Thermal-reliable 3D clock-tree synthesis considering nonlinear electrical-thermal-coupled TSV model

Shang, Yang; Zhang, Chun; Yu, Hao; Tan, Chuan Seng; Zhao, Xin; Lim, Sung Kyu

2013

Shang, Y., Zhang, C., Yu, H., Tan, C. S., Zhao, X., & Lim, S. K. (2013). Thermal-reliable 3D clock-tree synthesis considering nonlinear electrical-thermal-coupled TSV model. 2013 18th Asia and South Pacific Design Automation Conference (ASP-DAC), 693-698.

<https://hdl.handle.net/10356/96541>

<https://doi.org/10.1109/ASPDAC.2013.6509681>

© 2013 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published version is available at: [<http://dx.doi.org/10.1109/ASPDAC.2013.6509681>].

Downloaded on 26 Apr 2023 16:36:37 SGT

Thermal-reliable 3D Clock-tree Synthesis Considering Nonlinear Electrical-thermal-coupled TSV Model

Yang Shang, Chun Zhang, Hao Yu, Chuan Seng Tan

School of Electrical and Electronic Engineering
Nanyang Technological University
Singapore, Nanyang Ave 639798
e-mail: haoyu@ntu.edu.sg

Xin Zhao, Sung Kyu Lim

GTCAD Laboratory
Georgia Institute of Technology
Atlanta, North Ave 30332
e-mail: limsk@ece.gatech.edu

Abstract— 3D physical design needs accurate device model of through-silicon vias (TSVs). In this paper, physics-based electrical-thermal model is introduced for both signal and dummy thermal TSVs with the consideration of nonlinear electrical-thermal dependence. Taking thermal-reliable 3D clock-tree synthesis as a case-study to verify the effectiveness of the proposed TSV model, one nonlinear programming-based clock-skew reduction problem is formulated to allocate thermal TSVs for clock-skew reduction under non-uniform temperature distribution. With a number of 3D clock-tree benchmarks, experiments show that under the nonlinear electrical-thermal TSV model, insertion of thermal TSVs can effectively reduce temperature-gradient introduced clock-skew by 58.4% on average, and has 11.6% higher clock-skew reduction than the result under linear electrical-thermal model.

I. INTRODUCTION

With the provision of interconnection along vertical dimension, 3D integration has become a promising solution for continued scaling for high-performance computing systems. As multiple device-layers (tiers) can be vertically connected by through-silicon vias (TSVs), the latency of the long interconnection in 2D is substantially reduced [1, 2, 3, 4, 5, 6, 7, 8, 9]. At the same time, since the heat-dissipation path becomes far apart from heat-sink, there is severe temperature increase as well as higher temperature gradient for designs in 3D domain. As such, a robust physical design in 3D needs to consider the optimization from both electrical and thermal perspectives.

TSVs are the foundations of 3D integration with applications in inter-layer signal/clock connection, power distribution and also heat removal. Recent device modelings [3, 4] show that due to the existence of liner for isolation, TSVs work quite similarly to the nonlinear MOS-capacitance (MOSCAP) under different signal voltages and operating frequencies. As such, signal delay from TSV capacitance becomes non-negligible when the signal frequency at several gigahertz. Moreover, MOSCAP has nonlinear increase with temperature [4]. The signal delay induced by the nonlinear capacitance becomes even larger at high temperature, which can potentially degrade the signal distribution such as clock. Therefore, instead of modeling TSV under traditional linear electrical-thermal model as a resistor, it becomes important to consider its nonlinear effect for a thermal-reliable 3D design as a nonlinear temperature-dependent capacitor.

Since the performance of clock is sensitive to the delay difference at all sinks, which is known as *skew*, thermal-reliable clock-tree synthesis is one perfect example to study the impact of nonlinear electrical-thermal coupling of TSVs in 3D. In this paper, based on recent measurement results [3, 4], one nonlinear electrical-thermal model is developed for signal TSV. Moreover, based on the accurate multi-physics solver, one temperature-sensitivity function is developed with respect to thermal TSV density. Utilizing the developed

accurate TSV models, a thermal-reliable 3D clock-tree synthesis problem is formulated as a case study to analyze the impact and solution of the nonlinear electrical-thermal behaviors of TSVs. Specifically, the thermal TSVs are inserted [7] for the reduction of clock-skew under temperature gradient in 3D with the consideration of the signal TSV delay. A nonlinear programming-based algorithm is developed and solved for an optimal thermal TSV insertion to minimize the clock-skew. Experiment results show that with reasonable number of thermal TSVs allocated, the average clock-skew can be reduced by 58.4% for clock-tree benchmarks [10] in 3D design [5]. Furthermore, compared to the use of linear electrical-thermal model [2, 5, 11], our approach reduces 11.6% more clock-skew under the same thermal TSV density constraint, which validates the impact of the nonlinearity in TSV models.

