

A holistic analysis of circuit timing variations in 3D-ICs with thermal and TSV-induced stress considerations

Sravan K. Marella, Sanjay V. Kumar, and Sachin S. Sapatnekar

Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455

email:{sravan,sanjay, sachin}@umn.edu

Abstract—

In 3D ICs, TSV-induced thermal residual stress impacts transistor mobilities due to the piezoresistive effect. This phenomenon is coupled with other temperature effects on transistor parameters that are seen even in the absence of TSVs. In this paper, analytical models are developed to holistically represent the effect of thermally-induced variations on circuit timing. The analysis is based on a semianalytical formulation that is demonstrated to accurately capture the biaxial nature of TSV stress and its effect on delay.

Key Terms : 3D IC, Through Silicon Via, Static Timing Analysis, Finite Element Method

I. INTRODUCTION

As device dimensions shrink from one technology generation to the next and we approach the physical and economic limits of process shrinks, 3D IC technology offers an alternative that reduces critical wire lengths, achieves higher densities with existing technology nodes, and enables heterogeneous integration. However, a major issue with 3D ICs is that on-chip temperature variations can be significant as ever more circuitry is packed in per unit footprint. On-chip temperatures can affect the behavior of a 3D IC in several ways.

First, thermal effects can change the threshold voltage and carrier mobilities in a transistor. The former serves to speed up the circuit while the latter slows it down: one or the other effect may dominate at a specific temperature. As a result, a circuit may show either positive temperature dependence (PTD) where the delay decreases monotonically with temperature, negative temperature dependence (NTD) where it increases monotonically, or mixed temperature dependence (MTD), where it changes nonmonotonically [1].

Second, stress in the through-silicon vias (TSVs), which connect different wafers/dies in a 3D IC, may affect the timing behavior of a circuit in a 3D IC. The TSVs may be made of copper, tungsten, or polysilicon: of these, copper is the primary choice owing to its low resistivity. During the manufacturing process for 3D ICs, the TSV and silicon wafer undergo several thermal cycles before a final annealing step that embeds the TSV in the wafer. During annealing and subsequent cooling, the structure is subjected to a thermal ramp from about 250°C down to room temperature. Because of the difference in the coefficient of thermal expansion (CTE) of the copper TSV and the silicon, a residual thermal stress is induced in the region surrounding the TSV. According to piezoresistive theory of materials, this stress alters the electrical conductivity of devices

This work is supported in part by NSF 1017778 and 1162267 and SRC 2009-TJ-2234.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

IEEE/ACM International Conference on Computer-Aided Design (ICCAD) 2012, November 5-8, 2012, San Jose, California, USA
Copyright ©2012 ACM 978-1-4503-1573-9/12/11... \$15.00

and hence the mobility of the transistors. The amount of mobility variation and the sign of the change (increase or decrease) depends upon the stress levels, which are determined by the position of a transistor with respect to the TSV and upon the channel orientation of the transistor with the crystallographic axis.

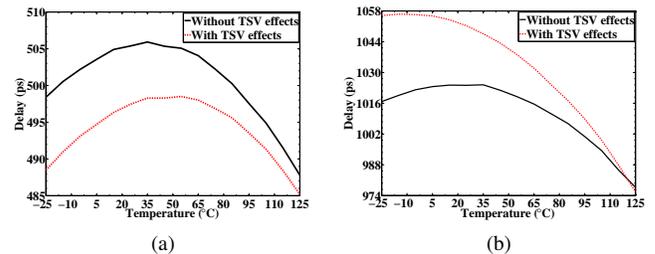


Fig. 1. Delay dependence of benchmarks (a) ac97_ctrl and (b) des for the cases where TSV effects are ignored and taken into account.

To understand the variation of delay with temperature in 3D circuits, a holistic analysis must be conducted, considering both of the above effects. This variation of delay with temperature is shown for two sample benchmark circuits in Figure 1. In each plot, the solid curve shows the trend without TSV effects, which shows MTD effects similar to those reported in [1] in both cases. In the presence of TSV stress, the temperature dependence is altered, as shown by the dotted curve. The delays change, and the nature of the dependence remains MTD for ac97_ctrl, but looks similar to PTD for des. Moreover, in one case, the delays increase and in the other, they decrease. Prior approaches [2]–[4] have considered TSV stress effects without incorporating the inherent effects of temperature on mobility and threshold voltage, and have assumed that the worst-case delay occurs at the lowest temperature: as seen above, this is not always true.

