A Layered Finite Element Method for Electromagnetic Analysis of Large-Scale High-Frequency Integrated Circuits

Dan Jiao, *Senior Member, IEEE*, Sourav Chakravarty, and Changhong Dai

***Abstract*—A high-capacity electromagnetic solution, layered finite element method, is proposed for high-frequency modeling of large-scale three-dimensional on-chip circuits. In this method, first, the matrix system of the original 3-D problem is reduced to that of 2-D layers. Second, the matrix system of 2-D layers is further reduced to that of a single layer. Third, an algorithm of logarithmic complexity is proposed to further speed up the analysis. In addition, an excitation and extraction technique is developed to limit the field unknowns needed for the final circuit extraction to a single layer only, as well as keep the right-hand side intact during the matrix reduction process. The entire procedure is numerically rigorous without making any theoretical approximation. The computational complexity only involves solving a single layer irrespective of the original problem size. Hence, the proposed method is equipped with a high capacity to solve large-scale IC problems. The proposed method was used to simulate a set of large-scale interconnect structures that were fabricated on a test chip using conventional Si processing techniques. Excellent agreement with the measured data has been observed from dc to 50 GHz.**

***Index Terms*—Electromagnetics, finite element method, high capacity, high frequency, on-chip circuits, three dimension.**

I. Introduction

As ON-CHIP designers move to faster clock frequencies enabled by process technology scaling with reduced feature sizes, electromagnetic analysis has drawn the attention of the on-chip design community. In 2001, a research team at Intel started to validate RLC-based parasitic extraction at tens of gigahertz. Significant mismatch between measurements and RLC models was observed at multigigahertz frequencies on 3-D interconnect bus structures [1]. In contrast, full-wave electromagnetic-based modeling accurately captured the measured behavior over the entire frequency band [1], [2]. The mismatch between RLC models and measurements was attributed to the decoupled E and H model employed in static modeling by extracting the capacitance and inductance independent of each other [1]. This finding demonstrated the importance of full-wave electromagnetic-based solutions in high-frequency IC design [3]. The importance has been further realized in today's low power design. In power efficient mobile chips, low power states and clock gating are gaining momentum as the main power saving mechanisms. In these architectures, entire blocks of circuits are switched on and off to achieve an optimal power-performance operating point. These power reduction techniques result in large processor current variations and fast transient droops and noises in the power supply network, which cannot be accurately captured by a static-based IR drop or transient droop analysis.

In addition to high-frequency digital IC design, electromagnetic analysis is also of paramount importance to analog, RF, and mixed-signal IC design. Integrated computing and communication calls for increasing levels of integration of RF, analogue, and digital systems. Integrating as many circuits as possible on the same die leads often to undesired coupling and sometimes to system failure. For instance, switching currents induced by logic circuits cause ringing in the power-supply rails and in the output driver circuitry. This, in turn, couples through the common substrate to corrupt sensitive analog signals on the same chip. Prevailing circuit-based signal integrity paradigms are reaching their limits of predictive accuracy when applied to high-frequency mixed-signal settings. An electromagnetic solution is indispensable to sustain the continual scaling and integration of digital, analog, mixed-signal, and RF circuitry.

However, high-frequency IC design imposes many modeling challenges to electromagnetic analysis. These challenges include conductor loss, large numbers of dielectric stacks, strong non-uniformity, the presence of substrate, large numbers of conductors, large aspect ratio, broadband, and 3-D complexity [3]. Almost every challenge increases the number of unknowns, and hence the problem size one needs to solve when tackling an IC problem. For instance, due to conductor loss, one has to discretize into conductors with very fine elements to capture rapid field variation within skin depth. This generates a large number of unknowns even for small on-chip structures. In addition to on-chip intricacy that increases the problem size, the need for full-chip analysis also stresses problem size. A full-chip analysis is often needed to capture the global electrical interactions between integrated circuits on the die, and between the die and the package. However, to date, the fastest integral equation solver needs $O(N \log N)$ operations and $O(N \log N)$ storage in dealing with N-unknown electrodynamic problems; the fastest partial-differential-equation based solvers scale as $O(N)$ in both memory requirement and CPU cost. This performance is generally regarded as the limit that one can

Manuscript received May 21, 2006; revised August 31, 2006. This work was supported in part by a grant from the Office of Naval Research under Award N00014-06-1-0716 and in part by a grant from Intel Corporation.

D. Jiao is with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47906 USA (e-mail: djiao@purdue.edu).

S. Chakravarty and C. Dai are with the Design and Technology Solutions, Intel Corporation, Santa Clara, CA 95052-8119 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAP.2006.889847

achieve in computational electromagnetics. Since N is a big number in IC analysis even for a circuitry of a modest size, the current performance of computational electromagnetic techniques is still insufficient when tackling a realistic IC design problem that can involve billions of unknowns. Therefore, it is of paramount importance to study and develop a high-capacity electromagnetic solution.

Having realized the importance of full-wave electromagnetic analysis in high frequency chip design, and the unique modeling challenges of IC problems, researchers in both circuit and fields have been working on developing innovative electromagnetic solutions [2]–[13]. However, most of the research efforts have been placed on enriching existing electromagnetic modeling techniques with new capabilities to address on-chip intricacy. Little work has been reported in open literature on high-capacity electromagnetic solutions, which can potentially tackle full-chip problems. In [2], we presented a novel, rigorous, and fast method for the full-wave modeling of large-scale high speed interconnect structures. In this method, a general interconnect structure is decomposed into a few structure seeds. In each structure seed, the original wave propagation problem is represented into a generalized eigenvalue problem. A novel mode-matching technique is developed to solve large-scale 3-D problems by using 2-D-like CPU time and memory. This method has shown great capability in modeling Manhattan-type large-scale interconnect structures. In this paper, we propose a layered finite element method to tackle both Manhattan- and non-Manhattan-type large-scale multilayered structures. Layout periodicity can be employed to further speed up the proposed method, but it is not a must to achieve the high capacity of the proposed method. We will elaborate this method in the following six sections: Section II problem statement, Section III layered finite-element scheme, Section IV reduction of the 3-D layered system matrix to a 2-D layered one, Section V reduction of the 2-D layered system matrix to a single-layer one, Section VI an algorithm of logarithmic complexity for further speed-up, Section VII excitation and extraction, and Section VIII performance analysis. Finally, we will demonstrate the accuracy and high capacity of the proposed method by a number of numerical and experimental results.

II. Problem Statement

Consider 3-D circuit problems shown in Fig. 1.

The circuit can be a RF CMOS device metallic system, a global on-chip interconnect structure, an RF IC circuit, or others. Inside these circuits, the electric field $\mathbf{E}$ satisfies the second-order vector wave equation

$$\nabla \times \left[\mu_r^{-1} \nabla \times \mathbf{E}\right] - k_0^2 \bar{\varepsilon}_r \mathbf{E} = -jk_0 Z_0 \mathbf{J} \quad \text{in } V \tag{1}$$

subject to certain boundary conditions. In (1), the bar over ε_{r} denotes a complex permittivity that comprises both permittivity and conductivity; μ_{r} is the relative permeability; k_0 and Z_0 are free-space wave number and impedance, respectively; $\mathbf{J}$ is the current source; V is the computational domain that encloses the circuit. The stack-growth direction is defined as z and used throughout this paper.

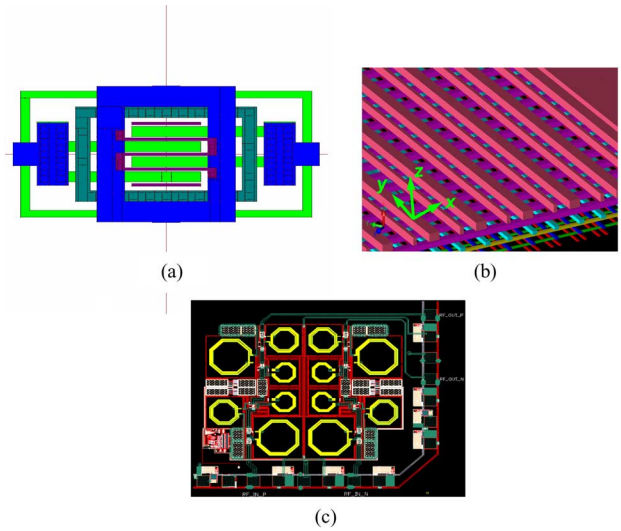


Fig. 1. 3-D circuit problems. (a) RF CMOS. (b) Global on-chip interconnects. (c) RF IC.