The rest of this paper is organized as follows. Section II presents the new clock-skew reduction problem in 3D. In Section III, the nonlinear electrical-thermal model of signal TSV and temperature-sensitivity function with respect to thermal TSV density are developed, respectively. Then, the nonlinear optimization of clock-skew reduction is studied in Section IV. Experiment results are shown in Section V with conclusion in Section VI.

II. PROBLEM FORMULATION

Fig. 1.: 3D clock-tree distribution network at different tiers, (a) Clock-tree with 14 TSV bundle locations (htree1); (b) Clock-tree with 28 TSV bundle locations (htree2); and (c) Layer configuration under non-uniform temperature distribution

Same as in 2D clock-tree synthesis [2, 11, 12], the clock-skew reduction is an important subject under study for the 3D clock-tree synthesis. Fig. 1 illustrates two typical four-layer 3D clock-trees. Different from the 2D clock-tree, TSVs are utilized to provide vertical connections. As such, in addition to traditional techniques such as buffer sizing [2], merging point adjustment [11], wire-length balancing [12],

etc., many new design factors have to be considered with the introduction of the vertical direction. Firstly, the signal TSV becomes a nonlinear MOSCAP with temperature dependence. Secondly, temperature gradient is more severe in 3D designs. As such, the temperature-dependent delay from TSVs in the clock-tree may become dominant when temperature gradient is large in 3D designs as there is a nonlinear dependence between delay and temperature.

Due to the aforementioned two design considerations in 3D clock-tree, the delay at each sink i of one clock-tree is considered as a nonlinear function of temperature distribution Γ : $D_i = f(\Gamma)$. Moreover, the insertion of thermal TSVs can change the temperature distribution and hence adjust the clock-skew S which is defined as the maximal difference between any two clock sinks: $S = \max : |D_i - D_j|$. As such, we have the following problem formulation of thermal-reliable 3D clock-tree synthesis for clock-skew reduction.

Problem 1: Given a pre-synthesized zero-skew 3D clock-tree with signal TSVs as inter-tier connections, the clock-skew S is minimized by allocating position and number of thermal TSVs under temperature distribution Γ with the consideration of nonlinear electrical-thermal models of both signal and thermal TSVs.

III. ELECTRICAL-THERMAL TSV MODELING

This section discuss how to build accurate electrical-thermal TSV models for 3D thermal-reliable clock-tree. As illustrated in Fig.2(a), TSV can provide both electrical connection between adjacent tiers as well as heat dissipation path to the heat-sink. Here we define the TSV used for electrical connection as *signal* TSV and the TSV used for heat dissipation as *thermal* TSV. Note that in order to avoid unwanted diffusion of metal-atom into the silicon substrate, *liner* material (SiO_2 or Si_3N_4) is used for isolation purpose during TSV fabrication in the BEOL (Back-End-Of-Line) process. As discussed in this paper, the liner material can significantly affect the electrical-thermal behavior of TSVs.

Fig. 2.: (a) Signal TSV and thermal TSV in 3D IC; (b) 3D view of TSV; and (c) Equivalent circuit of TSV

A. Signal TSV Modeling

As shown in Fig.2(a), signal TSVs make the connections between BEOL layers of adjacent tiers but are not connected to the heat-sink. In the previous 3D clock-tree synthesis [2, 5], via can be electrically modeled with simple RC-model and only R is temperature dependent [11]. But in 3D IC, TSV is different from the via due to the existence of liner, of which the impact can not be ignored. In fact, liner forms a MOS-capacitance (MOSCAP) in between signal TSV and the substrate. Such a nonlinear capacitance depends on not only the biasing voltage (V_{BIAS}), but also the temperature [4], and hence brings new design implications in 3D IC design such as clock-tree, which has not been considered in previous works [2, 5].

A typical signal TSV model is illustrated in Fig.2(c), in which the RC parameters are given by following equations

$$\frac{1}{C_T} = \frac{1}{C_{ox}} + \frac{1}{C_{dep}}; R_T = \frac{\rho h}{\pi r_{metal}^2} \quad (1)$$

where $C_{ox} = \frac{2\pi\epsilon_{ox}h}{\ln(\frac{r_{ox}}{r_{metal}})}$ is the liner capacitance, $C_{dep} = \frac{2\pi\epsilon_{si}h}{\ln(\frac{r_{dep}}{r_{ox}})}$ is the depletion capacitance of TSV, ρ is the resistivity of metal-material of TSV, h is the TSV height, ϵ_{ox} and ϵ_{si} are the dielectric constant of silicon oxide and silicon, and r_{metal} , r_{ox} and r_{dep} are the outer radius of TSV metal, silicon and depletion region, respectively as shown in Fig.2(b).