Stress in 3D IC structures has been studied using the finite element method (FEM) and through analytical methods [3], [5], although these works did not consider the impact of stresses on circuit delays. FEM simulations yield accurate estimation of stress levels around a TSV, but the computational cost of evaluation of the resulting stress data at different temperature corners for a given layout becomes quite prohibitive. FEM-based precharacterization approaches [6] are faster, but require significant storage to store the results of simulation on a grid due to the large number of points, and the fact that PTD/NTD/MTD requires such stresses to be stored at multiple temperature points. In contrast, an analytical approach lends to faster computation with no additional storage requirement since the stress at any point in the layout can be computed on-line.

In this work, we derive a complete analytical model for delay variation under stress. Our contributions are as follows:

- We incorporate both sets of thermal effects into a single analysis, capturing TSV stress effects, and thermally-driven mobility and threshold voltage variations. In contrast, prior works [3]–[5], [7] perform this analysis only at the lowest temperature in the range,

ignoring NTD/MTD effects.

- We model the biaxial nature of TSV stress in this work. As compared to prior analytical models [4], [7] that focus on uniaxial stresses, we will demonstrate that incorporating biaxial stresses substantially improves the accuracy of the analysis.
- The biaxial analytical solution is compared with actual FEA simulations, to determine empirical scaling factors that improve the accuracy of the 2D analytical models, in the useful range from and beyond the Keep-Out Zone (KOZ)¹
- On benchmark circuits, we demonstrate how the path delays in a circuit can change, depending on the relative locations of gates on the path and the TSVs. We show the magnitude of these changes and how they impact the critical path in a circuit.

II. TSV STRESS MODEL

We employ a two-dimensional (2D) plane stress analytical techniques in which the TSV is surrounded by an infinite material that is linearly elastic and isotropic, and is modeled as a cylinder of infinite length. These approximations are justified by prior stress analysis data [2], [5], which have shown the correlation between the stress patterns predicted by 2D analytical techniques and the actual 3D stress patterns. Thus, while 3D models are required to capture effects such as TSV fatigue, 2D models are sufficiently accurate to model effects of stress within the silicon, which is the subject of this work.

The stress field is represented as a tensor that comprises six unique stress components: three normal stresses (σ_{11} , σ_{22} , σ_{33}) and three shearing stresses (τ_{12} , τ_{23} , τ_{31}), where the subscripts 1, 2, and 3 correspond to the three orthogonal axes in any spatial coordinate system. In Cartesian coordinates, these correspond to the x , y , and z directions, while in cylindrical coordinates the axes are along radial (r), circumferential (θ), and axial (z) directions.

A stress tensor defined by only one normal stress component, other components being zero, corresponds to *uniaxial stress*; one defined by two normal stresses, other stresses being zero, is referred to as *biaxial stress*. Several previous approaches [4], [7] have assumed the stress to be uniaxial in nature, considering only the radial component, or used a variant of a uniaxial mobility model for biaxial stress.

Our work features an improved analytic model for TSV stress that also considers the impact of crystal orientation effects. We employ analytical expressions that show that the stress tensor has two normal components, implying that the stress is indeed biaxial. Moreover, unlike prior works [3]–[5] we capture the dependence of stress on temperature rather than computing mobility variation model only under a single thermal load condition (temperature corner). This enables us to build a unified temperature-dependent delay model that captures both TSV stress and TSV-independent NTD/PTD/MTD.

A. Stress in cylindrical coordinates

For 2D plane stress analysis around a TSV, of the normal stresses, only the radial (σ_{rr}) and circumferential ($\sigma_{\theta\theta}$) stresses are present, with the axial (σ_{zz}) stress being zero. Since the TSV is firmly embedded in the silicon, the shearing stress components corresponding to τ_{rz} , $\tau_{\theta z}$, $\tau_{r\theta}$ are zero [5], and only the normal stress components, σ_{rr} , $\sigma_{\theta\theta}$, and σ_{zz} must be considered. Using plane stress theory and under force equilibrium conditions, we obtain expressions for cylindrical stress which correspond to similar classical Lamé solution

¹The KOZ is the (often rectangular) region around the TSV within which no transistor is allowed to be placed, since the stresses are very high and can adversely affect transistor performance and reliability.

