SuPREME: Substrate and Power-delivery Reluctance-Enhanced Macromodel Evaluation

Tsung-Hao Chen, Clement Luk
Electrical and Computer Engineering
University of Wisconsin-Madison
{ tchen, lukc }@cae.wisc.edu

Charlie Chung-Ping Chen
Electrical Engineering
National Taiwan University
cchen@cc.ee.ntu.edu.tw

ABSTRACT

The recent demand for system-on-chip RF mixed-signal design and aggressive supply-voltage reduction require chip-level accurate analysis of both the substrate and power delivery systems. Together with the rising frequency, low-k dielectric, copper interconnects, and high conductivity substrate, the inductance effects raised serious concern recently. However, the increasing design complexity creates tremendous challenges for chip-level power-delivery substrate co-analysis. In this paper, we propose a novel and efficient reluctance-based passive model order reduction technique to serve these tasks. Our work, SuPREME(Substrate and Power-delivery Reluctance-Enhanced Macromodel Evaluation) not only greatly reduces the computational complexity of previous reluctance-based model order algorithms but is also capable of handling large number of noise sources efficiently. To facilitate the analysis of inductive substrate return paths and evaluate the high-frequency substrate coupling effects, we derive a novel RLKC substrate model from Maxwell's equations for the first time. Experimental results demonstrate the superior runtime and accuracy of SuPREME compared to the traditional MNA-based simulation.

1. INTRODUCTION

The strong need for cost reduction drives the demands for System-on-a-Chip (SoC), which frequently requires the integration the RF analog circuits with digital circuits. The supply voltages have been aggressively scaled down because of the demand for power reduction in mobile computation and the need to decrease heat-dissipation. As a result, the power delivery and substrate noise margin budgets have shrunk to less than 100 mV. Both trends require high-quality power-delivery and substrate design to avoid digital-analog noise coupling nightmare. As a result, extensive power and substrate simulations are required to ensure power-delivery quality.[1][2]

Unfortunately, the rising clock frequency for both analog and digital circuits and the adoption of low-k and high conductivity interconnects and high conductivity substrates require consideration of both self and mutual inductances. The long range inductive coupling effect makes the already difficult analysis even worse. Furthermore, substrate also serves as return path and have been ignored for many interconnect inductance analysis. With the increasing coupling between power-delivery and substrate, power-delivery substrate co-analysis will be crucial in the near future.[3]

One of the co-analysis difficulty is that most of the existing substrate models consider only resistanceor at most capacitance. The inductance substrate models are required for performing return path analysis. The lossy silicon effect has been addressed in some previous works[4][5][6]. Most of the substrate and power-delivery models may not compatible with each other because of the problem nature. Due to the regular structure, finite difference, finite element, or boundary elements models are often applied to substrate but not to power-delivery problems. As a result, it is necessary to build a consistent model, which can facilitate power-delivery substrate co-analysis.

Due to its compatibility with circuit simulators, the PEEC model[7] introduced the concept of partial inductance into the VLSI area and has been wildly used for interconnects inductance modeling. However, the partial inductance assumes the return path is at infinity, and thus magnetic coupling between faraway conductors are not negligible. This long-range feature enforces the system equation to embrace a dense matrix to express all partial inductance couplings. This results in a high complexity of inductance extraction and simulation. Together with the fact that direct truncation of the inductance matrix could result in unstable system models, several sparsification techniques have been proposed to reduce the density of partial inductance matrices and preserve stability. For example, shift-and-truncate method[8], Halo method[9], and the block diagonal method[10] all provided strategies to sparsify inductance matrices.

Recently, the concept of reluctance (inverse partial inductance matrix) has been revealed, and is looming as a trend and an alternative way for solving inductance problems based on the following reasons. Since reluctance demonstrates high locality and shielding effect similar to capacitance, it is better to sparsify reluctance matrices than inductance matrices. It was proven that the reluctance matrix is diagonally dominant and positive definite for equal-length buses. This guarantees that the negative off-diagonal elements can be safely deleted without sacrificing stability[11]. Later, with a window-selection strategy and a bisection subroutine, the window-based reluctance extraction was shown to be efficient, accurate, and stable for magnetic effect analysis[12].

To lessen the computation complexity, model order reduction techniques have widely been considered as another substitute of SPICE simulation. After several years of research efforts, the model order reduction techniques such as PVL[13] and PRIMA[14] have been successfully extended to consider inductance effects. With the success of the reluctance technique, model reduction society starts to incorporate reluctance into the model order reduction frame-

ICCAD'03, November 11-13, 2003, San Jose, California, USA.

work. Although the reluctance matrix has a strong benefit of sparsity, the explicit inversion of it may still be dense. This creates difficulty for the MNA(Modified Nodal Analysis) model reduction type framework such as PRIMA. Recently, by introducing the inverse of the inductance matrix, ENOR[15] provides an elegant method that incorporates the inverse inductance matrix in the NA(Nodal Analysis) matrix. Later, SMOR[16] enhanced some numerical accuracy of ENOR while solving the admittance matrix which is $(s_0\mathbf{C}+\mathbf{G}+\frac{\mathbf{\Gamma}}{s_0})$, where $\mathbf{\Gamma}=\mathbf{A}_l\mathcal{K}\mathbf{A}_l^T$ and s_0 is the specific frequency.

The drawback for the NA type of analysis is that the sparsity may be destroyed by the projection of reluctance matrix ($\mathcal{K}$) to the incident matrix ($\mathbf{A}_l$) space and the summation to the $\mathbf{G}$ and $\mathbf{C}$ matrices. As a result, the efficiency of matrix solvers may be significantly degraded. Furthermore, the need to select an expanding frequency s_0 can make the procedure complicated; there is no explicit guide line for making the selection. In the case when complex frequency expansion point is needed, the complexity will get much worse.

In this paper, we resolve the above issues in two parts. First, we propose a new reluctance-enhanced model order reduction analysis techniques, SuPREME (SUbstrate and Power-delivery Reluctance-Enhanced Macromodel Evaluation), which still perform model order reduction in the MNA framework. However, by using implicit inversion of $\mathcal{K}$, we intelligently avoid the explicit formulation of $\mathcal{K}^{-1}$ and hence preserve the performance and sparsity. As a result, SuPREME is fully compatible with PRIMA and the selection of s_0 can be avoided. We also developed an efficient passive reluctance-based multipoint expansion procedure. The accuracy of SuPREME is exactly same as that of PRIMA due to compatibility between them.