The existence of depletion region is due to the work-function difference between the metal-material of TSV and silicon substrate. This is also the reason of nonlinear TSV capacitance against biasing voltage and temperature. As shown in Fig.3, the C-V curve of TSV can be divided into accumulation region, depletion region and inversion region, which are separated by the flat-band voltage (V_{FB}) and threshold voltage (V_T). When working at higher frequency (i.e., $> 1MHz$), the inversion region can be subdivided into the deep-depletion region.

Fig. 3.: Typical C-V curve of TSV MOSCAP with temperature dependence

As the typical V_T of a copper TSV is around $-2V$, the signal TSV capacitance is usually located at inversion region for the application of a digital circuit with positive voltage swing. Further due to the normally high frequency clock signal in 3D clock-tree network, the deep-depletion region is actually of our concern. Note in the deep-depletion region, the TSV C-V curve tends to be flat with changing bias voltage V_{BIAS} .

However, the capacitor for signal TSV still shows nonlinear electrical-thermal coupling effect due to the nonlinear dependence of r_{dep} on temperature. In other words, there exists a nonlinear temperature dependence effect at deep-depletion region for signal TSV, which can be characterized using (2) based on real measurement results from fabricated testing TSVs [4]

$$R_T = R_0(1 + \alpha(T - T_0)); C_T = C_0 + \beta_1 T + \beta_2 T^2 \quad (2)$$

where T is the temperature of TSV, R_0 is the TSV resistance at room temperature T_0 , α is the measured temperature-dependent coefficients, C_0 is the capacitance of TSV (C_T) at zero temperature, and β_1 , β_2 are the first and second order temperature-dependent coefficients of C_T .

As such, we can observe that the delay contribution from nonlinear terms becomes more significant as temperature is increased. For example, when temperature approaches to $200^\circ C$, the first-order and second-order terms contribute similarly to the capacitor and further for delay. With the further consideration of temperature-dependent resistor, the delay of signal TSV can be significantly changed by the temperature variation as discussed later in this section.

B. Thermal TSV Modeling

The previous thermal TSV model also ignores the impact of liner. In fact the thermal conductivity of liner (SiO_2) is one hun-

ded times worse than the thermal conductivity of silicon substrate ($100W/m \cdot K$), which brings non-negligible thermal impact. As shown in Fig.4(a), thermal TSV with higher thermal-conductivity metal-material Cu ($400W/m \cdot K$) forms a high thermal conductivity channel through the 3D IC. However, the liner material still forms a wall for heat dissipation from metal to substrate, and the heat generated can only be transferred to the heat-sink from top and bottom surface of the thermal TSV. As the thermal conductivity of Si_3N_4 ($30W/m \cdot K$) is much larger than that of SiO_2 , Si_3N_4 is used for liner material for thermal TSVs in this paper.

Moreover, in the chip-level thermal analysis, one is not interested in the thermal behavior of one thermal TSV but the total impact of many thermal TSVs. In our approach, we model the thermal TSV in term of local density. As shown in Fig.4(b), thermal TSVs are locally inserted into regular unit chip-area (A) and η is the ratio of area occupied by thermal TSVs.

Fig. 4.: (a) 3D heat removal path with thermal TSV; and (b) 3D view of thermal TSVs insertion

Assuming that unit chip-area A is thermally isolated from the top and the surrounding, the heat generated can only be transferred vertically to the heat-sink at the bottom. As such, one can define thermal conductivity between A and heat-sink by

$$\sigma_{Total} = \eta \cdot \sigma_{TSV} + (1 - \eta)\sigma_0 \quad (3)$$

where σ_{TSV} and σ_0 are the thermal conductivity of thermal TSVs inserted and regular area, respectively.

One can obtain the temperature reduction function with thermal TSV density η as

$$\Delta T = T_0 - T_{TSV} = \frac{P \cdot l}{A\sigma_0} \cdot \frac{\eta}{\frac{\sigma_0}{\sigma_{TSV} - \sigma_0} + \eta} \quad (4)$$

where P is the heat power flowing from A to heat-sink, and l is the equivalent length of heat-transfer path.

Typically, as the number of thermal TSVs inserted is limited for the minimal area overhead, η is much smaller than $\sigma_0/(\sigma_{TSV} - \sigma_0)$. As a result, (4) can be approximated by a linear function of η . On the other hand, when the value of η is approaching or larger than $\sigma_0/(\sigma_{TSV} - \sigma_0)$, the temperature reduction is less sensitive to the inserted thermal TSVs. In other words, the temperature reduction impact of thermal TSVs starts to saturate.

As such, thermal TSVs can be allocated to reduce both the local temperature and the inter-layer temperature difference when inserted at different positions. The temperature-sensitivity function (with dependence on thermal TSV density) is thereby useful for guiding the optimization of thermal TSV insertion. Note that the insertion of thermal TSVs might create obstacles in the routing and occupy the logic placement resources. As such, in order to minimize the impact of thermal TSVs insertion, there is a constraint of the maximum allowable η .

