Thermal-aware Power Network Design for IR Drop Reduction in 3D ICs*

Zuowei Li[1], Yuchun Ma[1], Qiang Zhou[1], Yici Cai[1], Yu Wang[1], Tingting Huang[2], Yuan Xie[3]

[1]Tsinghua National Laboratory for Information Science and Technology
Department of Computer Science and Technology, University, Beijing, China
[2]National Tsinghua University, Taiwan, China
[3]Pennsylvania State University, USA
lizuoweirain@gmail.com myc@mail.tsinghua.edu.cn

Abstract–Due to the high integration on vertical stacked layers, power/ground network design becomes one of the critical challenges in 3D IC design. With the leakage-thermal dependency, the increasing on-chip temperature in 3D designs has serious impact on IR drop due to the increased wire resistance and increased leakage current. Power/ground (P/G) TSVs can help to relieve the IR drop violation by vertically connecting the on-chip P/G networks on different layers. However, most previous work only fulfills a margin of the full potential of PG TSVs planning since the P/G grids are restricted in a uniform topology. Besides, the overlook of resistance variation and leakage current will make the results less accurate. In this paper, we present an efficient thermal-aware P/G TSVs planning algorithm based on a sensitivity model with temperature-dependent leakage current considered. The proposed method can overcome the limitation of uniform P/G grid topology and make full use of P/G TSVs planning for the optimization of P/G network by allowing short wires to connect the P/G TSVs to P/G grids in non-uniform topology. Moreover, with resistance variation and increased leakage current caused by high temperature in 3D ICs, more accurate result can be obtained. Both the theoretical analysis and experimental results show the efficiency of our approach. Results show that neglecting thermal impacts on power delivery can underestimate IR drop by about 11%. To relieve the severe IR drop violation, 51.8% more P/G TSVs are needed than the cases without thermal impacts considered. Results also show that our P/G TSV planning based on the sensitivity model can reduce max IR drop by 42.3% and reduce the number of violated nodes by 82.4%.

I. INTRODUCTION

In recent years, through-silicon via (TSV) based 3D IC designs emerge as a promising solution because of many benefits such as wire length reduction, smaller form factors, and heterogeneous integration [1-2]. The inter-layer TSVs can carry signals (***signal vias***) or help heat dissipation (***thermal vias***), and connect the power/ground (P/G) networks (***P/G vias***). Considering the thermal issues in 3D designs, most of the power-hungry modules are placed near to the bottom heat sink. The imbalance of power distribution among layers results in quite different distributions of IR drops on different layers.

(a) layer1 (near to heat sink) (b)layer2

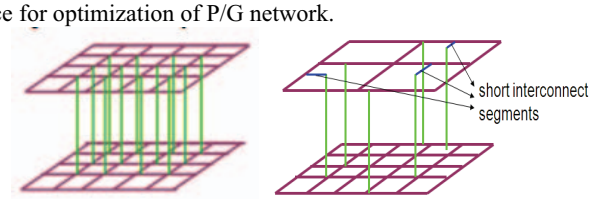
Figure 1. Voltage distribution in different layers*

Fig.1 shows the voltage distribution on two stacked layers without any P/G vias between them. We can see that the IR drop on layer1 is more critical than on layer2. However, the imbalance of voltage distribution on different layers may provide an opportunity to relieve IR drop violation by vertically connecting separated P/G grids with P/G vias. Instead of routing more P/G trunks on individual layers to avoid violations, P/G vias between layers can be used to construct a 3D P/G network in which the IR drop distribution on multiple layers can be optimized at the same time. Since TSVs go through the silicon layer, the distribution of P/G TSVs is influenced by not only the power distribution but also the whitespace distribution around circuit modules. Therefore, it is necessary to plan the P/G TSV distribution in the floorplanning stage and provide flexibility to optimize the P/G network.

Prior works on 3D P/G Network Design. Currently, only a few of works have explored 3D P/G network design. The related works can be divided into two categories by their optimization objectives. The first category focuses on the reduction of power supply noise [5-13]. For example, decoupling capacitors (decaps) can be used during placement to reduce the power supply noise [5-9]. In addition to using decaps to reduce P/G noise, the density of P/G TSVs can be adjusted to reduce the power supply noise [13]. The second category considers the IR drop in 3D ICs during early stages. A 3D floorplan and P/G network co-synthesis tool was proposed to explore the floorplanning and P/G network topology design with IR drop constraint considered [14]. In this paper, we further explore P/G grid design during early stages, so we focus on the optimization of IR Drop of P/G grid to guide grid design.

The limitation of Prior Work. The existing work on the P/G network in 3D ICs is based on the evenly distributed P/G TSVs or pre-determined distribution of P/G TSVs. Hence, TSV planning is not fully adjusted to help the optimization of power delivery. Since the number of P/G TSVs is limited due to the routing congestion and yield reduction, the lack of P/G TSV planning may limit the design flexibility and then cause high cost or even failure in design. Meanwhile, the topology of 3D P/G network also plays a significant role in power delivery. A good topology will effectively reduce the IR drop with less wiring cost.

Normally, there are two types of topology in 3D P/G network (Fig.2): (1) Uniform grid, in which the pitches of grids on all layers are the same. (2) Non-uniform grid, in which each layer has different P/G grid pitches. Previous work on 3D P/G network [13,14] restricts the uniform grid or aligned trunks on different layers by restricting P/G TSVs insertion only at the cross points between P/G networks on two adjacent layers. But due to the different IR drop distribution, the uniform distribution of grids may need more P/G wire resources. The limitation of inserting TSVs in uniform grids also narrows the space for optimization of P/G network.



