<table>
<thead>
<tr>
<th>Title</th>
<th>Reliable 3-D clock-tree synthesis considering nonlinear capacitive TSV model with electrical–thermal–mechanical coupling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>P. D., Sai Manoj; Yu, Hao; Yang Shang; Chuan Seng Tan; Sung Kyu Lim</td>
</tr>
<tr>
<td>Date</td>
<td>2013</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10220/18217">http://hdl.handle.net/10220/18217</a></td>
</tr>
<tr>
<td>Rights</td>
<td>© 2013 IEEE. This is the author created version of a work that has been peer reviewed and accepted for publication by IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, IEEE. It incorporates referee’s comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at: <a href="http://dx.doi.org/10.1109/TCAD.2013.2270285">http://dx.doi.org/10.1109/TCAD.2013.2270285</a>.</td>
</tr>
</tbody>
</table>
Reliable 3D Clock-tree Synthesis Considering Nonlinear Capacitive TSV Model with Electrical-thermal-mechanical Coupling

Sai Manoj P.D, Student Member, IEEE, Hao Yu, Member, IEEE, Yang Shang, Student Member, IEEE, Chuan Seng Tan, Member, IEEE, and Sung Kyu Lim, Senior Member, IEEE

Abstract—A robust physical design of 3D-IC requires investigation on through-silicon-via (TSV). The large temperature and stress gradients can severely affect TSV delay with large variation. The traditional physical model treats TSV as resistor with linear electrical-thermal dependence, which ignores the fundamental device physics. In this paper, a physics-based electrical-thermal-mechanical delay model is developed for signal TSVs in 3D-IC. With consideration of liner material and stress, a nonlinear model is established between electrical delay with temperature and stress. Moreover, sensitivity analysis is performed to relate the reduction of temperature and stress gradients with respect to dummy TSVs insertion. Taking the design of 3D clock-tree as a case study, we have formulated a nonlinear optimization problem for clock-skew reduction. By allocating dummy TSVs to reduce the temperature and stress gradients, the clock-skew introduced by signal TSVs and drivers can be minimized. A number of 3D clock-tree benchmarks are utilized in experiments. We have observed that with use of dummy TSV insertion, clock-skew can be reduced by 61.3% on average when the accurate nonlinear electrical-thermal-mechanical delay model is applied.

Index Terms—Through-silicon via (TSV), TSV stress, thermal TSV, nonlinear MOSCAP, temperature gradient, stress gradient, clock-skew reduction, electrical-thermal-mechanical coupling.

I. INTRODUCTION

3D integrated circuits (3D-ICs) have regained the interest for big bandwidth in the design of many-core microprocessor server. The utilization of through-silicon vias (TSVs) in 3D-IC can significantly reduce the latency and power dissipation in global interconnect such as memory buses and also clock-trees [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18]. However, one robust 3D-IC design requires a careful examination of TSVs in which electrical states such as delays are coupled from multiple physical domains. Firstly, thermal reliability is one primary concern since heat-sink has long distance for top layer. Moreover, dynamic voltage and frequency scaling (DVFS) of many-core can result in highly non-uniform power density. As such, there exists a large temperature gradient that can result in delay variation of TSV by electrical-thermal coupling. What is more, because coefficients of thermal expansion (CTE) of TSV material and substrate material are different, large mechanical stress can be introduced, which in turn leads to delay variation of driver by electrical-mechanical coupling. Since delay variation introduces skew for delay sensitive clock-tree design, a robust physical design in 3D-IC with TSVs thereby needs to consider optimization from coupled electrical, mechanical and thermal domains.

The design of clock-tree is primarily involved with reduction of delay difference at different sinks, known as skew [19], [20], [21], [22], [23], [24], [25], [26]. Compared to clock-tree in 2D-IC, the one in 3D-IC will experience much larger temperature and stress gradient both vertically and horizontally. As signal TSVs are deployed to route 3D clock-tree over the entire 3D chip, such non-uniform temperature difference can lead to a significant clock-skew by electrical-thermal coupling of signal TSVs [24], [25], [26]. Such an electrical-thermal coupling becomes nonlinear when liner material is considered. Moreover, the TSV induced stress also affects the mobility and delay of drivers, which further worsens the clock-skew over the entire 3D chip by electrical-mechanical coupling of drivers [10]. As such, the traditional clock-tree design methods [19], [20], [21] without considering temperature and stress gradients will become inaccurate and unreliable. The thermal-aware 3D clock-tree synthesis has been discussed in [4] considering thermal profile. A 3D embedding method was developed in [17] to reduce the wire-length. Further optimization in [11] is developed to reduce power and slew-rate. However, previous methods only conduct the clock network optimization based on linear or none electrical-thermal-mechanical coupling, ignoring the TSV physical model and hence is not accurate. As such, there is no specific problem formulated for clock-tree design based on the reduction of both thermal and stress gradients in 3D-IC.

In this paper, based on recent measurement results in [6], [8], [10], [12], [13], [14], a nonlinear electrical-thermal-mechanical delay model is developed for 3D clock-tree design. Further, sensitivity analysis is performed for dummy TSV density with respect to reduction of temperature or stress gradient. Based on accurate TSV
models, a reliable 3D clock-tree design is formulated by dummy TSVs insertion to balance the clock-skew by reducing temperature and stress gradient, with consideration of nonlinear electrical-thermal-mechanical delay model. A nonlinear programming based optimization is developed and implemented to determine the allocation of dummy TSVs. Experimental results show that with insertion of reasonable number of dummy TSVs, the average clock-skew can be reduced by 61.3% for clock-tree benchmarks in [27] in 3D design [11]. Compared to clock-tree design by linear delay model, our approach by nonlinear delay model reduces clock-skew further by 12.2% under same thermal conditions and same TSV density.

The rest of the paper is organized as follows. The reliable 3D clock-tree design problem formulation is discussed in Section II. Modeling of signal TSVs and drivers under electrical-thermal-mechanical coupling is discussed in Section III, and the dummy TSV model and the sensitivity study is discussed in Section IV. Section V discusses the nonlinear electrical-thermal-mechanical delay and skew model. Nonlinear optimization for reduction of clock-skew is presented in Section VI. Numerical experimental results for modeling and optimization are presented in Section VII. The paper is concluded in Section VIII.

II. 3D RELIABLE CLOCK-TREE DESIGN PROBLEM

Many approaches have been applied in 2D clock-tree designs to reduce clock-skew such as buffer sizing [4], merging point adjustment [22], and wire-length balancing [21] etc. Due to non-uniform power densities by many-core microprocessors, there is a large non-uniform temperature gradient. Moreover, a large stress gradient induced by TSVs during annealing. Accordingly, the 3D clock-tree will experience significant clock-skew under large temperature and stress gradient with a new problem formulation required.

A. TSV Fabrication

An accurate physical model of TSV needs a detailed investigation on its fabrication. As shown in Fig. 1, TSVs are used as vertical interconnections between stacked dies for providing electrical interconnect as well as heat dissipation. To perform these functions, the materials used for TSVs should have good electrical conductivity as well as thermal conductivity. Tungsten (W), poly-silicon and copper (Cu) can be considered as TSV fill materials. Due to low resistivity and cost, copper is the widely used material for TSV fill [10]. To fabricate TSVs, TSV etching is performed by deep ion reactive etching (DRIE), laser or chemical etching etc. After the etching is performed, the liner material is deposited to prevent ion particle diffusion. After forming liner layer, TSV material such as copper or tungsten is filled in the etched region at high temperature. After annealing to low temperature, the substrate is thinned and the current layer can be aligned and integrated with other layers. Due to the existence of liner and also the annealing, the physical TSV model becomes electrical, thermal and mechanical coupled.

B. Problem Formulation

3D clock-tree, as shown in Fig. 2, makes use of TSV for vertical interconnections, which can have significant delay. As such, in 3D clock-tree design, unlike 2D clock-tree synthesis, the impact of TSV also needs to be considered. Stacking of dies in vertical direction in 3D design increases the overall temperature and also temperature gradient due to the increased and non-uniform power density and heat-dissipation path. Due to the temperature gradient, the device characteristics also vary because of electrical-thermal coupling. The temperature distribution on a 3D clock-tree is shown in Fig. 2. Normally, RC-delay \( D_{RC} \) for the traditional TSV is modeled as RC-interconnect by a linear electrical-thermal coupling

\[
R_T = R_0(1 + \alpha \cdot \delta T); \quad D_{RC} = R_0C_0(1 + \alpha \cdot \delta T)
\]

where \( R_T \) is thermal-dependent resistance, \( R_0 \) represents resistance at room temperature, \( C_0 \) represents capacitance at room temperature, \( \delta T \) is the difference of the operating temperature \( T \), and room temperature \( T_0 \). In addition, note that \( \alpha \) is the temperature-dependent coefficient for resistance, whose value is experimentally determined.

However, due the existence of liner material around TSV-fill, it forms a nonlinear MOSCAP that can significantly affect TSV delay. According to the measurement results in [8], the contribution of 2nd-order temperature-dependent contribution in MOSCAP model is increased to 30% of the overall TSV capacitance at 150°C. As such, TSV needs to be characterized as a capacitor with nonlinear temperature-dependent instead of linear temperature-dependent resistor. What is more, TSVs exert...
mechanical stress on the silicon substrate due to mismatched CTEs. The impact of stress can affect mobility and delay of driver. Eventually, one needs a nonlinear electrical-thermal-mechanical coupled delay model in 3D-IC.

