Thermal-aware P/G TSV planning for IR drop reduction in 3D ICs

Zuowei Li a, Yuchun Ma a,*, Qiang Zhou a, Yici Cai a, Yuan Xie b, Tingting Huang c

a Department of Computer Science and Technology, Tsinghua University, Beijing
b Pennsylvania State University, USA
c National Tsinghua University, Taiwan

ARTICLE INFO

Keywords:
Thermal
Power/Ground network
IR drop
TSV
3D IC

ABSTRACT

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. Therefore, it is necessary to consider Power/Ground network design with thermal effects in 3D designs. Though Power/Ground (P/G) TSV can help to relieve the IR drop violation by vertically connecting on-chip P/G networks on different layers, most previous work restricts the uniform P/G grids so that the potential of P/G TSV planning has not been fully explored. In this paper, we present an efficient thermal-aware P/G TSV planning algorithm based on a sensitivity model with temperature-dependent leakage current considered. Non-uniform P/G grid topology is explored to optimize the P/G network by allowing short wires to connect the P/G TSVs to P/G grids. 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%.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Recently, 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 [1–4]. Due to the increased power density, the temperature distribution varies a lot between layers in 3D chips. The imbalance of power distribution among layers results in not only the various temperature distributions, but also quite different distributions of IR drops on different layers. The power grid on a layer is usually a metal mesh. It distributes power and ground voltages from pad locations to chip cells. Currents drawn from the package pads flow on the metal stripes to current loadings. This will cause some voltage drops (IR drops) on the branch because of the metal resistance. In 3D ICs, it tends to appear different IR drops on different layers. Under such circumstance, P/G TSVs can be allocated between adjacent layers to relieve the IR drop violations. According to different fabrication processes, as shown in Fig.1, the P/G-network structures in 3D ICs can be classified into three major schemes: (a) connected (P/G TSVs are used to connect different P/G networks on adjacent layers); (b) non-connected (P/G network on each layer is connected to the package IO directly by wire bonding, there is no directed connection between P/G networks); and (c) a combination of connected and non-connected P/G networks. It is obvious that (c) is the most flexible among the three structures.

In this work, we focus on the structure as shown in Fig.1(c). Therefore, the P/G network design in 3D ICs is no longer several separated 2D designs on an individual layer, but a 3D P/G network should be constructed with P/G TSVs so that the IR drop distribution on multiple layers can be optimized at the same time. Currently, only a few 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 floor-planning and P/G network topology design with IR drop constraint considered [14]. 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. In this paper, we further explore P/G TSV planning...
During early stages, so we focus on the optimization of IR Drop of P/G network design. Though thermal issue is the most critical issue in 3D designs, few of the previous work considered thermal effects during 3D P/G network optimization. The effect that the wire resistance will increase with the high local temperature has been analyzed in 2D P/G networks [15]. But unfortunately, leakage current, which is another thermal dependent design factor, is always neglected in P/G analysis. It is shown that at 130 nm technology node with supply voltage of 1.2–1.3 V, leakage current represents 10–30% of total power, and for 70 nm with supply voltage less than 1.0 V, over 50% of the power dissipation may be due to leakage current [16]. With much higher temperature on layers, leakage current is no longer negligible in 3D designs. Though the inter-dependent relations between leakage and temperature are thoroughly investigated in floorplanning [17,18], no work about the impact of leakage current on power delivery was proposed. Therefore, our work should be the first work that takes the thermal issues into the IR drop analysis in 3D ICs considering the increase of both wire resistance and leakage current due to high temperature.

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 cause high cost or even failure in designs. 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 topologies 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 different layers have different P/G grid pitches. Previous works on 3D P/G network [13,14] restrict the uniform grid or aligned trunks on different layers so that P/G TSVs are restricted at the cross points between P/G networks on two adjacent layers. But due to the various IR drop distributions, more P/G wire resources may be needed in the uniform grid mode. The limitation of inserting TSVs with uniform grids also limits the optimization of P/G network. In this paper, we connect P/G TSVs to P/G grids with 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 also the layout of the modules has much more flexibility.