(a) (b)

Figure 2. 3D uniform Grid(a) and non-uniform(b)

* This paper is supported by National Natural Science Foundation of China under Grant No. 60606007, No. 61076035, No.61176035 and TNList cross discipline Foundation.

978-1-4673-0772-7/12/$31.00 ©2012 IEEE

In this paper, we connect P/G TSVs to P/G grids with short interconnect segments or short wires so that the locations of P/G TSVs are much more flexible. Hence, not only the P/G network can be optimized, but the layout of the modules has much more flexibility.

As the continuing increase of power density, the temperature distribution varies a lot among layers in 3D chips. The wire resistance will be influenced by increased local temperature which has been analyzed in 2D P/G networks [15]. Besides the wire resistance, another thermal dependent design factor, leakage current is always neglected in P/G analysis. It is shown that at 130nm with supply voltage of 1.2-1.3V, leakage current represents 10-30% of active power, and for 70nm with supply voltage less than 1.0V, over 50% of a chip's power dissipation may be due to leakage current [16]. Leakage current is no longer negligible in 3D designs due to the high temperature. Though the inter-dependent relation between leakage and temperature are thoroughly investigated in both 2D and 3D floorplanning [17] [18], none work about the impact of leakage current on power delivery in 3D designs was proposed.

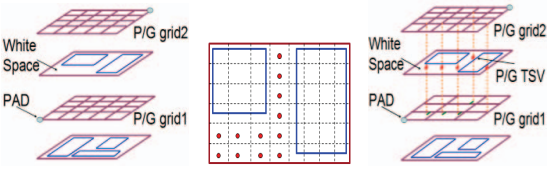
In this paper, we first study the impact of P/G TSV sensitivity on the total IR drop, considering the thermal impact on wire resistance and leakage current, and then further propose the co-optimization flow of P/G TSVs with the topology of P/G grids. More specifically, we make the following contributions:

- A sensitivity model evaluating the IR drop impact of P/G TSV insertion among layers is proposed to guide the P/G TSV planning. According to the sensitivity of each P/G TSV, we can optimize the P/G TSV insertion to amend the IR drop violations with least TSVs.
- Short wires are adopted to connect the P/G TSVs to P/G grids in non-uniform grids to ensure more flexible locations for P/G vias insertion. With this flexibility, the topologies of P/G grids can be optimized freely without much restriction.
- We take the thermal issues into IR drop analysis in 3D IC considering the thermal impact on both wire resistance and leakage current. Our analysis results show that the temperature impacts are no longer negligible in 3D designs.

The rest of the paper is organized as follows: Problem formulation is given in Sect. II. In Sect. III and IV, the thermal aware IR drop analysis and the P/G TSV planning approach are proposed. Experimental results are given in Sect. V. Finally, Sect. VI concludes this paper.

II. PROBLEM FORMULATION

If multiple layers are stacked, such as 2 layers (2 metal layers and 2 silicon layers) shown in Fig.3 (a), in order to help the IR drop dissipation, some P/G TSVs can be inserted in the whitespace between blocks. Some short interconnect segments connecting the P/G TSVs to P/G networks are expected if P/G TSVs are not aligned to P/G trunks (Fig.3 (c)). By dividing the silicon layer between two adjacent metal layers into virtual grids which are called sensitivity estimation grids in this paper, we can find out all the grids which have whitespaces between modules on this layer, and then we assume the center of these grids as the candidate positions which TSVs can be inserted into (Fig.3 (b)).



(a) (b) (c)

Figure 3. 3D P/G mesh with TSV

Thermal-aware 3D P/G TSV planning problem can be formulated as follows: given the feasible packing with modules on multiple layers, with the thermal issue considered, our goal is to optimize the P/G TSVs distribution and the topology of P/G network on multiple layers so that the IR drop can be minimized with least TSVs and P/G wires resource. To describe the problem, following notations listed in Table I will be used:

Table I. Notions used in this paper

n	Number of nodes.
m	Number of P/G TSVs.
$G \in R^{n\times n}$	Conductance matrix
$V \in R^{n\times 1}$	Voltage vector for all nodes.
$I \in R^{n\times 1}$	Current vector.
VDD	Supply voltage.

The P/G network on each metal layer can be described as a non-uniform mesh structure which is composed of a uniform mesh with certain size such as *100×100* and some short interconnect segments. The P/G TSVs between adjacent layers can be allocated at the whitespace around modules and be connected to P/G grids by short interconnect segments. With the whole P/G network, the IR drop of each node can be obtained by solving a linear system. In this paper, we use the resistive model and the static current source model for P/G networks. The objectives are as follows:

• The minimization of total IR drop:

For node *i* in the mesh, it has an IR drop with its corresponding voltage V_i, we denote it as $V_{drop_i} = VDD - V_i$, and then we can get a vector $V_{drop} = VDD - V$ for all nodes. Our goal is to minimize the total IR drop on the whole chip by planning the P/G networks with minimal P/G TSVs inserted. Since the quadratic sum may bring convenience to compute the sensitivity of inserting each P/G TSV, we define the total IR drop ϕ as the quadratic sum of difference between VDD and all nodes' voltage, which can be formulated as:

$$min\ \phi = {V_{drop}}^T \cdot V_{drop} \quad (1)$$

In this paper, we define the sensitivity of a P/G TSV as the derivative of ϕ to the conductance of this TSV as described in Sect. IV.