To solve (1), we formulate a numerical algorithm to obtain the $\mathbf{E}$ fields or $\mathbf{H}$ fields inside the computational domain at each discretized point, from which the design parameters of interest are obtained. Due to the computational complexity of on-chip circuit problems as stated in Section I, the resultant numerical system is generally extremely big even for a small circuitry. This prevents the direct use of existing computational electromagnetic techniques in guiding chip-level high-frequency IC design. We propose to tackle this problem by developing a layered finite-element method.

III. Layered Finite Element Scheme

In accordance with variational principle [14], the solution to the boundary value problem defined by (1) and its boundary conditions can be obtained by seeking the stationary point of the functional

$$\begin{aligned} F(\mathbf{E}) &= \frac{1}{2}\iiint_V \left[\mu_r^{-1}(\nabla \times \mathbf{E}) \cdot (\nabla \times \mathbf{E}) - k_0^2 \bar{\varepsilon}_r \mathbf{E} \cdot \mathbf{E}\right] dV \\ &\quad + \iiint_V jk_0 Z_0 \mathbf{J} \cdot \mathbf{E}\, dV \\ &\quad - \frac{1}{2}\iint_{S_o} \mathbf{E} \cdot \mathbf{P}(\mathbf{E})\, dS. \end{aligned} \tag{2}$$

In (2), S_o denotes the truncation boundary, which is the outermost region in the computational domain. $\mathbf{P}$ is an operator associated with the absorbing boundary condition placed on the truncation boundary. If the first-order absorbing boundary condition is used, (2) can be written as

$$\begin{aligned} F(\mathbf{E}) &= \frac{1}{2}\iiint_V \left[\mu_r^{-1}(\nabla \times \mathbf{E}) \cdot (\nabla \times \mathbf{E}) - k_0^2 \bar{\varepsilon}_r \mathbf{E} \cdot \mathbf{E}\right] dV \\ &\quad + \iiint_V jk_0 Z_0 \mathbf{J} \cdot \mathbf{E}\, dV \\ &\quad + \frac{1}{2}\iint_{S_o} jk_0 (\hat{n} \times \mathbf{E}) \cdot (\hat{n} \times \mathbf{E})\, dS. \end{aligned} \tag{3}$$

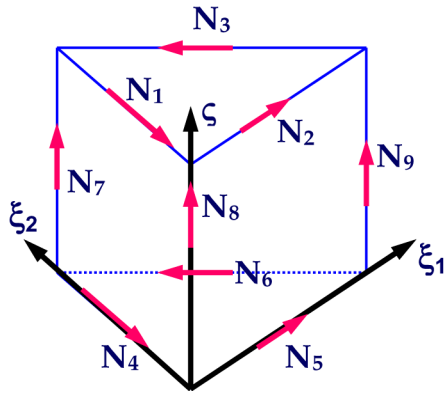


Fig. 2. Illustration of prism vector basis functions.

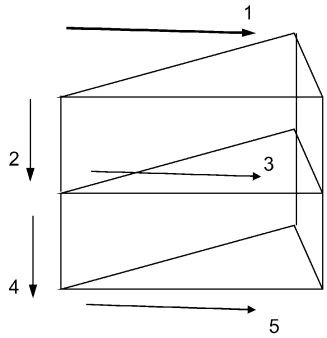


Fig. 3. Unknown ordering scheme.

For the boundaries that are shorted to ground, tangential $\mathbf{E}$ is explicitly enforced to be zero therein. Next, we perform discretization. The discretization is conducted for both dielectric regions and conducting regions. Disretizing into conductors allows for an accurate modeling of conductor loss. The triangular prism elements are used to discretize the computational domain. These elements are very suitable for the discretization of on-chip structures, which are multilayered structures. They extrude along the layer-growth direction while capturing the irregular geometry in the transverse cross section. This allows for the modeling of irregular on-chip circuits such as trapezoidal- and spiral-type interconnects. In this method, the layer-growth direction can be chosen the same as the natural layer-growth direction, which is the stack-growth direction. It can also be chosen from other directions to minimize the number of unknowns in the transverse cross section that is perpendicular to the layer-growth direction. For example, a Manhattan-type integrated circuit structure is layered by looking from any of x, y, and z directions. Therefore, the layer-growth direction can be chosen from any of the x, y, and z directions to render a transverse cross section that has the minimal problem size.

In each prism element, the electric field $\mathbf{E}$ is expanded into prism vector basis functions $\mathbf{N}$ [15]

$$\mathbf{E}^e = \sum_{i=1}^{n} u_i \mathbf{N}_i^e. \tag{4}$$

The superscript e denotes e element. In each prism element, there are nine vector bases as shown in Fig. 2.

These functions can be written as

$$\begin{aligned}
\mathbf{N}_1 &= \zeta(\xi_1\nabla\xi_2 - \xi_2\nabla\xi_1)\\
\mathbf{N}_2 &= \zeta(\xi_2\nabla\xi_3 - \xi_3\nabla\xi_2)\\
\mathbf{N}_3 &= \zeta(\xi_3\nabla\xi_1 - \xi_1\nabla\xi_3)\\
\mathbf{N}_4 &= (1-\zeta)(\xi_1\nabla\xi_2 - \xi_2\nabla\xi_1)\\
\mathbf{N}_5 &= (1-\zeta)(\xi_2\nabla\xi_3 - \xi_3\nabla\xi_2)\\
\mathbf{N}_6 &= (1-\zeta)(\xi_3\nabla\xi_1 - \xi_1\nabla\xi_3)\\
\mathbf{N}_7 &= \xi_1\nabla\zeta/|\nabla\zeta|\\
\mathbf{N}_8 &= \xi_2\nabla\zeta/|\nabla\zeta|\\
\mathbf{N}_9 &= \xi_3\nabla\zeta/|\nabla\zeta|.
\end{aligned} \tag{5}$$

In (5), ξ_1, ξ_2, and ξ_3 are area coordinates (also known as node basis functions ([14, pp. 80]), ζ is 0 at the bottom plane, and 1 at the upper plane. Clearly, as shown in Fig. 2, basis functions $\mathbf{N}_1$, $\mathbf{N}_2$, $\mathbf{N}_3$, $\mathbf{N}_4$, $\mathbf{N}_5$, and $\mathbf{N}_6$ reside on the surfaces of each layer, and hence their associated unknowns are called as surface unknowns throughout this paper, while the unknowns associated with basis functions $\mathbf{N}_7$, $\mathbf{N}_8$, and $\mathbf{N}_9$ are named as volume unknowns. From (5), one can also see clearly that the surface basis functions $\mathbf{N}_1$, $\mathbf{N}_2$, $\mathbf{N}_3$, $\mathbf{N}_4$, $\mathbf{N}_5$, and $\mathbf{N}_6$ are formed by multiplying ζ or $(1-\zeta)$ with edge basis functions $\mathbf{W}$ ([14, pp. 234–237]).