By considering all the above effects, delay at each clock sink, \( i \) of one 3D clock-tree needs to be modeled as nonlinear model of temperature and stress gradient \( \Gamma \). Thus, delay at sink \( i \) can be modeled as \( \delta D_i = f(\Gamma) \). Clock-skew \( S \) is defined as the maximal difference between two clock sinks. Note that by adding dummy TSVs, one can balance the temperature and stress gradient and can further reduce delay and skew. As such, we have the following problem formulation towards reliable 3D clock-tree design under temperature and stress gradient.

**Problem 1:** For a pre-synthesized zero-skew 3D clock-tree with \( N_S \) sinks, using signal TSVs for inter-tier connections, the clock-skew \( S \) needs to be estimated by considering position and number of TSVs, i.e. temperature and stress gradient \( \Gamma \), and by considering nonlinear electrical-thermal-mechanical coupling. A large number of dummy TSVs that can be inserted to minimize \( S \) under \( \Gamma \).

\[
S = \max (|\delta D_i - \delta D_j|; 0 \leq i, j \leq N_S)
\]  
(2)

where \( \delta D_i \) and \( \delta D_j \) are delays from sinks \( i, j \) respectively.

In order to solve this problem, electrical, thermal and mechanical couplings are initially studied in Section III. Then, a sensitivity analysis is performed to study the reduction of temperature gradient and stress gradient when adding dummy TSVs in Section IV. A nonlinear electrical-thermal-mechanical coupled delay model is derived to calculate clock-tree delay and skew in Section V. Finally, a nonlinear optimization is deployed to further minimize the clock-skew under non-uniform temperature and stress gradient in Section VI. Note the pre-synthesized zero-skew 3D clock-tree is based on the work [11] to consider wire-length and driver but without considering electrical-thermal-mechanical coupling.

### III. 3D NONLINEAR ELECTRICAL-THermal-MECHANICAL COUPLING

#### A. Signal TSV Model

The TSV utilized for inter-layer signal connection is called signal TSV, which connects clock-tree at two different layers. In the previous 3D clock-tree synthesis [22], [4], [11], vias are electrically modeled with simple RC-delay model with only resistance \( R \) as linearly temperature dependent and capacitance \( C \) as constant. Due to the existence of liner, there is a nonlinear dependence on temperature. Electrical-thermal model of a single signal TSV is thereby studied in this section. Note that the mechanical stress to signal TSVs is too small to be considered.

1) **Nonlinear MOSCAP Model**

Due to the existence of liner in TSVs, a MOS-Capacitor (MOSCAP) is formed between signal TSV and substrate. This nonlinear capacitance depends on biasing voltage \( V_{BIAS} \) as well as temperature [8]. The difference in work-function between metal-material of TSV and substrate \( \delta \) due to mismatched CTEs. The impact of stress can affect mobility and delay of driver. Eventually, one needs a nonlinear electrical-thermal-mechanical coupled delay model in 3D-IC.

![Fig. 3: (a) Signal TSV and dummy TSV in 3D-IC; (b) 3D view of TSV; and (c) Equivalent RC-circuit of TSV](image)

**TABLE I: Physical parameters used in TSV modeling**

<table>
<thead>
<tr>
<th>Notation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R_{metal} ), ( R_{depl} ), ( R_{ox} )</td>
<td>Radius of TSV metal, silicon and depletion region respectively</td>
</tr>
<tr>
<td>( V_{bias} )</td>
<td>Voltage bias</td>
</tr>
<tr>
<td>( C_{Th} )</td>
<td>Thermal capacitance of TSV</td>
</tr>
<tr>
<td>( C_{DEP} )</td>
<td>Depletion capacitance of TSV</td>
</tr>
<tr>
<td>( C_{TSV} )</td>
<td>Total capacitance of TSV</td>
</tr>
<tr>
<td>( C_{D} )</td>
<td>Unit capacitor of TSV</td>
</tr>
<tr>
<td>( C_{L} )</td>
<td>Capacitance of TSV load</td>
</tr>
<tr>
<td>( L )</td>
<td>Line length of TSV</td>
</tr>
<tr>
<td>( r_2 )</td>
<td>Radius of TSV metal</td>
</tr>
<tr>
<td>( S_1 ), ( S_2 )</td>
<td>Diameter of TSV metal and silicon respectively</td>
</tr>
<tr>
<td>( S_3 )</td>
<td>Diameter of TSV depletion region</td>
</tr>
<tr>
<td>( S_{TSV} )</td>
<td>TSV heat conductivity</td>
</tr>
<tr>
<td>( P )</td>
<td>Heat power flowing from chip to heat-sink</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>Stress gradient sensitivity in ( 1^{st} ) grid</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>Stress gradient sensitivity in ( 2^{nd} ) grid</td>
</tr>
</tbody>
</table>

**TABLE II: Electrical parameters used in TSV modeling**

<table>
<thead>
<tr>
<th>Notation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R_0 )</td>
<td>Interconnect resistance at room temperature ( T_0 )</td>
</tr>
<tr>
<td>( C_{Th} )</td>
<td>Thermal capacitance at zero-temperature</td>
</tr>
<tr>
<td>( r_{ox} ), ( r_{si} )</td>
<td>Diode constant of silicon oxide and silicon</td>
</tr>
<tr>
<td>( L_{cap}, C_{cap} )</td>
<td>Line capacitance and depletion capacitance of TSV</td>
</tr>
<tr>
<td>( R_{D}, C_{D} )</td>
<td>Unit buffer and load capacitances</td>
</tr>
<tr>
<td>( R_{DEP}, C_{DEP} )</td>
<td>Unit wire resistances</td>
</tr>
<tr>
<td>( R_D )</td>
<td>Unit buffer resistance without stress and with impact of stress</td>
</tr>
<tr>
<td>( \mu )</td>
<td>Mobility of charge carrier without stress</td>
</tr>
<tr>
<td>( \Delta \mu )</td>
<td>Mobility enhancement factor</td>
</tr>
<tr>
<td>( \Delta \mu )</td>
<td>Mobility enhancement factor</td>
</tr>
<tr>
<td>( \theta )</td>
<td>Stress gradient sensitivity in ( 1^{st} ) grid</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>Stress gradient sensitivity in ( 2^{nd} ) grid</td>
</tr>
</tbody>
</table>

**TABLE III: Thermal parameters used in TSV modeling**

<table>
<thead>
<tr>
<th>Notation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T_{Th} )</td>
<td>Room temperature and temperature ( T ) respectively</td>
</tr>
<tr>
<td>( R_{Th} )</td>
<td>Interconnect resistance at temperature ( T )</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>Temperature dependent coefficient for calculating resistance</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>First-order and second-order coefficients of ( C_{Th} )</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>Temperature gradient</td>
</tr>
<tr>
<td>( \beta_{Total} )</td>
<td>Total heat conductivity of the chip with TSVs inserted</td>
</tr>
<tr>
<td>( \Delta )</td>
<td>Heat conductivity of regular chip area</td>
</tr>
<tr>
<td>( \Delta )</td>
<td>Heat conductivity of regular chip area</td>
</tr>
</tbody>
</table>

**TABLE IV: Mechanical parameters used in TSV modeling**

<table>
<thead>
<tr>
<th>Notation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma )</td>
<td>Stress</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>Stress</td>
</tr>
<tr>
<td>( \mu )</td>
<td>Mobility enhancement factor</td>
</tr>
<tr>
<td>( \eta )</td>
<td>Stress gradient sensitivity in ( 1^{st} ) grid</td>
</tr>
<tr>
<td>( \eta )</td>
<td>Stress gradient sensitivity in ( 2^{nd} ) grid</td>
</tr>
<tr>
<td>( \eta )</td>
<td>Stress gradient sensitivity in ( 3^{rd} ) grid</td>
</tr>
</tbody>
</table>

**TABLE V: Temperature gradient sensitivity in \( 1^{st} \) grid**
depletion and inversion regions separated by flat-band voltage (\(V_{FB}\)) and threshold voltage (\(V_T\)). For higher frequencies (> \(1 \text{MHz}\)) the inversion region can be further divided into deep-depletion and inversion regions. As such, the RC-delay parameters of TSV with dependence on temperature can be modeled as

\[
R_T = \frac{\rho h}{\pi r_{metal}^2} + \frac{1}{C_T} + \frac{1}{C_{ox}} + \frac{1}{C_{dep}};
\]

with

\[
C_{ox} = \frac{2\pi \varepsilon_{ox} h}{\ln\left(\frac{r_{ox}}{r_{metal}}\right)}; \quad C_{dep} = \frac{2\pi \varepsilon_{si} h}{\ln\left(\frac{r_{dep}}{r_{ox}}\right)}
\]

where \(R_T, C_T\) represents the temperature dependent resistance and capacitance of TSV respectively. \(C_{ox}, C_{dep}\) are liner and depletion-region capacitances of TSV respectively. TSV height is represented by \(h\). \(\rho\) is the resistivity of metal-material of TSV. \(\varepsilon_{ox}\) and \(\varepsilon_{si}\) are dielectric constants of silicon dioxide and silicon respectively. \(r_{metal}, r_{ox}\) and \(r_{dep}\) are the outer radii of TSV metal, silicon and depletion-region respectively. Since the thermal conductivity of liner (\(\text{SiO}_2\)) is one hundred times lower than silicon substrate, liner prevents the dissipation of heat from substrate and results in hot-spot at signal TSVs. As shown by measured results in [8], the TSV capacitance can approaches liner capacitance at high temperature due to the existence of liner material. As such, the nonlinear electrical-thermal coupling of signal TSV can bring significant impact on delay when using TSVs in 3D clock-tree.

