Improved Hierarchical IR Drop Analysis in Homogeneous Circuits

Chengrui Zhang and Pingqiang Zhou
School of Information Science and Technology
ShanghaiTech University, Shanghai, P. R. China
E-mail: {zhangchr, zhoupq}@shanghaitech.edu.cn

***Abstract*—IR drop analysis plays an important role in chip design. However, this analysis usually takes long time. In this work, we propose an improved hierarchical method for IR drop analysis in homogeneous chips. Experimental results show that the more speedup can be achieved with larger number of homogeneous cores. For the chip with 8 homogeneous but not identical cores, our approach can obtain 5.21X speedup in the whole analysis with 99% accuracy.**

***Index Terms*—IR drop analysis, homogeneous circuit, hierarchical analysis, GMRES**

I. Introduction

With the increasing integration level of IC chips, the size of on-chip power supply network expands fast which poses stringent challenges to the IR drop analysis. Traditional IR drop analysis is based on MNA [1], which turns a power supply network into a linear system and then solves it to find the IR drop. Although this method is accurate, it suffers from large runtime and memory costs.

In the past decades, many approaches have been proposed to address this problem [2–6]. The IR drop analysis is accelerated by employing power grid reduction methods [2, 3] or hotspot detection methods [4, 5], however, both methods suffer accuracy loss side effect. Fortunately, M.Zhao at [6] enables the IR drop analysis acceleration by introducing hierarchical analysis (HA) with keeping high accuracy.

These methods aforementioned are effective and universal. However, they all ignore the fact that recent IC chips not only have large scale power supply networks, but also utilize homogeneous cores in their designs. For example, a CIM-based CNN processor comprises four homogeneous DSS-based CIM cores [7]. The commercial product Intel CPU Core i9-10900T integrates 10 independent and similar computing units. The homogeneous cores integrated on chip maybe share similar property, which provides an opportunity to decrease the runtime and memory costs.

This paper focuses on the IR drop (DC) analysis of homogeneous multi-core chips. In this paper, homogeneous cores are classified into three types 1) type 1 – cores with identical power supply network, 2) type 2 – cores with identical power supply network structure but slightly different resistance value, and 3) type 3 – cores with identical resistance value but slightly different power network structure. In this work, we propose an improved hierarchical analysis method (IHA) based on the prior HA method [6] to solve the IR drop analysis in all the three types of homogeneous multicore chips. In the prior HA method, the analysis flow consists of three steps (see Fig. 1): 1) local grid macro-modeling, 2) global grid solution, and 3) local grid solution. For a homogeneous chip, we improve the HA method by 1) reusing the macromodel generated for one local grid in step 1, and 2) adopting iterative method (such as GMRES) to solve each local grid in step 3. Our experiment results show that more speedup can be achieved with the increasing number of homogeneous cores. For the chip with 8 homogeneous but not identical cores, our approach can achieve 7.00X speedup in *Step 1* and 5.21X speedup in the whole analysis with 99% accuracy.

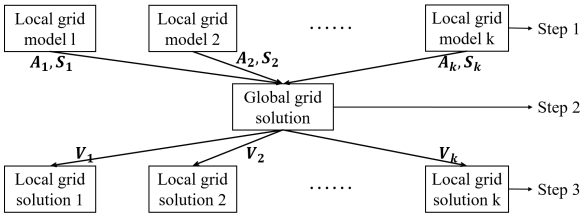

Fig. 1. Flow of hierarchical analysis [6].

II. Overview of Hierarchical Analysis

Before presenting the IHA approach, we present an overview of HA[6] in general. The HA method aims to overcome the capacity limitation of traditional IR drop analysis by the usage of macromodels. They design a hierarchical model to divide the power network into global grid and local grid. After macromodeling, each local grid is expressed as:

$$\underbrace{\begin{bmatrix} G_{11} & G_{12} \\ G_{12}^T & G_{22} \end{bmatrix}}_{G} \underbrace{\begin{bmatrix} U_{int} \\ V_{port} \end{bmatrix}}_{V} = \underbrace{\begin{bmatrix} J_{int} \\ J_{port} + I \end{bmatrix}}_{J} \tag{1}$$

where

- G_{11}, G_{22}: conductance matrices of internal nodes and port nodes respectively; G_{12}: conductance matrix between internal nodes and port nodes,
- U_{int}, V_{port}: voltage vectors of internal nodes and port nodes respectively,
- J_{int}, J_{port}: current sources connected at internal nodes and port nodes respectively,
- I: vector of currents through the interface;
- G, V, J: conductance matrix, voltage vector and current resources vector in the local grid respectively.