• The minimal wiring resources:

With different sizes of P/G meshes and the different wires connecting P/G TSVs with P/G grids, the metal resources used by P/G network might be quite different for various designs. Therefore, with the objective to reduce IR drop violations as much as possible, the wiring resources of P/G network A_{PG} should be minimized, where A_{PG} includes both the area of P/G trunks (A_{trunk}) and the area of short interconnect segments (A_{short}). We denote it as:

$$A_{PG} = A_{trunk} + A_{short} \quad (2)$$

III. THERMAL EFFECTS ON P/G NETWORK

The temperature variation on the power grid can cause significant changes in the interconnect resistances, and therefore can substantially increase the IR drops in the power grid. Even in 2D IC design, as the current densities of interconnects increase, the effects of self-heating become more significant and cannot be ignored anymore [15]. Hence, in 3D designs, with much higher temperature and large temperature difference between layers, the effects of temperature on the IR drop analysis is necessary to be considered. In addition, thermal-leakage dependency may cause the increase of leakage current. With more absorbed current, the IR drop distribution will be influenced a lot.

A. Leakage aware current model

In CMOS digital circuits, power dissipation consists of dynamic and leakage components. In current CMOS technologies, the sub-threshold leakage current I_{sub} is much larger than the other

leakage current components [20]. The sub-threshold leakage power/current increases exponentially with temperature increases.

$$\frac{I_{sub}}{I_{sub}{}^{0}} = \alpha \times e^{\beta \times (T_i - T_{base})} \quad (3)$$

Where T_i is the temperature of module m_i. α and β are empirical factors that have different values for different technologies. T_{base} is the reference temperature at which α, β and the initial $I_{sub}{}^{0}$ are defined. Typically, $\alpha = 1\times10^5\ W/m^2$, $\beta = 0.025$ for 130nm [21]. T_{base} is defined to 27°C with the initial $I_{sub}{}^{0}$ is about 5% of total current including both dynamic current and leakage current in the experimental environment.

Leakage is important not only in the standby mode but also in the active mode of operation. In fact, the leakage in the active mode (active leakage) is significantly larger due to higher die temperature in the active mode and the exponential temperature dependence of sub-threshold leakage [22]. The standby leakage may be made significantly smaller than the active leakage by changing the body bias conditions or by power-gating. In the active mode of operation (high temperature), sub-threshold leakage is the dominant component of leakage. Therefore, for IR drop estimation, with leakage current considered, the absorbed current on each pin may be computed as:

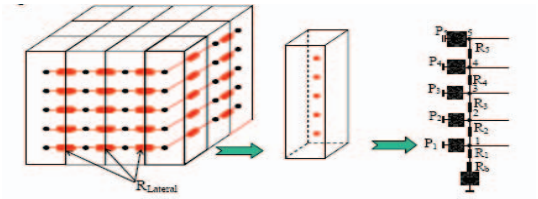
$$I = I_{dyn} + I_{sub} \quad (4)$$

B. Thermal-leakage dependent model

As shown in Fig.4, the 3D circuit stacking is divided by a two-dimensional array of tile stacks and each tile stack is composed of several vertically-stacked tiles. The lateral thermal resistances, $R_{lateral}$ are used to connect those tile stacks, a thermal resistor R_i is modeled for the i-th device layer, while thermal resistance of the bottom layer and silicon substrate is modeled as R_b as shown in Fig.4(c). Similar to Ref. [13, 23], a tile stack is modeled as a resistive network. The isothermal bases of room temperature are modeled as a voltage source. A current source is presented at every node to represent the heat sources. The system can be spatially discretized and be solved using the following equation to determine the steady-state thermal profile as a function of power profile:

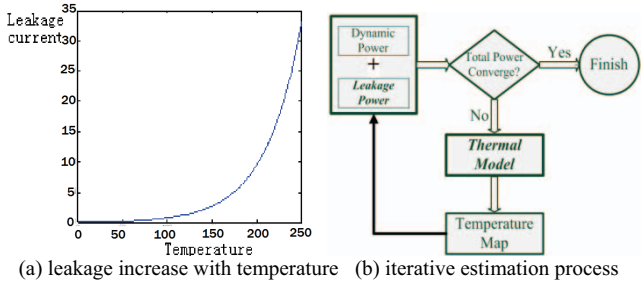
$$T = PA^{-1} \quad (5)$$

Where A is an $N\times N$ sparse thermal conductivity matrix. T and $P(T)$ are $N\times1$ temperature and power vectors.



(a) Tiles Stack Array (b) Single Tile Stack (c)Tile Stack Analysis

Figure 4. Resistive thermal model for a 3D IC [23]



(a) leakage increase with temperature (b) iterative estimation process

Figure 5. Relation between temperature and leakage current

Since we have leakage power considered, the total power includes not only the dynamic power, but also the leakage power.

$$P = P_{dynamic} + P_{leakage} \quad (6)$$

Where $P_{leakage} = I_{sub} \times VDD$.

On the other hand, leakage power will increase the total power consumption which in turn increase the power vector P in (5). Therefore, the thermal profile can be obtained by iteratively conducting thermal analysis and leakage power estimation until convergence as shown in Fig.5. This usually requires about 2-3 iterations.