### B. Driver Model

TSVs can further exert mechanical stress on the device layer. This mechanical stress has impact on the mobility and delay of drivers. As TSV density can be non-uniform across the chip, result in stress gradient and can further introduce delay variation or skew.

1) **Thermal-mechanical Coupling**

TSV material and device layer have different CTEs. In addition, they can have different temperatures during the time of operation. The different temperatures lead to difference in the amount of expansion of TSV and the substrate, resulting in mechanical stress. The mechanical stress exerted by multiple TSVs on substrate can be found by principle of superposition [10] by

\[
\sigma_i = -\frac{B \Delta T \Delta \alpha}{R_i^2} \frac{R_i}{r_i} \sum_{i=1}^{n} \sigma_i = -\frac{n B \Delta T \Delta \alpha}{2} \frac{R_i^2}{r_i} \quad \sigma = \eta A
\]

where \(\sigma_i\) is the stress from \(i^{th}\) TSV, \(B\) is the biaxial modulus, \(\Delta \alpha\) is CTE difference between TSV material and substrate as a constant, \(\Delta T\) is the annealing temperature difference, \(R_i\) is the radius of \(i^{th}\) TSV, \(r_i\) represents the distance of a transistor or a driver from center of TSV, and \(n\) represents the number of TSVs with a TSV density of \(\eta\) in area \(A\). For simplicity, all TSVs are considered to be of the same radius \(R\), and the drivers are approximated with the same distance \(r\) from center of TSVs. As such, we can observe an thermal-mechanical dependence to characterize the mechanical stress between TSV and the substrate. In thermal-mechanical coupling, the main focus is on exertion of stress with respect to temperature and TSV density.

2) **Electrical-mechanical Coupling**

Exerted mechanical stress from TSVs will affect the carrier mobility of drivers [10], [12]. Higher the amount of exerted stress from TSV, stress impact on carrier mobility will be high. This variation of mobility will affect the electrical delay of drivers. Variation in carrier mobility due to exerted mechanical stress can be given by [10], [12]

\[
\frac{\delta \mu}{\mu} = -\Pi \times \sigma; \quad m = -\Pi_x
\]

where \(\frac{\delta \mu}{\mu}\) is the ratio of mobility variation, \(\Pi\) is the tensor of piezo-resistive coefficients, and \(\sigma\) represents mechanical stress. Note that \(m\) represents the maximum value of \(\Pi\) among all the directions (\(x, y, z\)), in order to capture its most significant impact on transistor. For example, \(\Pi_x\) is used to represent the value of the maximum value in tensor \(\Pi\). Value of \(m\) indicates the enhancement factor along the direction that results in the maximum stress. It can be different for PMOS and NMOS devices, and can result in different amount of mobility variations [10], [12]. The ratio of mobilities with and without stress can be calculated as
dummy TSVs can balance density of TSV distribution, which also helps to reduce the stress gradient. What is more, adding dummy TSVs can balance the temperature gradient. As the amount the stress exerted increases, the ratio of mobility with and without stress also increases. This variation in mobility results in change of source resistance of one driver. The variation of source resistance with mobility can be given by

\[
\frac{\mu_s}{\mu} = 1 + \frac{\delta \mu}{\mu} = 1 + m \sigma
\]

(8)

where \(\mu_s\) and \(\mu\) represent mobility of charge carriers with and without impact of stress; \(\frac{\delta \mu}{\mu}\) is the mobility variation ratio.

As the amount the stress exerted increases, the ratio of mobility with and without stress also increases. This variation in mobility results in change of source resistance of one driver. The variation of source resistance with mobility can be given by

\[
R_D^{M} = \frac{R_D}{1 + \frac{\delta \mu}{\mu}} = \frac{R_D}{1 + m \sigma}
\]

(9)

where \(R_D^{M}\) is the driver resistance with impact of stress, and \(R_D\) is the driver resistance without impact of stress. As such, when the TSV density is different, the impact of exerted mechanical stress on the driver delay can be different, which further affects the clock-tree delay. Note that there is a thermal-mechanical coupling discussed above, which can further worsen the delay dependence on both thermal and mechanical couplings.

IV. 3D Dummy TSV Insertion and Sensitivity

In the previous section, modeling of signal TSV and driver is studied under electrical-thermal-mechanical coupling. The characterization of dummy TSVs is equally important to determine the relation between the reduction of thermal and stress gradient with the dummy TSV insertion density.

A. Dummy TSV Insertion Density

TSV utilized for inter-layer insertion but without signal connection is called dummy TSV. As dummy TSV is filled with metal-material Cu with good thermal conductivity of 400W/m.K, it can provide heat dissipation path vertically to balance the temperature gradient. What is more, adding dummy TSVs can balance density of TSV distribution, which also helps to reduce the stress gradient.

1) Reduction of Thermal Gradient

For a chip level thermal analysis, single dummy TSV impact is not effective. Dummy TSVs are modeled in terms of local density as shown in Fig. 5, where dummy TSVs occupy an area of \(\eta A\) on a regular chip area \(A\). Considering the vertical heat dissipation, the total thermal conductivity \(\lambda\) is given by

\[
\lambda = \eta \lambda_{TSV} + (1 - \eta) \lambda_0;
\]

\[
= (\eta + \delta \eta) \lambda_{TSV} + (1 - (\eta + \delta \eta)) \lambda_0; \delta n = \delta \eta A
\]

(10)

where initial thermal conductivity is \(\lambda_0\), initial TSV density is \(\eta\), change of TSV density is \(\delta \eta\), and change of TSVs is \(\delta n\) with respect to initial number of TSVs \(n\).

As such, the temperature gradient reduction with a change of \(\delta \eta\) TSV density is given by

\[
\delta T = T - T_0 = \frac{P \cdot l}{A \lambda_0} \cdot \frac{\delta \eta}{\lambda_{TSV} - \lambda_0} + \eta + \delta \eta
\]

(11)

where \(P\) is the heat power flowing from chip to heat-sink, and \(l\) is the length of heat-transfer path distance with a chip area of \(A\). From (11), one can observe that as \(\delta \eta\) approaches or becomes larger than \(\lambda/(\lambda_{TSV} - \lambda_0) + \eta\), the reduction in temperature due to dummy TSVs becomes saturated.

2) Reduction of Stress Gradient

Considering a square having four TSVs at its four corners, all transistors inside that particular square will experience stress from all four TSVs. The stress contour from TSVs and its impact on the neighbor transistors are shown in Fig. 4b. The stress on each of the transistor from different TSVs can be calculated by (6). What is more, it can be observed that there will be reduction of stress gradient with insertion of dummy TSVs at proper locations. The stress gradient reduction \(\delta \sigma\) caused by TSV density difference can be given as

\[
\delta \sigma = -\frac{B \Delta \alpha \Delta T}{2} \left( \frac{R_D}{T} \right)^2 \delta \eta = -\frac{B \Delta \alpha \Delta T}{2} \left( \frac{R_D}{T} \right)^2 \delta \eta A
\]

(12)

where \(\delta \eta\) represents the additional number of TSVs added and \(\delta \eta\) represents change in TSV density due to insertion of additional TSVs. When the density becomes more uniform, the stress gradient also becomes smaller.

B. Sensitivity of Temperature Gradient Reduction

As it can be observed from (11), at a certain dummy TSV density the reduction in temperature gradient tends to saturate. Sensitivity of temperature gradient reduction with respect to the dummy TSV density can be given by

\[
\frac{\partial T}{\partial \eta} = \frac{P \cdot l}{A \lambda_0} \cdot \frac{\eta_0}{(\eta_0 + \eta)^2}; \eta_0 = \frac{\lambda_0}{\lambda_{TSV} + \lambda_0}
\]

(13)

From (13), it can be clearly concluded that, when the dummy TSV density \(\eta\) is smaller than its saturation value, \(\eta_0\), i.e. \(\eta \ll \eta_0\), the sensitivity of temperature gradient with dummy TSV density remains almost constant; and as \(\eta \gg \eta_0\) the sensitivity approaches zero, implying that reduction of temperature tends to saturate. Thus, temperature sensitivity function with dependence on dummy TSV density can be used during the optimization of dummy TSV insertion.

