$C_{ij} = \sum_{\substack{\text{all BEs } \mu \\ \text{on conductor } i}} \sum_{\substack{\text{all BEs } \nu \\ \text{on conductor } j}} \left(\mathcal{G}^{-1}\right)_{\mu\nu} . \qquad (1)$$

When using a general-purpose routine for inverting the influence matrix, such as Gaussian elimination, the memory complexity scales as $\mathcal{O}(N^2)$ and the time complexity as $\mathcal{O}(N^3)$, where N denotes the number of boundary elements. This is only tolerable for relatively small N. If the BEM is to be applied to large ICs, both complexities need to be reduced substantially.

One way of overcoming the time complexity is the use of the fast multipole algorithm, implemented in the FASTCAP extractor by Nabors and White [2]. In the inversion step, when the BE charges are calculated for given BE potentials, this method does not treat the influence of distant BEs individually, but rather uses a multipole expansion of their combined effect. By using this algorithm, together with a generalized conjugate residual method, FASTCAP achieves a reduction of the time complexity to $\mathcal{O}(N \times m)$, where $m \propto N$ is the number of conductors.

An alternative scheme for reducing the computational complexity, which is particularly suited for application to integrated circuits, is based on the omission of all *direct* capacitive couplings for distances beyond a selectable cut-off value w. In this way only a band matrix needs to be stored and inverted. Experience shows that in typical IC interconnect structures, sufficiently precise capacitance values can be achieved with rather small cut-off distances. However, to avoid unphysical artifacts such as negative capacitance values, the cut-off must be introduced in an indirect way, by requiring that all the elements of the *inverse* matrix $\mathcal{G}^{-1}$ with mutual distance larger than w vanish. Then one furthermore has the advantage of reducing the capacitance model to be used in a simulation program.[1] An efficient algorithm that accomplishes this task, the so-called *hierarchical Schur algorithm*, has been developed [3], implemented [4] and refined [5] in the SPACE[2] extractor. The hierarchical Schur algorithm reduces the time complexity to $\mathcal{O}(Nw^4)$, while the memory requirement becomes $\mathcal{O}(\sqrt{N}w^4)$. Note that w is *independent* of the layout size.

In fact, the acceleration of the matrix inversion by the hierarchical Schur algorithm can be so large that the *calculation* of the matrix elements $\mathcal{G}_{\mu\nu}$, rather than the *inversion* of the matrix $\mathcal{G}$, becomes the bottleneck in execution speed. In one way or another, computation of $\mathcal{G}_{\mu\nu}$ requires the evaluation of a four-dimensional integral, *viz.* two-dimensional integrals of the electrostatic Green function with respect to both arguments. In the simplest case, the Green function is just the Coulomb potential. When a ground plane and a stratified

[1] This cannot be accomplished by "sparsifying" $\mathcal{G}$ itself, since the inverse of a sparse matrix is, in general, not sparse.

[2] SPACE is actually a full-featured layout to circuit extractor, of which 3D capacitance extraction is an integrated part. Starting from *e.g.* a GDS II layout SPACE directly produces a SPICE netlist ready for simulation.

ED&TC '97 on CD-ROM

ambient dielectric are present, however, the Green function takes the form of an infinite sum over image charges. For the Coulomb potential and polygonal boundary elements, one of the integrations can be done analytically [6]. Due to transcendental functions in the analytic formulas, however, on RISC machines numerical integration may actually be faster than the use of these formulas. No analytic formulas are known for the second integral. There is a vast literature on fast numerical integration formulas for triangular and quadrilateral domains [7], but even with very efficient integration rules the time needed for calculation of the matrix elements typically dominates the total execution time when the hierarchical Schur algorithm is used. It turns out that a lot can be gained by calculating the $\mathcal{G}_{\mu\nu}$ with a the aid of a multipole expansion. This scheme, where multipoles are used to evaluate individual influence matrix elements, is obviously very different from the use of multipole expansions in FASTCAP.