C. Thermal-aware wire resistance model

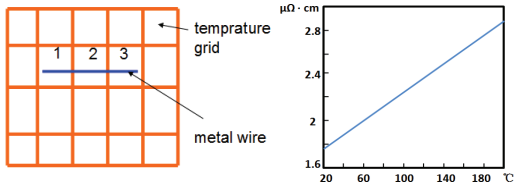
With the thermal analysis, we can evaluate the thermal effect on wire resistivity accordingly. Similar to the analysis method used in [15] for 2D ICs, for a metal wire with temperature profile $T(x)$ along its length, the resistivity $\rho(x)$ at point x will change linearly with temperature as:

$$\rho(x) = \rho_0(1 + \beta \cdot T(x)) \quad (7)$$

Where ρ_0 is the resistivity at reference temperature (27◦C for example) and β= 0.0039/◦C is the temperature coefficient of resistance. As shown in Fig.6 (a), the resistance of a wire in thermal grids can be figured out:

$$R_{wire} = \frac{\rho_1 l_1 + \rho_2 l_2 + \rho_3 l_3}{A} \quad (8)$$

Where $\rho(i) = \rho_0(1 + \beta \cdot T(i))$, for i=1, 2, 3. $T(i)$ is the temperature of $grid_i$, l_i is the length of metal wire in $grid_i$. A is sectional area of the wire. In the actual calculation, square resistance is used instead of A. Fig.6 (b) shows the linear relationship between the resistivity of a metal wire with the temperature.



(a) metal wire in thermal grids (b) temprature impact on resistivity

Figure 6. Metal wire under thermal efffects

IV. P/G TSV PLANNING

A. Sensitivity Model

Based on the regular mesh structure shown in Fig.3, P/G wires are modeled as resistors. A P/G pin in a hard module is modeled as a current source. Traditionally, the static analysis of a P/G network is formulated:

$$G \cdot V = I \quad (9)$$

Where G is the conductance matrix for the resistors, V is the vector of node voltages, and I is the vector of current loads. The dimensions of I and V are equal to the number of nodes in the P/G network. With the additional P/G TSVs, the conductance matrix can be expressed for two layer case as following:

$$G = \begin{pmatrix} G_1 + G_{tsv} & -G_{tsv} \\ -G_{tsv} & G_2 + G_{tsv} \end{pmatrix} \quad (10)$$

Where G_1 is the conductance matrix on layer$_1$ without TSVs and G_2 is the conductance matrix on layer$_2$ without TSVs, and G_{tsv} can be written as:

$$G_{tsv} = diag(\dots, 0, \dots, g_{tsv_1}, \dots, 0, \dots, g_{tsv_2}, \dots, g_{tsv_m}, \dots) \quad (11)$$

Where g_{tsv_i} is the conductance of tsv_i. It is similar for multiple layers by applying (10) pair-wisely between adjacent layers.

Once the P/G network without P/G TSV is created on each layer, it needs to be analyzed to find the optimal position to insert P/G TSV so that the total IR drop violation can be minimized. Intuitively, if there is a voltage difference between two vertically- aligned nodes on two adjacent layers, a P/G TSV might be useful to connect them so that their voltages would come to a balance and the node with lower voltage would be made up, consequently the total IR drop will

be reduced. Besides the difference of voltage values on adjacent layers, the matrix structure of the P/G network also influences the effect of individual PG TSVs.

In this paper, we adopt the sensitivity of each P/G TSV on the total IR drop to guide TSV insertion in white spaces [24]. We define the sensitivity of each P/G TSV on the total IR drop as follows:

$$Sensitivity(g_{tsv_i}) = \frac{\partial \phi}{\partial g_{tsv_i}} \tag{12}$$

Where ϕ is defined by (1), and we can easily get

$$\frac{\partial \phi}{\partial g_{tsv_i}} = 2 \cdot {V_{drop}}^T \cdot \frac{\partial V_{drop}}{\partial g_{tsv_i}} = -2 \cdot {V_{drop}}^T \cdot \frac{\partial V}{\partial g_{tsv_i}} \tag{13}$$

In formula (13), ${V_{drop}}^T$ is known, then in order to calculate the sensitivity of tsv_i, we must first compute the deviation of V to g_{tsv_i}.

According to the circuit equation using MNA (Modified Node Approach) in (9) we can calculate the derivative of V to g_{tsv_i} (i=1, 2,…,t) in above equation, then we have

$$\frac{\partial G}{\partial g_{tsv_i}} \cdot V + G \cdot \frac{\partial V}{\partial g_{tsv_i}} = \frac{\partial I}{\partial g_{tsv_i}} \tag{14}$$

Since $\frac{\partial I}{\partial g_{tsv_i}} = 0$, then we can get

$$\frac{\partial V}{\partial g_{tsv_i}} = -G^{-1} \cdot \frac{\partial G}{\partial g_{tsv_i}} \cdot V \tag{15}$$

Then we can express (13) as

$$\frac{\partial \phi}{\partial g_{tsv_i}} = 2 \cdot {V_{drop}}^T \cdot G^{-1} \cdot \frac{\partial G}{\partial g_{tsv_i}} \cdot V \tag{16}$$

Since we know that $\frac{\partial \phi}{\partial g_{tsv_i}}$ is a scalar, we can transpose both side of formula (16). Then we have:

$$\frac{\partial \phi}{\partial g_{tsv_i}} = 2 \cdot \left(\frac{\partial G}{\partial g_{tsv_i}} \cdot V\right)^T \cdot G^{-1} \cdot V_{drop} \tag{17}$$