Substituting (4) into (3), and taking the partial derivative of (3) with respect to unknown coefficients u_i yield the following matrix equation:

$$\mathbf{K}u = b \tag{6}$$

in which $\mathbf{K}$ and b are assembled from their elemental counterparts

$$\begin{aligned}
\mathbf{K}^e &= \mu_r^{-1}\langle\nabla\times\mathbf{N}_i, \nabla\times\mathbf{N}_j\rangle_V - k_0^2\bar{\varepsilon}_r\langle\mathbf{N}_i, \mathbf{N}_j\rangle_V\\
&\quad + jk_0\langle\hat{n}\times\mathbf{N}_i, \hat{n}\times\mathbf{N}_j\rangle_{S_o}\\
b^e &= -jk_0 Z_0 \langle\mathbf{N}_i, \mathbf{J}\rangle_V.
\end{aligned} \tag{7}$$

In (7), the inner product is defined as

$$\begin{aligned}
\langle\mathbf{a}, \mathbf{b}\rangle_V &= \iiint_V \mathbf{a}\cdot\mathbf{b}\, dV\\
\langle\mathbf{a}, \mathbf{b}\rangle_S &= \iint_S \mathbf{a}\cdot\mathbf{b}\, dS.
\end{aligned} \tag{8}$$

By ordering the unknowns layer by layer as shown in Fig. 3, we generate a banded matrix $\mathbf{A}$. Though a banded matrix, its solution can be highly computationally expensive when the number of unknowns is large. A direct solution generally requires a large amount of memory; an iterative solution can converge slowly, and is inefficient in the presence of multiple right-hand sides. To solve this problem, we first reduce the system matrix $\mathbf{A}$ to one that only involves 2-D surface unknowns in each layer, the detail of which is illustrated in the next section.

IV. Reduction of the 3-D Layered System Matrix to a 2-D Layered One

To form a matrix system that only involves 2-D surface unknowns in each layer, we need to eliminate volume unknowns. This can be achieved by using the procedure we proposed in

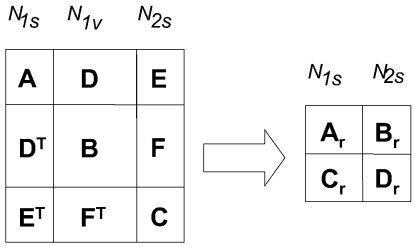


Fig. 4. Procedure of eliminating volume unknowns.

[16]. For instance, the volumetric unknowns in Layer 1, which is $\mathrm{N}_{1\,v}$, can be eliminated by using the procedure illustrated in Fig. 4. $\mathrm{N}_{1\,s}$ and $\mathrm{N}_{2\,s}$ are the surface unknowns on the top and bottom surfaces of Layer 1.

In Fig. 4, matrix $\mathbf{A}$ is formed between unknowns $\mathrm{N}_{1\,s}$ and $\mathrm{N}_{1\,s}$, $\mathbf{B}$ is formed between unknowns $\mathrm{N}_{1\,v}$ and $\mathrm{N}_{1\,v}$; $\mathbf{C}$ is formed between unknowns $\mathrm{N}_{2\,s}$ and $\mathrm{N}_{2\,s}$; $\mathbf{D}$ is formed between unknowns $\mathrm{N}_{1\,s}$ and $\mathrm{N}_{1\,v}$; $\mathbf{E}$ is formed between unknowns $\mathrm{N}_{1\,s}$ and $\mathrm{N}_{2\,s}$; and $\mathbf{F}$ is formed between unknowns $\mathrm{N}_{1\,v}$ and $\mathrm{N}_{2\,s}$. The relationship between the transformed matrices and the original matrices can be written as

$$\begin{aligned}\mathbf{A_r} &= \mathbf{A} - \mathbf{D}\mathbf{B^{-1}}\mathbf{D^T}\\ \mathbf{B_r} &= \mathbf{E} - \mathbf{D}\mathbf{B^{-1}}\mathbf{F}\\ \mathbf{C_r} &= \mathbf{B_r^T}\\ \mathbf{D_r} &= \mathbf{C} - \mathbf{F^T}\mathbf{B^{-1}}\mathbf{F}.\end{aligned} \tag{9}$$

Essentially, the volume unknowns are eliminated by using the relationship between the surface and volume unknowns. This relationship can be also used to recover volume unknowns from the surface unknowns.

From (9), apparently, in order to eliminate volume unknowns, one needs to fill in matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{E}$, and $\mathbf{F}$ for each layer. In addition, one has to evaluate $\mathbf{DB^{-1}D^T}$, $\mathbf{DB^{-1}F}$, and $\mathbf{F^TB^{-1}F}$ for each layer. The resultant computational cost can be very high when the number of layers is large as well as the number of unknowns in each layer. In this paper, a fast technique is proposed to eliminate the volume unknowns efficiently. This fast technique is achieved by deriving the following matrix properties.

a) Matrix $\mathbf{D}$ is the same for all the layers.
This is because matrix $\mathbf{D}$ is assembled from the following elemental matrix:

$$\mathbf{D}_{ij}^{e} = \mu_r^{-1}\langle \mathbf{W}_i, \nabla\xi_j\rangle_\Omega. \tag{10}$$

In (10), Ω denotes the region forming a triangular element, $\mathbf{W}$ is the edge basis function ([14, pp. 234–237]), and ξ is the node basis function ([14, pp. 80]). Since the permeability μ_{r} does not change in the realistic on-chip structures, $\mathbf{D}$ remains the same for all the layers. Therefore, matrix $\mathbf{D}$ only needs to be filled for one layer.

b) Matrices $\mathbf{F}$ and $\mathbf{D}$ are correlated

$$\mathbf{F} = -\mathbf{D^T} \tag{11}$$

As a result, the need of evaluating $\mathbf{DB^{-1}F}$ and $\mathbf{F^T}\,\mathbf{B^{-1}F}$ is eliminated and only $\mathbf{DB^{-1}}\,\mathbf{D^T}$ needs to be evaluated.

c) Matrix $\mathbf{A}$ is equal to matrix $\mathbf{C}$ in each layer

$$\begin{aligned}\mathbf{A}_{ij}^{e} &= \mathbf{C}_{ij}^{e}\\ &= \mu_r^{-1}\left[\frac{l}{3}\langle\nabla\times\mathbf{W}_i, \nabla\times\mathbf{W}_j\rangle_\Omega + \frac{1}{l}\langle\mathbf{W}_i, \mathbf{W}_j\rangle_\Omega\right]\\ &\quad - k_0^2\bar{\varepsilon}_r\frac{l}{3}\langle\mathbf{W}_i, \mathbf{W}_j\rangle_\Omega\end{aligned} \tag{12}$$

in which l is the layer thickness. Therefore, based on b) and c), the following equality holds true:

$$\mathbf{A_r} = \mathbf{D_r}. \tag{13}$$

d) Matrix $\mathbf{B}$ is linearly proportional to the layer thickness.
In fact, matrix $\mathbf{B}$ is assembled from the following elemental matrix:

$$\mathbf{B}_{ij}^{e} = -k_0^2\bar{\varepsilon}_r l\langle\xi_i, \xi_j\rangle_\Omega + \mu_r^{-1} l\langle\nabla\xi_i, \nabla\xi_j\rangle_\Omega. \tag{14}$$

Therefore, matrix $\mathbf{B}$ only needs to be formed and inverted for a layer of unit thickness. Others can be obtained by scaling accordingly.

e) For interconnect structures, matrix $\mathbf{B}$ only needs to be formed for each unique structure seed.

The concept of structure seeds in 3-D interconnect structures was first introduced in [2]. A structure seed has a unique cross section. For a 3-D bus structure of n orthogonal layers, the number of structure seeds is 2^n. This number is small. For instance, for an interconnect involving seven metal layers, this number is only 8 irrespective of the number of wires. If we choose the layer-growth direction to be either x or y [please refer to Fig. 1(b)], then each layer features the same dielectric configuration. Then, from (14), one can see clearly that matrix $\mathbf{B}$ of unit thickness is *different* only when the conductor configuration is different. Hence, matrix $\mathbf{B}$ only needs to be formed and inverted for each structure seed of unit thickness. The inverse of matrix $\mathbf{B}$ in each layer can then be readily obtained by linearly scaling the structure-seed-based inverse matrix with the layer thickness l.

