Relaxed Hierarchical Power/Ground Grid Analysis*

Yici Cai[1], Zhu Pan[1], Shelton X-D Tan[2], Xianlong Hong[1], Wenting Hou[3], Lifeng Wu[3]
[1]Department of Computer Science and Technology, Tsinghua University, Beijing, 100084, P.R.China
[2]Department of Electrical Engineering, University of California at Riverside, Riverside CA, USA
[3]Cadence Design Systems, CA, USA

Abstract- This paper proposes a novel hierarchical approach to the efficient analysis of large VLSI power/ground grids. Different from the existing hierarchical approach where sub-circuit equivalent models are sparsified with computation-intensive integer programming and the resulting modeling may lead to larger errors if the top circuit matrix has large condition number, the new approach employs an iterative (relaxation) procedure to explicitly compensate the errors and avoid introducing dense matrix caused by the circuit reduction. We also propose an efficient scheme for partitioning high performance center-bumped P/G grids. Experimental results demonstrate that the new algorithm is more accurate than the existing hierarchical method while delivering more speedup over the flat simulators.

I. Introduction

Signal integrity in the power/ground (P/G) bus is emerging as a red-hot problem in the nanometer regime VLSI chip designs. Driven by the increasing importance of it, many efficient linear circuit simulation techniques have been proposed for fast P/G grid analysis recently. Those methods include the hierarchical and macro-modeling based method0[2] , subspace projection based approach[3] , iterative approaches such as PCG method[4] , ADI method[5] , multi-grid methods[6] [7] [8] and the method using node reduction scheme[9] .

Hierarchical reduction and analysis is an efficient way to reduce the circuit size. With the "divide and conquer" scheme, it can scale to handle very large circuits as shown in 0, while one problem with the existing hierarchical method is that errors induced during the sparsification of reduced sub-circuit admittance matrices may lead to significant errors in the final solutions of the top level circuit as shown in the next section. For existing node reduction based approach[9] , only special structures like tree or chain sub-circuits can be reduced in an error-free manner. As a result, their reduction ratio, thus efficiency depends on the circuit structures.

In this paper, we propose a new hierarchical analysis method that mitigates the unbounded error problem in the existing hierarchical reduction algorithm. Our method can be viewed as a general node-reduction algorithm where errors are explicitly introduced to allow more general structure level reduction (thus sparsificaiton) of sub-circuit (thus its admittance matrix). An iterative relaxation process is proposed to compensate the errors introduced due to structure level reduction. Our motivation is inspired by the recent multigrid methods [6] [7] [8] , where a relaxation process is used to smooth the errors caused by the simplified and coarsened grid. We also introduce an adaptive partitioned method to divide today's centre-bumped P/G grids. Experimental results show that the relaxation can be done very efficiently as only one relaxation is typically required for most of P/G grid circuits to achieve satisfactory accuracy requirements, and the new algorithm is more accurate than the existing hierarchical method but delivers more speedup over corresponding flat simulators.

This paper is organized as follows. Section 2 gives a detailed analysis of unbounded error problem in the existing hierarchical approach. Section 3 presents the main idea of the new relaxed hierarchical P/G analysis. Section 4 presents the modeling and partition scheme for high-performance C4-based P/G grid. Experimental results are described in Section 5 and Section 6 concludes the paper.

II. Review of Hierarchical Analysis

A. Overview

For a linear RLCK network, its system questions can be formulated using modified nodal analysis (MNA) as follows:

$$\begin{pmatrix} C & 0 \\ 0 & L \end{pmatrix}\begin{pmatrix} \dot{V}(t) \\ \dot{I}(t) \end{pmatrix}+\begin{pmatrix} G & -A_l^T \\ A_l^T & 0 \end{pmatrix}\begin{pmatrix} V(t) \\ I(t) \end{pmatrix}=\begin{pmatrix} U(t) \\ 0 \end{pmatrix} \quad (1)$$