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 TSV planning with the topology optimization of the P/G grids. More specifically, we make the following contributions:

- Thermal aware IR drop analysis is proposed in 3D IC to support the P/G TSV planning. Both wire resistance and leakage current are taken into account due to the thermal impact. Our analysis results show that the temperature impacts are no longer negligible in 3D P/G TSV planning.
- With thermal aware IR drop analysis, a sensitivity model of P/G TSV insertion among layers is proposed to guide the P/G TSV planning. According to the sensitivity of IR drops to the changing conductance of a given 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 TSVs. With this flexibility, the topologies of P/G grids can be optimized freely without much restriction.

The rest of the paper is organized as follows: Problem formulation is given in Section 2. In Sections 3 and 4, the thermal aware IR drop analysis and the P/G TSV planning approach are proposed. Experimental results are given in Section 5. Finally, Section 6 concludes this paper.

2. Problem formulation

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 with 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 1 will be used.

We take two stacked layers (include two metal layers and two silicon layers) in a 3D IC for example (Fig.3(a)). In order to help the IR drop dissipation in this case, some P/G TSVs can be inserted in the whitespace between placed modules to connect...
two metal layers. Some short wires are expected to connect the P/G TSVs to P/G networks if P/G TSVs are not aligned to P/G trunks on each layer (Fig.3(c)). By dividing the silicon layer into virtual grids that are called sensitivity estimation grids in this paper, we can find out all the grids that contain whitespaces, and then we assume the center of these grids as the candidate positions at which TSVs can be inserted (Fig.3(b)).

The P/G network on each metal layer is composed of a uniform mesh 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:

2.1. Total IR drop minimization:

For node $i$ in the mesh, it has an IR drop with its corresponding voltage $V_i$, we denote it as $V_{drop} = VDD - V_i$, 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 $\phi$ as the quadratic sum of difference between VDD and the voltage values for all nodes, which can be formulated as

$$\min \phi = V_{drop}^T V_{drop}$$

In this paper, we define the sensitivity of a P/G TSV as the derivative of $\phi$ to the conductance of this TSV as described in Section 2.

2.2. Wiring resources minimization

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 $A_{PC}$ should be minimized, where $A_{PC}$ includes both the area of P/G trunks ($A_{trunk}$) and the area of short interconnect segments ($A_{short}$). We denote it as

$$A_{PC} = A_{trunk} + A_{short}$$

3. Thermal effects on P/G network

The temperature variation can cause significant changes of the interconnect resistances, and therefore can substantially increase the IR drops. Even in 2D IC design, as the current densities of interconnects increase, the effects of self-heating become more significant and cannot be ignored [15]. In addition, thermal-leakage dependency may cause the increase of leakage current, the IR drop distribution will be influenced a lot with more absorbed current. 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.

3.1. Thermal 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 $R_0$ as shown in Fig.4(c). Similar to [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:

$$T = PM^{-1}$$

where $M$ is a $N \times N$ sparse thermal conductivity matrix. $T$ and $P(T)$ are $N \times 1$ temperature and power vectors.

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Notions used in this paper.</td>
</tr>
</tbody>
</table>


![Fig. 3. 3D P/G mesh with TSV. (a) Two layer design; (b) TSVs in whitespace; (c) short wires.](Please cite this article as: Z. Li, et al., Thermal-aware P/G TSV planning for IR drop reduction in 3D ICs, INTEGRATION, the VLSI journal (2012), http://dx.doi.org/10.1016/j.vlsi.2012.05.002)
3.2. Thermal-leakage dependent model

In CMOS digital circuits, power dissipation consists of dynamic and leakage components [19]. In current CMOS technologies, the sub-threshold leakage current $I_{\text{sub}}$ is much larger than the other leakage current components [20]. The sub-threshold leakage power/current increases exponentially with temperature increases.

$$\frac{I_{\text{sub}}}{P_{\text{sub}}} = \alpha \times e^{\beta(T_i - T_{\text{base}})}$$