As an immediate result of the aforementioned factors, the computational cost of reducing the 3-D system matrix to a 2-D layered one only involves solving $\mathbf{DB^{-1}}\,\mathbf{D^T}$ for each structure seed. Since $\mathbf{B}$ and $\mathbf{D}$ are extremely sparse matrices, and generally there are only a few structure seeds for on-chip interconnect structures, the reduction can be performed very efficiently.

V. Reduction of the 2-D Layered System Matrix to a Single-Layer One

With all the volume unknowns eliminated, we obtain a system matrix that only involves 2-D surface unknowns in each layer. If the number of layers is only a few, we can stop there and solve the reduced system matrix as a whole without further reduction. However, in reality, we often encounter a large number of layers. For example, in a realistic on-chip interconnect structure, one can encounter a large number of layers by either segmenting along the x or y direction. Therefore, further reduction is often

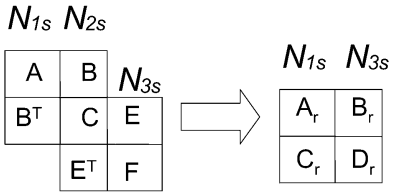


Fig. 5. Matrix cascading.

needed. Hence, we continue to reduce the dimension of the system matrix to the size that one can handle with available computational resources. For instance, if one is only able to handle ten layers, we reduce the system to involve only ten layers; if one can only handle one layer, we reduce the system matrix all the way down to the one that only involves single-layer unknowns.

The reduction process is conducted by matrix cascading. Fig. 5 illustrates a matrix cascading process, in which N_{1s}, N_{2s}, and N_{3s} denote the top surface unknowns of layer 1, layer 2, and layer 3, respectively. Cascading layer 1 with layer 2 is equivalent to eliminating N_{2s} unknowns. The relationship between the submatrices in the reduced matrix and the original ones can be written as

$$\begin{aligned}\mathbf{A_r} &= \mathbf{A} - \mathbf{B}\mathbf{C^{-1}}\mathbf{B^T}\\ \mathbf{B_r} &= -\mathbf{B}\mathbf{C^{-1}}\mathbf{E}\\ \mathbf{C_r} &= \mathbf{B_r^T}\\ \mathbf{D_r} &= \mathbf{F} - \mathbf{E^T}\mathbf{C^{-1}}\mathbf{E}.\end{aligned} \tag{15}$$

The reduction in (15) can be achieved efficiently by using symmetric backward Gaussian elimination [17]. To be specific, the unknowns to be eliminated are first reordered to the bottom. They are then eliminated one by one by symmetric backward Gaussian elimination. Since the cascading is performed on the field-based matrix in each layer, the tangential field continuity is guaranteed at each interface. Therefore, different from circuit-port based cascading, the cascading procedure proposed here is rigorous.

Now, assuming one is interested in layer i, the matrix in each layer before the ith layer can be cascaded together to form matrix $\mathbf{P^-}$. Similarly, those matrices after the ith layer can be cascaded together to form matrix $\mathbf{P}^+$. If one is equipped with sufficient resources to solve three layers, he can stop here: solving the reduced matrix system shown in Fig. 6 to obtain the solution. If not, he can continue to cascade the three layers together to form one that only involves single-layer unknowns. The multiple-layer matrix cascading can be implemented either serially or in parallel.

Serial Implementation: First we cascade layer 1 with layer 2 using the procedure shown in Fig. 5, we obtain matrix $\mathbf{A}_{12}$, which is a square matrix relating unknowns $\mathbf{N}_{1s}$ to $\mathbf{N}_{3s}$. The interface $\mathbf{N}_{2s}$ unknowns is eliminated in this process. We then cascade the $\mathbf{A}_{12}$ with the submatrix in layer 3 to form matrix $\mathbf{A}_{123}$, which is a square matrix relating unknowns $\mathbf{N}_{1s}$ to $\mathbf{N}_{4s}$. In this process $\mathbf{N}_{3s}$ is eliminated. We continue this procedure until we reach the $(i-1)$th layer, the resultant matrix is $\mathbf{P^-}$, which relates $\mathbf{N}_{1s}$ to $\mathbf{N}_{\mathrm{is}}$, where $\mathbf{N}_{\mathrm{is}}$ are the top surface unknowns of layer i. Similarly, we cascade from layer $i+1$ to layer N to obtain matrix $\mathbf{P}^+$. Clearly the computational cost is

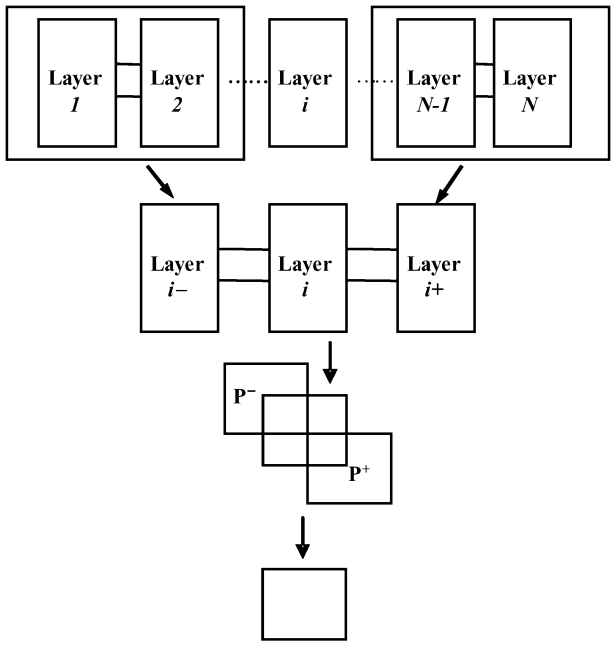


Fig. 6. Reduction of the 2-D layered system to a single-layer one.

the number of cascading multiplied by the time cost in each cascading. The former is equal to the number of layers minus one, while the latter is the cost of a single-layer unknown elimination. Therefore if one is capable of solving one-layer matrix, he is capable of solving all the N layers. However, due to the large number of layers, this implementation could be slow despite the high capacity it can achieve. A fast version can be obtained by parallel implementation.

Parallel Implementation: Assuming there are L layers, we divide them into L/2 groups. Each group consists of adjacent two layers. Each group has no overlapping layers. We then assign these L/2 groups to L/2 machines, each of which cascades two matrices. Thus we obtain L/2 matrices simultaneously. We then subdivide these matrices into L/4 groups, each of them again contains a pair of matrices without overlapping. We then assign them to L/4 machines, which return L/4 cascaded matrices simultaneously. We repeat this procedure until we have only two matrices left. Clearly, with the above parallel implementation, the computational cost is only $\log_2 L$ multiplied by a single-layer unknown elimination cost. This performance can be further improved by exploring more advanced parallel algorithms.

VI. An Algorithm of Logarithmic Complexity for Further Speed-Up

If the on-chip structure is periodic with L layers, $\log_2 L$ complexity can be achieved even with a single machine. First, we use the approaches in Sections IV and V to form the matrix for one period. We then cascade two of this matrix to form the matrix for two periods. We then cascade two of the two-period matrix to form the matrix for four periods. We continue this procedure until we reach the length of the structure. Clearly, for a structure which has L layers, by using the aforementioned approach, one only needs to cascade $\log_2 L$ times to reach the required length. This algorithm of logarithmic complexity drastically speeds up the analysis of periodic on-chip structures. This

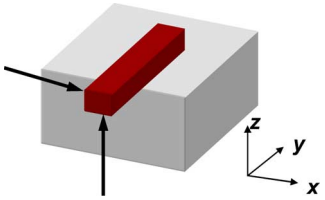


Fig. 7. Excitation and Extraction.

algorithm is also used to handle multiple layers resulted from the discretization of thick silicon and conductors. For instance, the thick silicon substrate often constitutes a numerical challenge to a partial-differential-equation based solver because of the large number of volume unknowns resulted from its discretization. With this technique, it does not constitute a challenge any more because one can account for its contribution to the rest of the system in $\log_2 L$ operations and single-layer storage. This algorithm of logarithmic complexity, to certain extent, resembles the one used for treating deep cavities with a constant cross section [18].

