Power Grid Reduction Based on Algebraic Multigrid Principles

Haihua Su Emrah Acar Sani R. Nassif
IBM Austin Research Lab
11400 Burnet Rd.
Austin, TX 78758
{haihua,emrah,nassif}@us.ibm.com

ABSTRACT

With the scaling of technology, power grid noise is becoming increasingly significant for circuit performance. A typical power grid circuit contains millions of linear elements, making noise analysis and verification challenging in terms of both run time and memory. We propose a power grid reduction scheme based on algebraic multigrid principles, in which the coarser-level grid and the restriction operators are constructed automatically from the circuit matrices. This method is suitable for large-scale power grid transient and AC analysis. Experimental results show an order of magnitude speed-up over flat analysis in addition to practical tradeoffs for accuracy, CPU time and memory usage.

Categories and Subject Descriptors

B.8.2 [**Performance and Reliability**]: Performance Analysis and Design Aids

General Terms

Algorithms

Keywords

Algebraic Multigrid, Power Grid Noise, Transient Analysis, AC Analysis

1. INTRODUCTION

In general, multigrid methods consist of two complementary components [1, 3]: Relaxation (smoothing), which reduces the high frequency error components using an iterative solver, and coarse grid correction, which reduces the low frequency error components. It involves mapping the problem to some coarser grid (restriction), solving the mapped smaller problem, and mapping the solution back to the original fine grid (interpolation). Figure 1 illustrates a recursive V-cycle [1] of the multigrid method with three nested iterations. At the bottom level, the exact solution is usually obtained from a direct solver.

DAC 2003, June 2–6, 2003, Anaheim, California, USA.

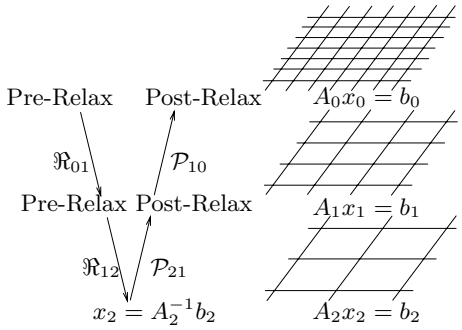

Figure 1: The V-cycle of the multigrid method.

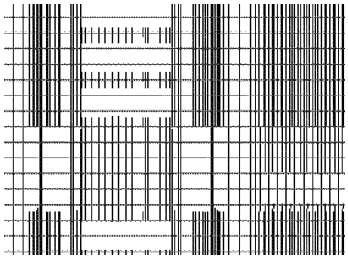

Figure 2: Typical irregularity of power grids (bottom two layers).

Two kinds of multigrid methods, standard and algebraic, have been proposed in the literature. A typical Standard Multigrid (SMG) method uses uniform coarsening and linear interpolation in coarse grid correction and relies on the choice of the relaxation operator (iterative solver) to reduce the error components not well approximated by coarse grid correction. On the other hand, Algebraic Multigrid (AMG) methods fix the number of iterations during smoothing and apply coarsening and interpolation to reduce the error components not well reduced by smoothing. A general-purpose AMG solver is very efficient for DC analysis. However, for transient and AC analysis, the cost in every time or frequency step almost equals that of the DC solve. Therefore, it is not applicable for transient and AC analysis. The method in [4, 5] belongs to a SMG-reduction technique, which requires keeping track of power grid geometry for each multigrid level. The cost of maintaining this geometrical information is large for irregular grids, especially ones with lower metal layers extracted (Fig. 2), making SMG-based methods impractical for real industrial designs.

We propose an AMG-based reduction scheme which is much more practical than prior approaches. We construct the restriction and interpolation operators directly from the circuit Modified Nodal Analysis (MNA) matrices[6]. No geometrical information is maintained, therefore it enables significant memory reduction. Similar to [4, 5], our scheme assumes smooth voltage variation and ignores the relaxation steps in Fig. 1, therefore it is also a direct solution method. We show two important applications of the proposed technique in reducing the runtime complexity of Transient and AC analysis of the power grid network. Experimental results show that problem size can be reduced by more than 90% without sacrificing much accuracy.