The Eq.(1) can be transformed into :

978-1-7281-6235-5/20/$31.00 ©2020 IEEE

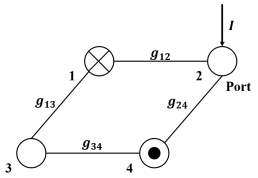

Fig. 2. Example circuit. Node 1 connects to a voltage source, Node 4 connects to a current source, I represents the external current.

$$I = \underbrace{(G_{22} - G_{12}^T G_{11}^{-1} G_{12})}_{Port\ admittance\ matrix\ A} V_{port} + \underbrace{(G_{12}^T G_{11}^{-1} J_{int} - J_{port})}_{S} \tag{2}$$

Once the macromodels for all the local grids are generated, the entire circuit is represented only by the global grid. Then, after using MNA[1], the global grid is shown as follows:

$$\begin{bmatrix} G_{00} & G_{01} & G_{02} & \cdots & G_{0k} \\ G_{01}^T & A_1 & G_{12} & \cdots & G_{1k} \\ G_{02}^T & G_{12}^T & A_2 & \cdots & G_{2k} \\ \vdots & & & & \vdots \\ G_{0k}^T & G_{1k}^T & G_{2k}^T & \cdots & A_k \end{bmatrix} \begin{bmatrix} V_0 \\ V_1 \\ V_2 \\ \vdots \\ V_k \end{bmatrix} = \begin{bmatrix} I_0 \\ -S_1 \\ -S_2 \\ \vdots \\ -S_k \end{bmatrix} \tag{3}$$

where

- global node index is labeled as 0,
- I_0: vector of currents that flow out of the global nodes,
- G_{ij}: conductance matrix between partition i and j,
- S_i, V_i, A_i: constant vector, voltage vector and port admittance matrix of partition i respectively, where $i \in [1, k]$.

Therefore, global grid solution can be obtained (i.e. the solution of port nodes V_{port}). Then, local grids solutions are:

$$U_{int} = G_{11}^{-1}(J_{int} - G_{12} V_{port}) \tag{4}$$

From the above scheme, the voltages and currents in the entire power grid can be solved in the following steps:

- *Step 1*: Generate the hierarchical model, find A_i, S_i for all local grids.
- *Step 2*: Generate modified nodal equations of global grid with all A_i, S_i, and obtain global grid voltages by (3).
- *Step 3*: For each local grid, obtain internal nodes voltages using the ports voltages by (4).

III. Improved Hierarchical Analysis Approach

The HA method reduces running time significantly and provides accurate solution. However, it is not efficient in homogeneous chips and has the constrain that the local grids should not contain voltage sources, which may be contained in homogeneous chips. Section II.B, II.C will introduce IHA method to handle these two drawbacks respectively.

A. Improved Hierarchical Analysis with Voltage Source

In this part, we expand the HA method to chips with voltage sources. Consider the local grid circuit shown in Fig. 2, which contains one voltage sources V_{dd} in node 1 and one current source J in node 4. Node 2 is port node which receives current I from global grid. Here we assume the resistance between voltage source and node 1 is zero.

First, we generate modified nodal equations for this circuit.

$$\begin{bmatrix} G_1 & -g_{12} & -g_{13} & 0 & 1 \\ -g_{12} & G_2 & 0 & -g_{24} & 0 \\ -g_{13} & 0 & G_3 & -g_{34} & 0 \\ 0 & -g_{24} & -g_{34} & G_4 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \\ I_v \end{bmatrix} = \begin{bmatrix} 0 \\ I_{port} \\ 0 \\ J \\ V_{dd} \end{bmatrix} \tag{5}$$

Then, based on [2], we can focus on the unknown node voltages by eliminating V_1, I_v which are directly related to voltage source. The elimination result is shown below.

$$\begin{bmatrix} G_3 & -g_{34} & 0 \\ -g_{34} & G_4 & -g_{24} \\ 0 & -g_{24} & G_2 \end{bmatrix} \begin{bmatrix} V_3 \\ V_4 \\ V_2 \end{bmatrix} = \begin{bmatrix} 0 + g_{13} V_{dd} \\ J \\ I_{port} + g_{12} V_{dd} \end{bmatrix} \tag{6}$$

Therefore, the improved hierarchical model is:

$$\begin{bmatrix} G_{11}^{'} & G_{12}^{'} \\ G_{12}^{'T} & G_{22}^{'} \end{bmatrix} \begin{bmatrix} V_{int} \\ V_{port} \end{bmatrix} = \begin{bmatrix} J_{int} + J_{int,v} \\ I_{port} + J_{port} + J_{port,v} \end{bmatrix} \tag{7}$$

$$\begin{aligned} I_{port} =& (G_{12}^{'T} G_{11}^{'-1} (J_{int} + J_{int,v}) - (J_{port} + J_{port,v})) \\ &+ (G_{22}^{'} - G_{12}^{'T} G_{11}^{'-1} G_{12}^{'}) V_{port} \end{aligned} \tag{8}$$

Where $J_{int,v}$ and $J_{port,v}$ represent the voltage-caused current effect to internal nodes and port nodes respectively. In this way, voltage source can be contained into the HA method and the dimension of conductance matrix is reduced slightly.