C. Implications to 3D Clock Tree

Next, we discuss the implications of accurate TSV modeling to the delay and skew calculation in the 3D clock-tree. A temperature-

dependent and scalable Elmore delay model is constructed for one typical 3D clock distribution network with signal TSVs in Fig.5.

Fig. 5.: Delay model of clock circuit with nonlinear electrical-thermal coupled signal TSV

With the consideration of nonlinear electrical-thermal coupling for signal TSV in (2), the signal delay in Fig.5 is calculated as

$$\begin{aligned} \tau = & R_{in}\alpha\beta_2T^3 + R_{in}[(1 - \alpha T_0)\beta_2 + \alpha\beta_1]T^2 \\ & + [\alpha(\tau_0 + R_{in}C_0) + (1 - \alpha T_0)R_{in}\beta_1]T \\ & + (1 - \alpha T_0)(R_{in}C_0 + \tau_0) \end{aligned} \quad (5)$$

with

$$\begin{aligned} R_{in} = & \frac{R_D}{S_D} + S_{W1}R_{W1} + \frac{R_T}{2S_T} \\ \tau_0 = & \frac{1}{2}(S_{W1}^2R_{W1}C_{W1} + S_{W2}R_{W2}S_LC_L) \\ & + (\frac{R_T}{S_T} + S_{W1}R_{W1} + \frac{S_{W2}R_{W2}}{2})(S_{W2}C_{W2} + S_LC_L) \\ & + \frac{R_D}{S_D}(S_{W1}C_{W1} + S_{W2}C_{W2} + S_LC_L + S_DC_P) \end{aligned} \quad (6)$$

where R_{in} is the total resistance looking from C_T to the input and τ_0 is the delay of circuit without C_T ; S_D and S_L are the size scaling-factor of driving and loading transistors or buffers; R_D , C_P and C_L are the accordingly unit buffer resistance and capacitance; S_{W1} and S_{W2} are the length scaling-factor of input and output wire connected to TSVs; R_{W1} and R_{W2} are the accordingly unit length wire resistance; and S_T is the number scaling-factor of TSVs. Note that each TSV has the same capacitance C_T and resistance R_T . Note here the delay impact of wires and buffers are also taken into consideration, where their temperature dependent model follows [2].

As we can observe from (5), for a 3D clock-tree distribution network with TSVs, the delay becomes a nonlinear function with temperature

$$D_{TSV} = k_0 + k_1T + k_2T^2 + k_3T^3 \quad (7)$$

which is significantly different from the 2D case in which the delay is only linearly dependent on temperature $\tau = d'_0 + k'_0T$. This is because in 3D the TSV is mainly modeled as nonlinear temperature dependent capacitor, while the wire is mainly modeled as linear temperature dependent resistor [2]. As a result, the electrical-thermal nonlinear coupling from signal TSV may significantly increase the clock-skew due to the large temperature gradient in 3D IC. The proper design by applying thermal TSVs for heat-removal and further to balance the clock-skew thereby becomes one important approach to be explored for 3D clock-tree.

IV. NONLINEAR OPTIMIZATION OF SKEW REDUCTION

Due to the nonlinear electrical-thermal coupling, the clock-skew reduction for 3D clock-tree becomes nonlinear optimization problem. In this section, we introduce one nonlinear programming based algorithm for thermal TSV insertion to minimize the thermal induced clock-skew for 3D clock-tree network.

Note that in this paper, for a clock-tree with C sinks, the clock-skew S is defined as the maximum delay difference between any two sinks i and j :

$$S = \max : |D_i - D_j|, 0 \leq i, j \leq C \quad (8)$$

where D_i and D_j denote the delays of i and j from clock source respectively.

A. Nonlinear Optimization

At micro-architecture level, each tier in the 3D IC can be divided into $M \times N$ grids. When the clock-tree passes the i -th grid g_i , the delay contributed by g_i can be calculated by the developed electrical-thermal model. Generally the contribution of 3^{rd} order term in (5) is negligible, thus the delay function can be simplified as

$$\tau_i = \begin{cases} d_0 + k_0 T_i + k_1 T_i^2, & \text{signal TSV exists} \\ d'_0 + k'_0 T_i & \text{otherwise} \end{cases} \quad (9)$$

Note the linear-temperature-dependent delay of horizontal metal wires and buffers in (9) are also counted here for accurate model of clock-skew. Then one clock-tree branch C_k is defined as the set of grids $C_k = \{g_i | \text{branch } k \text{ passes } g_i\}$. As such, the delay of one clock-tree branch becomes the summation of delays from all grids

$$D_k = \sum_{i \in C_k} \tau_i. \quad (10)$$

Note that although the exact temperature changes dynamically at runtime, the overall temperature distribution tend to follow certain patterns with steady-state profile. Therefore, the delay is calculated based on expected steady-state temperature gradient, which will be introduced with more details in Section V.