We can first solve an auxiliary linear equation:

$$G \cdot X = V_{drop} \tag{18}$$

And we can get

$$X = G^{-1} \cdot V_{drop} \tag{19}$$

And (17) can be expressed as below:

$$\frac{\partial \phi}{\partial g_{tsv_i}} = 2 \cdot \left(\frac{\partial G}{\partial g_{tsv_i}} \cdot V\right)^T \cdot X \tag{20}$$

According to the form of G discussed previously (formula (10)), then we can get

$$\frac{\partial G}{\partial g_{tsv_i}} = \begin{pmatrix} \cdots & & \cdots & & \cdots \\ & 1 & \cdots & -1 & \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ & -1 & \cdots & 1 & \\ \cdots & & \cdots & & \cdots \end{pmatrix} \tag{21}$$

Where 1 appears at (a, a) and (b, b), -1 appears at (a, b) and (b, a). a and b are two end nodes of tsv_i. Finally we can get the sensitivity of tsv_i, which can be expressed as:

$$\frac{\partial \phi}{\partial g_{tsv_i}} = 2 \cdot \left((V_a - V_b) \cdot X_a + (V_b - V_a) \cdot X_b\right) \tag{22}$$

Where V_a and V_b are the voltage values of two end nodes of tsv_i. X_a and X_b are the a-th and b-th element of vector X.

It is shown from formula (22) that the sensitivity of one TSV is associated with the voltage difference between two end nodes of the TSV. But there is still another factor such as X also influences the sensitivity of the TSV. We can find From (18) that the term X is defined by the topology of P/G mesh in terms of G^{-1}. Formula (22) also tells us that it is not true that the more TSVs, the better IR drop. If the sensitivity value of tsv_i is positive and we add a TSV to the position of tsv_i, the total IR drop ϕ will increase, and power integrity will become worse. So we can't insert TSV where the sensitivity is positive. On the other hand, if the sensitivity value of TSV_i is negative and the larger its absolute value is, the more ϕ will be reduced after insertion.

Besides the power distribution, another constraint of PG TSV insertion during floorplan stage is that only whitespace can be used to allocate TSVs since there might not be spaces left for additional TSVs in macros or IP blocks. As discussed in section II, we divide the silicon layer between two adjacent metal layers into virtual grids, and then we can find out all the grids which have whitespaces between modules, we assume the center of these grids as the candidate positions which TSVs can be inserted into. We can simply get the sensitivity of a grid based on the average value of sensitivities of TSVs in the grid. The size of sensitivity estimation grid used in this paper is *50×50*. After we get the sensitivity of each candidate TSV position, we surely know where needs TSVs most. Then we sort the grids by the sensitivities and insert certain number of TSVs with the most negative sensitivities to the grids. As the TSVs inserted, the sensitivity and capacitance of each grid will be updated for the next round of TSV insertion. The TSV insertion with sorted sensitivities will be iteratively processed until no further decrease of ϕ/n can be made. The P/G TSV planning flow is described below:

Algorithm 1. P/G TSV planning
Initialize parameters ε_1, ε_2;
$\phi = old\_\phi = 0$;
Sweep the whitespaces using sensitivity estimation grids;
Add short interconnect segments to each layer if possible;
Calculate the sensitivity of each candidate TSV;
While ($\phi/n > \varepsilon_1$ *and* $fabs(\phi - old\_\phi)/n > \varepsilon_2$) Do
Select a grid with most negative sensitivities and with enough capacity for one TSV;
Insert a TSV into this grid;
Solve linear system $G \cdot V = I$;
$old\_\phi = \phi$;
Calculate the total IR drop ϕ;
Update the sensitivity of each candidate TSV;
End.

B. *Optimization of P/G grid topology*

Besides the insertion of P/G TSVs, the topology of P/G grids also influences the total IR drop. As we discussed previously, a good P/G network topology may have a good effect on power delivery with less wiring cost, or even bring us a faster convergence to P/G TSV planning. Furthermore, the insertion of P/G TSVs cannot eliminate the violated nodes if the topology of P/G grid is non-desired and the IR drop is too severe. However, the improvement of topology on P/G grid can effectively control the violated nodes. A violated node is defined as that the node's voltage is less than its threshold voltage. We define $V_{vio} = V_{min} - V$ and $\varphi = {V_{vio}}^T \cdot S \cdot V_{vio}$ to evaluate the total violation of the P/G network. S is defined as:

$$S = \frac{1}{2} diag(1 + sgn(V_{vio})) \text{ where } sgn(x) = \begin{cases} 1 & if\ x \geq 0 \\ -1 & else \end{cases}$$

The P/G grid size on each metal layer discussed in this paper varies from *50×50* to *200×200*. We define *InitSize* as *50×50* and *MaxSize* as *200×200*. To meet the IR drop constraints with less wiring sources, we start the exploration from *InitSize* to *MaxSize*. The P/G grid topology optimization flow is described below:

Algorithm 2. P/G grid topology optimization
Construct the P/G mesh with *InitSize* size for each layer
Initialize parameters τ_1, τ_2;
$\varphi = old\_\varphi = 0$;
While ($\varphi/n > \tau_1$ *and* $fabs(\varphi - old\_\varphi)/n > \tau_2$) Do
P/G TSV plan;
Calculate V after inserting P/G TSVs;
$old\_\varphi = \varphi$;
Calculate the violation and IR drop on each layer and the total violation φ;
Increase the size of the layer which has the worst IR drop and its size does not reach *MaxSize;*
End.