(4)

where $T_i$ is the temperature of module $m_i$, $\alpha$ and $\beta$ are the empirical factors that have different values for different technologies. $T_{\text{base}}$ is the reference temperature at which $\alpha$, $\beta$ and the initial $I_{\text{sub}}$ are defined. Typically, $\alpha = 1 \times 10^5$ $W/\text{m}^2$, $\beta = 0.025$ for 130 nm [21]. $T_{\text{base}}$ is defined to 27°C with the initial $I_{\text{sub}}$ 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

$$I = I_{\text{dyn}} + I_{\text{sub}}$$

(5)

Since we have leakage power considered, the total power considered in (3) includes not only the dynamic power, but also the leakage power.

$$P = P_{\text{dynamic}} + P_{\text{leakage}}$$

(6)

where $P_{\text{leakage}} = I_{\text{sub}} \times V_{\text{DD}}$.

On the other hand, leakage power will increase the total power consumption that will in turn increase temperature in (4). 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.

3.3. Thermal-aware wire resistance model

Lack of accurate evaluation of thermal effect on the IR drop in the power grid may lead to over-design; or worse, underestimates the IR drop due to increased local temperature. It is shown in [15] that without considering temperature impact, the worst IR drop analysis can have error up to 10% in 2D ICs. Hence the thermal impacts on wire resistance are no longer negligible in P/G analysis in 3D ICs with much higher temperature.

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

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

(7)

where $\rho_0$ is the resistivity at reference temperature (27°C for example) and $\beta = 0.0039/\text{°C}$ is the temperature coefficient of resistance. As shown in Fig.6(a), the resistance of a wire in
thermal grids can be figured out:

\[ R_{\text{wire}} = \frac{\rho_0 l_1 + \rho_1 l_2 + \rho_2 l_3}{A} \]  

where \( \rho(i) = \rho_0(1 + \beta \cdot T(i)) \), for \( i = 1, 2, 3 \). \( T(i) \) is the temperature of grid, \( A \) is the 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.

4. P/G TSV planning

4.1. 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:

\[ G \cdot V = I \]  

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:

\[ G = \begin{pmatrix} G_{11} + G_{tsv} & -G_{tsv} \\ -G_{tsv} & G_{22} + G_{tsv} \end{pmatrix} \]  

where \( G_{11}, G_{22} \) is the conductance matrix build by short wires and branches on layer1 and layer2 respectively. \( G_{tsv} \) is the conductance matrix built by TSVs between layer1 and layer2 and the short wires connecting TSVs to grids, it can be written as

\[ G_{tsv} = \text{diag}(g_{tsvi}, g_{tsvi}, \ldots, g_{tsvi}, \ldots, g_{tsvi}, \ldots) \]  

where \( g_{tsvi} \) is the conductance of TSVi, \( m \) is the number of P/G TSVs. 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 voltage would come to a balance and the node with lower voltage would be made up, consequently the total IR drop will be reduced. But due to the complexity of the P/G deliver, besides the difference of voltage values on adjacent layers, the matrix structure also influences the effect of P/G TSVs.

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

\[ \text{Sensitivity}(g_{tsvi}) = \frac{\partial \phi}{\partial g_{tsvi}} \]  

where \( \phi \) is defined by (1), and we can easily get

\[ \frac{\partial \phi}{\partial g_{tsvi}} = 2V_{\text{drop}} \cdot \frac{\partial V}{\partial g_{tsvi}} = -2V_{\text{drop}} \cdot \frac{\partial V}{\partial g_{tsvi}} \]  

In formula (13), as \( V_{\text{drop}} \) is known, in order to calculate the sensitivity of TSVi, we must first compute the deviation of \( V \) to \( g_{tsvi} \).