Second, to enable efficient inductive substrate analysis, we derive the RLKC substrate models from the Maxwell's equations. The substrate model is fully compatible with general circuit simulators and model order reduction algorithms as well. As a result, power-delivery and substrate can be easily simulated together by efficient model order reduction techniques such as SuPREME.

The experimental results show that SuPREME is very efficient. It demonstrates over 50X runtime improvement over traditional MNA time domain simulation. The frequency response of our RLKC substrate model also shows that high frequency inductive effects need to be taken into consideration in the near future.

2. MODEL ORDER REDUCTION METHOD WITH RELUCTANCE

A linear circuit containing RLC elements can be represented as the following set of Modified Nodal Analysis (MNA) circuit equations:

$$\begin{bmatrix}\mathbf{G} & \mathbf{A}_l^T\\ -\mathbf{A}_l & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{v}_n\\ \mathbf{i}_l\end{bmatrix}+\begin{bmatrix}\mathbf{C} & \mathbf{0}\\ \mathbf{0} & \mathcal{L}\end{bmatrix}\frac{d}{dt}\begin{bmatrix}\mathbf{v}_n\\ \mathbf{i}_l\end{bmatrix}=\begin{bmatrix}-\mathbf{A}_i\mathbf{I}_s\\ \mathbf{0}\end{bmatrix}, \quad (1)$$

where $\mathbf{G}$ and $\mathbf{C}$ are the conductance and capacitance matrices respectively, $\mathbf{A}_l$ and $\mathbf{A}_i$ represent the adjacency matrices of inductors and independent current sources, $\mathbf{v}_n$ and $\mathbf{i}_l$ denote vectors of nodal voltages and inductance current variables, and $\mathcal{L}$ is the inductance matrix containing self and mutual inductance information.

Model order-reduction methods [14][17][18] generate an analytic macro model, say macro-model, which is a compact description of original circuits by matching their moments or poles. Equation (1) can be written in Laplace domain:

$$\begin{bmatrix}\mathbf{G} & \mathbf{A}_l^T\\ -\mathbf{A}_l & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{v}_n\\ \mathbf{i}_l\end{bmatrix}+s\begin{bmatrix}\mathbf{C} & \mathbf{0}\\ \mathbf{0} & \mathcal{L}\end{bmatrix}\begin{bmatrix}\mathbf{v}_n\\ \mathbf{i}_l\end{bmatrix}=\begin{bmatrix}-\mathbf{A}_i\mathbf{I}_s\\ \mathbf{0}\end{bmatrix}. \quad (2)$$

To illustrate the idea of moment matching, we expand both sides of the Equation (2) in a Taylor series around frequency $s=0$. Rearranging the terms, we get

$$\left(\begin{bmatrix}\mathbf{G} & \mathbf{A}_l^T\\ -\mathbf{A}_l & \mathbf{0}\end{bmatrix}+s\begin{bmatrix}\mathbf{C} & \mathbf{0}\\ \mathbf{0} & \mathcal{L}\end{bmatrix}\right)\begin{bmatrix}m_0^v+m_1^v s+m_2^v s^2+\cdots\\ m_0^i+m_1^i s+m_2^i s^2+\cdots\end{bmatrix}=\begin{bmatrix}-\mathbf{A}_i(u_0+u_1 s+u_2 s^2+\cdots)\\ \mathbf{0}\end{bmatrix}, \quad (3)$$

where m_k^v m_k^i and u_k are the coefficients of the k^{th} term in the Taylor series, which are also known as the k^{th} moments of $\mathbf{v}_n$, $\mathbf{i}_l$, and $\mathbf{I}_s$ respectively. The basic idea of moment matching is to calculate finite number of moments in the left-hand-side in terms of the known moments in the right-hand-side, and use the obtained moments to approximate the whole frequency-domain spectrum of a circuit. In PRIMA, a special case of the above equation, sources are assumed to be impulse functions attached to ports to preserve the I/O transfer characteristics, and hence only u_0 is present in the right-hand-side of (3).

The voltage and current moments can be calculated by solving the following procedure:

$$\begin{bmatrix}\mathbf{G} & \mathbf{A}_l^T\\ -\mathbf{A}_l & \mathbf{0}\end{bmatrix}\begin{bmatrix}m_0^v\\ m_0^i\end{bmatrix}=\begin{bmatrix}-\mathbf{A}_i u_0\\ \mathbf{0}\end{bmatrix};$$
$$\begin{bmatrix}\mathbf{G} & \mathbf{A}_l^T\\ -\mathbf{A}_l & \mathbf{0}\end{bmatrix}\begin{bmatrix}m_k^v\\ m_k^i\end{bmatrix}=\begin{bmatrix}-\mathbf{A}_i u_k\\ \mathbf{0}\end{bmatrix}-\begin{bmatrix}\mathbf{C} & \mathbf{0}\\ \mathbf{0} & \mathcal{L}\end{bmatrix}\begin{bmatrix}m_{k-1}^v\\ m_{k-1}^i\end{bmatrix}. \quad (4)$$

In order to avoid numerical errors, an orthonormalization process is often used to span the same subspace as spanned by the finite moments. Using the orthogonal bases $\mathbf{V}$, which is a Krylov subspace, as a projection matrix and perform a congruent transformation, the original system equation (1) can be reduced to a small-dimentional one:

$$\widetilde{G}\widetilde{x}+\widetilde{C}\frac{d}{dt}\widetilde{x}=\widetilde{b}, \quad (5)$$

where

$$\widetilde{G}=\mathbf{V}^T\begin{bmatrix}\mathbf{G} & \mathbf{A}_l^T\\ -\mathbf{A}_l & \mathbf{0}\end{bmatrix}\mathbf{V},\ \widetilde{C}=\mathbf{V}^T\begin{bmatrix}\mathbf{C} & \mathbf{0}\\ \mathbf{0} & \mathcal{L}\end{bmatrix}\mathbf{V},\ \widetilde{b}=\mathbf{V}^T\begin{bmatrix}-\mathbf{A}_i\mathbf{I}_s\\ \mathbf{0}\end{bmatrix}.$$

Since the dimension of Equation (5) is very small, the time-domain simulation for this equation is not crucial to the total runtime. The major effort of order reduction methods is to solve (4) and construct the projection basis. There are two reasons that order reduction method is more efficient than the MNA time-domain solution. First, procedure (4) only has to factorize the conductance matrix, while the MNA has to decompose the summation of the conductance and susceptance matrices, which is much denser than the former case and introduces more fill-ins that kill the performance. Second, (4) only has to perform backward and forward substitutions for a few times, which the MNA has to do that for every time-step with the denser lower and upper triangular matrices.