C. Sensitivity of Stress Gradient Reduction

From (14), one can observe that the stress gradient reduction depends on the TSV density. The sensitivity of stress gradient reduction with respect to TSV density is given by
\[
\frac{\partial \sigma}{\partial \eta} = -\frac{B \Delta \alpha \Delta T A}{2} \left( \frac{R}{r} \right)^2. \tag{14}
\]

The stress gradient reduction sensitivity may look as independent from TSV density, but depends on radius of TSV and area, which has impact on TSV density. For a particular TSV density, within given area, TSVs greater than a particular radius cannot be inserted. When the radius of TSV becomes very small compared to the distance, the stress gradient tends to saturate early compared to TSVs with larger radius, due to smaller mechanical stress from TSV. All these sensitivity analysis will be deployed in the optimization of clock-skew reduction in later part of the paper.

V. 3D NONLINEAR DELAY MODEL

The effect of temperature and stress on electrical parameters can be utilized for a detailed delay analysis when including signal TSV. Moreover, one can further study the delay sensitivity with respect to dummy TSV insertion density as well. In this section, the delay modeling in 3D-IC is developed based on electrical-thermal coupling, electrical-mechanical coupling and eventually electrical-thermal-mechanical coupling. Its sensitivity is also derived with respect to dummy TSV density. As shown in Fig. 6, the 3D delay model for clock-tree includes driver, 2D wire and 3D signal TSV.

A. Delay Model with Electrical-thermal Coupling

Temperature has a major impact on signal TSV nonlinear MOSCAP capacitance according to (3). By considering nonlinear electrical-thermal coupling from signal TSV as in (5), the signal delay \( D_{TSV1} \) for clock-tree shown in Fig. 6 is calculated as

\[
D_{TSV1} = R_{in}\alpha\beta_2 T^3 + R_{in}[(1 - \alpha T_0)\beta_2 + \alpha\beta_1]T^2 + \alpha(D_0 + R_{in}C_0) + (1 - \alpha T_0)R_{in}\beta_1 T + (1 - \alpha T_0)(R_{in}C_0 + D_0)
\]

\[
R_{in} = \frac{R_D}{S_D} + S_{w1}R_{w1} + \frac{R_T}{2S_T}; \tag{15}
\]

\[
D_0 = \frac{1}{2}(S_{w1}^2R_{w1}C_{w1} + S_{w2}R_{w2}SLCL) + \frac{R_T}{S_T} + \frac{R_D}{S_D}(S_{w1}C_{w1} + S_{w2}R_{w2}S_{w2}SLCL + S_{DP}C_P)
\]

where \( R_{in} \) is the total resistance counted from TSV capacitor, \( C_T \) is the total capacitance counted from the input, \( C_L \) is the load capacitance, \( \alpha \) is temperature dependent coefficient of \( R_T, \beta_1, \beta_2 \) are the first order and second order temperature dependent coefficients of \( C_T \), and \( T \) represents the temperature. Note that \( D_0 \) is the delay of the circuit shown in Fig. 6b without TSV; and all other parameters can be found from Table II and Table III.

In the calculation of delay with electrical-thermal coupling, the impacts of horizontal metal wires and buffers on delay are also considered. The nonlinearity in the delay mainly arises from nonlinear temperature-dependent TSV capacitor, while horizontal metal wire is modeled as linear temperature-dependent resistor. Majority of the delay is contributed by the signal TSV.

B. Delay Model with Electrical-mechanical Coupling

In this section, the impact of mechanical stress on the transistors or drivers is further considered. The mechanical stress from TSVs has non-negligible impact on the driver resistance according to (9). The clock-tree delay \( D_{TSV2} \) with electrical-mechanical coupling is then calculated as

\[
D_{TSV2} = \frac{D_\sigma}{1 + m\sigma} + D_w;
\]

\[
D_\sigma = \frac{R_D}{S_D}[C_P S_D + S_{w1}C_{w1} + S_T C_0 + S_{w2}R_{w2} + S_{LTCL}];
\]

\[
D_w = S_{w1}R_{w1}\left[\frac{S_{w1}C_{w1}}{2} + S_T C_0 + S_{w2}R_{w2} + S_{LTCL}\right]
\]

\[
+ S_{w2}R_{w2}\left[\frac{S_{w2}R_{w2}}{2} + \frac{R_0}{S_T}\right] + R_0 C_0
\]

\[
\tag{16}
\]

where \( D_\sigma \) and \( D_w \) are stress dependent delay and independent delay, respectively. \( R_D \) is the driver resistance without impact of stress, and \( R_0, C_0 \) are the temperature independent TSV resistance and capacitance respectively. All other parameters can be found from Table II, III and IV. One can observe that \( D_\sigma \) is composed of stress dependent and stress independent components. Majority of stress affected part is to the driver.

C. Delay Model with Electrical-thermal-mechanical Coupling

Till now, the delay models have been studied by considering electrical-thermal and electrical-mechanical coupling one by one. Considering all the couplings, one can develop an accurate TSV modeling to the delay and skew in 3D clock-tree design. The clock-tree delay \( D_{TSV} \) considering electrical-thermal-mechanical coupling is given by

\[
D_{TSV} = D_c + D(T) + D(\sigma) + D(T, \sigma)
\]

\[
D_c = D_0 - D_\sigma + R_0 C_0(1 + \alpha T_0)
\]

\[
- \frac{R_0(S_{w2}R_{w2} + S_LCL)\alpha T_0}{S_T};
\]

\[
D(\sigma) = \frac{R_D}{S_D(1 + m\sigma)}[S_D C_P + S_{w1}C_{w1} + S_T C_0 + S_{LTCL} + S_{w2}R_{w2}]; \tag{17}
\]

\[
D(T) = R_0\alpha\beta_2 T^3 + R_0[(1 - \alpha T_0)\beta_2 + \alpha\beta_1]T^2
\]

\[
+ [R_0\beta_1(1 - \alpha T_0) + \alpha R_0 C_0 + S_{w1}R_{w1}S_T\beta_1
\]

\[
+ R_0\alpha(S_{w2}R_{w2} + S_{LTCL})]T;
\]

\[
D(T, \sigma) = \frac{R_D}{S_D(1 + m\sigma)}[S_T\beta_1 T + S_T\beta_2 T^2] \tag{16}
\]
where $D_c$ is a constant delay composed of $D_0$ and $D_o$ given in (15) and (16); $D(T)$ is the delay as function of temperature only; $D(\sigma)$ represents the delay as function of stress only; and $D(T, \sigma)$ is the delay as function of both temperature and stress. All other parameters can be found from Table II, Table III and Table IV.

As a summary, delay is quite in contrast with 2D case that has only linear dependence on temperature. The main reason for nonlinearity arises from the nonlinear signal TSV MOSCAP capacitance with respect to temperature.

D. Delay Sensitivity with Respect to Temperature and Stress Gradient

Till now, the impact of different couplings on delays and its dependence with stress and temperature have been studied. With one further step, we show the delay sensitivity with respect to temperature gradient and stress gradient.

The impact of the temperature and stress gradient on delay as variation or clock-skew can be given by

$$S = S_c + D(\delta T) + D(\delta \sigma) + D(\delta T, \delta \sigma)$$

with 

$$S_c = D_c + [R_0r_1 + \alpha R_0C_0 + S_{w1}R_{w1}S_T\beta_1 + R_0\alpha(S_{w2}C_{w2} + S_LC_L)]T_0 + R_0\beta_2T_0^2;$$

$$D(\delta T) = [(R_0\beta_1 + \alpha R_0C_0 + S_{w1}R_{w1}S_T\beta_1 + R_0\alpha(S_{w2}C_{w2} + S_LC_L)) + 2R_0(\beta_2 + \alpha \beta_1)]T_0 + R_0\alpha_2T_0^3;$$

$$D(\delta \sigma) = -\frac{R_D}{S_D(1 + m\sigma_0)^2}[(S_DC_P + S_{w1}C_{w1} + S_TC_0 + S_LC_L + S_{w2}C_{w2}) + 2T_0\beta_1 + 2T_0^2\beta_2]m\delta\sigma;$$

$$D(\delta T, \delta \sigma) = -\frac{R_D}{S_D(1 + m\sigma_0)^2}[(S_T\beta_1 + 2T_0)\delta T + S_T\beta_2(\delta T)^2]m\delta\sigma;$$

$$T = T_0 + \delta T; \quad \sigma = \sigma_0 + \delta \sigma.$$  

Here, $\delta T$ and $\delta \sigma$ represent temperature and stress gradients, respectively. $S_c$ is the temperature and stress gradient independent coefficient, with $D_c$ representing the constant delay given in (17). $D(\delta T)$ and $D(\delta \sigma)$ represent dependence on temperature and stress gradients respectively. $D(\delta T, \delta \sigma)$ represents dependence on both temperature and stress gradients. $T_0$ and $\sigma_0$ represent initial room temperature and stress respectively. All other parameters can be found from Table II, Table III and Table IV.

From (18), the temperature gradient has major impact on the clock-skew sensitivity compared to stress gradient. The nonlinear terms from temperature and its gradient has major impact on clock-skew than that from stress gradient, which is verified by the experiment. When inserting dummy TSVs to reduce the temperature and stress gradient, one can observe the corresponding clock-skew reduction, which can be deployed for the optimization flow as discussed in next section.