The following power distribution network model is used throughout this paper:

- The on-chip power distribution network (grid) is modeled as a resistive mesh with via resistors connecting metal layers.
- The loads (functional blocks) are modeled as distributed time-varying current sources in parallel with parasitic capacitors connected between power and ground. To simplify the analysis, we split every parasitic capacitor into two grounded capacitors.
- The decoupling capacitors (decaps) are modeled as single lumped capacitors connected between power and ground. Similarly, to simplify the analysis, a decap is split into two grounded capacitors.
- The top-level metal is connected to a package modeled with inductors or RL elements connected to ideal constant voltage sources. For package impedance resonance analysis, a more detailed RL(K)C model can be used.

2. CONSTRUCTION OF RESTRICTION AND INTERPOLATION MATRICES

Using this model, the entire power distribution system becomes a large-scale linear circuit. The transient behavior of the system can be described using MNA equations as follows:

$$G_0 x_0 + C_0 \dot{x_0} = u_0 \tag{1}$$

where x_0 is an N-dimensional real vector of node voltages and inductor currents (a simple reformulation can remove voltage source current variables from x_0); G_0 (size $N \times N$) is the grid conductance matrix; C_0 (size $N \times N$) includes the decoupling capacitance, load parasitic capacitance and package inductance terms, and u_0 includes the loads and voltage sources. G_0 is symmetric, while C_0 is a diagonal matrix because all capacitors are grounded and there is no mutual inductance in our power grid model. N can be on the order of millions for a normal power grid circuit. Our objective is to reduce such a large sparse system equations to a smaller system representing the coarse-level grid subject to a smooth approximation error [7]. The smaller set of equations should describe an equivalent circuit of the original system.

For one-level of multigrid reduction, the restriction operator can be constructed as an $N \times M$ matrix R_{01}, where $M < N$. Therefore the equation at the coarse level becomes

$$G_1 x_1 + C_1 \dot{x_1} = u_1 \tag{2}$$

where

$$G_1 = R_{01}^T\ G_0\ R_{01} \quad \text{and} \quad C_1 = R_{01}^T\ C_0\ R_{01}, \tag{3}$$

$$x_1 = R_{01}^T x_0 \quad \text{and} \quad u_1 = R_{01}^T u_0. \tag{4}$$

As is done in most AMG method, the interpolation operator is chosen as $P_{10} = R_{01}^T$. Then we interpolate from the coarse-level solution to the fine level using

$$x_0 = P_{10}^T x_1 = R_{01} x_1 \tag{5}$$

By performing the interpolation operation, the voltage on a node p (i.e. $x_0(p)$) in the fine grid can be determined as a linear combination of nodal voltages of its neighborhood.

In general, variables representing important boundary conditions should be preserved. For our power grid model, these variables include all ideal voltage source nodes, all nodes in the top-level metal layer that are connected to package/pins, all package inductance or RL-in-series branches and all nodes in the bottom-level metal layer that are connected to critical loads. The equation for these boundary nodes/branches are put at the beginning of the original G and C matrices, which makes them easy to preserve. For the rest of the variables, AMG grid reduction algorithm can be applied to determine the coarse-level grid points.

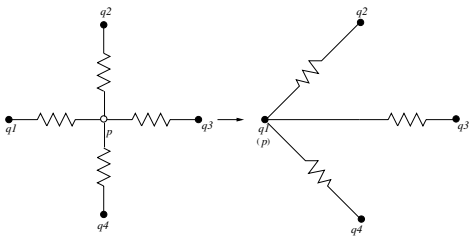

Figure 3: Shorting node p to its strongly connected neighbor $q1$.

The coarse grid has to be chosen to represent smooth errors and has to be able to interpolate these errors onto the fine grid. It is shown in [7] that smooth error varies slowly in the direction of strong connections. In our case, a strong connection between node p and q in G means a relatively large conductance value at the off-diagonal entries (p, q) and (q, p), compared to the diagonal entries at (p, p) and (q, q). Therefore, we choose

$$mes_{pq} = (g_{pq}/G_p + g_{pq}/G_q)/2 \tag{6}$$

as a measure of connection between node p and q, where G_p and G_q are self conductance at node p and q. If $mes_{pq} > \psi$, node q is chosen in the coarse grid and p in the fine grid and will be interpolated as $x(p) = x(q)$, where ψ is a threshold chosen to control the reduction rate and accuracy. This is equivalent to shorting node p to q when the resistor connected between them is small. The corresponding restriction matrix for shorting node p to $q1$ in Fig 3 becomes:

$$R_{5\times 4} = \begin{array}{c} \begin{matrix} q1 & q2 & q3 & q4 \end{matrix} \\ \begin{bmatrix} 1 & & & \\ & 1 & & \\ 1 & & & \\ & & 1 & \\ & & & 1 \end{bmatrix} \end{array} \begin{matrix} q1 \\ q2 \\ p \\ q3 \\ q4 \end{matrix} \tag{7}$$

This reduction scheme iteratively removes relatively smaller resistors in the grid, therefore the number of nonzeros in the coarse-level matrix $R^T G R$ decreases.