To reduce the clock-skew, the thermal TSVs can be inserted at desired grid to control the local temperature reduction and thus balance the delay at each clock sink. As discussed in Section B, the temperature reduction depends linearly on the allocated thermal TSV density x_i as well as the local power density P_i :

$$T_i^{new} = T_i - \gamma P_i x_i \quad (11)$$

where γ is the *thermal sensitivity* capturing $\Delta T / \Delta x$.

Substituting equations (9) and (11) into (10), the clock-tree branch delay D_k becomes a quadratic function of inserted thermal TSV density x_i :

$$D_k = c_k + \mathbf{f}_k^T \mathbf{x} + \frac{1}{2} \mathbf{x}^T H_k \mathbf{x} \quad (12)$$

where column vector $\mathbf{f}_k$ and diagonal matrix H_k represent linear and quadratic coefficients, respectively.

As a result, the 3D clock-tree skew reduction problem can be detailed as to minimize the delay variance over all clock-tree branches C_k

$$\min : f(\mathbf{D}) = \frac{1}{C-1} \sum_{k=1}^C (D_k - \bar{D})^2. \quad (13)$$

where the average delay is also a quadratic function of $\mathbf{x}$.

$$\bar{D} = \frac{1}{C} \sum_{k=1}^C D_k = \bar{c} + \bar{\mathbf{f}}^T \mathbf{x} + \frac{1}{2} \mathbf{x}^T \bar{H} \mathbf{x} \quad (14)$$

By substituting the above thermal TSV density $\mathbf{x}$ dependent delay into (13), the original problem can be rewritten with one quadratic-polynomial function

Problem 2:

$$\begin{aligned} \min : f(\mathbf{x}) &= \frac{1}{C-1} \sum_{k=1}^C (\hat{c}_k^2 + 2\hat{c}_k \hat{\mathbf{f}}_k^T \mathbf{x} \\ &+ \mathbf{x}^T (\hat{\mathbf{f}}_k \hat{\mathbf{f}}_k^T + \hat{c}_k \hat{H}_k) \mathbf{x} \\ &+ \hat{\mathbf{f}}_k^T \mathbf{x}^T \mathbf{x} \hat{H}_k \mathbf{x} + \frac{1}{4} \mathbf{x}^T \hat{H}_k \mathbf{x} \mathbf{x}^T \hat{H}_k \mathbf{x}) \end{aligned} \quad (15)$$

Fig. 6.: (a) Nonlinear temperature-dependent capacitance of one signal TSV; and (b) Nonlinear temperature-dependent delay of signal TSV bundles with number of 2, 4, 8 and 10

where $\hat{c}_k = c_k - \bar{c}$, $\hat{\mathbf{f}}_k = \mathbf{f}_k - \bar{\mathbf{f}}$ and $\hat{H}_k = H_k - \bar{H}$.

In addition, since dummy thermal TSVs occupy die area and become obstacles for signal routing, one needs to constrain the total number of thermal-TSVs inserted by

$$\mathbf{l}b \leq \mathbf{x} \leq \mathbf{u}b \quad (16)$$

where $\mathbf{l}b$ is determined by the foundry process limitation, and $\mathbf{u}b$ is determined by temperature reduction sensitivity function as well as the maximum allowed chip overhead introduced in Section B.

B. Conjugate-gradient Solving

Based on the formulated problem in equation (15), the clock-skew minimization problem becomes finding thermal TSV density insertion scheme $\mathbf{x}$ that numerically minimizes the value of $f(\mathbf{x})$ given constraints in equation (16). As one efficient technique for nonlinear optimization problem, the well-known conjugate gradient method with line search [13] is implemented to find the desired solution.

First, to remove the inequality constraint, the original problem is relaxed with Lagrange penalty factor and is reformulated by *Problem 3:*

$$\min : f^*(\mathbf{x}) = f(\mathbf{x}) + \lambda \cdot h^2(\mathbf{x}) \quad (17)$$

where

$$h(\mathbf{x}) = \begin{cases} 0, & \mathbf{l}b \leq \mathbf{x} \leq \mathbf{u}b \\ \rho \gg 0, & \text{otherwise} \end{cases} \quad (18)$$

Intuitively, the conjugate gradient method iteratively searches along the gradient drop reduction to find the $\mathbf{x}$ which minimizes $f^*(\mathbf{x})$. At each iteration, the algorithm selects the successive direction vector as a conjugate version of the successive gradient obtained as the method progresses. Specifically, the next search direction vector $\mathbf{d}_{k+1}$ is decided by adding to the current negative gradient vector

$$\mathbf{d}_{k+1} = -\nabla f^*(\mathbf{x}_k)^T + \frac{\mathbf{g}_{k+1}^T \mathbf{g}_{k+1}}{\mathbf{g}_k^T \mathbf{g}_k} \mathbf{d}_k \quad (19)$$

as a linear combination of the previous direction vector.