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

\[ \frac{\partial G}{\partial g_{tsvi}} \cdot V + G \cdot \frac{\partial V}{\partial g_{tsvi}} = \frac{\partial V}{\partial g_{tsvi}} \]  

(14)

Since \( \frac{\partial l}{\partial g_{tsvi}} = 0 \), then we can get

\[ \frac{\partial V}{\partial g_{tsvi}} = G^{-1} \cdot \frac{\partial G}{\partial g_{tsvi}} \cdot V \]  

Then we can express (13) as

\[ \frac{\partial \phi}{\partial g_{tsvi}} = 2 \cdot V_{\text{drop}} \cdot G^{-1} \cdot \frac{\partial G}{\partial g_{tsvi}} \cdot V \]  

(16)

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

\[ \frac{\partial \phi}{\partial g_{tsvi}} = 2 \left( \frac{\partial G}{\partial g_{tsvi}} \cdot V \right)^T \cdot G^{-1} \cdot V_{\text{drop}} \]  

(17)

We can first solve a linear equation:

\[ G \cdot X = V_{\text{drop}} \]  

(18)

And we can get

\[ X = G^{-1} \cdot V_{\text{drop}} \]  

(19)

And (17) can be expressed as below:

\[ \frac{\partial \phi}{\partial g_{tsvi}} = 2 \left( \frac{\partial G}{\partial g_{tsvi}} \cdot V \right)^T \cdot X \]  

(20)

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

\[ \frac{\partial G}{\partial g_{tsvi}} = \begin{pmatrix} \cdots & 1 & \cdots & -1 \\ \vdots & \vdots & \vdots & \vdots \\ -1 & 1 & \cdots & \cdots \end{pmatrix} \]  

(21)

where 1 appears at \((a, a)\) and \((b, b)\), \(-1\) appears at \((a, b)\) and \((b, a)\), \(a\) and \(b\) are the two end nodes of \( g_{tsvi} \). Finally we can get the sensitivity of \( g_{tsvi} \), which can be expressed as

\[ \frac{\partial \phi}{\partial g_{tsvi}} = 2 \cdot (V_a - V_b) \cdot X_a + (V_b - V_a) \cdot X_b \]  

(22)

where \( V_a \) and \( V_b \) are the voltage values of two end nodes of \( g_{tsvi} \), and \( X_a \) are the \( a \)-th and \( b \)-th element of vector \( X \).

It is shown from formula (22) that the sensitivity of IR drops \( \phi \) to the conductance of a given TSV is associated with the voltage difference between two end nodes of the TSV. But there is still another factor such as \( X \) which also influences the sensitivity. 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 the TSVs, the better the IR drop. If the sensitivity value of \( g_{tsvi} \) is positive due to the additional connecting wires to the P/G grid and we add a TSV to that position of \( g_{tsvi} \), the total value of \( \phi \) will increase, and power integrity will become even worse. So we cannot insert TSV where the sensitivity value is positive. On the other hand, if the sensitivity value of \( g_{tsvi} \) is negative and the larger its absolute value is, the more \( \phi \) 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 2.
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 \(50 \times 50\). After we get the sensitivity of each candidate TSV position, we surely know where TSVs are needed the most. Then we sort the grids by the sensitivities and insert TSVs with the most negative sensitivities to the grids. The max number of TSVs can be inserted through a grid is figured out by the whitespace area of this grid and the size of TSV. As the TSVs inserted, the sensitivity and capacitance of the grids 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 \(\phi/n\) can be made. The P/G TSV planning flow is described below:

**Algorithm 1. P/G TSV planning**

Initialize parameters \(e_1, e_2\);
\(\phi = \text{old}_\phi = 0\);
Sweep the whitespaces using sensitivity estimation grids;
Add short interconnect segments to each layer if possible;
Calculate the sensitivity of \(\phi\) to conductance of each candidate TSV;
While \((\phi/n > e_1 \text{ and } \text{fabs}(\phi - \text{old}_\phi)/n > e_2)\) Do
Select a grid with most negative sensitivities and with enough capacity for one TSV;
Insert a TSV into this grid;
Solve linear system \(G \cdot V = I\);
\(\text{old}_\phi = \phi\);
Calculate the total IR drop \(\phi\);
Update the sensitivity of \(\phi\) to conductance of each candidate TSV;
End.