About the moments for sources, u_k, in (3), it can be calculated by the Laplace transform and some algebraic operations. It was also proven that finite time piece-wise-linear (PWL) sources have zero negative-order moments and no moment shifting is needed[18].

2.1 Handling Reluctance

Reluctance matrix $\mathcal{K}$ is the inverse of the partial inductance matrix $\mathcal{L}$. It was shown that reluctance matrix has better locality and can be extracted by the windowing technique [11][12][19]. In contrast to the dense inductance matrix $\mathcal{L}$, the reluctance matrix $\mathcal{K}$ can be very sparse and captures the magnetic coupling effect accurately. In the following discussion, the symbol $\mathcal{K}$ represents the sparse reluctance matrix, not the actual inverse of the dense partial inductance matrix. In stead of inverting the dense matrix $\mathcal{L}$, which is very time consuming, the sparse reluctance matrix $\mathcal{K}$ can be extracted by the algorithm presented in [12]. In case the reluctance matrix is known instead of the inductance matrix, the system equation (2) can be rewritten into

$$\begin{bmatrix} \mathbf{G} & \mathbf{A}_l^T \\ -\mathbf{A}_l & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{v}_n \\ \mathbf{i}_l \end{bmatrix} + s \begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathcal{K}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{v}_n \\ \mathbf{i}_l \end{bmatrix} = \begin{bmatrix} -\mathbf{A}_i \mathbf{I}_s \\ \mathbf{0} \end{bmatrix} \quad . \quad (6)$$

Since the only change in this equation is that replacing $\mathcal{L}$ with $\mathcal{K}^{-1}$, we can easily rewrite the moment-calculation iteration in (4) as follows:

$$\begin{bmatrix} \mathbf{G} & \mathbf{A}_l^T \\ -\mathbf{A}_l & \mathbf{0} \end{bmatrix} \begin{bmatrix} m_k^v \\ m_k^i \end{bmatrix} = \begin{bmatrix} -\mathbf{A}_i u_k \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathcal{K}^{-1} \end{bmatrix} \begin{bmatrix} m_{k-1}^v \\ m_{k-1}^i \end{bmatrix} . \quad (7)$$

The LHS of this equation remains the same, which means that we are solving the same matrix for both $\mathcal{L}$ and $\mathcal{K}$ approaches. The matrix multiplication in the RHS of (4) can be simply calculated by

$$\begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathcal{L} \end{bmatrix} \begin{bmatrix} m_{k-1}^v \\ m_{k-1}^i \end{bmatrix} = \begin{bmatrix} \mathbf{C} m_{k-1}^v \\ \mathcal{L} m_{k-1}^i \end{bmatrix} = \begin{bmatrix} \mathbf{C} m_{k-1}^v \\ \mathcal{K}^{-1} m_{k-1}^i \end{bmatrix} \quad . \quad (8)$$

While the upper part of this vector remains the same, we can obtain the lower part by solving $\mathcal{K}$. The Cholesky decomposition can be applied to solve this matrix since $\mathcal{K}$ is shown to be symmetric and positive definite; the cost of solving this matrix would be low because of its sparsity. Orthonormalizing the moment vectors, we can span the projection matrix $\mathbf{V}$ and perform congruent transformation as in Equation (5). $\widetilde{G}$ and $\widetilde{b}$ remain the same, and $\widetilde{C}$ can be obtained as follows:

$$\widetilde{C} = \mathbf{V}^T \begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathcal{K}^{-1} \end{bmatrix} \mathbf{V} = \begin{bmatrix} V_1^T \mathbf{C} V_1 \\ V_2^T \mathcal{K}^{-1} V_2 \end{bmatrix} \quad , \quad (9)$$

where V_1 and V_2 are the upper and lower parts of $\mathbf{V}$ respectively. Instead of performing a matrix multiplication $x = \mathcal{L}V_2$, we only have to solve the system $\mathcal{K}x = V_2$ to obtain the moments. Since $\mathcal{K}$ is chosen to be sparse, the complexity of solving it is low. The Cholesky-decomposed matrix of $\mathcal{K}$ in previous step can be reused in this step.

Comparison with SNOR and EMOR

By substituting the bottom set of equation in (6) into the top one and eliminating the current variable $\mathbf{i}_l$, ENOR[15] and SMOR[16] are able to run the model order reduction method with reluctance elements. However, both of them have to solve the matrix $(\mathbf{C}s_0 + \mathbf{G} + \frac{\mathbf{\Gamma}}{s_0})$, where $\mathbf{\Gamma} = \mathbf{A}_l \mathcal{K} \mathbf{A}_l^T$ and s_0 is the specific frequency that the Taylor series is expanded around. This matrix is actually harder to solve because of the reasons listed below. Compared to them, the advantages of our proposed method are also listed.

1. The proposed algorithm does not need to solve $(\mathbf{C}s_0 + \mathbf{G} + \frac{\mathbf{\Gamma}}{s_0})$ but it factorizes $\begin{bmatrix} \mathbf{G} & \mathbf{A}_l^T \\ -\mathbf{A}_l & \mathbf{0} \end{bmatrix}$ and $\mathcal{K}$ separately. It is known that the run-time of the matrix factorization is determined by the number of fill-ins, and the number of fill-ins is determined by the degree of connectivity of the matrix. Thus summation of these three matrices, $(\mathbf{C}s_0 + \mathbf{G} + \frac{\mathbf{\Gamma}}{s_0})$, would be much denser than two separated ones, and also much more difficult to solve. Not to mention that matrix $\mathbf{C}$ is not involved in this procedure.