Based on the search direction vector, the step-size α_k can be optimally decided through the line search to minimize the the function $f^*(\mathbf{x}_k + \alpha \mathbf{d}_k)$. As the result, the vector $\mathbf{x}$ is updated as:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k. \quad (20)$$

The algorithm completes when $|\mathbf{x}_{k+1} - \mathbf{x}_k|$ is less than certain error bound, or the maximum iteration number is reached.

Practically, to avoid trapped in local minimum, the problem is solved with different randomly generated initial values $\mathbf{x}_0$. The minimal value among all these solutions is chosen as the final result.

Fig. 7.: Temperature reduction effect in 4-tier 3D IC with thermal TSVs under different power densities ($P, W/mm^3$)

Fig. 8.: 3D clock-tree of r5 after thermal TSV (black dots) insertion to balance clock-skew for (a)Tier 0; (b)Tier 1; (c)Tier 2 and (d)Tier 3

V. EXPERIMENTAL RESULTS

In this section, we first present the device-level results for signal and thermal TSVs modeling, and then discuss the thermal-TSV based clock-skew reduction for 3D clock-tree benchmarks. All programs are implemented by MATLAB optimization package on Linux. All results are computed on an Intel Xeon server with 3.47GHz clock frequency and 48GB of RAM.

The electrical analysis of signal TSVs are based models in (2) and (7). The thermal analysis of thermal TSVs is based on (4) verified by COMSOL multi-physics simulator [14]. The benchmarks are generated based on [5] as our starting zero-skew 3D clock-tree at room temperature.

A. Nonlinear Electrical-thermal Coupling of Signal TSV

As shown in Fig.6(a), the temperature dependent resistance and capacitance of each signal TSV in 2 is scaled ¹ from the measurement

¹We use TSV of 40 μm height, which is more reasonable in current fabrication process

TABLE I

: Coefficients of temperature dependent TSV model given in Equation (2)

Parameter	R_0 ($m\Omega$)	α ($/K$)	C_0 (fF)	β_1 (fF/K)	β_2 (fF/K^2)
Values	44	0.00125	88.8	0.0667	0.0014

TABLE II

: Coefficients of temperature dependent TSV delay given in Equation (7)

TSV type	k_0 (ps)	k_1 (ps/K)	k_2 (ps/K ²)	k_3 (ps/K ³)
T2	18.0	0.069	0.0002	0
T4	36.66	0.14	0.0004	0
T8	70.39	0.27	0.0008	0
T10	88.75	0.34	0.001	0

results reported in [4], which is summarized in Table I. Note that for reliability concern, a bundle of redundant signal TSVs are used for the signal distribution. This can further increase the capacitance of the signal TSVs. Fig. 6(b) shows the study of signal TSV bundles T2, T4, T8 and T10 for number of 2, 4, 8 and 10 TSVs.

Based on the delay model derived in Section III, we can calculate the delay induced by TSVs at different temperatures. The temperature-dependent delays of one or multiple TSVs are illustrated in Fig.6(b). As a comparison, the delay obtained from linear model is also shown in dot lines, which is generated by neglecting the 2nd and higher order terms in (7). It is shown that the delay difference between linear and nonlinear model grows with temperature and TSV bundle number. For clearance, all the TSV delay coefficients are listed in Table II as well. Note that to obtain the intrinsic delay of TSV, the length of both input and output wires to TSVs are assumed to be zero, and both source and load buffer transistors are the same size with $\frac{R_D}{S_D} = 100\Omega$ and $S_D C_P = S_L C_L = 2fF$. One can observe that the large temperature gradient in 3D is amplified by the nonlinear electrical-thermal coupling. As such, one can observe that for the T8-bundle at 120C°, the signal TSV delay can be as large as 100ps, which is 67% of half clock-cycle for a 3.3GHz multi-processor.

B. Temperature Reduction of Thermal TSV

A 4-tier 3D IC is constructed with 40 μm thickness in the top three tiers and 200 μm for the bottom one, and hence the overall chip height is 320 μm . Each thermal TSV has a diameter of 15 μm and a linear thickness of 200nm. The heat-sink is also added to the bottom of substrate as an equivalent distributed thermal conductance ($1.24 \times 10^5 W/(K \cdot m^2)$). Moreover, each tier is assigned with the same power density as the heat source. The initial temperature distribution at each tier is obtained without thermal TSVs. Then thermal TSVs are placed in the grids to obtain the new temperature distribution, and hence the temperature reduction distribution can be obtained accordingly.

As we can see from Fig.7, the temperature reduction first increases linearly with the inserted thermal TSV density then saturated at certain level. This results correlate well with (4). The linear-fitting curve is also shown in Fig.7 and the maximum inserted thermal TSV density observed is 400/mm².