where *C, L* and *G* are sub-matrices for capacitors, inductors and conductors and *U(t)* is the input vector. By using numerical integration methods like Backward Euler, Eq.(1) can be transformed into a linear algebraic equations **Ax = b,** where

$$A=\begin{pmatrix} C/h+G & -A_l^T \\ A_l^T & L/h \end{pmatrix},\quad b=\begin{pmatrix} U(t)+(C/h)V(t-h) \\ (L/h)I(t-h) \end{pmatrix} \quad (2)$$

and *h* is the fixed time step. Suppose that we partition the circuit into two circuits, one is small sub-circuit *I* and another is rest of the circuit *R*. In between, there are some boundary nodes *B* connecting the two circuits. As a result, the circuit equation set can be rewritten in the following form:

$$\begin{bmatrix} A^{II} & A^{IB} & 0 \\ A^{BI} & A^{BB} & A^{BR} \\ 0 & A^{RB} & A^{RR} \end{bmatrix}\begin{bmatrix} x^I \\ x^B \\ x^R \end{bmatrix}=\begin{bmatrix} b^I \\ b^B \\ b^R \end{bmatrix} \quad (3)$$

* Supported by "the National Natural Science Foundation of China (NSFC) 60176016 and 90307017" and "Hi-Tech Research & Development (863) Program of China 2002AA1Z1460".

where A^{II} is the internal matrix associated with internal variable vector x^I, A^{IB} and A^{BI} are the connection matrices between internal nodes x^I and boundary nodes x^B. To reduce the sub-matrix A^{II}, block Gaussian elimination can be performed and Eq.(3) becomes:

$$\begin{bmatrix} A^{BB*} & A^{BR} \\ A^{RB} & A^{RR} \end{bmatrix} \begin{bmatrix} x^B \\ x^R \end{bmatrix} = \begin{bmatrix} b^{B*} \\ b^R \end{bmatrix} \quad (4)$$

$$A^{BB*} = A^{BB} - A^{BI}(A^{II})^{-1}A^{IB}, \quad b^{B*} = b^B - A^{BI}(A^{II})^{-1}b^I \quad (5)$$

So, instead of analyzing Eq.(3), one can analyze size-reduced parent circuit matrix Eq.(4). Once the parent circuit is analyzed, the solution of internal circuits can be obtained. The benefit is that large circuits, which can't be handled before due to super-linear time complexity of LU-based or other methods, can be analyzed now. Also parallel computing can be exploited to speed up the analysis of sub-circuits.

Notice that the sub-matrix A^{BB*} is typically a full matrix due to inversion of A^{II}, which will lead to a dense matrix for the parent circuit in Eq.(4). As a result, sparsification is performed in 0where off-diagonal elements of small value are rounded off to zero using integer programming method subject to some local error bounds. Unfortunately, the simple truncation based sparsification method may induce larger errors when parent circuit equation Eq.(4) is solved as shown in the next subsection.

B. Error Analysis of Simple Truncation-based Sparsification Method

In this subsection, we show that simply throwing away small values of a circuit matrix may lead to large simulation errors even though the errors are bounded locally.

For a linear system **Ax = b**, if we introduce some errors in matrix A designated as δA with $||\delta A|| / ||A||$ as the relative error, then the relative error in solution x, $||\delta x|| / ||x||$, will be [10] :

$$\frac{\| \delta x \|}{\| x \|} \le \frac{\| A^{-1} \| \| A \| \frac{\| \delta A \|}{\| A \|}}{1 - \| A^{-1} \| \| A \| \frac{\| \delta A \|}{\| A \|}} \quad (6)$$

where $\| T \|$ is any norm definition of a vector or matrix T. $\| A^{-1} \| \| A \|$ is the condition number of the matrix A. As a result, the relative error in matrix A will approximately be amplified by the condition number in the solution x. Thus, the location truncation errors may lead to larger errors in solution x if we have a large condition number for matrix A.