2. The term $\frac{\mathbf{\Gamma}}{s_0}$ has s_0 as the denominator, which means s_0 can be any number but not zero. Zero frequency (DC) is very important. Desired circuit simulation may not contain only AC signals, but some quiet ones. Not matching the moment around 0 results in loss of accuracy for DC and limits its application. Applications such as power-grid and substrate analyses contain a big portion of DC signal. Interconnection such as a bus usually has some bits with transitions and some without. If the frequency-domain information for DC is not accurate, the simulation results for those quiet lines may look noisy although the actual responses are not.

3. Factorization for matrix $\begin{bmatrix} \mathbf{G} & \mathbf{A}_l^T \\ -\mathbf{A}_l & \mathbf{0} \end{bmatrix}$ can also be used to solve the DC solution, which is useful for most of the simulation problems. No extra effort is required.

Table 1 summarizes the proposed algorithm.

The SuPREME algorithm
Input: a circuit equation as (1);
a desired number of moments m;

Find projection matrix:

1 Calculate the moments for input sources u_k, $k = 0 \sim m-1$;

2 Let $\hat{\mathbf{G}} = \begin{bmatrix} \mathbf{G} & \mathbf{A}_l^T \\ -\mathbf{A}_l & \mathbf{0} \end{bmatrix}$, and $\mathbf{b}_k = \begin{bmatrix} -\mathbf{A}_i u_k \\ \mathbf{0} \end{bmatrix}$;

3 Decompose $\hat{\mathbf{G}}$ and $\mathcal{K}$;

4 $\mathbf{m}_0 = \hat{\mathbf{G}}^{-1}\mathbf{b}_0$, $\alpha_0 = \frac{1}{||\mathbf{m}_0||}$, and $\hat{\mathbf{r}}_0 = \mathbf{r}_0 = \alpha_0 \mathbf{m}_0$.

5 For $k = 0 : m-1$

6
$$\mathbf{r}_k = \hat{\mathbf{G}}^{-1} \left(\prod_{j=0}^{k-1} \alpha_j \mathbf{b}_k - \begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathcal{K}^{-1} \end{bmatrix} \mathbf{r}_{k-1} \right) ,$$
where the matrix multiplication can be done using (8). Note that $\mathcal{K}^{-1}\mathbf{r}_{k-1}^i$ is implicitly solved by the factorized LU; we never explicitly perform the inversion.

7
$$\hat{\mathbf{r}}_k = \mathbf{r}_k - \sum_{j=0}^{k-1} (\hat{\mathbf{r}}_j^T \mathbf{r}_k) \hat{\mathbf{r}}_j \quad \leftarrow \text{orthogonalize}$$
$$\alpha_k = \frac{1}{||\hat{\mathbf{r}}_k||}, \ \hat{\mathbf{r}}_k = \alpha_k \hat{\mathbf{r}}_k, \text{ and } \mathbf{r}_k = \alpha_k \mathbf{r}_k \quad \leftarrow \text{normalize}$$

8 End For

9 $\mathbf{V} = \{\hat{\mathbf{r}}_0, \hat{\mathbf{r}}_1, \cdots, \hat{\mathbf{r}}_{m-1}\}$ is the projection matrix.

Reduce the system:

1 Calculate $\widetilde{G} = \mathbf{V}^T \hat{\mathbf{G}} \mathbf{V}$.

2 Calculate $\widetilde{C} = \mathbf{V}^T \begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathcal{K}^{-1} \end{bmatrix} \mathbf{V}$ by (9).
$\mathcal{K}$ is again implicitly solved by the factorized LU.

3 Calculate $\widetilde{B} = \mathbf{V}^T \begin{bmatrix} -\mathbf{A}_i \\ \mathbf{0} \end{bmatrix}$.

3 The reduced system becomes $\widetilde{G}\widetilde{x} + \widetilde{C}\dot{\widetilde{x}} = \widetilde{B}\mathbf{I}_s$.

Table 1: The model order reduction algorithm with reluctance

2.2 Multipoint Expansion

The previous discussion expands the input sources with Taylor series and matches system moments at $s = 0$. This would be adequately accurate for low frequency components of the circuit responses. In case we want to have better accuracy for higher frequencies, we can expand the input sources and responses with Taylor series at $s = s_0$, where s_0 is the desired frequency. Defining a new variable $z = s - s_0$, Equation (3) becomes

$$\left(\begin{bmatrix} \mathbf{G} & \mathbf{A}_l^T \\ -\mathbf{A}_l & \mathbf{0} \end{bmatrix} + s\begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathcal{L} \end{bmatrix}\right)\begin{bmatrix} \hat{m}_0^v + \hat{m}_1^v z + \hat{m}_2^v z^2 + \cdots \\ \hat{m}_0^i + \hat{m}_1^i z + \hat{m}_2^i z^2 + \cdots \end{bmatrix} = \begin{bmatrix} -\mathbf{A}_i(\hat{u}_0 + \hat{u}_1 z + \hat{u}_2 z^2 + \cdots) \\ \mathbf{0} \end{bmatrix}, \quad (10)$$

where $\hat{m}$ and $\hat{u}$ are coefficients of Taylor series of system responses and input sources respectively. Pre-multiplying both sides of (10) by $\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathcal{K} \end{bmatrix}$, we have

$$\left(\begin{bmatrix} \mathbf{G} & \mathbf{A}_l^T \\ -\mathcal{K}\mathbf{A}_l & \mathbf{0} \end{bmatrix} + s\begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}\right)\begin{bmatrix} \hat{m}_0^v + \hat{m}_1^v z + \hat{m}_2^v z^2 + \cdots \\ \hat{m}_0^i + \hat{m}_1^i z + \hat{m}_2^i z^2 + \cdots \end{bmatrix} = \begin{bmatrix} -\mathbf{A}_i(\hat{u}_0 + \hat{u}_1 z + \hat{u}_2 z^2 + \cdots) \\ \mathbf{0} \end{bmatrix}. \quad (11)$$