B. Improved Hierarchical Analysis in Homogeneous Chips

When dealing with homogeneous chips, we assume that the topology structure and resistance value differences between each core are small. The global grid needs to be properly designed in order to provide enough power for each local grid.

Consider a circuit with k similar independent cores, and each core contains n nodes. The k similar cores are modeled as local grid and other parts are modeled as global grid. To start with, the improvement in *Step 1* is the reuse of matrix A, S in (3). Average conductance matrix $\overline{G}$ and average current resources vector $\overline{J}$ will be generated through:

$$\overline{G} = \frac{1}{k}\sum_{i=1}^{k} G_i, \quad \overline{J} = \frac{1}{k}\sum_{i=1}^{k} J_i, \quad \overline{V} = \frac{1}{k}\sum_{i=1}^{k} V_i \tag{9}$$

Then, $\overline{G}, \overline{J}$ will be stamped into Eq.(1)(2) to obtain $\overline{A}, \overline{S}$ to replace all $A_j, S_j, j \in \{1, 2...k\}$ in (3), without generating A_j, S_j for local grids. This strategy can achieve almost k times speedup in macromodeling and reduce the memory usage.

After the modification in *Step 1*, the global grid voltage V_i, $i \in \{0, 1, ..k\}$ can be solved by (3) in *Step 2*, and $\overline{V}$ is calculated by (9). However, the accuracy will be affected by the differences between each core. Therefore, an improvement in *Step 3* is needed to improve accuracy.

In *Step 3*, after obtaining $\overline{V}, \overline{G_{11}^{-1}}, \overline{G_{12}}$ and $\overline{J_{int}}$ by (1)(3)(9). We first solve $\overline{U_{int}}$ by (4) as a initial solution. Then, we apply a Krylov-subspace iteration algorithm called 'Generalized minimal Residual Method (GMRES)'[8] to find the accurate solution for all local grids by the usage of this initial solution $\overline{U_{int}}$ and the real G, J of each core by (4). Since the difference of each local grid is small, the initial solution will be closed to the accurate solution, which is suitable for iteration.

978-1-7281-6235-5/20/$31.00 ©2020 IEEE

TABLE I
IHA IN TYPE 1 HOMOGENEOUS CHIPS

Benchmark	*Sp_all*	*Sp_1*	*Sp_3*	*MaxE(mV)*	*MAE(mV)*
M1	2.04X	2.08X	1.83X	0	0
M2	4.56X	4.75X	3.66X	0	0
M3	5.88X	6.02X	5.88X	0	0
M4	7.17X	7.36X	7.17X	0	0

TABLE II
IHA IN TYPE 2 HOMOGENEOUS CHIPS

Benchmark	*Sp_all*	*Sp_1*	*Sp_3*	*MaxE(mV)*	*MAE(mV)*
M1	1.93X	2.10X	1.26X	0.658	7.12e-2
M2	3.30X	3.87X	1.82X	0.935	1.18e-1
M3	4.64X	6.04X	2.03X	1.019	1.16e-1
M4	5.21X	7.00X	2.25X	1.184	1.09e-1

TABLE III
IHA IN TYPE 3 HOMOGENEOUS CHIPS

Benchmark	*Sp_all*	*Sp_1*	*Sp_3*	*MaxE(mV)*	*MAE(mV)*
M1	1.62X	1.97X	0.82X	1.82	8.9e-2
M2	2.60X	3.85X	0.94X	1.59	9.1e-2
M3	3.73X	5.92X	1.24X	1.56	7.8e-2
M4	4.03X	7.64X	1.12X	1.56	6.8e-2

IV. Experimental Results

A. Experimental Setup

Since the existing benchmark *IBM/THU Power Grid Benchmark* do not contain homogeneous structures, we design a new benchmark with homogeneous structure based on *IBM Power Grid Benchmark*. The new benchmark contains 4 homogeneous multi-core chips (*M1–M4*), which contain 2, 4, 6, 8 *ibmpg1 Vdd_net* power supply network respectively. Each network contains $12K$ nodes and *VDD* is $1.8V$. All experiments are performed in Intel(R) Core(TM) i7-7500U CPU @ 2.70GHz 2.90 GHz and 8.00GB RAM.