Table 1 lists a number of mesh-structured RC linear circuits with different complexities. We truncate the small off-diagonal elements for each row of the matrix such that the accumulated error for the row is smaller than the 0.5% of the diagonal element value of the row. This truncation strategy is similar to the row/column based sparsification method0. The second column reports the maximum errors we find when the truncated matrices are solved compared with the exact solutions.

Table 1. Relative Errors in Solutions Under Fixed Local Error

Circuit Name	Maximum relative-error
Ckt40 x 40	5.2%
Ckt50 x 50	4.6%
Ckt100 x 100	0.25%
Ckt200 x 200	7.9%
Ckt300 x 300	13.6%
Ckt400 x 400	9.7%

From this table, we can clearly see simple truncation of small off-diagonal elements may lead to significant errors (as large as 13.6% error) in the solutions of the linear equation systems even when the errors are bounded locally.

III. New Relaxed Hierarchical Analysis Approach

As shown in the previous chapter that simple truncation of small off-diagonal values of circuit admittance matrices may lead to large errors. In order to solve this problem and avoid calculating the sub-matrix A^{BB*} and b^{B*} in Eq.(4), we introduce an relaxation process to explicitly compensate the errors due to hierarchical circuit reduction.

Specifically, at time step *k+1*, instead of solving Eq.(3), we solve the following equation sets sequentially:

Where $x^I_{k+1,p}$,$x^B_{k+1,p}$,$x^R_{k+1,p}$ are the solution vectors for sub-circuit, boundary node and rest of the nodes in the parent circuit at time step *k+1* and relaxation step *p*. When Eq. (7)

$$\begin{bmatrix} A^{BB} & A^{BR} \\ A^{RB} & A^{RR} \end{bmatrix} \begin{bmatrix} x^B_{k+1,p} \\ x^R_{k+1,p} \end{bmatrix} = \begin{bmatrix} b^B_{k+1} - A^{BI}x^I_{k+1,p-1} \\ b^R_{k+1} \end{bmatrix} \quad (7)$$

$$A^{II}x^I_{k+1,p} = b^I_{k+1} - A^{IB}x^B_{k+1,p} \quad (8)$$

is solved first time at time step *k+1*, set $x^I_{k+1,p-1} = x^I_k$.

Eq.(7)-(8) will be solved several times as indexed by relaxation index *p* and the whole process can be viewed as a relaxed hierarchical analysis process where we break the whole circuit into two parts and solve them separately under relaxed boundary conditions. Such relaxation process can be viewed as structure-level reduction for the parent circuit: Instead of building the dense sub-matrix A^{BB*}, we stamp original A^{BB} sub-matrix into the parent circuit. The connections between the sub-circuit and parent circuit are reduced. In other words, we perform very aggressive structure reduction such that sub-matrices A^{IB} and A^{BI} are truncated in Eq.(7) and errors are compensated via a relaxation process.

This idea is better illustrated in Fig. 1 where *Va* and *Vb* are the voltages for boundary node *a* and internal node *b* in a sub-circuit. We assume this happens at time step *k+1* and relaxation iteration step *p* where the RLCK network is a resistor-only network after numerical integration approximation (time-discretization).

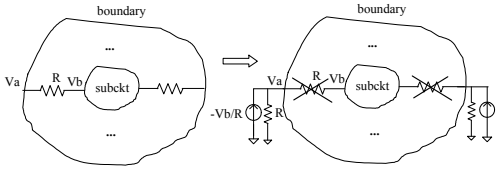


Fig. 1. The relaxed equivalent circuit for a boundary resistor

To remove the connections between an internal circuit and its parent circuit, we need to remove all the resistive branches between the two circuits. For instance, we need to remove branch R between node a and b. Such a reduction will introduce errors. To compensate the errors, two voltage controlled sources (Vb/R for at node a and Va/R at node b) are introduced at each node a and b as shown in Figure 1 for node a. If the voltage is coming from the yet-to-be-solved voltage at time step k+1 and relaxation step p, the compensation leads to no error. As a result, we use latest known voltage from previous time step k or previous relaxation step p-1 as the controlling voltages. This is equivalent to remove the off-diagonal stamp of resistor R and add a compensation current at the right-hand side of the circuit equations as shown in Eq.(7).