VI. 3D CLOCK-SKEW REDUCTION BY NONLINEAR OPTIMIZATION

The nonlinear electrical-thermal-mechanical coupling transforms the 3D clock-skew reduction problem into a nonlinear optimization problem. In this part, a nonlinear programming based algorithm is developed for insertion of dummy TSVs to reduce the clock-skew. Before optimization is performed, sensitivity of clock-skew with respect to dummy TSV density needs to be discussed.

The nonlinear optimization of 3D clock-tree in this paper is performed at micro-architecture level by dividing each layer into $M \times N$ grids. If one TSV passes through a grid $g_i$, the delay contributed by that grid needs to be calculated by the developed coupled electrical-thermal-mechanical model. Generally, temperature has much higher impact than stress. Moreover, the linear delay from horizontal metal wires and buffers are also considered in (19). As such, based on (18), the skew from individual grid $i$ becomes

$$S_i = \begin{cases}  
S_c + [w_1\delta T_i + w_2(\delta T_i)^2 + w_3(\delta T_i)^3] + [s_0 + s_1\delta T_i + s_2(\delta T_i)^2]\delta \sigma; & \text{3D signal TSVs;} \\
3z_0 + z_1\delta T_i : 2D \text{ wires;}
\end{cases}$$

with 

$$w_1 = 2R_0(\beta_2 + \alpha \beta_1)T_0 + [R_0\beta_1 + \alpha R_0C_0 + S_{w1}R_{w1}S_T\beta_1 + R_0\alpha(S_{w2}C_{w2} + S_LC_L)];$$

$$w_2 = 2R_0\alpha_2T_0 + R_0(\beta_2 + \alpha \beta_1); w_3 = R_0\beta_2;$$

$$w_3 = R_0\beta_2;$$

$$S_i = \begin{cases}  
S_c + [w_1\delta T_i + w_2(\delta T_i)^2 + w_3(\delta T_i)^3] + [s_0 + s_1\delta T_i + s_2(\delta T_i)^2]\delta \sigma; & \text{3D signal TSVs;} \\
3z_0 + z_1\delta T_i : 2D \text{ wires;}
\end{cases}$$

with 

$$w_1 = 2R_0(\beta_2 + \alpha \beta_1)T_0 + [R_0\beta_1 + \alpha R_0C_0 + S_{w1}R_{w1}S_T\beta_1 + R_0\alpha(S_{w2}C_{w2} + S_LC_L)];$$

$$w_2 = 2R_0\alpha_2T_0 + R_0(\beta_2 + \alpha \beta_1); w_3 = R_0\beta_2;$$
and
\[ s_0 = -\frac{mR_D}{S_D(1 + m\sigma_0)^2} \left[ (S_D C_P + S_w L C_w1 + S_T C_0 + S_L C_L + S_w C_w2) + S_T \beta T_0 + S_T \beta_2 T_0^2 \right]; \]
\[ s_1 = -\frac{mR_D}{S_D(1 + m\sigma_0)^2} [S_T \beta_1 + 2S_T \beta_2 T_0]; \]
\[ s_2 = -\frac{mR_D}{S_D(1 + m\sigma_0)^2} S_T \beta_2; \]
\[ z_0 = R_0 C_0; \quad z_1 = R_0 C_0 \alpha \]
where \( w_i, s_i \) are the skew coefficients in presence of TSV; \( z_0, z_1 \) are the skew coefficients in the absence of TSV; \( \delta T_i \) and \( \delta \sigma_i \) represent temperature and stress gradients in the \( i \)th grid respectively. \( S_i \) is the temperature and stress independent coefficient of clock-skew and given in (18). Other parameters can be found from Table II, Table III and Table IV.

A. Clock-skew Sensitivity with Respect to Dummy TSV Density

The sensitivity of the clock-skew with respect to dummy TSV density plays an important role during the optimization. From (19), the sensitivity of clock-skew in the \( i \)th grid can be derived as follows

\[ \frac{\partial S_i}{\partial \eta_i} = (S^i_{T,T} + S^i_{T,T}) \cdot \frac{\partial T}{\partial \eta} + (S^i_{\sigma,T} + S^i_{\sigma,T}) \cdot \frac{\partial \sigma}{\partial \eta}; \]  
with 
\[ S^i_{T,T} = w_1 + 2w_2 \delta T_i + 3w_3 (\delta T_i)^2; \]
\[ S^i_{\sigma,T} = (s_1 + 2s_2 \delta T_i) \delta \sigma_i; \]
\[ S^i_{\sigma,\sigma} = s_0; \quad S_{\sigma,T} = s_0 + s_1 \delta T_i + s_2 (\delta T_i)^2. \]

Here, \( S^i_{T,T} \) and \( S^i_{\sigma,T} \) represent the temperature and temperature-stress gradient coefficients for temperature sensitivity in the \( i \)th grid; \( S^i_{\sigma,\sigma} \) and \( S^i_{\sigma,T} \) represent the stress and stress-temperature gradient coefficients for stress sensitivity in the \( i \)th grid; and \( s_1, w_1 \) are the skew coefficients in the presence of TSV and are given in (19). The clock-skew sensitivity depends on the temperature and stress gradient sensitivities, which are eventually related to dummy TSV density, and can be determined from (13) and (14).

Based on the calculation of clock-skew sensitivities, the updated temperature and stress by the updated dummy TSV density \( \eta_i \) in the \( i \)th grid can be given by

\[ T_i^{\text{new}} = T_i + \gamma_P P_i \eta_i; \quad \sigma_i^{\text{new}} = \sigma_i + \gamma_\sigma \eta_i; \]  
(23)
where \( \gamma_P \) and \( \gamma_\sigma \) represent temperature and stress sensitivities in the \( i \)th grid with respect to the dummy TSV density, and are given by \( \partial T_i / \partial \eta_i \) and \( \partial \sigma_i / \partial \eta_i \) determined from (13) and (14), respectively.

Moreover, \( T_i, \sigma_i \) represent temperature and stress in the \( i \)th grid; and \( P_i, \eta_i \) is the heat power density and TSV density in the \( i \)th grid. Based on the updated values of temperature and stress, by updating TSV density, the skew and its sensitivity values in (19) and (22) can be updated.

B. Nonlinear Optimization

Clock-tree branch \( B_k \) is a set of grids that branch \( k \) passes through, \( B_k = \{ g_i \mid \text{branch } k \text{ passes } g_i \} \), with \( g_i \), representing the \( i \)th grid. Therefore, the clock-skew of a clock-tree branch is the sum of skews from all the grids it passes through

\[ S = \sum_{i \in B_k} S_i \]  
(24)
where \( S_i \) represents the skew from the \( i \)th grid among set \( B_k \). Based on the derived sensitivity and skew function, a nonlinear optimization can be performed as follows.

Substituting (19), (22) and (23) into (24), the clock-tree branch skew \( S \) converts to a quadratic function of inserted dummy TSV density \( \eta_i \). By considering clock-skew from each grid, one can represent clock-skew into a matrix form as

\[ S = c + f^T \mathbf{x} + \frac{1}{2} \mathbf{x}^T H \mathbf{x} \]  
(25)
with
\[ c = S_c; \quad f = \left( f_0, f_1 \right)^T; \quad \mathbf{x} = \left( \eta_0, \eta_1 \right)^T; \quad H = \left( \begin{array}{c} \left( S^i_{T,T} + S^i_{T,T} \right) \gamma_P^i + (S^i_{\sigma,T} + S^i_{\sigma,T}) \gamma_\sigma^i \end{array} \right); \]
\[ f_i = \begin{cases} 0 & \text{if } i \in B_k; \\ \delta T_i & \text{else}; \end{cases} \]
\[ H = \left( \begin{array}{c} H_{0,0} \quad H_{0,0} \quad \cdots \quad H_{0,N^2} \\ H_{1,0} \quad H_{1,1} \quad \cdots \quad H_{1,N^2} \\ \vdots \quad \vdots \quad \ddots \quad \vdots \\ H_{(M^2 \times N)^0} \quad H_{(M^2 \times N)^1} \quad \cdots \quad H_{(M^2 \times N),N^2} \end{array} \right) \]
\[ H_{i,j} = \left( \begin{array}{c} (6w_3 \delta T_i \gamma_P^i + 2w_2 \gamma_P^i + s_1 \gamma_\sigma^i + 2s_2 \delta T_j) \\ (s_1 \gamma_P^j + 2s_2 \delta T_j \gamma_\sigma^j) \end{array} \right); \]
\[ \text{if } i, j \in B_k; \]
\[ 0 \text{ else}. \]
Here, $c$ represents the zero-order coefficient of clock-skew; $f$, $H$ represent the linear and nonlinear coefficients of clock-skew; $\mathbf{x}$ represents the dummy TSV density vector; $N_s$ represents total number of sinks; and $M \times N$ is the total number of grids; $\gamma^T_r$ and $\gamma^T_b$ are temperature and stress gradient sensitivities in $i^{th}$ grid given by (23).

Since clock-skew is the difference in delay between two clock sinks, the problem thus becomes to minimize the skew variance over all clock-tree branches $B_k$, i.e. to minimize variance of $S$ in (25)

$$\min : f(S) = \frac{1}{N_s - 1} \sum_{k=1}^{N_s} (S - \bar{S})^2 \quad (26)$$

where $S$ represents the clock-skew for $B_k$ clock-tree branches, and $\bar{S}$ represents the average skew of $S$ for for $N_s$ sinks.