C. Thermal TSV Insertion for Clock Skew Reduction

This section verifies the effectiveness of thermal-TSV insertion for reducing 3D clock-skew. HotSpot [15] is used to extract the temperature distribution at each location. To eliminate the application-specific bias, the temperature distribution is calculated as the average over all SPEC2000 benchmarks. Although the temperature distribution can change at runtime, using the average distribution profile is a common practice in thermal-aware clock tree synthesis [2]. At architecture level, a four-tier 3D IC is built with each tier one Alpha-2 processor. The IBM clock-tree benchmarks r1-r5 [10] are synthesized to 4-tier 3D clock-tree using the method in [5]. The 3D clock-tree of

TABLE III

: Benchmarks comparison of clock-skew reduction for linear and non-linear models

Type	htree1 (14 Signal TSVs)						
	Orig	Lin	Impr%	Time(s)	Nonlin	Impr%	Time(s)
T2	15.34	10.02	34.7%	14.29	2.59	83.1%	57.95
T4	26.44	8.67	67.2%	14.19	4.48	83.1%	57.90
T8	47.42	12.10	74.5%	14.58	8.14	82.8%	58.81
T10	58.42	15.10	74.2%	15.35	10.19	82.6%	58.84
Mean	-	-	62.6%	14.60	-	82.9%	58.38
Type	htree2 (28 Signal TSVs)						
	Orig	Lin	Impr%	Time(s)	Nonlin	Impr%	Time(s)
T2	23.48	8.69	63.0%	13.98	3.57	84.8%	56.95
T4	43.97	12.40	71.8%	14.02	5.38	87.8%	57.12
T8	82.76	16.18	80.4%	13.92	9.35	88.7%	58.87
T10	103.1	17.69	82.8%	13.93	11.44	88.9%	57.58
Mean	-	-	74.5%	13.96	-	87.5%	57.63
Type	r1 (45 Signal TSVs)						
	Orig	Lin	Impr%	Time(s)	Nonlin	Impr%	Time(s)
T2	30.50	18.40	39.7%	41.5	15.34	49.7%	106.6
T4	61.87	36.29	41.3%	29.3	27.50	55.6%	158.4
T8	121.1	71.39	41.6%	31.9	57.10	52.8%	170.5
T10	152.7	91.10	40.3%	37.2	74.26	51.4%	108.1
Mean	-	-	40.7%	35.0	-	52.4%	135.9
Type	r2 (60 Signal TSVs)						
	Orig	Lin	Impr%	Time(s)	Nonlin	Impr%	Time(s)
T2	35.13	23.60	32.8%	134.0	20.20	42.5%	389.0
T4	69.75	48.31	30.7%	102.8	37.90	45.7%	393.8
T8	134.7	94.67	29.7%	106.5	74.00	45.1%	705.8
T10	169.0	119.3	29.4%	139.3	93.82	44.5%	325.6
Mean	-	-	30.7%	120.7	-	44.5%	453.6
Type	r3 (75 Signal TSVs)						
	Orig	Lin	Impr%	Time(s)	Nonlin	Impr%	Time(s)
T2	32.36	20.92	38.4%	220.7	19.5	39.7%	749.5
T4	64.80	41.02	36.7%	170.8	34.3	47.1%	451.0
T8	125.6	80.29	36.1%	177.8	66.9	46.7%	745.4
T10	157.7	100.7	36.2%	231.0	85.8	45.6%	436.7
Mean	-	-	36.9%	200.1	-	44.8%	595.7
Type	r4 (90 Signal TSVs)						
	Orig	Lin	Impr%	Time(s)	Nonlin	Impr%	Time(s)
T2	31.68	18.63	41.2%	211.8	17.63	44.0%	890.6
T4	64.57	38.30	40.7%	232.2	30.10	53.4%	557.8
T8	126.8	75.38	40.6%	233.4	69.80	45.0%	707.5
T10	159.8	93.10	41.7%	327.6	80.00	50.1%	564.7
Mean	-	-	41.1%	251.2	-	48.1%	680.2
Type	r5 (90 Signal TSVs)						
	Orig	Lin	Impr%	Time(s)	Nonlin	Impr%	Time(s)
T2	35.00	21.5	38.6%	665.6	19.90	42.6%	1963
T4	68.40	39.0	43.0%	695.0	33.62	50.9%	1716
T8	131.0	77.9	40.5%	725.2	65.20	50.2%	1694
T10	164.1	97.0	40.9%	938.5	80.13	51.2%	1750
Mean	-	-	40.8%	756.1	-	48.7%	1781
Overall	-	-	46.8%	-	-	58.4%	-

r5 is illustrated in Fig.8 with TSVs marked in each tier in solid dots. In addition, two simple 3D clock-tree examples (htree1 and htree2) are generated for level-1 and level-2 H-trees as shown in Fig.1. The whole 3D chip is divided into 64x64 grids for thermal TSV insertion and the maximal thermal TSV density is limited to be lower than 7% of the local grid area. In addition, different signal TSV-bundles T2, T4 T8 and T10 are deployed with number of 2, 4, 8 and 10 TSVs.