If only one relaxation iteration (Eq.(7)-(8)) is carried out, the errors of such circuit reduction process will depend on the voltage difference between present time step and the previous time step, which will be much smaller than the simple truncation scheme as the voltage difference between two consecutive time after compensation will be small also. In contrast, the errors due to the simple truncations will depend on the absolute values of the node voltage for each port. Also we can perform multiple relaxation iterations at each time step to further improve the accuracy at each time step.

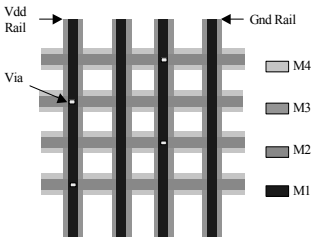
IV. Modeling of High-Performance P/G Grids and Efficient Partition Scheme

A. Modeling of C4-PADs P/G Grid

The high-performance VLSI on-chip P/G grids are commonly hierarchical structure with several metal layers and C4 pads by using flip-chip package.

Fig. 2 shows a typical P/G grid layout, which consists of 5 metal layers with vertical metal 1(M1) and metal 3(M3) and horizontal M2 and M4. The C4 pads are connected to metal 5(M5) on the top (absent in the figure) that are in turn connected to the M1-M4 layers. In each metal layer, the VDD and GND rails alternate. The alternating vertical VDD and GND rails are connected together using alternating horizontal metal runners. Vias are placed at the appropriate crossings of the horizontal and vertical rails [11] .

The simulation model of the P/G grid can be simply modeled as a general-structured grid consisting of many RLC for possible K (for mutual inductance) sections extracted from chip layouts as shown in Fig. 3, devices on the substrate are modeled as independent current source, which connect to the lowest metal layer.



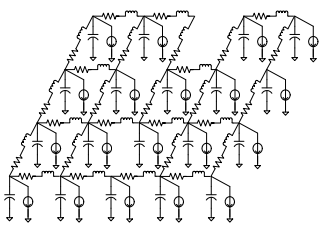
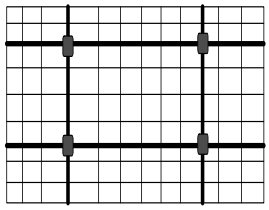


Fig.2. A physical model of a portion of P/G network

Fig.3. A general-structured P/G grid consisting of RLC sections

B. Efficient Partition Scheme

In this work, we target P/G grids with C4 pads for high-performance chips and take advantage of such center-bumped P/G structures. Specifically, we first treat the vias linking two meshes on low and upper levels as a voltage source. Then we partition the low-level fine P/G grid to a number of smaller sub-circuits as shown in Fig. 4. And we can further partition the upper-level coarse P/G grid into some sub-circuits according to C4 pads.



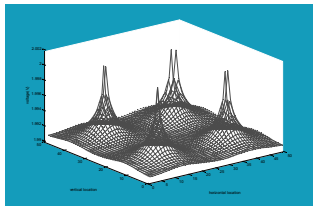


Fig. 4. Efficient Partition scheme

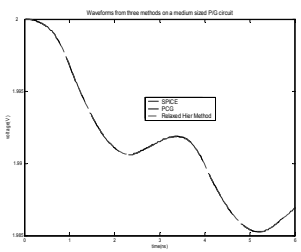
Fig. 5. The lowest voltage distribution in a circuit

The dark rectangles in Fig. 4 stand for voltage source. The wider P/G wires links them and become the boundaries of the small sub-circuits. Such geometry-based partitioning is based on the observation that the boundary nodes typically experience the smaller voltage fluctuation. Fig. 5 shows the typical voltage distributions of a center-bumped P/G grid where the lowest voltages are seen at the boundary nodes.