The algorithm for nested iteration and calculating the restriction matrices R_{ij} from the original conductance matrix G_0 (size $N \times N$) can be summarized as follows:

1 Set $i = 0$, choose the reduction threshold ψ and number of reduction levels L.

2 Set $j = i + 1$.

3 Initialize R_{ij} as an $N \times N$ sparse matrix, initialize a size N array $KEEP$ and set zero.

4 Mark every boundary node/branch k as "keep", i.e. set KEEP(k)=1 and set $R(k,k) = 1$. Note that this corresponds to the first K rows of matrix R and array $KEEP$ because they are the first K rows of G_i.

5 Go through each row (node) p in G_i as follows. If node p is marked as neither "keep" or "remove", i.e. KEEP(p)==0, then calculate mes_{pq} for every nonzero entry $G_i(p,q)$ $(q > p)$ using Eqn. (6), find the maximum q_m. If the $mes_{pq_m} > \psi$, set $R(q_m, q_m) = 1$, $R(p, q_m) = 1$ and mark node q_m as "keep", node p as "remove", i.e. set $KEEP(q_m) = 1$ and $KEEP(p) = -1$, and then mark all neighboring nodes of p and q_m as "keep".

6 Remove all column r in R_{ij} if $KEEP(r) == -1$.

7 Compute $G_j = R_{ij}^T\ G_i\ R_{ij}$; i = i+1; if $i == L$, then stop; otherwise, goto step 2.

3. APPLICATIONS IN POWER GRID ANALYSIS

Our proposed reduction scheme is directly applicable to fast transient and AC analysis.

3.1 Transient Analysis

By applying the Backward Euler integration formula [6] to Eq. (1), we have:

$$(G + C/h)x(t + h) = u(t + h) + x(t)C/h \tag{8}$$

where h is the time step for the transient analysis. h is usually kept constant so that matrix $A = G + C/h$ is independent of time. Since C is diagonal (as is discussed in Section 2), our proposed AMG reduction algorithm can be directly applied to matrix A with smooth approximation error. Since A is fixed, such a reduction is performed only at the *first* time step. For the coarsest level matrix A_L, only one single initial factorization is required. The following time steps require only a forward/backward solution at the coarsest-level and then a mapping back to the original grid:

$$x = Rx_L = R_{01}R_{12}R_{23}\cdots R_{L-1,L}x_L \tag{9}$$

The chain of matrix multiplications $R = \prod_{i=0}^{L-1} R_{i,i+1}$ also needs to be calculated only at the first time step.

3.2 AC Analysis

It is often interesting to analyze the impedance of the power grid and the package together. A recently published package-level power distribution network impedance (resonance) analysis work [2] performs Fourier transforms of the time domain voltage and current waveforms. However, such a time-domain based method is very time-consuming, especially when the package is taken into account because the transient analysis needs to be performed for thousands of clock cycles before reaching the package resonance frequency. Alternatively, we perform AC analysis on our reduced power grid and the package directly in the frequency domain.

In AC analysis, the circuit MNA equation becomes

$$(G_0 + j\omega C_0)x = (G_0 + j\omega C_0)(x_R + jx_I) = Re(u_0) + Im(u_0) \tag{10}$$

where $x = x_R + jx_I$ is a complex vector of node voltages and inductor currents. u_0 contains complex current and voltage sources. Let the real vector $X = [x_R\ \ x_I]^T$, we can rewrite Eqn (10) as

$$\begin{bmatrix} G_0 & -\omega C_0 \\ \omega C_0 & G_0 \end{bmatrix} \begin{bmatrix} x_R \\ x_I \end{bmatrix} = \begin{bmatrix} Re(u_0) \\ Im(u_0) \end{bmatrix} \tag{11}$$

For some frequency range $[\omega_{low}, \omega_{high}]$ of interest, Eqn (11) can be solved with a desired frequency step. The interesting frequency range for power grid and package analysis is usually from $10MHz$ to $5GHz$. Typically, such an AC analysis is formidable since the matrix size in Eqn. (11) is double that of the transient matrix and it requires one factorization at every frequency step. Our proposed technique reduces the complexity of the system matrices, hence making such an analysis feasible.