As such, $f(S)$ becomes a quadratic function of $\mathbf{x}$, given by

$$\bar{S} = \frac{1}{N_s} \sum_{k=1}^{C} S = \tilde{c} + \tilde{f}^T \mathbf{x} + \frac{1}{2} \mathbf{x}^T H \mathbf{x}. \quad (27)$$

where $c$ and $\tilde{c}$ represent the zero-order coefficient of clock-skew and its mean; $f$ and $\tilde{f}$ represent the linear coefficient vector of clock-skew and its mean; and $H$ and $\tilde{H}$ represent the nonlinear coefficient matrix of clock-skew and its mean.

Substituting (27) in (26), the original problem can be rewritten as one polynomial function by

Problem 2:

$$\min : f(\mathbf{x}) = \frac{1}{N_s - 1} \sum_{k=1}^{N_s} (\tilde{c}^2 + 2\tilde{c}f^T \mathbf{x} \quad (28)$$

+ $\mathbf{x}^T (\tilde{f}^T \tilde{f} + \tilde{H} \mathbf{x}) + \tilde{H} \mathbf{x} \mathbf{x}^T \tilde{H} \mathbf{x}$)

where $\tilde{c} = c - \bar{c}$, $\tilde{f} = f - \bar{f}$ and $\tilde{H} = H - \bar{H}$. All these values represent the deviations from their means.

Though the clock-skew depends on TSV density matrix $\mathbf{x}$, neither large number of TSVs nor small number of TSVs can be inserted, because of design constraints below

$$lb \leq \mathbf{x} \leq ub \quad (29)$$

where lower bound $lb$ is determined by foundry process such as minimum metal density; and upper bound $ub$ is determined by the maximum allowed overhead with respect to signal routing.

C. Conjugate-gradient Solving

Now the objective to minimize clock-skew becomes minimizing (28). This can be done by finding the optimum value of the dummy TSV density vector $\mathbf{x}$, that also satisfies constraints given in (29). Conjugate-gradient method with line-search [28] can be an efficient method to solve this non-linear equation with given constraints.

To remove inequalities in problem 2, Karush-Kuhn-Tucker (KKT) optimization method along with Lagrange penalty factor, $\xi$ is used to reformulate original problem by

Problem 3:

$$\min : f^*(\mathbf{x}) = f(\mathbf{x}) + \xi h^2(\mathbf{x}) \quad (30)$$

with

$$h(\mathbf{x}) = \begin{cases} 0, & lb \leq \mathbf{x} \leq ub \\ \varphi > 0, & \text{otherwise} \end{cases} \quad (31)$$

where $f^*(\mathbf{x})$ is the objective function by considering boundary conditions, i.e., removing inequalities of $f(\mathbf{x})$; $\xi$ is the Lagrange penalty factor, which can be determined as stationary point of $f(\mathbf{x})$; and $\varphi$ is a weighting parameter.

Conjugate-gradient method iteratively searches for value of $\mathbf{x}$ that can minimize $f^*(\mathbf{x})$ along the search gradient vector $g_k$, which points to the direction where lies the greatest rate of variation of objective function $f^*(\mathbf{x})$. To achieve a converged solution, in each iteration $k$, the search direction vector $d_k$, which indicates the direction where the objective variable has to be varied, moves in a negative gradient direction to minimize the variation. Thus, a new search direction vector $d_{k+1}$ can be obtained by linear addition of the previous search direction vector $d_k$ with the current negative search gradient vector $g_k$.

Therefore, the next search direction vector becomes

$$d_{k+1} = -\nabla f^*(\mathbf{x}_k)^T + \frac{g_{k}^T}{g_k} g_k + \frac{1}{g_k} g_k + d_k; \quad g_k = -\nabla f^*(\mathbf{x}_k). \quad (32)$$

where $d_{k+1}$ is the search direction vector. Note that $g_k$, $g_k^T$ are search gradient vectors of $\mathbf{x}_k$, determined from the slope of search vector. Based on the search direction vector $d_k$, and gradient $g_k$, the optimal value for step-size $\alpha_k$, is decided to minimize the function $f^*(\mathbf{x}_k + \alpha_k d_k)$.

The new vector $\mathbf{x}$ can be updated based on the previous direction vector $d_k$ by

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k d_k; \quad \alpha_k = \frac{g_k^T g_k}{g_k^T f(\mathbf{x}) g_k}. \quad (33)$$

This iterative search for the optimum value of TSV density vector $\mathbf{x}$ stops when the difference in successive approximations of $\mathbf{x}_k$ reaches certain threshold. To perform with the faster convergence and avoid the local minimum, the problem is solved by starting with some randomly generated approximation of $\mathbf{x}_i$. Once the final value for $\mathbf{x}$ is reached, then it can satisfy (30) and density constraints based on stress and temperature gradients, leading to the reduction of the clock-skew.

VII. EXPERIMENTAL RESULTS

The electrical analysis of signal TSVs is performed based on (3) and (5), and the mechanical analysis is based on (6) and (9). COMSOL multi-physics simulator [29] is used to verify the results. A 4-layer 3D-IC for clock-tree design is constructed with each layer having a thickness of $40\mu m$ and bottom layer having a thickness of $200\mu m$. They are stacked vertically, resulting in total height of $320\mu m$. The height of TSV is same as the thickness of layer. The heat-sink is placed at the bottom with a distributed thermal conductance of $1.24 \times 10^5 W/(K.m^2)$. The IBM clock-tree benchmarks [27] are used for synthesizing 3D clock-tree design using the method in [11]. Each layer in 3D-IC design is considered to be one Alpha-2 processor, stacked one above other forming four layers. After forming a zero-skew initial clock-tree in
3D-IC, the dummy TSVs are inserted and clock-skew is calculated with consideration of nonlinear electrical-thermal-mechanical coupling under temperature and stress gradients. The power traces and temperature profiles are generated with input of SPEC2000 [30] by GEM5 [31] and 3D-ACME [32]. The experiments are run on Intel core-i5 2400 CPU with 3.1 GHz clock frequency and 8 GB RAM. All optimization procedures are programmed in C++.

A. Device Modeling

In this subsection, device modeling for signal TSV, driver and impact of dummy TSV insertion are discussed.

1) Signal TSV

Variation of signal TSV capacitance and delay with respect to temperature is discussed in this part. The signal TSV is modeled by considering the nonlinear temperature dependent capacitance as described in (5). For a 3D clock-tree shown in Fig. 2, using TSV of height 40µm, diameter of 15µm with a resistance of 44mΩ at room temperature. Based on the measured results reported in [8], the values for coefficients of temperature dependent parameters $\alpha$, $C_0$, $\beta_1$ and $\beta_2$ in (1) and (5) are used as: 0.00125/K, 88.8fF, 0.0067fF/K and 0.0014fF/K², respectively. In addition, for reliability consideration, a bundle of TSVs are used for signal distribution instead of a single TSV. TSV Bundle is formed by grouping a few number of TSVs, which are named as T2, T4, T8 and T10 to represent 2, 4, 8 and 10 TSVs in each bundle, respectively.

The nonlinear variation of the signal TSV MOSCAP based on (5) is shown in Fig. 7a. It can be observed from Fig. 7a, at high temperatures the TSV capacitance varies nonlinearly due to the existence of liner material. The same is explained mathematically in (4) and (5). For example, one signal TSV capacitance at room temperature $25^\circ C$ is $87fF$, at $75^\circ C$ is nearly $93fF$, and at $150^\circ C$ is $113fF$, which shows a nonlinear growth. Experiments with process variations in TSV are carried out by varying its capacitance with a maximum of 10%. The according variation in delay of signal TSVs is however, less than 3%, and hence is negligible when compared to thermal or mechanical impact.

To obtain pure signal TSV delay, length of input and output 2D wires to signal TSV are assumed to as small as possible. An inverter in 22nm CMOS process is used as buffer with following settings: $R_D = 1000\Omega$ and $SP_{CP} = SL_{C_k} = 2fF$. The nonlinear effects of temperature on RC-delay at different temperatures for different TSV bundles T2, T4, T8 and T10 are shown in Fig. 7b. Though the nonlinear temperature dependent MOSCAP contributes to a significant amount of delay, the use of signal TSV bundles can help in reduction of temperature, thereby reducing the overall skew.

It can be observed from Fig. 7b, that delay for T8-bundle at 120$^\circ C$ reaches nearly 100ps, which is 67% of the half-clock cycle for a 3.3 GHz multi-processor. For a normal temperature of 75$^\circ C$, the delay for TSV T8-bundle is nearly 60ps; and at the maximum temperature of 200$^\circ C$, the delay for T8-bundle reaches nearly 140ps, which is nearly 46% of a clock cycles of 3.3GHz multi-processor. This delay is of serious concern if no cooling is applied.

These results are consistent with discussion in Section III-A. It also can be observed that if TSV is modeled as the traditional linear coupled model, then the calculated delay will be less. For example for T10-bundle, at a temperature of nearly 125°C, the delay with the nonlinear coupled model is nearly 130ps whereas with the traditional linear coupled model the delay is 120ps. This difference can bring big impact on 3D clock-tree designs.