Because designers usually put more pads and increase the number of vias linking two meshes on low and upper levels in the area that must be supplied more current, this partitioning scheme is adaptive.

V. Experiment Results

The proposed simulation algorithm has been implemented in C. All the experimental results are collected on a SUN V880 workstation with 750MHz Ultra Sparc CPU and 2GB memory. Preconditioned conjugate gradient (PCG) method is used to solve the sub-circuit. The number of time steps is assigned 120 for one clock cycle and VDD is 2.0V.



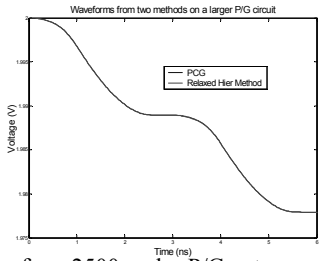


Fig. 6. Accuracy comparison for a 2500 nodes P/G net

Fig. 7. Accuracy comparison for a large P/G net

First, we compare our new method with Spice3f4[12] and flat PCG method in terms of accuracy. The three algorithms are used to solve a P/G network with about 2500 nodes. The three waveforms of a randomly chosen node are shown in Fig. 6. We also use our method and PCG to solve a larger circuit with about 1M nodes, and the result is in Fig. 7. The maximum error is below 0.035% for both cases.

Table 2. Run Time Comparison Results with PCG and SPICE3

Circuit Name	#Node Num	#Branch Num	CPU Time (sec)			Speedup Over (X)	
			Relaxed Hier	PCG	SPICE3	PCG	SPICE3
Ckt1	2.5k	4.7k	0.71	1.01	15.62	1.42	22.3
Ckt2	10k	19.4k	2.93	4.53	197.46	1.55	67.4
Ckt3	40k	78.8k	13.90	21.82	2797.76	1.57	201.3
Ckt4	250k	496k	122.65	271.25	N/A	2.21	N/A
Ckt5	1M	1.92M	491.62	1305.0	N/A	2.65	N/A
Ckt6	4M	7.66M	2002.46	5412.61	N/A	2.70	N/A
Ckt7	6.25M	12.3M	3151.94	N/A	N/A	N/A	N/A
Ckt8	16M	31.8M	8022.90	N/A	N/A	N/A	N/A

Table 2 compares the performance in terms of CPU time of the proposed new algorithm "Relaxed Hier", PCG and SPICE3 on a number of large C4 P/G grids, which have the similar structures from real extracted industrial chips. It can be seen that the new method can easily scale to handle P/G grids with 16 million nodes in less than 2.5 hours. We also observe at least two orders of magnitudes speedup over SPICE3 for medium sized P/G circuits. But more important is that comparing with the flat PCG, which is also used for solving sub-circuits, we have about 2x speedup over flat PCG which is significant compared with hierarchical approach in0, where hierarchical approach without parallel computing has led to performance degradation (speed up is less one) due to larger overhead coming from solving sub-circuit modeling and sub-circuit solving. As a result, the proposed new hierarchical method indeed leads to more speedup over existing hierarchical method.

Table 3. Accuracy Comparison Results with and without Relaxation

Circuit Name	#Node Num	Average Absolute Error (V)		Average Relative Error (%)	
		No Relax	Relax	No Relax	Relax
Ckt1	2.5k	0.12965	0.0031	6.483	0.155
Ckt2	10k	0.19904	0.0038	9.952	0.190
Ckt3	40k	0.21503	0.0047	10.752	0.235
Ckt4	250k	0.21516	0.0043	10.758	0.215
Ckt5	1M	0.10790	0.0056	5.395	0.280
Ckt6	4M	0.18266	0.0044	9.133	0.220

Relative Error = 100%*Absolute Error/VDD. VDD=2.0v

Table 3 compares the errors caused by the new relaxed hierarchical method under conditions with and without relaxation process. From Table 3, it is clear that without the relaxation process to compensate the errors due to structure reduction, the solution errors will be significant (more than 10%). With relaxation process, solution errors for all the cases are below 0.5%, which is much better than the 2% errors seen in 0.