2 Multipole expansion

To calculate the influence matrix elements we need the Green function $G(\vec{r}_o,\vec{r}_c)$, which gives us the potential at the observation position $\vec{r}_o$ due to a unit point charge at the charge position $\vec{r}_c$. The most general case that we consider is a circuit above a ground plane at $z=0$, embedded in a dielectric that is uniform in x and y direction. In this case the Green function can be written as an infinite sum over image charges,

$$G(\vec{r}_o,\vec{r}_c)=\sum_{j=0}^{\infty}\frac{A_j}{\sqrt{[z_o-(\sigma_j z_c+d_j)]^2+[x_o-x_c]^2+[y_o-y_c]^2}}\,, \tag{2}$$

as mentioned above. For the calculation of the weights A_j, signs $\sigma_j=\pm1$, and distances d_j see *e.g.* Reference [8]. The influence matrix elements are given by

$$\mathcal{G}_{\mu\nu}=\int d^2r_o\int d^2r_c\, f_\mu^{(o)}(\vec{r}_o)G(\vec{r}_o,\vec{r}_c)\,f_\nu^{(c)}(\vec{r}_c)\,. \tag{3}$$

Various choices for shape functions $f^{(c)}$ and weight functions $f^{(o)}$ are commonly used. In the Galerkin method the shape function $f_\nu^{(c)}$ and the weight function $f_\nu^{(o)}$ are equal; here we use a $f_\nu^{(c)}$ constant on BE ν, zero elsewhere, and normalized such that $\int d^2r\, f_\nu^{(c)}(\vec{r})=1$. The collocation method saves some work by omitting the integration over the observation BE (*i.e.* by choosing for the weight function $f_\mu^{(o)}$ a delta function at the center of BE μ), at the expense of reduced precision and an unphysical, though generally slight, asymmetry of the influence matrix.

When evaluating the elements of the influence matrix we shall interchange the summation in (2) with the integrations in (3). In each term the integration is then of the form

$$\int d^2r_o\int d^2r_c\, f_o(\vec{r}_o)\,\frac{1}{|\vec{r}_o-\vec{r}_c|}\,f_c(\vec{r}_c)\,, \tag{4}$$

where the translation d and sign σ have been shifted from the potential into the shape function f_c (not indicated explicitly for notational simplicity). Thanks to special properties of the Coulomb potential, the multipole expansion is here equivalent to Taylor expansions around the centers $\vec{R}_o$ and $\vec{R}_c$ of the supports of f_o and f_c, respectively. Performing such a double Taylor expansion the integral (4) becomes, with $\vec{r}_o'\equiv\vec{r}_o-\vec{R}_o$ etc.,

$$\begin{aligned}\int d^2r_o'\int d^2r_c'\,&\frac{f_o(\vec{R}_o+\vec{r}_o')\,f_c(\vec{R}_c+\vec{r}_c')}{|\vec{r}_o'+(\vec{R}_o-\vec{R}_c)-\vec{r}_c'|}\\ &=\sum_{i_o=0}^{\infty}\sum_{i_c=0}^{\infty}(-1)^{i_c}\left\{\frac{1}{i_o!}\int d^2r_o' f_o(\vec{R}_o+\vec{r}_o')(\vec{r}_o')^{i_o}\right\}\odot^{i_o}\\ &\quad\left[\left(\frac{\partial}{\partial\vec{R}_o}\right)^{i_o+i_c}\frac{1}{|\vec{R}_o-\vec{R}_c|}\right]\odot^{i_c}\\ &\quad\left\{\frac{1}{i_c!}\int d^2r_c' f_c(\vec{R}_c+\vec{r}_c')(\vec{r}_c')^{i_c}\right\}\,.\end{aligned} \tag{5}$$