In our method, the restriction matrix $R = \prod_{i=0}^{L-1} R_{i,i+1}$ can be constructed by performing the reduction on the G_0 matrix. Since matrix G_0 is fixed for all frequencies, such a reduction only needs to execute once. We reduce matrix C_0 using the same matrix R and justify this as follows. C_0 is a diagonal matrix which contains package inductors and grounded capacitors. All inductor branches have been preserved as boundary conditions during the construction of R, so the reduction won't affect these branches. When two nodes with grounded capacitors are shorted during the reduction, applying $R^T C_0 R$ will automatically add up the two capacitances to one of the two nodes that is kept in the coarse-level grid, hence this reduction will conserve the total load capacitance of the grid. Therefore, at the coarsest level L, Eqn. (11) becomes

$$\begin{bmatrix} R^T G_0 R & -\omega R^T C_0 R \\ \omega R^T C_0 R & R^T G_0 R \end{bmatrix} \begin{bmatrix} x_{RL} \\ x_{IL} \end{bmatrix} = \begin{bmatrix} R^T Re(u_0) \\ R^T Im(u_0) \end{bmatrix} \tag{12}$$

Similarly, we can obtain the fine-level solution by interpolation:

$$x_R = Rx_{RL} \ \text{ and } \ x_I = Rx_{IL} \tag{13}$$

The magnitude and phase of every node can then be found. For package and power grid analysis, a significant amount of reduction (allowing larger approximation error) for AC analysis is attractive because the AC analysis of even the coarsest-level grid can provide close estimates of resonance frequencies. A more accurate AC analysis can be performed on the fine-level grid focusing the region near these resonance frequencies, rather than sweeping a wider range of frequency. The AC analyses explain important frequency behaviour of the power grid and the package, which is of practical interest.

4. EXPERIMENTAL RESULTS

We have implemented our AMG based reduction algorithm in C++ and performed both transient and AC analysis on several large-scale power grid circuits. We ran these experiments on an Intel Pentium-III 700MHz machine with 4GB memory. All testcases were in a 0.13 micron CMOS technology with seven metal layers.

Table 1 lists the reduction of total number of nodes and total number of nonzeros (NNZ) in the MNA matrix with twenty iterations. For both cases, less than 10% of nodes remains at the coarsest level grid and the number of nonzeros in the coarsest level matrix decrease similarly.

	Node num	bottom-level node num	Redc	NNZ	bottom-level NNZ	Redc
T1	0.4M	17K	%95.8	1.9M	160K	%91.6
T2	1.3M	83K	%93.6	5.6M	770K	%86.3

Smoothing threshold = 0.2, 20 nested iterations

Table 1: Grid reduction results.

	total CPU time (*min*)			peak memory	
	RAMG	AMG	time steps	**RAMG**	AMG
T1	31.8	1146.0	1000	476M	360M
T2	80.7	1694.4	500	1.7G	1.0G

Threshold = 0.2, 20 nested iterations

Table 2: Transient simulation results.

We performed transient analysis using our method (named as RAMG) and compared the performance to a general-purpose AMG solver in Table 2. Neither case was able to run using direct solvers due to memory limitations. We ran T1 with 1000 time steps and T2 500 steps. The results show a speedup of 36X and 21X for each testcase, with a memory increase of 32% and 70% respectively. The slow run time of the AMG solver is due to the fact that it is an iterative solver and does not re-use the reductions at every time step.

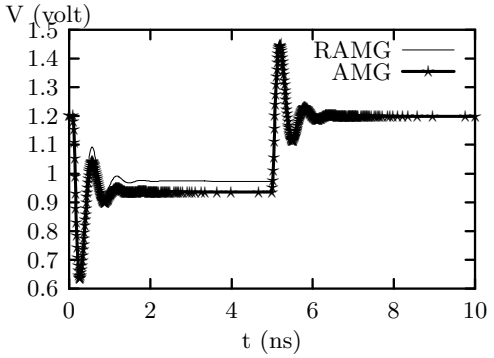

Figure 4: Comparison of waveforms.