Substituting $s = z + s_0$ and performing moment-matching process similar to (4), we get the following recurrence relation:

$$\begin{bmatrix} \mathbf{G} + s_0\mathbf{C} & \mathbf{A}_l^T \\ -\mathcal{K}\mathbf{A}_l & s_0\mathbf{I} \end{bmatrix}\begin{bmatrix} \hat{m}_0^v \\ \hat{m}_0^i \end{bmatrix} = \begin{bmatrix} -\mathbf{A}_i\hat{u}_0 \\ \mathbf{0} \end{bmatrix};$$
$$\begin{bmatrix} \mathbf{G} + s_0\mathbf{C} & \mathbf{A}_l^T \\ -\mathcal{K}\mathbf{A}_l & s_0\mathbf{I} \end{bmatrix}\begin{bmatrix} \hat{m}_k^v \\ \hat{m}_k^i \end{bmatrix} = \begin{bmatrix} -\mathbf{A}_i\hat{u}_k \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}\begin{bmatrix} \hat{m}_{k-1}^v \\ \hat{m}_{k-1}^i \end{bmatrix}. \quad (12)$$

Note that the moments calculated by (12) are equivalent to those matched from Equation (10). Hence we can use the new orthonormal basis to project the original system equation (1) and obtain (5) and (9).

Model order reduction methods using congruence transformations are proven to be passivity-preserved and stable, as long as the system satisfies that $D + D^T$ is a non-negative matrix [14][18][16], where $D = \begin{bmatrix} \mathbf{G} & \mathbf{A}_l^T \\ -\mathbf{A}_l & \mathbf{0} \end{bmatrix} + s\begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathcal{L} \end{bmatrix}$ is the system matrix. Since the moments obtained from (12) match the original system equation (1), the projection process is the same as in [18]; the passivity is still preserved for the above model order reduction method with reluctance elements.

3. AN APPLICATION TO SUBSTRATE AND POWER-DELIVERY CO-ANALYSIS

In this section, we propose a new substrate model taking the magnetic effect into consideration. The four Maxwell equations are listed for later derivation.

$$\nabla \times \mathbf{E} = -\partial\mathbf{B}/\partial t \quad (13)$$
$$\nabla \times \mathbf{H} = \mathbf{J} + \partial\mathbf{D}/\partial t \quad (14)$$
$$\nabla \cdot \mathbf{D} = \rho \quad (15)$$
$$\nabla \cdot \mathbf{B} = 0 \quad (16)$$

3.1 Substrate RC model

Outside the diffusion/active areas and contact areas, the substrate can be treated as consisting of uniformly-doped semiconductor-material layers of varying doping densities[1]. Ignoring the magnetic effect, taking divergence of both sides of (14), and using the null identity ($\nabla \cdot (\nabla \times \mathbf{A}) = 0$), we have

$$\epsilon \cdot \frac{\partial}{\partial t}(\nabla \cdot \mathbf{E}) + \frac{1}{\rho}\nabla \cdot \mathbf{E} = 0 \quad . \quad (17)$$

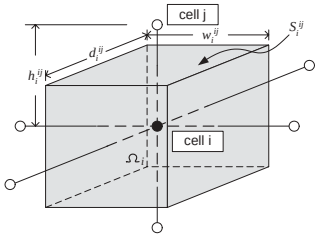

Figure 1: A control volume for a cell in the substrate

Let $\nabla \cdot \mathbf{E} = k$, where $k = \rho/\epsilon$ can be derived from Gauss' law (15). Integrating k over a volume Ω_i around node i as shown in Figure 1 and applying divergence theorem, we get

$$\int_{\Omega_i} k d\Omega = \int_{\Omega_i} \nabla \cdot \mathbf{E} d\Omega = \int_{S_i} \mathbf{E} dS \quad , \quad (18)$$

where and S_i is the surface area around cube i. The integral of Equation (18) can be approximated as

$$\sum_j E_i^{ij} \cdot S_i^{ij} = \sum_j E_i^{ij} \cdot w_i^{ij} d_i^{ij} = k \cdot \Omega_i \quad , \quad (19)$$

and hence

$$\nabla \cdot \mathbf{E} = k = \frac{1}{\Omega_i}\sum_j E_i^{ij} \cdot w_i^{ij} d_i^{ij} \quad . \quad (20)$$

Substituting (20) into (17) and using

$$E_i^{ij} = (V_i - V_j)/h_i^{ij} \quad . \quad (21)$$

Equation (17) becomes

$$\sum_j \left[\frac{(V_i - V_j)}{R_i^{ij}} + C_i^{ij}\left(\frac{\partial V_i}{\partial t} - \frac{\partial V_j}{\partial t}\right)\right] = 0 \quad , \quad (22)$$

where

$$R_i^{ij} = \rho\frac{h_i^{ij}}{w_i^{ij} d_i^{ij}} \quad , \text{ and} \quad (23)$$
$$C_i^{ij} = \epsilon\frac{w_i^{ij} d_i^{ij}}{h_i^{ij}} \quad . \quad (24)$$

The RC model of Equation (22) is shown in Figure 2(a).

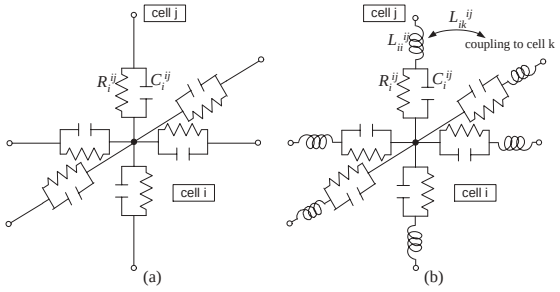

Figure 2: (a) RC model and (b) RLC model for a cell in substrate

3.2 Substrate RLKC model

In case the magnetic effect is not negligible, the electric field intensity contains two parts according to the Helmholtz's theorem.

$$\mathbf{E} = -\bigtriangledown V - \frac{\partial \mathbf{A}}{\partial t} \quad , \tag{25}$$

where $\mathbf{A}$ is the magnetic vector potential. Under the quasi-static assumption, $\mathbf{A}$ can be obtained from the solution of a vector Poisson's equation.