2) Driver

TSV can exert stress with impact on the mobility and delay of driver. The impact of mechanical stress is mainly on the devices that are on substrate. This mechanical stress affects the driver resistance by enhancing its mobility. In the following, the impacts of thermal-mechanical and electrical-mechanical effects on driver are presented, respectively.

a) Thermal-mechanical Impact

The mechanical stress from TSV is caused by difference of CTEs between TSV and substrate. Different layers of 3D-IC have different temperatures and hence TSVs and substrate will be at different temperature, resulting in stress gradient. In this study, all TSVs are considered to have a diameter of 15µm and a density of 400/mm². The exerted mechanical stress from TSV on device is given by (6).

By considering the stress from all TSVs, the exerted mechanical stress on a driver that is placed inside a square surrounded by TSVs is shown in Fig. 8a. It can be observed that, the stress and mobility remains nearly uniform out of a particular distance, which defines the keep-out-zone; but for area inside there is a significant variation in stress and mobility observed. In our experiment, a keep-out-zone of 3µm is considered.

The coupled impact of TSV density and temperature gradient on stress is shown in Fig. 8b. When temperature gradient increases and at high TSV density, the amount of exerted stress on the substrate will be high. Thus, to reduce TSV stress on substrate, temperature gradient has to be reduced as well. What is more, the amount of stress can be varied when the TSV density is different. Let us consider a temperature gradient of 150°C, the stress at a TSV density of 200/mm² is 28.93MPa, and the stress at a TSV density of 400/mm² is 57.87MPa, indicating stress has impact
from temperature gradient as well as TSV density.

b) Electrical-mechanical Impact

The electrical-mechanical impact on driver is studied in this part. As the amount of exerted mechanical stress varies, the deformation in lattice structure also varies, resulting in variation of carrier mobility.

For the purpose of illustration, a single TSV having a diameter of 15µm is considered as the source of stress, the variation in mobility and delay due to the exerted stress with different distance is discussed here. Considering a keep-out-zone of 3µm, inside which variation of mobility and delay with distance for a 22nm metal gate PMOS and a NMOS placed on substrate is shown in Fig. 9a. The delay of one according driver is shown in Fig. 9b. Note that all simulations are carried out in a SPICE simulator [33].

Moreover, one can have the following observations from Fig. 9a. The amount of stress exerted varies with the distance. In addition, with same amount of stress, the variation in hole mobility is higher than electron mobility. For example, for a distance of 2µm, there is a variation of nearly 4% and 1.8% in mobilities of holes and electrons, respectively.

What is more, the variation of delay of the driver in the presence of mechanical stress is shown in Fig. 9b. It can be observed that as the distance increases, the variation in delay decrease as the exerted stress decreases with distance. There is a decrease of just 1.6% delay at a distance of 4µm whereas nearly 3.7% at distance of 1µm. So if the keep-out-zone is increased, there will be less stress experience with small variation of mobility and delay.

3) Dummy TSV

The reduction of temperature and stress gradient with insertion of dummy TSVs is presented in this section.

In this study, each layer is provided with the same power density of $P$, which serves as the heat source. Study is performed for different values with $P = \{6, 80, 115\} \text{W/m}^2$. In experiment, temperature at each layer is collected without dummy TSVs initially. Note that the liner material of dummy TSV is Si$_3$N$_4$, of which the thickness is 200nm. Based on the formed temperature distribution, dummy TSVs are inserted with density upper and lower bounds discussed in Section VI-C.

The temperature reduction in each layer with insertion of dummy TSVs is shown in Fig. 10. It can be observed that the reduction in temperature initially increases but tends to saturate as a particular limit is reached. This is consistent with the discussion in IV-A. It can also be observed that reduction in temperature for bottom layer, i.e., layer-3 is less than other layers. As it is close to heat-sink, insertion of dummy TSV does not make big difference compared other layers. In addition, the maximum inserted dummy density observed is 400/µm$^2$ with the maximum temperature reduction of nearly 50°C in layer-0 i.e. the top layer.

b) Stress Gradient Reduction

Note that stress and stress gradient depends on both TSV density and temperature gradient. Stress gradient at different temperature gradients and TSV densities can be found in Fig. 11. As shown in Fig. 11a, a block B having a constant...
TSV density of $400/mm^2$ is considered. Another block A has a temperature gradient of $180^\circ C$ and its TSV density can be varied. The stress gradient between the two blocks is shown in Fig. 11b. It can be observed that as TSV density is increased, the stress gradient tends to decrease. The stress gradient reduction for same setup but with a temperature gradient of $250^\circ C$ is also plotted.

B. Delay with Electrical-thermal-mechanical Coupling

The delay of clock-tree by considering nonlinear electrical-thermal-mechanical coupling is presented in this subsection. Considering a zero-skew 3D clock-tree shown in Fig. 6b, the variation of delay on a single signal TSV with drivers in different technologies are presented in Fig. 12.

In Fig. 12, delay values are shown for a single TSV of $15\mu m$ diameter having buffers at both ends placed at $3\mu m$ distance from TSV (Fig. 6a). One can observe that one signal TSV adds up the delay by nearly 4 times, which indicates that the TSV delay is comparable to the minimum sized driver delay. As signal TSV is modeled as a nonlinear MOSCAP with nonlinear temperature-dependent, the delay increases with temperature significantly when electrical-thermal coupling is considered. It can be observed that the delay can increase by nearly 15% for all technologies when temperature is increased at $200^\circ C$. Then, with the electrical-mechanical coupling considered, the delay with stress gradient at a reference temperature of $75^\circ C$ is calculated, which is found to be of $9\%$ lower than the delay without insertion of TSV as stress enhances mobility. By considering all these effects, there is approximately 10% delay variation introduced at $200^\circ C$.

C. Skew Reduction by Nonlinear Optimization

The nonlinear optimization of clock-skew reduction for 3D clock-tree design by the insertion of dummy TSVs is presented in this subsection. With the help of 3D-ACME, a 3D-IC thermal simulator based on Hotspot [34], temperature distribution at each layer can be obtained. Moreover, average temperature based on SPEC2000 benchmarks are taken to avoid application specific temperature distribution [4]. To reduce the area overhead caused by dummy TSVs, the dummy TSV insertion density is limited to $7\%$ of the total area. Note that all benchmarks are pre-synthesized to 4-tier 3D clock-tree by [11] considering the wire-length and buffer.

As clock-skew is the difference in delay between two sinks, the variation of clock-skew with TSVs of one 3D H-tree is shown in Fig. 13. It can be observed from Fig. 13, that the initial clock-skew without insertion of dummy TSVs is higher than the clock-skew after insertion of dummy TSVs. For example, for a row-grid 16 and column-grid 24, having dummy TSV inserted results in a clock-skew reduction of nearly 17ps.

For the same 3D H-tree example discussed above, Fig. 14 shows the clock-skew for different TSV bundles presented. Here, orig represents the value of clock-skew without insertion of dummy TSVs. It can be observed that by modeling electrical-thermal and electrical-mechanical impacts during clock-skew reduction, one can observe the clock-skew reduction of 51 to 53%. With the consideration of all the impacts, i.e., electrical-thermal-mechanical coupled model, there can be a reduction of nearly 64% in the overall clock-skew, which is desired.

Next, a 3D clock-tree from the IBM benchmark r5 is studied. The 3D clock-tree after insertion of dummy TSVs is shown in Fig. 15 with TSVs in each layer indicated by solid dots. Dummy TSV insertion is performed under nonlinear optimization presented in Section VI such that clock-skew is reduced under an optimized dummy TSV insertion. One can observe that a large number of TSVs are inserted in the top layer, i.e., tier 0 compared to other layers since the top layer is farthest one from the heat-sink.

Finally, the comparison of clock-skew for different benchmarks before and after insertion of dummy TSVs is shown in Table VI with a detailed summary, which shows...
Moreover, insertion of dummy TSVs is utilized to balance electrical-thermal-mechanical delay model is developed. Moreover, TSVs exert stress on drivers which modify mobility nonlinearly with temperature due to liner material. What is for inter-layer signal connections act as MOSCAPs varying using through-silicon-vias (TSVs). The signal TSVs utilized under insertion of dummy TSVs, listed under result in clock-skew reduction by 61.3% on average, of nonlinear electrical-thermal-mechanical-thermal impacts presented in Table VI.

Fig. 15: 3D clock-tree after insertion of dummy TSV (Black dots) with balanced clock-skews for: (a) Tier 0 (b) Tier 1 (c) Tier 2 (d) Tier 3

the impact of 3D electrical-thermal-mechanical coupled delay model and also the insertion of dummy TSVs to reduce gradient. Clock-skew values are reported in pico-seconds. The runtime is in seconds when performing optimization of insertion of dummy TSVs based on linear and nonlinear modeling. Different clock-tree benchmarks and their corresponding numbers of buffers and TSVs are presented in Table VI. The delay models with consideration of nonlinear electrical-thermal-mechanical-thermal impacts result in clock-skew reduction by 61.3% on an average, listed under nonlin column, compared to clock-skew without insertion of dummy TSVs, listed under Orig column. Note that the reduced clock-skew by linear modeling is listed under Lin column with 49.1% clock-skew reduction compared to Orig. The runtime on average for nonlinear optimization is 611s and is 238s for linear optimization. Although it seems to be time consuming for the nonlinear method, the effective reduction in clock-skew is 12.2% larger in the nonlinear method than the linear method.