C. *Overall flow for co-optimization*

Based on the previous approaches, the overall flow to construct an optimal P/G network with P/G TSVs can be described as: First,

construct the P/G grid with initial size without TSVs for each layer, and then the initial thermal analysis with leakage power considered and the resistance of each metal wire is calculated. We then create sensitivity estimation grids on each silicon layer and sweep the whitespace between blocks for each layer. The whitespace information can be attached to each grid. Next, the sensitivity of each grid is calculated to guide the TSV insertion. In this step, we actually use the midpoint of a grid instead of the whole grid. After we insert these TSVs to the P/G network, thermal analysis is applied again. Then we calculate the total IR drop under the updated thermal profile with TSVs between layers. The iterative process will be continued until voltages of all nodes are upon the threshold voltage or no more decrement of the total IR drop is detected. Fig.7 shows the overall flow briefly:

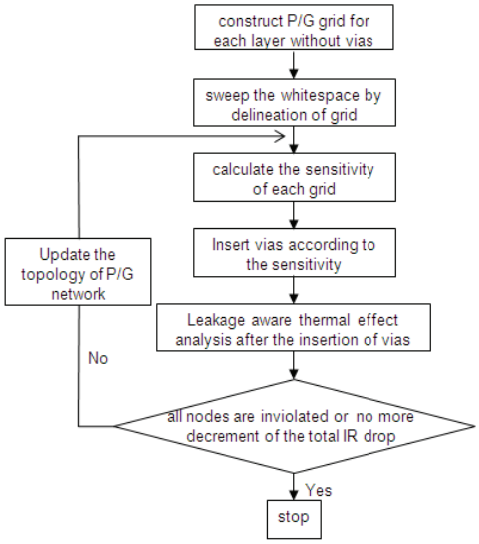


Figure.7 Overall flow of co-optimization

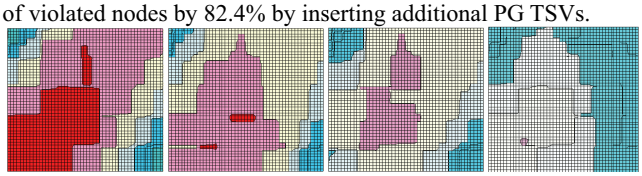
V. EXPERIMENTAL RESULTS

Extensive experiments have been performed to evaluate our proposed method. A set of MCNC benchmarks are considered. These benchmarks have complexities ranging from 33 to 300 modules. The modules in these benchmarks are placed on 4 layers in a 3D stacked design. Our algorithms have been implemented in C++. All experiments are performed on a PC with 3.0 GHz processor and 2 GB of memory.

A. *Effect of TSV sensitivity model*

In this set of tests, we define *100×100* mesh for the P/G network on each layer. To show the correctness and efficiency of our sensitivity model, we devise some other P/G TSV insertion strategies to compare with: 1) *Uniform_Grid*: Insert TSVs evenly in the cross nodes of the P/G grids between every two adjacent layers. The available positions may be limited due to the obstacles caused by circuit modules. 2) *Max_Diff*: Insert TSVs at the location that has the max difference of voltage of every two adjacent layers. It is a heuristic method with a fast run-time, but as we can see from our sensitivity model that the difference of voltage between two adjacent layers is not the only factor to effect on the total IR drop. So this method may not obtain the optimized results with the TSV planning.

An IR drop comparison in layer2 of ami33 between non-TSV insertion, *Uniform_Grid*, *Max_Diff*, and our optimized approach is shown in Fig.8. Two power pads with 1.8V are put at the top left corner and right bottom corner separately; the dark zone denotes the violation of voltage. We can see that if we don't make use of the P/G TSVs (a), many violations will occur, and it will be improved after inserting TSVs with *Uniform_Grid* approach (i.e., P/G TSVs are evenly distributed in the P/G area). But the violations still spread widely (b). If we conduct the TSV insertion according to the difference of voltage of adjacent layers, there will have further improvement in the power delivery(c). But our approach can obtain better balance of IR drop and the violation area is much smaller (d). For (b), (c) and (d), we use the same number of TSVs. The results are consistent with the theoretical analysis that the voltage different is not the only factor to influence the effects of TSV insertion. Our sensitivity model can guide the TSV inserted to the most effective position. The first two main columns shows the effects of our approach without the thermal effect considered. It shows that our approach can reduce Max IR drop by 42.3% and reduce the number of violated nodes by 82.4% by inserting additional PG TSVs.



(a) (b) (c) (d)

Figure 8: IR drop under different TSV-insertion method: (a) IR drop without insertion of P/G TSVs ;(b) IR drop with insertion of P/G TSVs by *Uniform_Grid* ;(c)IR drop analysis after inserting TSV by *Max_Diff* approach ;(d) IR drop with insertion of P/G TSVs by our method (the parameters in Algorithm 1 as $\varepsilon_1 = 1 \times 10^{-6}$, $\varepsilon_2 = 1 \times 10^{-7}$)

B. *Thermal impacts*

In this set of tests, we want to demonstrate the effect of thermal to IR drop. Therefore, two different algorithms with/without thermal considered are compared in Table II. The maximal temperature on the chip without TSV planning can reach 310°C in n300. From the first and third main columns, we can compare the thermal impact when no TSVs are used. It is shown that 11.5% increase of Max IR drop and 10.7% increase of Average IR drop will occur by taking thermal effect into account on P/G network , which means that neglecting thermal effects will lead as high as 11% error on IR drop estimation.