$$\mathbf{A} = \frac{\mu_0}{4\pi}\int_\Omega \frac{\mathbf{J}}{r} d\Omega = \frac{\mu_0}{4\pi}\int_a \int_l \frac{\mathbf{J}\cdot dl}{r} da \quad , \tag{26}$$

where Ω is a control volume, $\mathbf{J}$ is the current density, l is the length of the control volume, and a is the area of its cross section. Considering the effect of time-varying magnetic field, Equation (21) must be rewritten. In order to calculate E_i^{ij}, we calculate the average magnetic vector potential by integrating (26) over the volume of cell i and dividing it by the volume. Equation (21) thus becomes

$$\begin{aligned} E_i^{ij} &= \frac{(V_i - V_j)}{h_i^{ij}} - \frac{\partial}{\partial t}\sum_k \frac{\mu_0}{4\pi}\frac{1}{\Omega_i}\int_{a_i^{ij}}\int_{a_k^{ij}}\int_{h_i^{ij}}\int_{h_k^{ij}} \frac{J_k^{ij} dl\cdot dl}{r} dada \\ &= \frac{(V_i - V_j)}{h_i^{ij}} - \frac{\partial}{\partial t}\sum_k \frac{\mu_0 I_k^{ij}}{4\pi\Omega_i a_k^{ij}}\int_{a_i^{ij}}\int_{a_k^{ij}}\int_{h_i^{ij}}\int_{h_k^{ij}} \frac{dl\cdot dl}{r} dada \\ &= \frac{(V_i - V_j)}{h_i^{ij}} - \frac{1}{h_i^{ij}}\sum_k \frac{\mu_0}{4\pi a_i^{ij} a_k^{ij}}\int_{a_i^{ij}}\int_{a_k^{ij}}\int_{h_i^{ij}}\int_{h_k^{ij}} \frac{dl\cdot dl}{r} dada \frac{\partial I_k^{ij}}{\partial t} \\ &= \frac{(V_i - V_j)}{h_i^{ij}} - \frac{1}{h_i^{ij}}\sum_k \mathcal{L}_{ik}^{ij}\frac{\partial I_k^{ij}}{\partial t} \quad , \end{aligned} \tag{27}$$

where k means the k^{th} node in the circuit, J_k^{ij} and I_k^{ij} denote the current density and the current of the k^{th} node running in the direction of $\vec{ij}$ respectively, r is the distance between nodes i and k, and $\mathcal{L}$ is the inductance. Using E_i^{ij} derived in Equation (27) for (20), and substituting $\bigtriangledown\cdot\mathbf{E}$ into Equation (17), we get

$$\sum_j \left[\frac{(V_i - V_j) - \sum_k \mathcal{L}_{ik}^{ij}\frac{\partial I_k^{ij}}{\partial t}}{R_i^{ij}} + C_i^{ij}\frac{\partial}{\partial t}\left((V_i - V_j) - \sum_k \mathcal{L}_{ik}^{ij}\frac{\partial I_k^{ij}}{\partial t}\right)\right] = 0 \ , \tag{28}$$

where

$$R_i^{ij} = \rho\frac{h_i^{ij}}{w_i^{ij} d_i^{ij}} \quad , \tag{29}$$

$$C_i^{ij} = \epsilon\frac{w_i^{ij} d_i^{ij}}{h_i^{ij}} \quad , \text{ and} \tag{30}$$

$$\mathcal{L}_{ik}^{ij} = \frac{\mu_0}{4\pi a_i^{ij} a_k^{ij}}\int_{a_i^{ij}}\int_{a_k^{ij}}\int_{h_i^{ij}}\int_{h_k^{ij}} \frac{dl\cdot dl}{r} dada \quad . \tag{31}$$

Equations (29) and (30) are exactly same as (23) and (24). Equation (28) is similar to (22); the only difference being that the voltage drop $(V_i - V_j)$ in the RC cases shifts to the voltage drop minus the $\mathcal{L}di/dt$ drop. Therefore, the RC model in Figure 2(a) can be modified and becomes a new RLC model, which is shown in Figure 2(b).

Note that the inductance equation (31) is the same as the partial inductance formula commonly used in the interconnect inductance extraction process. Due to the long-range nature of the partial inductance, it will lead to a large dense inductance matrix if we try to extract the mutual coupling between every inductance element. Suppose we have a $n \times n \times n$ 3-D grid, the number of self inductors in this grid in one direction is n^3, and the number of total coupling will be n^6. It is impossible to extract and simulate a model with complexity so high. Therefore, we apply the window-based reluctance extraction technique to simulate the magnetic effect in this model.

3.3 Window-limited Reluctance Extraction

Due to the strong locality of reluctance elements, it was shown that extracting only a few number of neighbors can approach the accuracy of full partial inductance matrix very well. Thus, a windowing technique was proposed to extract the reluctance elements[11][19]. [12] presented a more generalized windowing policy to extract reluctance from not only equal-length conductors but also any geometry.

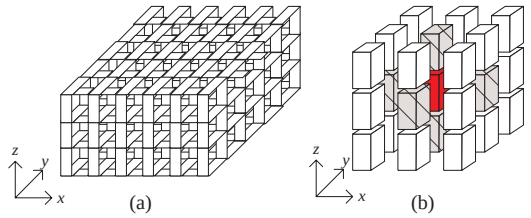

Figure 3: Windowing for reluctance extraction

The whole substrate is discretized into small volumes represented by the center node in Figure 2(b). Each RLC branch in Figure 2(b) is a volume filament, which represents the parasitic between two adjacent node(volume). The complete discretization for the whole substrate can be visualized as in Figure 3(a). Note that the volume filament in the figure is actually defined as the volume of the cell, not just a thin filament; there is no space between parallel volume filaments.

The volume filaments can categorized into three different directions(xyz). Since the mutual inductance coupling exists only between the same directional volume filaments, we can extract the reluctance elements for one direction each time. Figure 3(b) shows an example that illustrates the windowing policy we use. In this example, the z-direction volume filaments are being processed. For each volume filaments, we have to include its neighboring volume filaments that is within some small window, extract the small partial inductance matrix, and calculate the reluctance values. From the discussion of [12], the principle of choosing the neighboring conductors is to choose the most significant inductive couplings. Obviously, the most strongly coupled volume filaments are the six neighbor of the aggressor volume. As shown in Figure 3(b), the aggressor is colored as a dark solid cubic, and we select its six neighbors shaded in gray.