In this equation $\vec{a}^{\,m}$ denotes the mth tensorial power of a vector $\vec{a}$, and $A\odot^m B$ the m-fold contraction of two tensors A, B of ranks $n_A, n_B\ge m$; in components

$$\begin{aligned}(A\odot^m B)_{i_1,\ldots,i_{n_A-m},j_{m+1},\ldots j_{n_B}}&=\\ \sum_{k_1,\ldots,k_m}&A_{i_1,\ldots,i_{n_A-m},k_m,\ldots,k_1}B_{k_1,\ldots,k_m,j_{m+1},\ldots j_{n_B}}\,.\end{aligned} \tag{6}$$

The terms between braces in Equation (5) are the multipole moments; note that their structure is the same for charge- and observation-point. The Taylor series of $1/|\vec{r}-\vec{R}|$ in powers of $\vec{r}$ only converges if $r<R$. Therefore, the multipole expansion cannot be used if the center-to-center distance $|\vec{R}_o-\vec{R}_c|$ is smaller than the convergence radius

$$R_{\text{conv.}}\equiv\max_{f_o(\vec{R}_o+\vec{r}_o')\neq0}r_o'\;+\;\max_{f_c(\vec{R}_c+\vec{r}_c')\neq0}r_c'\,. \tag{7}$$

A number of advantages of the multipole expansion are immediately obvious:

1. Only elementary integrations are needed to evaluate the multipole moments. This can easily been done analytically, and the resulting formulas do not contain transcendental functions that are time-consuming to compute.
2. Once the multipole moments have been calculated, they can be used over and over again for each term in the summation over image charges. They do not change when the charge domain is shifted, and the multipole moments of the original charge distribution and its mirror image only differ by a sign in some components. In the sum over images of the charge domain it only rests to contract the multipole moments with the derivatives of the inverse center-to-center distance. If the multipole moments are kept in memory, they can furthermore be used again for different combinations of charge- and observation-domain.

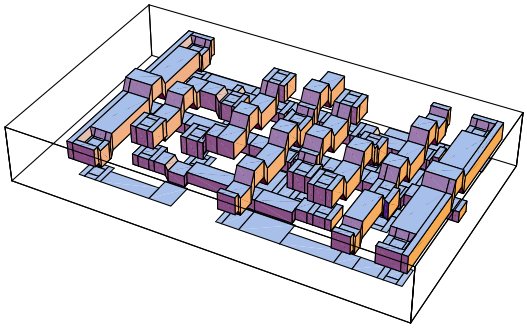

Figure 1: A static SRAM cell. The structure has a minimum distance of 0.3μ and a maximum distance of 3.9μ from a conducting ground plane; it is embedded in a 5μ thick SiO_2 layer. The footprint of the structure measures 8μ × 8μ.

3. The contractions in (5) require little CPU time when both multipole orders i_o and i_c are small. The contractions may become expensive for high-order multipoles, but, since each multipole order adds a factor $1/|\vec{R}_o - \vec{R}_c|$, the contribution of high-order multipoles is small except when the center-to-center distance is close to the convergence radius. The number of multipoles to be included can be chosen according to the ratio $|\vec{R}_o - \vec{R}_c|/R_{\text{conv}}$; if it is large, already the monopole approximation suffices.
4. Numerical integration routines need to compute the square root function time and time again when evaluating the Coulomb potential in the integrand. With the multipole expansion, on the other hand, *one* square root calculation suffices per charge-image/observation pair. The remaining operations are just multiplications and additions. On RISC machines, this helps in reducing the execution time.

A closely related utilization of a multipole-expansion has been presented by Andersson [9]. However, this author employed the more conventional expansion in terms of spherical harmonics. In that case very many time-consuming evaluations of trigonometric functions are needed, and we found that on our computer the multipole-expansion then leads to an increase, instead of a decrease in CPU time. Therefore, it is crucial to use the Cartesian implementation of the multipole expansion presented here.