VII. Excitation and Extraction

Here, we give a simple example to illustrate the excitation and extraction scheme. Consider a wire sitting above a ground plane as shown in Fig. 7. If the layer-growth direction is chosen to be x, we use a current source orientated in either y or z direction; If the layer-growth direction is chosen to be y, we use a current source orientated in either x or z direction; if the layer-growth direction is chosen to be z, we use a current source located in x-y plane. The purpose is to associate field unknowns involved in the excitation and extraction to those remaining in the final matrix system. For all the other unknowns, their corresponding right-hand sides are zero. Therefore, the matrix reduction process illustrated in Sections IV and V does not involve the modification of the right-hand side at all, which is efficient. Multiple columns of current filaments can be used from the wire to the ground. Multiple rows can also be used. But they are all placed in the layers or layer of interest. The right-hand sides corresponding to the field unknowns associated with the current filaments become

$$b_i = -jk_0 Z_0 Id \tag{16}$$

in which I is the current and d is the length of the current filament. When we inject current into one port, we leave other ports open. We then sample the voltage generated at each port. The voltage can be evaluated by performing a line integral of the electric field from the port to the ground. Thus, we obtain one column of the impedance matrix $\mathbf{Z}$. We then inject current into another port. We can obtain another column of $\mathbf{Z}$ matrix. We continue this procedure by injecting current into each port in turn. Finally, we obtain the entire $\mathbf{Z}$ matrix. From the $\mathbf{Z}$ matrix, one can easily obtain both $\mathbf{S}$- and $\mathbf{Y}$-parameter matrices. It should be noted that, different from the general RLC-based interconnect modeling process in which the extraction stage is separated from the simulation stage, here one can obtain both extraction and simulation results within one run. One can either obtain the S-parameter model of the interconnects as aforementioned, or load the interconnect with current sources, and obtain voltages directly from the proposed method. When the ground plane is placed far away from the structure of interest, instead of using a probe that goes all the way from the ground to the structure, a short probe that does not start from the ground is used. In doing so, the port current becomes unknown. It is extracted from the port voltages sampled at multiple points. With port currents and voltages known, the Z-, Y-, and S-parameters can be extracted.

VIII. Performance Analysis

The memory usage of the proposed method is modest compared to the conventional finite element method. Maximally, it only requires the storage of a single-layer matrix formed by surface unknowns irrespective of the original problem size. Therefore the proposed method possesses a high capacity to deal with very large scale electromagnetic problems.

The CPU run time can be analyzed for step I and step II in the content of both serial implementation and parallel implementation.

Step I (Reduction of the 3-D Layered System Matrix to a 2-D Layered One): Assuming the number of layers is L and the number of volume unknowns per layer is N_{v1}, the lower bound of the CPU cost for eliminating all the volume unknowns is apparently $O(\mathrm{LN}_{v1})$. However, the proposed method can achieve it in $\mathrm{MO}(N_{v1})$ operations, in which M is the number of structure seed, which is generally much less than L. If implemented in parallel, since the elimination of the volume unknowns in different layers is completely decoupled, each of them can be assigned to a single processor, and no communication is needed between the processors. Therefore, the CPU cost is just $O(N_{v1})$, the cost of a single-layer sparse matrix inversion.

Step II (Reduction of the 2-D layered System Matrix to a Single-layer One): Assuming the number of layers is L, the total number of surface unknowns is N_s, the CPU cost of the serial implementation can be estimated as

$$\mathrm{Time} \sim L \times f\left(\frac{N_s}{L}\right) \tag{17}$$

in which $f(N_s/L)$ is the cost of a single-layer surface-unknown elimination. The function f depends on the computational complexity of the matrix solver used to solve the single layer matrix. For instance, if an advanced sparse matrix solver is used, f can be a linear function

$$f(x) \sim x.$$

If an iterative solver such as the conjugate gradient method is used

$$f(x) \sim x^2.$$

If a direct solver such as the LU decomposition is used

$$f(x) \sim x^3.$$

In contrast, the CPU cost of a conventional method is

$$\text{Time} \sim f(N_s) \tag{18}$$

assuming it uses the same matrix solver as used in the proposed method for a fair comparison. Therefore, it can be seen clearly from (17) and (18) that the speed up of step II is determined by the speed of the conventional method. The slower the conventional method is, the faster the proposed method is. For instance, if the conventional method uses LU decomposition to solve a matrix, the speed-up of step II is L^2. If the conventional method uses an advanced sparse matrix solver that scales linearly with the number of unknowns, then the serial implementation of step II wouldn't save any CPU time. However, sparse solvers generally cannot scale linearly especially when the number of unknowns is large. Therefore, one can still gain better efficiency by implementing step II serially. Furthermore, when multiple right-hand sides exist, due to the reduced size of the final system matrix, the forward and backward substitution time is much less compared to a conventional direct solver. In addition, one benefits from the modest memory usage of the proposed method. The CPU cost of the parallel implementation of step II is $log\ _2 L$ multiplied by a single-layer elimination cost

$$\text{Time} \sim \log_2 L \times f\left(\frac{N_s}{L}\right). \tag{19}$$

This performance can be further improved by exploring more advanced parallel algorithms. Therefore, compared to serial implementation, the parallel implementation drastically speeds up the analysis. It should be noted that the numerical procedure of the proposed method facilitates the parallel implementation because of the decoupled nature of subproblems, and hence zeroing the communication between processors.

IX. Examples

To validate the proposed method, we simulated a set of interconnect structures that were fabricated on a test chip using conventional Si processing techniques [1]. High resolution cross-sectional scanning electron microscopy and optical microscopy were used to measure the relevant dimensions of the fabricated structures. Parasitics signals were removed from the measured S-parameters using a de-embedding approach [1].

The first test structure is of 300 mu width. It involves a 10-mu wide strip in metal 2(M2) layer, one ground plane in metal 1 (M1) layer, and one ground plane in metal 3 (M3) layer. The distance of this strip to the M2 returns at the left- and right-hand sides are 50 mu, respectively. The strip is of a length of 2000 mu. The discretization is done in x-z plane, and extruded along the y direction. The 2000 mu length is subdivided into 40 layers. The 2-D surface matrix is only formed for one layer. The algorithm