3.4 Power-delivery and Substrate Co-analysis

On a VLSI chip power is transferred through many complicated circuit structures. A power-delivery structure example is shown in Figure 4. From the power supply through the PCB, packaging, I/O pins, C4-bumps, and on-chip interconnect to the transistors, every portion of the circuit in the power-delivery path plays a crucial role for the quality of power delivery. All of them need to be carefully modeled and designed.

Usually, power-delivery structure is modeled with RC or RLC lumped elements. Thus a power-grid model looks like

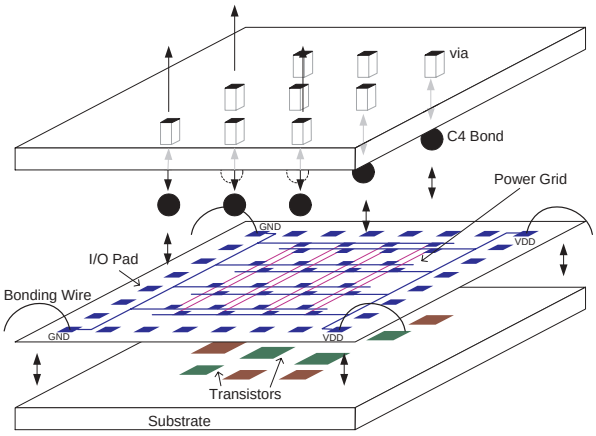

Figure 4: Power-delivery structure on the substrate

a multi-layer RL mesh. Between the power and ground meshes, there are independent current sources extracted or estimated from transistor behavior, and capacitors that model on chip decoupling capacitances. As a result of [3], a power-grid model without taking substrate into consideration actually over-estimates the voltage fluctuation on power-delivery structures. Therefore, we combine the RLC power-grid model with the RLC substrate model proposed as test cases, and perform the model order reduction with reluctance elements. The results are shown in the following section.

4. SIMULATION RESULT

We implemented SuPREME in C/C++ programming language. We also implement the MNA-based simulator that can deal with reluctance elements. In order to have fair comparison, both MNA and SuPREME use the same state-of-art sparse matrix solver. The simulations are run on an Intel Pentium IV 1.4GHz system with RedHat 7.2 Linux operation system.

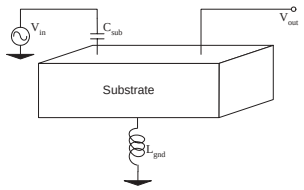

Figure 5: Substrate model

In order to demonstrate the need for inductive substrate model, we first setup an experiment shown in Figure 5. A $0.5 \times 0.5 \times 0.4mm^3$ substrate was modeled and simulated. The back-plane inductor to ground, L_{gnd}, was set to $1nH$. With 1V input voltage, we perform AC analysis from 10GHz to 100GHz; the result is shown in Figure 6. (a) and (b) show the magnitude and phase of the frequency responses respectively. From this figure, some observations can be made. First, the substrate noise increases dramatically as the frequency goes high, and becomes a significant effect after tens of giga-hertz. Second, the RC and RLC substrate models begin to have different responses after few giga-hertz; the magnitudes differ by about 27% at 30GHz and 39% at 100GHz. While the rise time of a transistor is about 10ps nowadays, the frequency components are roughly located in this region. The inductive effect is no more negligible for substrate analysis.

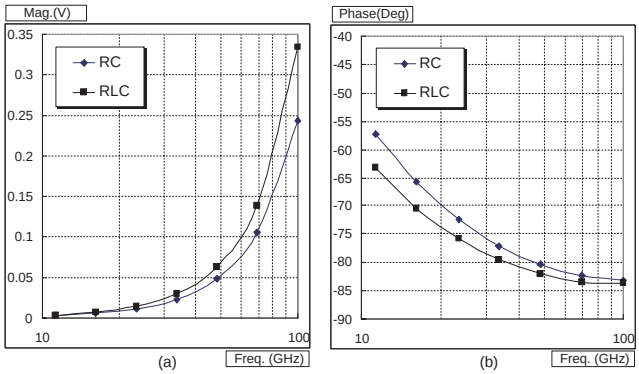

Figure 6: Frequency responses for RC and RLC substrate models: (a)Magnitude response (b)Phase response

Figure 7 and Table 2 show the simulation results for power-grid substrate co-analysis. Figure 7 shows the waveform comparison of the SuPREME and the MNA-based exact solution and the difference between them. The results demonstrate the superior runtime and accuracy over traditional MNA-based simulation. The speedup for the circuit with 40923 nodes is 44.2x. As the circuit size becomes larger, more significant speedup can be expected. Figure 8 shows runtime comparison of SuPREME and MNA. From the log-scale diagram, the runtime of SuPREME is almost linear and has orders of magnitude speedup, while our MNA-based simulator is superlinear.

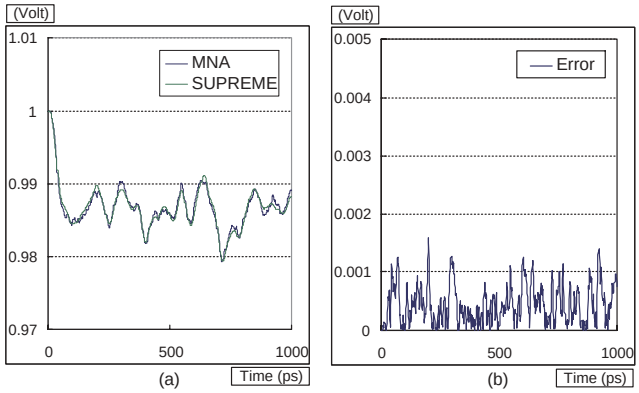

Figure 7: (a) Comparison of waveforms of the MNA exact solution and SuPREME (b) Difference between MNA and SuPREME for a power-grid substrate co-analysis

5. CONCLUSION

In this paper, we first propose a model order reduction technique that can handle the reluctance elements. Different from the existing ENOR and SMOR algorithms, our method does not have to solve the admittance matrix, $(s_0\mathbf{C}+\mathbf{G}+\frac{\mathbf{\Gamma}}{s_0})$.