4.2. 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 \(V_{\text{vio}} = V_{\text{min}} - V\) and \(\phi = V_{\text{vio}} \cdot S \cdot V_{\text{vio}}\) to evaluate the total violation of the P/G network. \(S\) is defined as

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

The P/G grid density on each metal layer discussed in this paper varies from \(50 \times 50\) to \(200 \times 200\). We define \(\text{InitSize}\) as \(50 \times 50\) and \(\text{MaxSize}\) as \(200 \times 200\). To meet the IR drop constraints with less wiring sources, we start the exploration from \(\text{InitSize}\) to \(\text{MaxSize}\). Besides the P/G grid density, the location of the package pads greatly influences the IR drops in the entire power distribution network. But in this work, these locations are not considered as variables in the P/G grid topology optimization algorithm since we focus on the effect of trunks, short wires and TSVs on the IR drops. The optimization flow is described below:

**Algorithm 2. P/G grid topology optimization**

Construct the P/G mesh with \(\text{InitSize}\) size for each layer,
Initialize parameters \(\tau_1, \tau_2\);
\(\phi = \text{old}_\phi = 0\);
While \((\phi/n > \tau_1 \text{ and } \text{fabs}(\phi - \text{old}_\phi)/n > \tau_2)\) Do
P/G TSV plan;
Calculate \(V\) after inserting P/G TSVs;
\(\text{old}_\phi = \phi\);
Calculate violation and IR drop on each layer and total violation \(\phi\);
Increase the size of the layer which has the worst IR drop and its size does not reach \(\text{MaxSize}\);
End.

4.3. 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.
5. Experimental results

Extensive experiments have been performed to evaluate our proposed method. A set of MCNC benchmarks is 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.

5.1. Effect of TSV sensitivity model

In this set of tests, we define $100 \times 100$ 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 drops 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 1.8 V 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 in Table 2 show 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 P/G TSVs.

5.2. Thermal impacts

Based on the thermal model analyzed in Section 3, 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 2. The temperature on the chip without TSV planning can reach $310 \degree C$ 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. $100 \times 100$ 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

![Fig. 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 $c_1 = 1 \times 10^{-5}$, $c_2 = 1 \times 10^{-7}$).]

<table>
<thead>
<tr>
<th>Circuit</th>
<th>Without thermal analysis</th>
<th>With thermal analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before via planning</td>
<td>After via planning</td>
</tr>
<tr>
<td></td>
<td>Max IR drop (v)</td>
<td>Avg. IR drop (v)</td>
</tr>
<tr>
<td>ami33</td>
<td>0.269 0.093 12409</td>
<td>0.139 0.089 1497</td>
</tr>
<tr>
<td>ami49</td>
<td>0.208 0.049 3433</td>
<td>0.137 0.049 196</td>
</tr>
<tr>
<td>n100</td>
<td>0.301 0.084 15538</td>
<td>0.154 0.083 2544</td>
</tr>
<tr>
<td>n200</td>
<td>0.277 0.081 13035</td>
<td>0.172 0.081 3073</td>
</tr>
<tr>
<td>n300</td>
<td>0.243 0.067 10042</td>
<td>0.150 0.065 2278</td>
</tr>
<tr>
<td>Avg.</td>
<td>0.260 0.075 10891</td>
<td>0.150 0.073 1918</td>
</tr>
</tbody>
</table>

Please cite this article as: Z. Li et al., Thermal-aware P/G TSV planning for IR drop reduction in 3D ICs, INTEGRATION, the VLSI journal (2012), http://dx.doi.org/10.1016/j.vlsi.2012.05.002
still reduce the Max IR drop from 0.290 to 0.158 that 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.

5.3. 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 $t_1 = 1 \times 10^{-5}$, $t_2 = 1 \times 10^{-5}$. Fig. 9 shows the P/G TSV distribution obtained by our approach in four layers of ami33. Each white point denotes several TSVs in the near point position. TSVs between blocks are inserted between two adjacent layers. Since all benchmarks we used in this paper have four layers (four metal layers and four silicon layers), there exist three silicon layers that have P/G TSV in the whitespace.