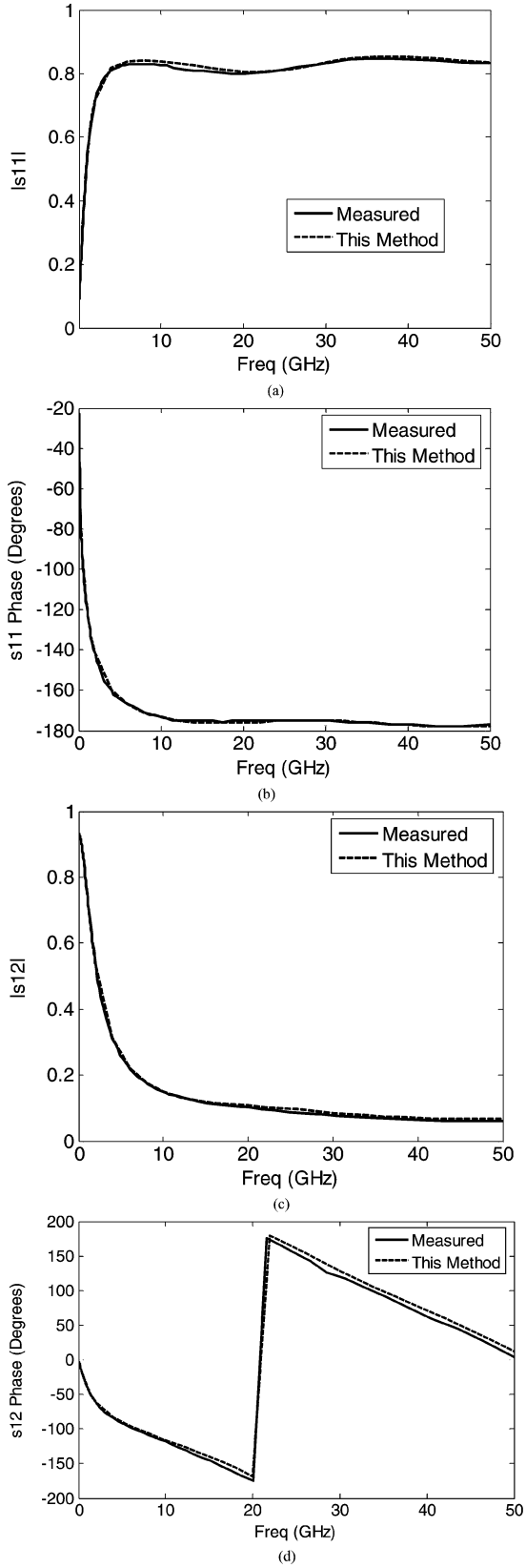


Fig. 8. S-parameters of an on-chip interconnect structure of length 2000 mu. (a). S11 magnitude. (b) S11 phase. (c). S12 magnitude. (d) S12 phase.

of logarithmic complexity stated in Section VI is then used to cascade all the layers to form one matrix that only involves the surface unknowns on the near end and the far end planes. The vertical current filament is placed between the ground and the strip at the near end and the far end to excite the structure and extract the circuit parameters. The back plane in M1 layer and the top plane in M3 layer are both discretized to model the conductor loss accurately. The simulated S-parameters are shown in Fig. 8 in comparison with the measured data. Excellent agreement is observed from dc to 50 GHz.

The second structure is a crosstalk structure. Two wires are placed in the center of M2 layer. One is of 1.1 mu width, and the other is of 2.07 mu width. The spacing between these two wires is 2.0 mu. The distance to M2 returns at the left- and right-hand sides is 10.1 mu. The solid planes in M1 and M3 layers in the first test structure are replaced by 146 parallel returns, respectively. These returns are 1.05-mu wide, and 1-mu apart. They are shorted to the ground at the near and the far end. The structure is 2000 mu long. Like the first structure, only one structure seed is involved. Hence, the 2-D surface matrix is formed for one layer and cascaded by using the algorithm of logarithmic complexity to form the final system matrix that only involves surface unknowns at the near end and far end. In Fig. 9(b), we compare measured and simulated crosstalk. Clearly the agreement is good in both magnitude and phase. The current distribution at 2 GHz at the near end of the structure is depicted in Fig. 9(a).

The third structure again is a crosstalk structure. However, the M1 and M3 metal layers are populated by orthogonal returns. These returns are 1-mu wide each and 1-mu wide apart. The length of the structure is 2000 mu. The crosstalk is measured between two wires embedded in the M2 layer. One is of 2.1 mu width; the other is of 1.1 mu width. The spacing between these two wires is 1.95 mu. The distance to the M2 returns of both wires is 10.3 mu. The 2-D surface matrices are formed for two layers. One has the orthogonal returns present, while the other does not. These two matrices are then cascaded to form the matrix of a period. The one-period matrix is then cascaded to form the matrix that covers 2000-mu length, and only involves near-end and far-end surface unknowns. The current filaments are placed between the ground and the M2 wires at the near and far end to extract circuit parameters. Fig. 10 shows the comparison of the simulated crosstalk in comparison with the measured data, which reveals an excellent agreement. If one uses a standard finite element method to solve the problem, due to the densely populated orthogonal returns in M1 and M3 layers, one has to solve over 3.043 million unknowns. In contrast, the proposed method solves the same problem rigorously using only 2270 unknowns.

In the aforementioned examples, the layer-growth direction is chosen to be y. In other words, the structure is segmented along y. The layer-growth direction can also be chosen the same as the stack-growth direction. This is useful to accommodate irregular geometries in the x-y plane. We simulated a spiral inductor to demonstrate this capability. The geometry of the spiral inductor is shown in Fig. 11(a). Its diameter (OD) is 1000 mu. The wire is 50 mu wide (w $=$ 50 mu) and 15 mu thick. The port separation, PS, is 50 mu. The inductor is backed by two package planes.

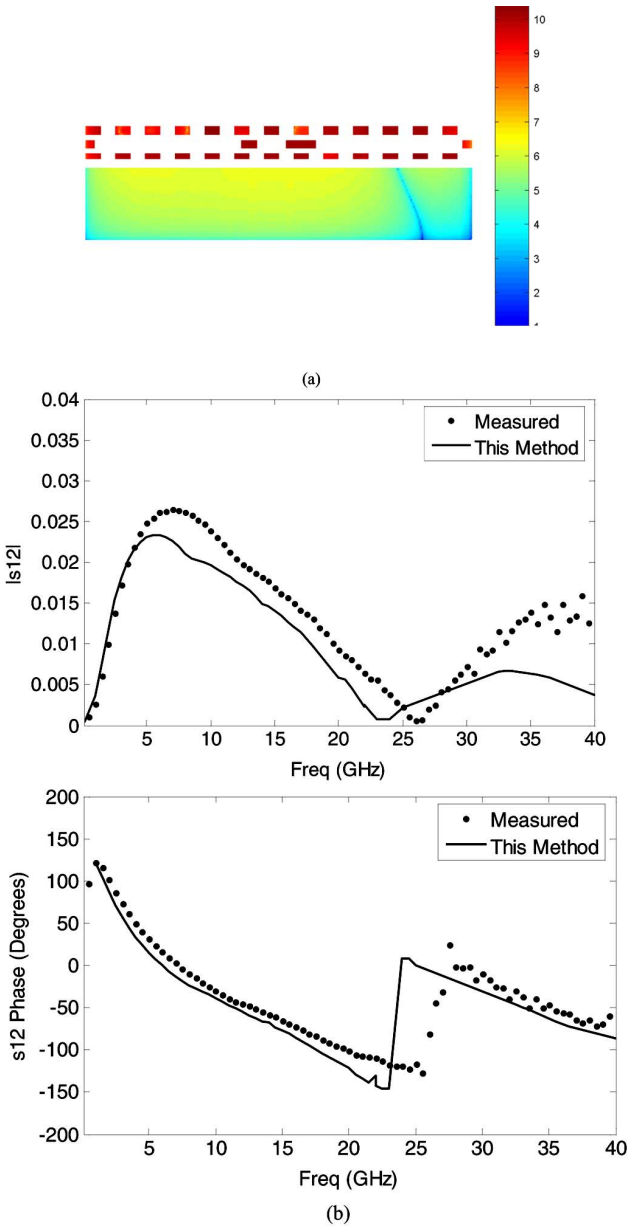


Fig. 9. Simulation of an on-chip interconnect structure with parallel returns. (a) Current distribution at 2 GHz. (b) Cross talk.

The backplane is 15-mu thick. Fig. 11(b) shows the Y parameter and Q value simulated by the proposed method in comparison with those simulated by commercial electromagnetic simulator HFSS. Excellent agreement can be observed. Q value becomes negative because the inductor in fact becomes a capacitor at high frequencies.