VIII. CONCLUSION

There is an emerging need for robust 3D-IC design when using through-silicon-vias (TSVs). The signal TSVs utilized for inter-layer signal connections act as MOSCAPs varying nonlinearly with temperature due to liner material. What is more, TSVs exert stress on drivers which modify mobility and delay. Therefore, the non-uniform temperature and stress introduce large delay variation. In this paper, nonlinear electrical-thermal-mechanical delay model is developed. Moreover, insertion of dummy TSVs is utilized to balance temperature and stress gradients for 3D clock-tree with clock-skew reduction. A nonlinear programming problem is formulated to determine the optimum dummy TSV density guided by sensitivity of clock-skew reduction. A number of 3D clock-tree benchmarks are used to verify the model and also the optimization. The results show a reduction of clock-skew by 61.3% on average when nonlinear electrical-thermal-mechanical delay model is applied.

<table>
<thead>
<tr>
<th>Type</th>
<th>Orig</th>
<th>Lin</th>
<th>Imp</th>
<th>Time (s)</th>
<th>Nonlin</th>
<th>Imp</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>15.70</td>
<td>17.97</td>
<td>26.45</td>
<td>891.08</td>
<td>18.69</td>
<td>24.21</td>
<td>899.21</td>
</tr>
<tr>
<td>T2</td>
<td>31.50</td>
<td>35.16</td>
<td>45.96</td>
<td>909.25</td>
<td>37.02</td>
<td>45.96</td>
<td>909.21</td>
</tr>
<tr>
<td>T3</td>
<td>64.57</td>
<td>76.21</td>
<td>93.92</td>
<td>912.39</td>
<td>70.89</td>
<td>93.92</td>
<td>912.38</td>
</tr>
<tr>
<td>T4</td>
<td>64.57</td>
<td>76.21</td>
<td>93.92</td>
<td>912.39</td>
<td>70.89</td>
<td>93.92</td>
<td>912.38</td>
</tr>
<tr>
<td>T5</td>
<td>64.57</td>
<td>76.21</td>
<td>93.92</td>
<td>912.39</td>
<td>70.89</td>
<td>93.92</td>
<td>912.38</td>
</tr>
<tr>
<td>T6</td>
<td>64.57</td>
<td>76.21</td>
<td>93.92</td>
<td>912.39</td>
<td>70.89</td>
<td>93.92</td>
<td>912.38</td>
</tr>
<tr>
<td>T7</td>
<td>64.57</td>
<td>76.21</td>
<td>93.92</td>
<td>912.39</td>
<td>70.89</td>
<td>93.92</td>
<td>912.38</td>
</tr>
<tr>
<td>T8</td>
<td>64.57</td>
<td>76.21</td>
<td>93.92</td>
<td>912.39</td>
<td>70.89</td>
<td>93.92</td>
<td>912.38</td>
</tr>
<tr>
<td>T9</td>
<td>64.57</td>
<td>76.21</td>
<td>93.92</td>
<td>912.39</td>
<td>70.89</td>
<td>93.92</td>
<td>912.38</td>
</tr>
<tr>
<td>T10</td>
<td>64.57</td>
<td>76.21</td>
<td>93.92</td>
<td>912.39</td>
<td>70.89</td>
<td>93.92</td>
<td>912.38</td>
</tr>
<tr>
<td>Mean</td>
<td>64.57</td>
<td>76.21</td>
<td>93.92</td>
<td>912.39</td>
<td>70.89</td>
<td>93.92</td>
<td>912.38</td>
</tr>
</tbody>
</table>

TABLE VI: 3D clock-skew reduction by linear and nonlinear delay models

REFERENCES


Chuan Seng Tan (S’00-M’07) received the B.Eng. in electrical engineering from the University of Malaya, Kuala Lumpur, Malaysia, in 1999, the M.Eng. degree in advanced materials from the National University of Singapore, Singapore, under the Singapore-Massachusetts Institute of Technology (MIT) Alliance Program, in 2001, and the PhD degree in electrical engineering from MIT, Cambridge, in 2006. He joined Nanyang Technological University, Singapore, in 2006, as a Lee Kuan Yew Postdoctoral Fellow, and since July 2008, he is a holder of a prestigious postdoctoral fellow of the National University of Singapore. His research interests are semiconductor process technology and device physics. He is currently working on process technology of 3-D integrated circuits (3-D ICs). Dr. Tan was the recipient of the Applied Materials Graduate Fellowship for 2003–2005.

Sung Kyu Lim (S’94-M’00-SM’05) received the B.S., M.S., and Ph.D. degrees from the Computer Science Department, University of California, Los Angeles (UCLA), in 1994, 1997, and 2000, respectively. In 2001, he joined the School of Electrical and Computer Engineering, Georgia Institute of Technology, where he is currently an Associate Professor. His current research includes the architecture, circuit, and physical design for 3-D ICs and 3-D system-in-packages. Dr. Lim received the Design Automation Conference (DAC) Graduate Scholarship in 2003, and the National Science Foundation Faculty Early Career Development ( CAREER) Award in 2006. He was on the Advisory Board of the ACM Special Interests Group on Design Automation (SIGDA) from 2003 to 2008 and received the ACM SIGDA Distinguished Service Award in 2008.

Sai Manoj P.D. (S’13) received the B.Tech and M.Tech. degree from JNTU Anantapur, India in 2010 and IIIT, Bangalore, India in 2012 respectively. He joined Nanyang Technological University in 2012 and is currently working towards the Ph.D. degree with School of Electrical and Electronics Engineering. His research interests are 3D-ICs, Network-on-Chips and Low power system design. He is also recipient of A. Richard Newton Young Research Fellow Award in Design Automation Conference 2013.

Hao Yu (M’06) received the B.S. degree from Fudan University, China; and Ph. D degree from the Electrical Engineering Department, University of California, USA. He was a senior research staff at Berkeley Design Automation. Since October 2009, he has been an Assistant Professor at the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. His primary research interests are 3D-IC and RF-IC at nano-tera scale. He has 94 peer-reviewed IEEE/ACM publications. Dr. Yu received Best Paper Award from the ACM TODAES’10, Best Paper Award nominations in DAC06, ICCAD’06, ASP-DAC’12, Best Student Paper (advisor) Finalist in SiRF’13, and Inventor Award’08 from semiconductor research cooperation. He is associate editor and technical program committee member for a number of journals and conferences.

Yang Shang (S’11) received the B.S. and M.S. degree in Electrical and Electronic Engineering both from Nanyang Technological University, Singapore in 2005 and 2009, respectively. He is currently working towards the Ph.D. degree with School of Electrical and Electronics Engineering at Nanyang Technological University. His research interests are Non-volatile-memory Device Model and Simulation and Meta-material based 60GHz–beyond Phase-arrayed Receiver Design.

Anantapur (M’06) received the B.S. degree from SAI Institute of Technology, Anantapur, India in 2010 and IIIT, Bangalore, India in 2012 respectively. He Joined Nanyang Technological University in 2012 and is currently working towards the Ph.D. degree with School of Electrical and Electronics Engineering. His primary research interests are 3D-ICs, Network-on-Chips and Low power system design. He is also recipient of A. Richard Newton Young Research Fellow Award in Design Automation Conference 2013.

Chuan Seng Tan (S’00-M’07) received the B.Eng. in electrical engineering from the University of Malaya, Kuala Lumpur, Malaysia, in 1999, the M.Eng. degree in advanced materials from the National University of Singapore, Singapore, under the Singapore-Massachusetts Institute of Technology (MIT) Alliance Program, in 2001, and the PhD degree in electrical engineering from MIT, Cambridge, in 2006. He joined Nanyang Technological University, Singapore, in 2006, as a Lee Kuan Yew Postdoctoral Fellow, and since July 2008, he is a holder of a prestigious postdoctoral fellow of the National University of Singapore. His research interests are semiconductor process technology and device physics. He is currently working on process technology of 3-D integrated circuits (3-D ICs). Dr. Tan was the recipient of the Applied Materials Graduate Fellowship for 2003–2005.

Sung Kyu Lim (S’94-M’00-SM’05) received the B.S., M.S., and Ph.D. degrees from the Computer Science Department, University of California, Los Angeles (UCLA), in 1994, 1997, and 2000, respectively. In 2001, he joined the School of Electrical and Computer Engineering, Georgia Institute of Technology, where he is currently an Associate Professor. His current research includes the architecture, circuit, and physical design for 3-D ICs and 3-D system-in-packages. Dr. Lim received the Design Automation Conference (DAC) Graduate Scholarship in 2003, and the National Science Foundation Faculty Early Career Development ( CAREER) Award in 2006. He was on the Advisory Board of the ACM Special Interests Group on Design Automation (SIGDA) from 2003 to 2008 and received the ACM SIGDA Distinguished Service Award in 2008.