Fig. 4 shows the Vdd voltage waveform for several clock cycles at one of the loads in T1. The voltage difference at the first peak is around 25mV out of the Vdd level of 1.2V, which is below 5%.

Table 3 shows the percentage of node reduction, total CPU time and average voltage waveform accuracy with respect to various thresholds ψ for testcase $T2$. We measure the waveform at a node near the center of T2, and compute the average percentage error over time with respect to the waveform obtained from the AMG solver. Both percentage of node reduction and total CPU time show that, for this testcase, some optimal threshold resulting in less CPU time and more node reduction lies between 0.1 and 0.3 (exclusive). The waveform accuracy is below 3% for all cases.

We ran AC analysis for a testcase T3 with seven metal layers and 0.15 millions of nodes in the power grid and compared the results with an exact solution obtained from an exact solver over the frequency range from $10MHz$ to $5GHz$ with a total of 28 frequency sampling points. Table 4 shows the average CPU time per frequency point for various reduction threshold using our method "RAMG" and the exact method. We can see significant speedup of our method over the exact method.The magnitude at a pair of nodes connecting one of the loads in the bottom metal layer is computed

ψ	reduction	CPU time (*min*)	avg err
0.3	87.2%	220.9	1.68%
0.2	93.6%	80.7	2.38%
0.15	92.8%	71.0	2.79%
0.1	91.9%	97.0	2.83%

Table 3: Performance v.s. threshold.

T3	average CPU time per frequency step (*sec*)				
RAMG	ψ=0.1	ψ=0.2	ψ=0.3	ψ=0.4	ψ=0.5
	2.53	2.17	5.38	24.4	95.3
Exact	350.2				

Table 4: AC simulation results for testcase T3.

and shown in Fig 5. For the case of $\psi = 0.5$, where 44% nodes remain at the coarsest level, the curve is almost identical to the exact one, showing the resonance frequency at around 500MHz.

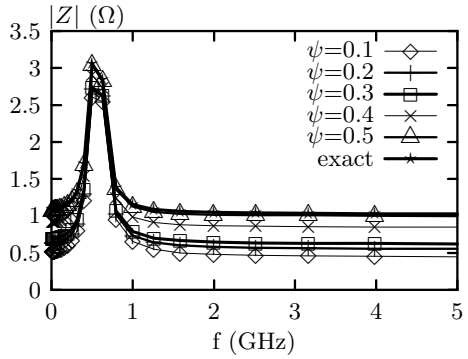

Figure 5: Impedance magnitude.

5. CONCLUSION

In this paper, we presented an effective power grid reduction approach based on AMG principles. This method offers superior speed and accuracy tradeoffs for large-scale power grid analysis. The significant order of magnitude reduction in runtime and memory for AC and transient analyses for such large systems is instrumental to shortening the grid design and optimization process.

6. REFERENCES

[1] W. L. Briggs. *A Multigrid Tutorial*, 1987.

[2] J. Fang, Z. Chen, Y. Chen, D. Xue, and J. Zhao. Integrated Electromagnetic and Circuit Simulation for Signal Integrity Analysis of High-Speed Electronic Packages. In *Proc. of 1997 Government Microcircuit Applications Conference*, pages 347–350, Las Vegas, NV, March 1997.

[3] W. Hackbusch. *Multi-Grid Methods and Applications*. Springer-Verlag Berlin, Berlin, 1985.

[4] J. N. Kozhaya, S. R. Nassif, and F. N. Najm. A Multigrid-Like Technique for Power Grid Analysis. *IEEE Trans. CAD*, October 2002.

[5] S. R. Nassif and J. N. Kozhaya. Fast Power Grid Simulation. In *Proc. Design Automation Conference*, pages 156–161, Los Angeles, CA, June 2000.

[6] L. T. Pillage, R. A. Rohrer, and C. Visweswariah. *Electronic and System Simulation Methods*. McGraw-Hill, New York, NY, 1995.

[7] J. W. Ruge and K. Stuben. *Handbook of Mathematics and Computational Science*. S. McCormick, Ed., 1987.