Finally, we simulated a large-scale on-chip pentium4 M2-M8 power grid structure as shown in Fig. 12(a). The grid is 10 000 mu long and 200 mu wide. Before simulating this example, we tested the accuracy of this method in power grid analysis by comparing its IR drop results at dc against a resistance (R)-based IR drop analysis. Since R-based analysis generates a large number of resistances that are beyond the capability of a conventional SPICE-type circuit simulator, a 200 mu $\times$ 400 mu block was sampled from the large-scale

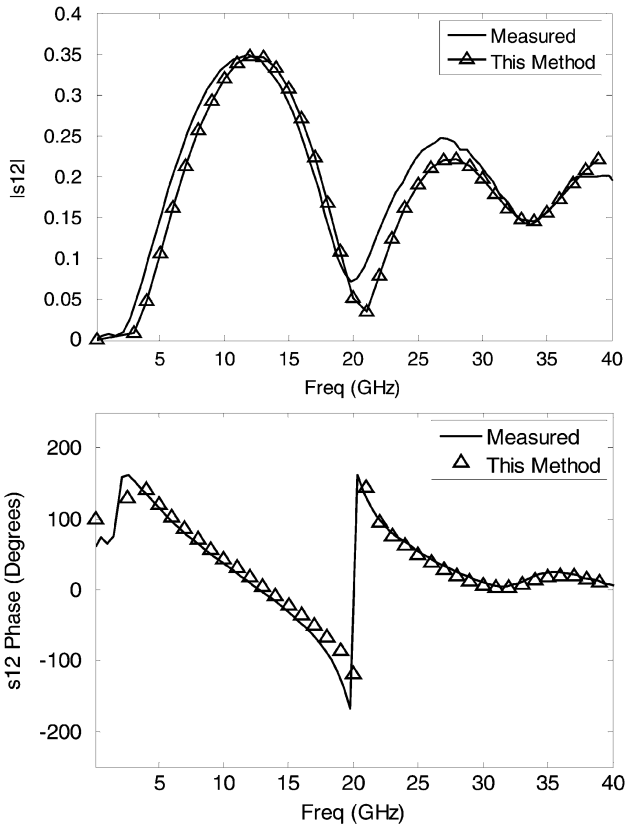


Fig. 10. Crosstalk of a 3-D interconnect structure with orthogonal returns.

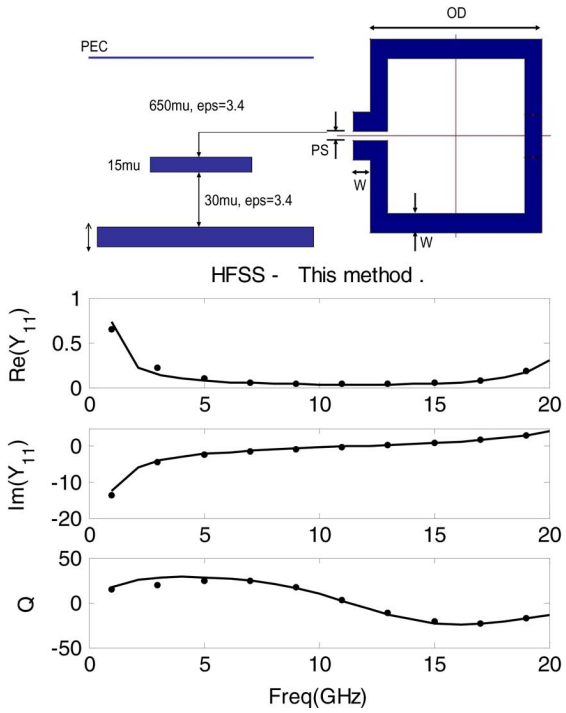


Fig. 11. Simulation of a RF spiral inductor. (a) Geometry. (b) Y-parameters and Q value.

power grid for the purpose of validation. Eight C4 bumps are landing at M8 wide metals, and 6-pair current sources are attached at the bottom metal layer M4. Fig. 12(b) shows

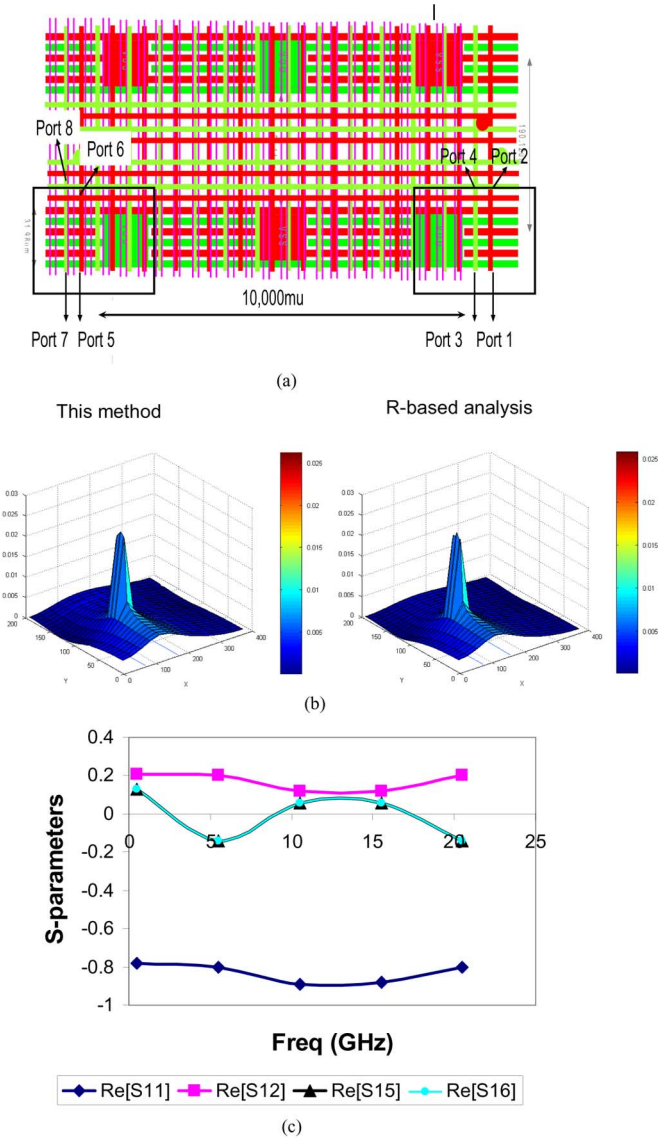


Fig. 12. Simulation of an on-die power grid. (a) Geometry. (b) VSS voltage droop at dc. (c) S-parameters.

the VSS voltage droop simulated by the proposed method in comparison with those obtained by R-based analysis. Excellent agreement can be observed. We then performed the dynamic analysis of the entire structure. Eight ports were sampled on the grid as shown in Fig. 12(a). Fig. 12(c) shows the calculated S parameters. Despite the large number of unknowns, the peak memory usage is only 738-Mbytes for this example.

X. Conclusion

In this paper, we proposed a layered finite element method for high-frequency modeling of large-scale three-dimensional on-chip circuit structures. This method is capable of solving an orders-of-magnitude smaller system to rigorously obtain the solution of the original big problem. The system matrix of the original 3-D problem is first reduced to that of 2-D layers. For on-chip interconnect structures, the computational cost of this reduction is modest, only involving the solution of a few 2-D structure seeds. The matrix system of 2-D layers is then further

reduced to that of a single layer. This reduction only involves single-layer unknowns irrespective of the original problem size. As a result, the proposed method possesses a high capacity to solve large-scale interconnect problems. Equally important, the entire procedure is numerically rigorous without making any theoretical approximation. In addition, it solves Maxwell's coupled E-H equations, and hence features uncompromised electromagnetic accuracy. Its accuracy and capacity are demonstrated by numerical and experimental results.

ACKNOWLEDGMENT

The authors would like to thank M. J. Kobrinsky at Intel Corporation for providing measured data, J. He (Intel) and R. Chao (then Intel, now Taiwan National Chiao Tung University) for providing HFSS data.

REFERENCES

[1] M. J. Kobrinsky, S. Chakravarty, D. Jiao, M. C. Harmes, S. List, and M. Mazumder, "Experimental validation of crosstalk simulations for on-chip interconnects using S-parameters," *IEEE Trans. Advanced Packaging*, vol. 28, no. 1, pp. 57–62, Feb. 2005, [see also *IEEE Trans. Compon., Packaging and Manufacturing Technol.*, Part B: Advanced Packaging].

[2] D. Jiao, M. Mazumder, S. Chakravarty, C. Dai, M. J. Kobrinsky, M. C. Harmes, and S. List, "A novel technique for full-wave modeling of large-scale three-dimensional high-speed on/off-chip interconnect structures," in *Proc. Int. Conf. Simulation of Semiconductor Processes and Devices*, Sep. 3–5, 2003, pp. 39–42.