# of nodes	# of elements	# of sources	MNA runtime	SUPRIME runtime
486	1009	125	15.81	1.10
3858	8433	977	283.09	10.81
10438	23073	2367	1073.78	31.86
40923	91243	10322	4978.77	112.60
162043	362883	40842	-	464.16
363363	814923	91562	-	1078.54

Table 2: Runtime(sec) comparison between MNA exact solution and SuPREME

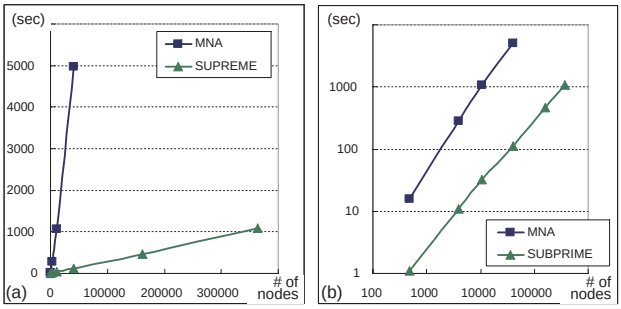

Figure 8: Runtime comparison (a) linear scale (b) log scale

Instead, we explicitly solve matrices $\mathbf{G} + \mathbf{C}$ and $\mathcal{K}$. Our method is fully compatible with PRIMA and hence is passive and stable. Second, in order to perform the power-grid and substrate co-analysis considering inductive effect, we derive a RLKC lumped circuit model for the substrate for the first time. With the window-based reluctance and model order reduction techniques, we are able to efficiently simulate the inductive substrate model. Our algorithm, SuPREME not only greatly reduces the computational complexity of previous reluctance-based model order algorithm but is also capable of handling large number of noise sources efficiently. Experimental results demonstrate the superior runtime and accuracy over traditional MNA-based simulation.

6. ACKNOWLEDGEMENT

This work is partially supported by NSF Grant CCR-0093309 and 0204468, Intel Corporation, and Faraday Technology Corporation. We also thank Mr. Sachin Garg for his help and comment.

7. REFERENCES

[1] B. R. Stanisic, N. K. Verghese, R. A. Rutenbar, L. R. Carley, and D. J. Allstot. Addressing substrate coupling in mixed-mode ic's simulation and power distribution synthesis. *IEEE Journal of Solid-State Circuits*, 29(3), March 1994.

[2] L. M. Silveira. Characterizing substrate coupling in deep-submicron designs. *IEEE Design & Test of Computers*, March-April 2002.

[3] R. Panda, S. Sundareswaran, and D. Blaauw. On the interaction of power distribution network with substrate. In *ISLPED*, August 2001.

[4] H. Ymeri, B. Nauwelaers, K. Maex, S. Vandenberghe, and D. De Roest. New analytic expressions for mutual inductance and resistance of coupled interconnects on lossy silicon substrate. *Digest of Papers of Silicon Monolithic Integrated Circuits in RF Systems*, pages 192 –200, 2001.

[5] Y. Massoud and J. White. Simulation and modeling of the effect of substrate conductivity on coupling inductance. In *Electron Devices Meeting*, pages 491 –494, December 1995.

[6] Minqing Liu, Tiejun Yu, and W. W.-M Dai. Fast 3-D inductance extraction in lossy multi-layer substrate. In *ICCAD*, pages 424 –429, 2001.

[7] A.E. Ruehli. Inductance calculatioin in a complex integrated curcuit environment. *IBM Journal of Research and Development*, pages 470–481, September 1972.

[8] B. Krauter and L.T. Pileggi. Generating sparse partial inductance matrices with guaranteed stability. In *ICCAD*, November 1995.

[9] K.L. Shepard and Z. Tian. Return-limited inductances: A practical approach to on-chip inductance extraction. *IEEE Trans. on Computer Aided Design of Integrated Ciurcuits and Systems*, 19(4):425–436, April 2000.

[10] K. Gala, V. Zolotov, R. Panda, B. Young, J. Wang, and D. Blaauw. On-chip inductance modeling and analysis. In *DAC*, June 2000.

[11] H. Ji, A. Devgan, and W. Dai. Ksim: A stable and efficient rkc simulator for capturing on-chip inductance effect. In *ASPDAC*, June 2001.

[12] Tsung-Hao Chen, Clement Luk, and Charlie Chung-Ping Chen. INDUCTWISE: Inductance-Wise Interconnect Simulator and Extractor. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 22(7), July 2003.

[13] P. Feldmann and R. W. Freund. Efficient linear circuit analysis by padé approximation via the lanczos process. *IEEE Trans. on Computer Aided Design of Integrated Ciurcuits and Systems*, 14, May 1995.

[14] A. Odabasioglu, M. Celik, and L.T. Pillage. Prima: Passive reduced-order interconnect macromodeling algorithm. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 17(8):645–654, August 1998.

[15] B. N. Sheehan. Enor: Model order reduction of rlc circuits using nodal equations for efficient factorization. In *DAC*, June 1999.

[16] H. Zheng and L. T. Pileggi. Robust and passive moel order reduction for circuits containing susceptance elements. In *ICCAD*, November 2002.

[17] J.M. Wang and T.V. Nguyen. Extended krylov method for reduced order analysis of linear circuits with multiple sources. In *DAC*, June 2000.

[18] Y. Cao, Y.M. Lee, T.H. Chen, and C.C.P. Chen. Hiprime: Hierarchical and passivity reserved interconnect macromodeling engine for rlkc power delivery. In *DAC*, June 2002.

[19] H. Zheng, B. Krauter, M. Beattie, and L. Pileggi. Window-based susceptance models for large-scale rlc circuit analysis. In *DATE*, 2002.

[20] D.K. Su, M.J. Loinaz, S. Masui, and B.A. Wooley. Experimemtal results and modeling techniques for substrate noise in maxed-signal integrated circuits. *IEEE Journal of Solid State Circuits*, 38(4), April 1993.

[21] R. Gharpurey and R. G. Meyer. Modeling and analysis of substrate coupling in integrated circuits. *IEEE Journal of Solid State Circuits*, 31(3), March 1996.