The overall results of P/G network analysis with thermal and sensitivity model, coupled with P/G topology optimization is shown in Table 3. The wire resources in the table include the trunks that construct P/G grids and the short wires to connect the P/G TSVs to the P/G grid. As we can see from Table 3, the results show that the co-optimization flow can save 18.5% of the total P/G wire resources with 44.7% additional

Please cite this article as: Z. Li, et al., Thermal-aware P/G TSV planning for IR drop reduction in 3D ICs, INTEGRATION, the VLSI journal (2012), http://dx.doi.org/10.1016/j.vlsi.2012.05.002
P/G TSVs inserted. Compared to the uniform grid in Table 4, the non-uniform grid topology analyzed in Section 1 and we used here 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 4 that a small amount of short wires can bring a significant saving (about 20.7%) of total P/G resources.

6. 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 that 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 in detailed placement stage instead of early design stages.

Acknowledgments

This paper is supported by the National Natural Science Foundation of China under Grant nos. 60606007, 61076035, and 61176035 and the TNList cross discipline Foundation.

References


Zuowei Li received his B.S. degree in Computer Science and Technology from Wuhan University of Science and Technology, Wuhan, China, in 2009. Currently, he is working toward the Master degree in Computer Science and Technology at Tsinghua University, Beijing, China, and conducting research in the EDA Laboratory. His current research interests include 3D power grid analysis and optimization. His team was awarded the first place winner in the TAU Power Grid Simulation Contest in 2011 and second place winner in the TAU Power Grid Transect Simulation Contest in 2012, which are sponsored by IBM Corporation.

Yuchun Ma received B.S. degree from X‘ian Jiaotong University in 1999 and received her doctoral degree with honors at Tsinghua University in the summer of 2004, and continued her research as a postal doctor in Tsinghua University till the summer of 2006. Currently, she is an associate professor in Tsinghua University. In 2005, she worked as a visiting scholar in UCLA for a year. Her research interests include physical design and synthesis for VLSI, 3D IC, etc.

Qiang Zhou received B.S. degree in Computer Science and Technology from University of Science and Technology of China in 1983, received M.S. degree in Computer Science and Technology from Tsinghua University in 1986, and received Ph.D. in Control Theory and Control Engineering from Chinese University of Mining and Technology in 2002. He has been a professor in the Department of Computer Science and Technology, Tsinghua University, Beijing, China. His research interests include VLSI layout theory and algorithms.

Yici Cai received B.S. degree in Electronic Engineering from Tsinghua University, Beijing, China in 1983, received M.S. degree in Computer Science and Technology from Tsinghua University in 1986, and received Ph.D. in Computer Science, University of California, Berkeley, China in 2007. She has been a professor in the Department of Computer Science and Technology, Tsinghua University, Beijing, China. Her research interests include design automation for VLSI integrated circuits algorithms and theory, power/ground distribution network analysis and optimization, high performance clock synthesis, and low power physical design.
Yuan Xie received the B.S. degree in electronic engineering from Tsinghua University, Beijing, China, and the M.S. and Ph.D. degrees in electrical engineering from Princeton University, Princeton, NJ. He is an Associate Professor with the Department of Computer Science and Engineering, Pennsylvania State University, University Park. He has published more than 100 technical papers in the areas of computer architecture, design automation, VLSI design, and embedded systems. Prof. Xie was a recipient of the SRC Inventor Recognition Award in 2002, NSF CAREER Award in 2006, and IBM Faculty Award in 2009. He also received the Best Paper Award in ASP-DAC 2008.

TingTing Hwang (M’90) received the B.S. degree in political science from the National Taiwan University, Taipei, Taiwan, in 1981, and the M.S. and Ph.D. degrees in computer science from Pennsylvania State University, University Park, in 1986 and 1990, respectively. From 1990 to 1996, she was a Vice-Professor with the Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan, where she is currently a Full Professor of computer science. She is the author of over 80 research papers in related areas. Her research interests include logic synthesis/optimization and high-level synthesis.