The TSV planning is based on sensitivity model under the same P/G grid topology (i.e. *100×100 P/G mesh*). Since TSVs inserted between layers can also help the heat dissipation, by comparing the temperature before and after P/G TSV insertion, we find that the relatively small number of P/G TSVs reduce the temperature by only about 16% which remains the temperature still very high. When the thermal effect is considered, the increased current coupled with increased wire resistance make the P/G TSV planning a little bit difficult, which will cause 51.8% more P/G TSVs than the case without thermal considered. But our approach can still reduce the Max IR drop from 0.290 to 0.158 which is about 45.5% reduction. It is proved that taking no account of the temperature during P/G analysis will lead to over-design or the risks of violated IR drops due to increased local temperatures. So the effects of thermal on the IR drop analysis are necessary to be considered in 3D ICs.

C. *Overall Results*

P/G network resource is another consideration in 3D IC design. Our co-optimization flow can optimize the topology with P/G TSVs inserted in terms of both IR drop constraints and the optimization of wire resources. In our P/G grid topology optimization algorithm, we set the parameters in Algorithm 2 as $\tau_1 = 1 \times 10^{-5}$, $\tau_2 = 1 \times 10^{-5}$. The overall results of P/G network analysis with thermal and sensitivity model, coupled with P/G topology optimization is shown in Table III. The wire resources in the table include the trunks which construct P/G grid and the short interconnect segments to connect the P/G TSVs to the P/G grid. As we can see from the table, the area

Table II Comparisons of Max IR drop, Average IR drop, number of TSVs among different approaches with/without thermal analysis

Circuit	Without thermal analysis							With thermal analysis									
	Before via planning			After via planning				Before via planning				After via planning					
	Max IR drop(v)	Avg. IR drop(v)	Violated nodes	Max IR drop(v)	Avg. IR drop(v)	Violated nodes	TSV number	Max IR drop(v)	Avg. IR drop(v)	Violated nodes	Temp (°C)	Max IR drop(v)	Avg. IR drop(v)	Violated nodes	Leakage ratio	TSV number	Temp (°C)
ami33	0.269	0.093	12409	0.139	0.089	1497	47	0.294	0.098	12942	180	0.142	0.095	2509	27.0%	63	152
ami49	0.208	0.049	3433	0.137	0.049	196	8	0.253	0.055	5013	234	0.139	0.054	503	35.3%	28	206
n100	0.301	0.084	15538	0.154	0.083	2544	59	0.330	0.094	17064	274	0.168	0.092	3021	38.4%	80	215
n200	0.277	0.081	13035	0.172	0.081	3073	103	0.297	0.093	14253	222	0.179	0.091	3807	35.6%	141	190
n300	0.243	0.067	10042	0.150	0.065	2278	61	0.275	0.076	10922	310	0.161	0.073	2884	40.6%	112	263
Avg.	**0.260**	0.075	**10891**	**0.150**	0.073	**1918**	**56**	**0.290**	0.083	**12039**	245	**0.158**	0.081	**2545**	35.38%	**85**	205

Table III Overall results of P/G network analysis with thermal and sensitivity model, coupled with P/G topology optimization.

Circuit	via planning without topology optimization								via planning with topology optimization							
	Max IR drop(v)	Avg. IR drop(v)	Violated nodes	A_{trunk} (mm^2)	A_{short} (mm^2)	A_{PG} (mm^2)	TSV number	Run time (s)	Max IR drop(v)	Avg. IR drop(v)	Violated nodes	A_{trunk} (mm^2)	A_{short} (mm^2)	A_{PG} (mm^2)	TSV number	Run time (s)
ami33	0.142	0.095	2509	2.996	0.255	3.251	63	22.62	0.128	0.051	5	2.551	0.155	2.706	108	43.37
ami49	0.139	0.054	503	15.556	2.203	17.759	28	11.91	0.120	0.048	0	13.592	0.921	14.513	48	20.43
n100	0.168	0.092	3021	0.995	0.066	1.061	80	32.20	0.131	0.051	12	0.781	0.029	0.81	113	45.50
n200	0.179	0.091	3807	1.147	0.031	1.178	141	56.38	0.133	0.055	17	0.855	0.029	0.884	175	68.58
n300	0.161	0.073	2884	1.309	0.064	1.373	112	44.85	0.130	0.050	8	1.115	0.045	1.16	171	19.29
Avg.	**0.158**	0.081	**2545**	4.401	0.524	**4.924**	**85**	33.59	**0.128**	0.051	8	3.779	0.236	**4.015**	**123**	39.43

Table IV Comparisons of wire resources between Uniform grid topology and Non-uniform grid topology

Circuit	Uniform grid topology optimization			Non-uniform grid topology optimization				
	A_{PG} (mm^2)	TSV number	Run time (s)	A_{trunk} (mm^2)	A_{short} (mm^2)	A_{PG} (mm^2)	TSV number	Run time (s)
ami33	3.292	25	20.13	2.551	0.155	2.706	108	43.37
ami49	17.866	13	15.29	13.592	0.921	14.513	48	20.43
n100	1.193	17	17.54	0.781	0.029	0.81	113	45.50
n200	1.317	41	24.16	0.855	0.029	0.884	175	68.58
n300	1.632	45	8.29	1.115	0.045	1.16	171	19.29
Avg.	**5.06**	**28**	17.08	3.779	0.236	**4.015**	**123**	39.43

of short wires (A_{short}) represent a very small portion of A_{PG}. The result in table III also shows that the co-optimization flow can save 18.5% of the total P/G wire resources with 44.7% additional P/G vias inserted. Compared to the uniform grid in table IV, the non-uniform grid topology we used can reduce IR drop more efficiently with less wire resources. It is also indicated that short wires can provide much flexibility to the whole design (i.e. More P/G TSVs are available to be inserted) but with very little costs on wires. We can see from table IV that a small amount of short wires can bring a significant saving (about 20.7%) of total P/G resources.