3 Example: SRAM cell

As a practical example we consider the SRAM cell shown in Figure 1. An impression of the error introduced by employing the multipole expansion for the calculation of the influence matrix elements is given in Figure 2. Here the relative frequency with which an element $\mathcal{G}_{\mu\nu}$ occurs is plotted versus the minimum distance ratio $|\vec{R}_o - \vec{R}_c|/R_{\text{conv}}$ of all the corresponding charge-image/observation pairs of BEs, and versus the relative error of $\mathcal{G}_{\mu\nu}$ that results when this matrix element is evaluated by multipole expansion to quadrupolar order. We see that a threshold distance ratio

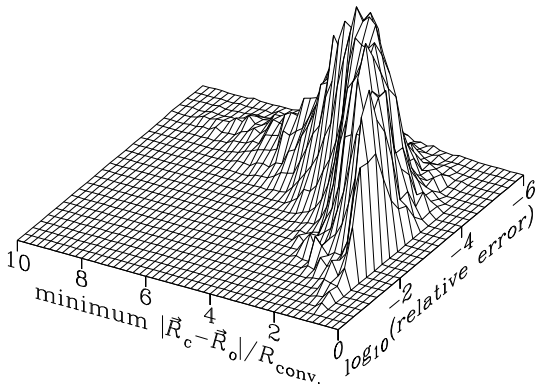

Figure 2: Relative frequency of influence matrix elements, plotted as function of the BE-distance (normalized by the convergence radius of the multipole expansion) and of the relative error introduced by the multipole expansion. The data correspond to the SRAM cell of Figure 1. The Galerkin method is used, and the highest included multipoles are quadrupoles.

Table 1: Total execution time (on a HP 9000/735 computer) for capacitance extraction of the SRAM cell using the SPACE extractor, with a ground plane at z=0, a oxide layer extending from z=0 to z=5μ, and air above. The data in the second row were obtained with a threshold distance ratio of 2, and quadrupoles as highest included multipoles. The window size w of the hierarchical Schur algorithm is 4μ.

	collocation	Galerkin
numerical integration	435 sec.	4450 sec.
multipole expansion	177 sec.	201 sec.

$|\vec{R}_o - \vec{R}_c|/R_{\text{conv}} = 2$ suffices when a precision of one percent is required; for smaller distance ratios conventional numerical integration should be used. Indeed, with this distance threshold and quadrupoles included, the error in the final result for the short-circuit capacitance matrix due to the multipole expansion lies below 0.5 percent.

While the multipole expansion thus does not impair the precision of the calculation, it has a very substantial influence on the execution time. The total CPU times (including mesh generation, I/O etc.) for capacitance extraction of the SRAM cell are shown in Table 1. The acceleration amounts to about a factor 2 for the collocation method, and a factor 22 for the Galerkin method. The multipole expansion also reduces the execution time ratio between the Galerkin method on one hand, and the collocation method on the other hand from about 10 to about 1.1. This makes the more precise and more robust Galerkin method much more attractive.

When the ground plane and/or the dielectric discontinuity due to the embedding SiO_2 layer are omitted, the speed gain obtained by application of the multipole expansion is less

Table 2: As Table 1, but in vacuum (*i.e.* with the ground plane and the dielectric SiO_2 layer removed).

	collocation	Galerkin
numerical integration	166 sec.	430 sec.
multipole expansion	143 sec.	160 sec.

dramatic. The reason is that then there is at most one mirror image of charges, not an infinite number of such image charges. Thus the possibility to re-use once calculated multipole moments is strongly reduced. Nevertheless, an acceleration of the calculation by factor 1.2 to 2.7 is still achieved, as shown in Table 2.

We end this section by a short performance comparison with FASTCAP. For the SRAM example with ground plane, but without dielectric medium, the execution times of SPACE is longer than that of FASTCAP: using the mesh for the cell as generated by SPACE, consisting of 880 BEs, and an explicitly specified ground plane with 400 BEs,[3] we found that FASTCAP needs 30 seconds and SPACE 142 seconds (with the collocation method also used by FASTCAP). The higher memory requirements of FASTCAP, however, prohibit its use for larger ICs. SPACE, with its $\mathcal{O}(\sqrt{N})$ storage complexity, has much less stringent limitations. A 144-transistor test circuit, with a mesh consisting of 42,664 BEs, was extracted by SPACE in 32 minutes using only 22 Mbyte of core memory, whereas FASTCAP needed more than the 210 Mbyte available on our machine. With that amount of core, FASTCAP (version 2.0) can handle at most about 20,000 BEs.