models reported in [5], [8]. The stress components at a point (r, θ) in the silicon, relative to the center of the TSV, are given by:

$$\sigma_{rr} = -\sigma_{\theta\theta} = C \left(\frac{D_{Cu}}{2r} \right)^2 \quad (1)$$

$$\sigma_{zz} = \tau_{r\theta} = \tau_{z\theta} = \tau_{rz} = 0$$

$$\text{where } C = \frac{E_{Cu}\alpha_{Cu}(T_{ref} - T) - \left(\frac{1+\nu_{Si}}{1+\nu_{Cu}} \right) E_{Cu}\alpha_{Si}(T_{ref} - T)}{(1 - 2\nu_{Cu}) + \left(\frac{1+\nu_{Si}}{1+\nu_{Cu}} \right) \frac{E_{Cu}}{E_{Si}}}$$

Here E_{Cu} , α_{Cu} , ν_{Cu} are the Young's Modulus, CTE, and Poisson's ratio of copper, E_{Si} , α_{Si} , ν_{Si} are the corresponding quantities for silicon. The term D_{Cu} represents the diameter of the TSV and remains a constant for a particular implementation technology. The operating temperature is denoted by T , and the reference temperature, T_{ref} is the copper annealing temperature (250°C). The physical constants for copper and silicon used in this work are listed in Table I.

TABLE I
PHYSICAL CONSTANTS FOR STRESS COMPUTATION

	E (GPa)	CTE (ppm/°C)	ν
Copper	111.5	17.7	0.343
Silicon	162.0	3.05	0.28
SiO ₂	71.7	0.51	0.16
BCB	3	40	0.34

A positive stress component is referred to as *tensile* and a negative component as *compressive*. From (1), we infer that:

- The stress is biaxial in nature since it has nonzero components σ_{rr} and $\sigma_{\theta\theta}$. In contrast, in the uniaxial formulation used in [4], [7], only σ_{rr} is present; all other stress components are zero.
- The radial stress (σ_{rr}) is tensile and the circumferential stress ($\sigma_{\theta\theta}$) is compressive. The uniaxial formulation however predicts that the stress is always tensile around the TSV.
- At a fixed distance r from the center of the TSV the stress components in the silicon, within the range of operating temperatures for a normal circuit (which are well below T_{ref}), the stress components have a larger magnitude at lower temperatures.

B. Stress in Cartesian coordinate systems

Biaxial case: Although the stress equations (1) have been expressed in the cylindrical coordinate system, IC design uses Manhattan geometries and it is convenient to transform these to the Cartesian coordinate system. This will facilitate the piezoresistivity calculations described in Section III. Using the transformations $x = r \cos \theta$ and $y = r \sin \theta$, as in [5] and with cylindrical-to-Cartesian tensor transformations, the following expressions are obtained from Equation (1):

$$\begin{aligned} \sigma_{xx} &= -\sigma_{yy} = C \left(\frac{D_{Cu}}{2} \right)^2 \frac{x^2 - y^2}{(x^2 + y^2)^2} = \sigma_{rr} \cos 2\theta \\ \tau_{xy} &= C \left(\frac{D_{Cu}}{2} \right)^2 \frac{xy}{(x^2 + y^2)^2} = \sigma_{rr} \sin 2\theta \\ \sigma_{zz} &= \tau_{yz} = \tau_{zx} = 0. \end{aligned} \quad (2)$$

As defined earlier, σ_{xx} , σ_{yy} , and σ_{zz} are the three normal stresses in Cartesian coordinate axis, and τ_{xy} , τ_{yz} , τ_{zx} are the shearing stress components. The angle θ corresponds to the angle made by the transistor with the TSV and σ_{rr} is given by Equation (1).

Uniaxial case: We show expressions for the approximate uniaxial case for completeness, and so that we can compare it with the correct biaxial 2D formulation. For the uniaxial formulation [4], [7], as

mentioned earlier, $\sigma_{\theta\theta} = 0$. The corresponding Cartesian co-ordinate stress tensors can be obtained similarly in terms of σ_{rr} and θ as:

$$\begin{aligned}\sigma_{xx} &= \sigma_{rr} \cos^2 \theta; \sigma_{yy} = \sigma_{rr} \sin^2 \theta; \tau_{xy} = \frac{\sigma_{rr}}{2} \sin 2\theta; \\ \sigma_{zz} &= \tau_{yz} = \tau_{zx} = 0.\end{aligned}\quad (3)$$

Comparison: This leads to the following observations:

- For the biaxial formulation, the stress along x and y directions are opposite (compressive/tensile) in nature. For the uniaxial case, the $\cos^2 \theta$ and $\sin^2 \theta$ terms in σ_{xx} and σ_{yy} imply that the stresses along the x and y directions are both tensile.
- Unlike cylindrical coordinates, there is a nonzero shearing (τ_{xy}) stress component in Cartesian coordinates. This value for the uniaxial case is half the magnitude of that in the biaxial case.
- The magnitudes and signs of stress components in the biaxial and uniaxial formulations differ, and the corresponding relative errors in mobility variation are quantified in Section III.
- As in cylindrical coordinates, the stress components are *linear functions* of the temperature T due to their dependence on the factor C .