[3] D. Jiao, C. Dai, S.-W. Lee, T. R. Arabi, and G. Taylor, "Computational electromagnetics for high-frequency IC design," in *Proc. IEEE Int. Symp. Antennas and Propagation* Transl.:invited paper, 2004, pp. 3317–3320.

[4] D. Gope and V. Jandhyala, "PILOT: A fast algorithm for enhanced 3-D parasitic extraction efficiency," in *Proc. IEEE 12th Topical Meeting on Electrical Performance of Electronic Packaging (EPEP)*, 2003, pp. 337–340.

[5] Z. H. Zhu, B. Song, and J. K. White, "Algorithm in fastimp: A fast and wideband impedance extraction program for complicated 3-D geometries ," in *Proc. 40th ACM/IEEE Design Automation Conf.*, 2003, pp. 712–717.

[6] A. C. Cangellaris, "Towards full-chip Analysis with Electromagnetic Accuracy," presented at the IEEE 12th Topical Meeting on Electrical Performance of Electronic Packaging (EPEP), 2003.

[7] L. Daiel, A. Sangiovanni-Vincentelli, and J. White, "Using conduction modes basis functions for efficient electromagnetic analysis of on-chip and off-chip interconnects," *Proc. DAC*, pp. 563–566, 2001.

[8] A. Rong, A. C. Cangellaris, and L. Dong, "Comprehensive broadband electromagnetic modeling of on-chip interconnects with a surface discretization-based generalized PEEC model," in *Proc. IEEE 12th Topical Meeting on Electrical Performance of Electronic Packaging (EPEP)*, 2003, pp. 367–370.

[9] A. E. Yilmaz, J. M. Jin, and E. Michielssen, "A parallel FFT-accelerated transient field-circuit simulator," *IEEE Trans. MTT*, vol. 53, pp. 2851–2865, Sep. 2005.

[10] S. Kapur and D. E. Long, "Large-scale full-wave simulation ," *DAC*, pp. 806–809, 2004.

[11] W. C. Chew, "Toward a more robust and accurate fast integral solver for microchip applications," in *Proc. IEEE 12th Topical Meeting on Electrical Performance of Electronic Packaging (EPEP)*, 2003, p. 333.

[12] D. Lukashevich, A. C. Cangellaris, and P. Russer, "Broadband electromagnetic analysis of interconnects by means of TLM and Krylov model order reduction," in *Proc. IEEE 14th Topical Meeting on Electrical Performance of Electronic Packaging*, 2005, pp. 355–358.

[13] Z. G. Qian, J. Xiong, L. Sun, I. T. Chiang, W. C. Chew, L. J. Jiang, and Y. H. Chu, "Crosstalk analysis by fast computational algorithms," in *Proc. IEEE 14th Topical Meeting on Electrical Performance of Electronic Packaging*, 2005, pp. 367–370.

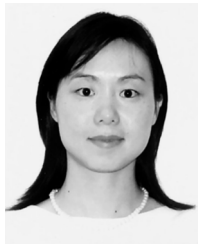
[14] J. M. Jin, *The Finite Element Method in Electromagnetics*, 1st ed. New York: Wiley, 1993.

[15] R. D. Graglia, D. R. Wilton, A. F. Peterson, and I. Gheorma, "Higher order interpolatory vector bases on prism elements," *IEEE Trans. Antennas Propag.*, vol. 46, no. 3, pp. 442–450, Mar. 1998.

[16] D. Jiao, S. Chakravarty, C. Dai, and S. W. Lee, "Surface-based finite element method for large-scale 3-D circuit modeling," in *Proc. 14th Topical Meeting on Electrical Performance of Electronic Packaging*, Oct. 24–26, 2005, pp. 347–350.

[17] C. S. Desai and J. F. Abel, *Introduction to the Finite Element Method*. New York: Van Nostrand Reinhold, 1972.

[18] J. Liu and J. M. Jin, "A special higher-order finite element method for scattering by deep cavities," *IEEE Trans. Antennas Propog.*, vol. 48, pp. 694–703, May 2000.



Dan Jiao (S'00–M'02–SM'06) received the Ph.D. degree in electrical engineering from the University of Illinois at Urbana-Champaign, in October 2001.

From 2001 to September 2005, she was a Senior CAD Engineer, Staff Engineer, and Senior Staff Engineer in the Technology CAD Division at Intel Corporation, Santa Clara, CA. In September 2005, she joined Purdue University, West Lafayette, IN, as an Assistant Professor in the School of Electrical and Computer Engineering. She has authored two book chapters and over 60 papers in refereed journals and international conferences. Her current research interests include high frequency digital, analogue, mixed-signal, and RF IC design and analysis, high-performance VLSI CAD, modeling of micro- and nano-scale circuits, computational electromagnetics, applied electromagnetics, fast and high-capacity numerical methods, fast time domain analysis, scattering and antenna analysis, RF, microwave, and millimeter wave circuits, wireless communication, and bio-electromagnetics.

Dr. Jiao received the 2006 Jack and Cathie Kozik Faculty Start-up Award, which recognizes an outstanding new faculty member in Purdue ECE. In 2004, she received the Best Paper Award from Intel's annual corporate-wide technology conference (Design and Test Technology Conference) for her work on generic broadband model of high-speed circuits. In 2003, she won the Intel Logic Technology Development (LTD) Divisional Achievement Award in recognition of her work on the industry-leading BroadSpice modeling/simulation capability for designing high-speed microprocessors, packages, and circuit boards. She was also awarded the Intel Technology CAD Divisional Achievement Award for the development of innovative full-wave solvers for high frequency IC design. In 2002, she was awarded by Intel Components Research the Intel Hero Award (Intel-wide she was the tenth recipient) for the timely and accurate two- and three- dimensional full-wave simulations. She also won the Intel LTD Team Quality Award for her outstanding contribution to the development of the measurement capability and simulation tools for high frequency on-chip cross-talk. She was the winner of the 2000 Raj Mittra Outstanding Research Award given her by the University of Illinois at Urbana-Champaign. She has served as a reviewer for many IEEE journals and conferences.

Sourav Chakravarty received the B.Tech. degree in electronics and telecommunication from the Regional Engineering College, Kurukshetra, India, in 1992, the M.E. degree in telecommunications from Jadavpur University, Calcutta, India, in 1997, and the Ph.D. degree in electrical engineering from the Pennsylvania State University, University Park, in 2001.

From 1992 to 1995, he was a Senior Antenna Design Engineer at Superline Microwave Pvt. Ltd., in Bangalore, India. He worked as a Research Assistant in the Electromagnetic Communication Laboratory, Pennsylvania State University, from 1997 to 2001. He is currently a Staff CAD Engineer at Intel Corporation, Hillsboro, OR. His research interests include computational electromagnetics with emphasis on probabilistic optimization techniques and the applications of MoM and FDTD techniques to predict delay and crosstalk in interconnects.

Changhong Dai received the B.S. degree in physics from Beijing University, Beijing, China, in 1985, and the M.S. and Ph.D. degrees in materials sciences and engineering from Stanford University, Stanford, CA, in 1992 and 1995, respectively.

He joined the Technology CAD Division of Intel Corporation, Santa Clara, CA, in 1995, as a Senior CAD Engineer. Since 1995, he has been an R&D Engineer or Manager for model and CAD tool development for circuit analysis and physical design, in the areas of circuit reliability, transistor modeling for circuit simulation, interconnect modeling for parasitic extraction, static and full wave simulation of interconnects and RF devices, and power and power delivery modeling for IC product design. He is currently a Director with the Technology and Manufacturing Group of Intel Corporation, with the responsibility of directing the development of Core CAD Technologies that enables Intel processing technology and product design. His current responsibility covers the full spectrum of CAD tool and infrastructure development with a focus of bridging the processing technology and chip design.