4 Conclusion and outlook

We have shown that by using a multipole method for the evaluation of the influence matrix elements, the total time needed for 3-D capacitance extraction with the Galerkin boundary element method and the hierarchical Schur algorithm can be reduced by about a factor 20, as compared to the case when all influence matrix elements are evaluated by numerical integration. For the collocation method the acceleration amounts to about a factor 2. This gain is only possible if the number of calls to time-consuming, transcendental functions is kept low, which requires a Cartesian formulation of the multipole expansion. The smaller the cut-off distance w can be chosen, the more does the *calculation* of the influence matrix dominate the execution time, and the larger is the acceleration due to the multipole expansion.

The gain in speed is largest when many image charges, due to a ground plane and dielectric discontinuities, must be taken into account. In the SRAM example discussed in this paper, only two different dielectric layers (SiO_2 and air) were used. When more dielectric layers, and consequently more image charges are present, we expect the acceleration due to the multipole expansion to improve further.

Finally we would like to point out that the method can be applied to other problems than capacitance extraction only; examples from the field of IC design are the calculation of substrate resistances [10] and thermal conduction properties.

[3] FASTCAP has no built-in facility for ground plane and dielectric layers, so that these must be specified explicitly.

5 Acknowledgement

This research was supported by the Dutch Foundation for Technical Sciences (STW), in the context of the DIMES research school for microelectronics. We thank F. Beeftink, P. Elias and A. van Genderen for useful discussions.

References

[1] P.D. Patel, *Calculation of Capacitance Coefficients for a System of Irregular Finite Conductors on a Dielectric Sheet*, IEEE Trans. Microwave Theory Tech., MTT-19, No. 11, November 1971, pp. 862-869.

[2] K. Nabors and J. White, *FastCap: A Multipole Accelerated 3-D Capacitance Extraction Program*, IEEE Trans. Computer-Aided Design, CAD-10, No. 11, November 1991, pp. 1447-1459.

[3] P. Dewilde and E. Deprettere, *Approximate Inversion of Positive Matrices with Applications to Modelling*, in: *Modelling, Robustness and Sensitivity Reduction in Control Systems*, NATO ASI Series, edited by Ruth F. Curtain, pp. 211-238 (Springer, Berlin 1987).

[4] N.P. van der Meijs and A.J. van Genderen, *An Efficient Finite Element Method for Submicron IC Capacitance Extraction*, Proc. 26th Design Automation Conference, Las Vegas, June 1989, pp. 678-681.

[5] A. J. van Genderen and N. P. van der Meijs, *A Frontal Computation Scheme for the Schur Algorithm to Efficiently Solve Large Boundary-Element Problems*, CompEuro 1992 Proceedings Computer Systems and Software Engineering, the Hague, the Netherlands, May 1992, pp. 568-573.

[6] D.R. Wilton, S.M. Rao, A.W. Glisson, D.H. Schaubert, O.M. Al-Bundak, and C.M. Butler, *Potential Integrals for Uniform and Linear Source Distributions on Polygonal and Polyhedral Domains*, IEEE Transactions on Antennas and Propagation, AP-32, No. 3, March 1984, pp. 276-281.

[7] For a bibliography see *e.g.* P.J. Davis and P. Rabinowitz, *Methods of Numerical Integration* (Academic Press, New York 1975).

[8] P. Dewilde and Zhen-Qiu Ning, *Models for Large Integrated Circuits* (Kluwer, Boston 1990).

[9] T. Andersson, *Method of Moments and the Use of Multipole Expansion*, J. Electromagnetic Waves and Applications, **5** (1991), 1237.

[10] T. Smedes, N. P. van der Meijs and A. J. van Genderen, *Extraction of Circuit Models for Substrate Cross-talk*, Proc. of the ICCAD 1995, San Jose, November 1995, pp. 199-206.