B. Results in type 1 homogeneous chips

In this part, we use both HA and IHA methods to analysis those benchmarks (i.e. type 1 homogeneous chips). The results are shown in Table I. *Sp_all* denotes the runtime speedup of IHA compared to HA. *Sp_1, 3* denote the runtime speedup of IHA over HA in *Step 1, 3*. They are calculated by:

$$Sp\_1,3 = \frac{Runtime\ in\ Step\ 1,3\ with\ HA}{Runtime\ in\ Step\ 1,3\ with\ IHA} \tag{10}$$

Mean Absolute Error (MAE) and *Max Error* (MaxE) are defined in (11)(12). N is the number of nodes, $\widehat{y_i}, y_i$ are the voltage value from IHA and *HSpice* respectively in node i.

$$MAE = \frac{\sum_{i=1}^{N} \|\widehat{y_i} - y_i\|}{N} \tag{11}$$

$$MaxE = max(\|\widehat{y_i} - y_i\|), i = 1\ to\ N \tag{12}$$

Table I shows that compared to HA method, IHA can achieve up to 7.17X speedup without compromising any accuracy.

C. Results in type 2,3 homogeneous chips

Since the homogeneous cores in chips may not exactly be identical, in this part, perturbation is added to these identical cores in Section IV.B to design type 2 homogeneous chips. The perturbation changes resistance value in these cores. For each core, 10% resistors are chosen randomly as perturbation source by being multiplied by $1+\alpha$, where α shows a random distribution at the range of $[-5\%, 5\%]$. Therefore, these cores have identical structures and similar resistance value (i.e. type 2 homogeneous chips).

The simulation results are shown in Table II. We find that in this case IHA achieves up to 5.21X speedup and achieves 99% accuracy at the same time. Furthermore, *Sp_3, Sp_all* and *MaxE* can be controlled by the selection of convergence conditions in GMRES. More strict convergence condition improves accuracy but results in the reduction of *Sp_3, Sp_all*, and vice versa.

Since all the resistance value in *ibmpg1* is less than 1Ω, therefore to mimic the possible structure change in a homogeneous circuit, we randomly choose 0.1% resistors from the local grid of each core, and set their resistance values to be *1000*Ω to create open branches (i.e. type 3 homogeneous chips). The simulation result is shown in Table III. IHA can achieve up to 4X speedup while keeping the error within 1.1%.

Table I–III show that *Sp_1* nearly remains unchanged for the three cases. Besides, the increase of *Sp_1* consistently matches the increase of the number of cores. However, *Sp_3* in Table II, III is lower than *Sp_3* in Table. I, which is caused by two reasons. One is the difference between each core. Comparing TABLE. II, III, we find that structure change has larger impact than value change. Another reason is the local grid size, the iterative GMRES algorithm does not have enough superiority than direct solving method for small scale matrices.

V. Conclusion

In this work, we have proposed an improved hierarchical analysis for IR drop analysis of homogeneous chips. We improve the Hierarchical Analysis and make it suitable for homogeneous chips. Our experiments show that for the circuit with 8 homogeneous but not identical cores, IHA can achieve 5.21X speedup in the whole analysis with 99% accuracy.

References

[1] C.-W. Ho *et al.*, "The Modified Nodal Approach to Network Analysis," *TCAS*, vol. 22, no. 6, pp. 504–509, 1975.

[2] J. N. Kozhaya *et al.*, "A Multigrid-like Technique for Power Grid Analysis," *TCAD*, vol. 21, no. 10, pp. 1148–1160, 2002.

[3] W. Ye *et al.*, "Power Grid Reduction by Sparse Convex Optimization," in *ISPD*, 2018, pp. 60–67.

[4] Z. Xie *et al.*, "PowerNet: Transferable Dynamic IR Drop Estimation via Maximum Convolutional Neural Network," in *ASP-DAC*, 2020, pp. 13–18.

[5] Y.-C. Fang *et al.*, "Machine-learning-based Dynamic IR Drop Prediction for ECO," in *ICCAD*, 2018, pp. 1–7.

[6] M. Zhao *et al.*, "Hierarchical Analysis of Power Distribution Networks," *TCAD*, vol. 21, no. 2, pp. 159–168, 2002.

[7] J. Yue *et al.*, "14.3 A 65nm Computing-in-Memory-Based CNN Processor with 2.9-to-35.8 TOPS/W System Energy Efficiency Using Dynamic-Sparsity Performance-Scaling Architecture and Energy-Efficient Inter/Intra-Macro Data Reuse," in *ISSCC*, 2020, pp. 234–236.

[8] Y. Saad, "A Flexible Inner-outer Preconditioned GMRES Algorithm," *SIAM J. Sci. Comput*, vol. 14, no. 2, pp. 461–469, 1993.

978-1-7281-6235-5/20/$31.00 ©2020 IEEE