VI. Conclusion

This paper has proposed a novel hierarchical simulation algorithm for efficient analysis of large high-performance on-chip VLSI power/ground (P/G) grids. We first showed that simple truncation-based matrix sparsification used in the existing hierarchical analysis may lead to substantial errors in the solutions even the truncation errors are bounded locally. Instead, the new algorithm explicitly compensates the errors introduced by sparsification or structure reduction of sub-circuits using a relaxation process. Such relaxed hierarchical approach can allow more aggressive sparsification or P/G complexity reduction and thus lead to more speedup without significant loss of accuracy. We also have proposed an efficient scheme for partitioning high-performance center-bumped P/G grids. Experimental results showed that the new algorithm is more accurate than the existing hierarchical method while delivering more speedup over the flat simulators.

VII. References

[1] M. Zhao, R. V. Panda, S. S. Sapatnekar and D. Blaauw, "Hierarchical analysis of power distribution networks", *IEEE Trans. On Computer Aided Design*, vol. 21, no. 2, pp. 159–168, Feb. 2002.

[2] Y. Cao, Y. Lee, T. Chen and C. Chen, "HiPRIME: Hierarchical and Passivity Reserved Interconnect Macromodeling Engine for RLKC Power Delivery", In Proceeding of *IEEE/ACM Design Automation Conference*, pp. 379-384, New Orleans, Jun. 2002.

[3] J. M. Wang and T. V. Nguyen, "Expended Krylov Subspace Method for reduced order analysis of linear circuits with multiple sources", In Proceeding of *IEEE/ACM Design Automation Conference*, pp. 247-252, Los Angeles, Jun. 2000.

[4] T. Chen and C. Chen, "Efficient Large-Scale Power Grid Analysis Based on Preconditioned Krylov-Subspace Iterative Methods", In Proceeding of *IEEE/ACM Design Automation Conference*, pp. 559-562, Las Vegas, Jun. 2001.

[5] Y. Lee and C. Chen, "Power grid transient simulation in linear time based on transmission-line-modeling alternating-direction-implicit method", In Proceeding of *IEEE/ACM International Conference on Computer Aided Design*, pp. 75-80, San Jose, Nov. 2001.

[6] S. R. Nassif and J. N. Kozhaya, "Fast power grid simulation", In Proceeding of *IEEE/ACM Design Automation Conference*, pp. 156-161, Los Angeles, Jun. 2000.

[7] Z. Zhu, B. Yao and C. Chen, "Power Network Analysis Using an Adaptive Algebraic Multigrid Approach", In Proceeding of *IEEE/ACM Design Automation Conference*, pp. 105-108, Anaheim, Jun. 2003.

[8] H. Su, E. A. Sani and S. R. Nassif, "Power Grid Reduction Based on Algebraic Multigrid Principles", In Proceeding of *IEEE/ACM Design Automation Conference*, pp. 109-112, Anaheim, Jun. 2003.

[9] X. Tan and C. Shi, "Fast power-ground network optimization using equivalent circuit modeling", In Proceeding of *IEEE/ACM Design Automation Conference*, pp. 550-554, Las Vegas, Jun. 2001.

[10] G. Golub and C. Van Loan, "*Matrix Computation"*, 3rd Edition, Johns Hopkins University Press, 1996.

[11] D. Acharyya and J. Plusquellic, "Impedance profile of a commercial power grid and test system", In Proceeding of *IEEE International Test Conference*, pp. 709-718, Charlotte, NC, USA, Sept. 2003.

[12] http://bwrc.eecs.berkeley.edu/Classes/IcBook/SPICE