Table III compares the clock-skew in pico-second before and after thermal TSV insertion for all benchmarks with different bundle number. The runtime of thermal TSV insertion based on linear/nonlinear TSV models is given in second. Compared to the case without thermal TSV insertion (i.e., *Orig* column), the thermal TSV insertion algorithm based on nonlinear electrical-thermal model in (2) (i.e., *Nonlin* column) reduces clock-skew by 58.4% on average. In addition, the thermal TSV insertion result considering only the linear part of the signal TSV model (i.e., the second order coefficient in (2) is set to zero) is also presented in the *Lin* column with 46.8% clock-skew reduction on average. Although more time is spent in solving the nonlinear optimization problem, it is just done once at design time, which we think is still worth the 11.6% clock skew reduction that affects the runtime behavior of the 3D system. We should note that the difference in performance between linear model and non-linear model are subjected to the ratio of TSV delay in the whole circuit, and it is independent on

the circuit size.

VI. CONCLUSION

Due to the existence of liner for isolation, TSV behaves as a MOSCAP with nonlinear electrical-thermal dependence. With the further consideration of high power-density and low heat-removal ability in 3D, there exists non-negligible delay variation or skew in 3D clock-tree distribution with TSVs. In this paper, physics-based electrical-thermal models for both signal and (dummy) thermal TSVs are provided with the consideration of nonlinear temperature dependence. As such, one nonlinear programming problem is formulated to reduce clock-skew via thermal TSVs insertion for the thermal-reliable 3D clock-tree synthesis. With a number of clock-tree benchmarks, experiments show that under realistic nonlinear TSV models, insertion of thermal TSV can effectively reduce the clock-skew by 58.4% on average, which is also 11.6% higher clock-skew reduction on average than using the linear model.

ACKNOWLEDGMENTS

This work is partially sponsored by Singapore MOE TIER-2 ARC5/11 project and MOE TIER-1 RG26/10 project.

REFERENCES

- [1] J. Cong and Y. Zhang, "Thermal-driven multilevel routing for 3d ics," in *IEEE/ACM ASP-DAC*, 2005.
- [2] J. Minz, X. Zhao, and S. K. Lim, "Buffered clock tree synthesis for 3d ics under thermal variations," in *IEEE/ACM ASP-DAC*, 2008.
- [3] T. Bandyopadhyay, R. Chatterjee, D. Chung, M. Swaminathan, and R. Tummala, "Electrical modeling of through silicon and package vias," in *IEEE 3DIC*, 2009.
- [4] G. Katti and et al., "Temperature dependent electrical characteristics of through-si-via (tsv) interconnections," in *IITC*, 2010.
- [5] X. Zhao, J. Minz, and S. K. Lim, "Low-power and reliable clock network design for through-silicon via (tsv) based 3d ics," *IEEE Trans. on Components, Packaging, and Manufacturing Technology*, vol. 1, no. 2, pp. 247–259, feb 2011.
- [6] Y. Xie, G. H. Loh, B. Black, and K. Bernstein, "Design space exploration for 3d architectures," *ACM J. on Emerging Technologies in Computing Systems*, vol. 2, no. 2, pp. 65–103, Apr. 2006.
- [7] B. Goplen and S. Sapatekar, "Thermal via placement in 3d ics," in *IEEE/ACM ISPD*, 2005.
- [8] S. Basir-Kazeruni, H. Yu, F. Gong, Y. Hu, C. Liu, and L. He, "Speco: Stochastic perturbation based clock tree optimization considering temperature uncertainty," *Elsevier Integration, the VLSI Journal*, vol. 46, no. 1, pp. 22–32, 2013.
- [9] H. Yu, J. Ho, and L. He, "Allocating power ground vias in 3d ics for simultaneous power and thermal integrity," *ACM Trans. on Design Automation of Electronic Systems*, vol. 14, pp. 41:1–41:31, June 2009.
- [10] "Ibm clock tree benchmarks," <http://vlsicad.ucsd.edu/GSRC/bookshelf/Slots/BST/>.
- [11] M. Cho, S. Ahmed, and D. Pan, "TACO: Temperature aware clock-tree optimization," in *IEEE/ACM ICCAD*, 2005.
- [12] J. Cong, A. Kahng, C. Koh, and C. A. Tsao, "Bounded-skew clock and steiner routing," *ACM Trans. on Design Automation of Electronic Systems*, vol. 3, no. 3, pp. 341–388, 1998.
- [13] D. Luenberger and Y. Ye, *Linear and nonlinear programming*. Springer Verlag, 2008, vol. 116.
- [14] "Comsol multiphysics simulation tool," <http://www.comsol.com/products/heat-transfer/>.
- [15] "Hotspot: <http://lava.cs.virginia.edu/hotspot/>."