VI. CONCLUSIONS

By considering the 3D P/G TSV planning and thermal effects while P/G analysis, the power delivery network can be optimized in terms of IR drop distribution. The sensitivity model proposed in this paper gives the decent guide of the P/G via planning. Both the analysis and experimental results show the effectiveness of the proposed method. Especially the analysis of thermal impacts shows the thermal effects should not be ignored any more in 3D P/G synthesis. Since power integrity is still a critical issue in 3D designs, there are many other works need to be considered such as the P/G TSV planning considering the noise reduction and the co-optimization with decap insertion and wire sizing.

VII. REFERENCES

[1] J. Cong and Y. Zhang. Thermal-Driven Multilevel Routing for 3-D ICs. Proc. of the Asia South Pacific Design Automation Conference, pages 121–126, Jan. 2005.

[2] B. Goplen and S. Sapatnekar. “Thermal Via Placement in 3D ICs,” Proc. of ISPD, pp 167-174, Apr. 2005.

[3] E. Wong, J. Minz, and Sung Kyu Lim. Effective thermal via and decoupling capacitor insertion for 3d system-on-package. Proceedings of Electronic Components and Technology Conference, pages 1795–1801, June 2006.

[4] B. Goplen and S. S. Sapatnekar, “Placement of 3D ICs with thermal and interlayer via considerations,” Proc. DAC, pp. 626–631, 2007.

[5] J.M. Minz, E. Wong, M. Pathak, and S.K. Lim. Placement and routing for 3D system-on-package designs. IEEE Transactions on Components and Packaging Technologies, 29(3):644–657, Sept. 2006.

[6] E. Wong, J. Minz, and Sung Kyu Lim. Multi-objective module placement for 3-d system-on-package. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 14(5):553–557, May 2006.

[7] E. Wong, J. Minz, and Sung Kyu Lim. Effective thermal via and decoupling capacitor insertion for 3d system-on-package. Proceedings of Electronic Components and Technology Conference, pages 1795–1801, June 2006.

[8] J. R. Minz, S. K. Lim, and C.-K. Koh, "3D module placement for congestion and power noise reduction," Proc. GLSVLSI, pp. 458–461, 2005.

[9] P.Zhou , K.Sridharan , Sachin S. Sapatnekar, "Congestion-aware power grid optimization for 3D circuits using MIM and CMOS decoupling capacitors", Proc.ASPDAC, pp.179-184, 2009.

[10] S. Sapatnekar, "Addressing thermal and power delivery bottlenecks in 3D circuit", Proc.ASPDAC, pp.423-428, 2009.

[11] E. Wong, J. Minz, and S. K. Lim, “Decoupling capacitor planning and sizing for noise and leakage reduction,” Proc. ICCAD, pp. 395–400, 2006.

[12] G. Huang et al., “Power delivery for 3D chip stacks: Physical modeling and design implication,” Proc. EPEP, pp. 205–208, 2007.

[13] Hao Yu, Joanna Ho, and Lei He. Simultaneous power and thermal integrity driven via stapling in 3d ics. Proc.ICCAD, pp. 802–808, 2006.

[14] Paul Falkenstern ,Yuan Xie, Yao-Wen Chang, Yu Wang, "Three-Dimensional Integrated Circuits (3D IC) Floorplan and Power/Ground Network Co-synthesis" Prof.ASPDAC, 2010.

[15] Z.Yu, D.F.Wong, "Thermal-aware IR drop analysis in large power grid", Proc.ISQED, pp.194-199, 2008.

[16] H. Su, S. Nassif et al.. “Full Chip Leakage Estimation Considering Power Supply and Temperature Variations”. In Proc. ISLPED, pp. 78–83, 2003.

[17] P. Zhou; Y. Maet al. "3D-STAF: scalable temperature and leakage aware floorplanning for three-dimensional integrated circuits Computer-Aided Design", Proc. ICCAD 2007. pp.590–597, 2007.

[18] A Gupta, ND Dutt, FJ Kurdahi, KS Khouri, “LeAF: A Leakage-Aware Floorplanner”, Proc.ASPDAC, pp.274-279, 2007.

[19] Fallah, F. and M. Pedram, Standby and active leakage current control and minimization in CMOS VLSI circuits. IEICE Transactions on Electronics, 2005. E88-C(4): p. 509--519.

[20] Semiconductor Industry Association, International Technology Roadmap for Semiconductors, http://public.itrs.net/.

[21] W. Huang et al, “The need for a Full-Chip and Package Thermal Model for Thermally Optimized IC Designs”, in Proc. Int. Symp. Low Power Electron. Design, Aug. 2005,. pp. 245–250

[22] S. Lee and T. Sakurai, “Run-time power control scheme using software feedback loop for low-power real-time applications,” Proc. of ASPDAC, pp.381-386, 2000.

[23] J.Cong, J.Wei, and Y.Zhang, “A Thermal-Driven Floorplanning Algorithm for 3D ICs”, in Procceedings of ICCAD, pp.306-313, 2004.

[24] Jiri Vlach, Kishore Singhal , “Computer Methods for Circuit Analysis and Design”, Van Nostrand Reinhold Company, pp.152-170, 1983.