C. Impact of the crystal orientation

The crystal orientation refers to the Miller index of the silicon crystal. The principal crystallographic axes create a coordinate system that corresponds to the [100], [010], and [001] directions. Within this system, the orientation of a wafer is defined as the direction normal to the plane of the silicon wafer. The (100) orientation is the dominant paradigm (although other orientations such as (111) may also be used) and our exposition will focus on this case; it turns out that this orientation also generally experiences lower stresses. Due to symmetry, the (100), (010), and (001) orientations are equivalent.

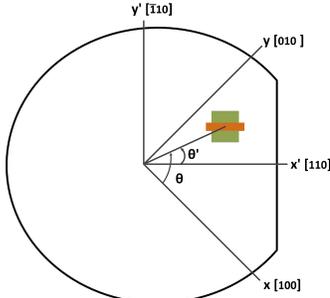


Fig. 2. Coordinate axes in (100) Si with a wafer flat orthogonal to the [110]. The transistor channel here is perpendicular to the [110] axis i.e., $\phi' = \pi/2$.

The orientation of transistors on a wafer is determined relative to the wafer flat, as shown in Fig. 2: transistors may be parallel or perpendicular to this feature. Therefore, a rotated coordinate space with a new x' -axis that is perpendicular to the wafer flat is a convenient frame of reference. This x' -axis is in the [110] direction, and therefore, the [100]–[010] axes must be rotated by 45° [9], [10].

By examination, a rotation by 45° causes the axial direction to move along the transverse direction. We can thus easily deduce the biaxial stress tensors in these coordinates from Equations (2) to be:

$$\sigma_{x'x'} = -\sigma_{y'y'} = \tau_{xy}; \tau_{x'y'} = -\sigma_{xx}\quad (4)$$

Using the physical constants in Table I and Equation (4), stress contours of $\sigma_{x'x'}$ and $\tau_{x'y'}$ are plotted and shown in Figure 3. The stress patterns for the biaxial formulation are seen to be tensile and compressive in mutually perpendicular directions. This results from the $\cos 2\theta$ [$\sin 2\theta$] term in σ_{xx} [τ_{xy}] in Equation (2).

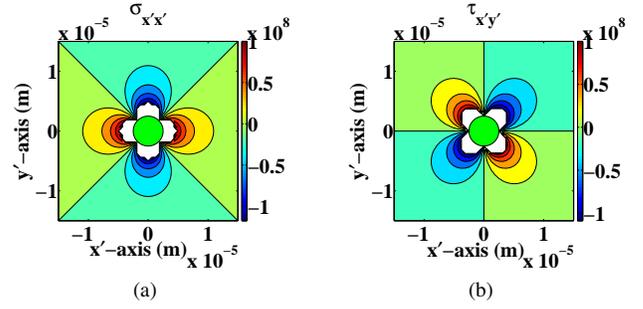


Fig. 3. Stress (in Pa) contour fields in the [110]–[110] axes. (a) $\sigma_{x'x'}$ stress contour field. (b) $\tau_{x'y'}$ stress contour field.

In contrast, for the uniaxial case used in previous papers, since $\sigma_{\theta\theta}$ is set to 0, the stress components are unchanged under rotation, i.e.,

$$\sigma_{x'x'} = \sigma_{xx}; \sigma_{y'y'} = \sigma_{yy}; \tau_{x'y'} = \tau_{xy}.\quad (5)$$

D. Comparison with finite element simulation

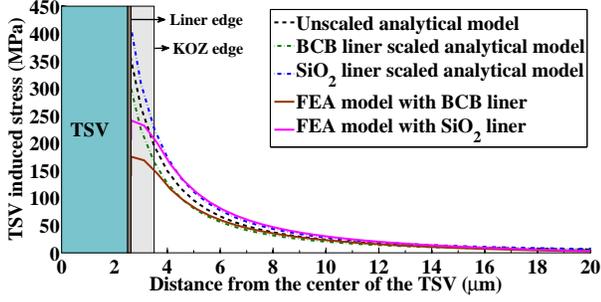
To validate the effectiveness of the closed-form 2D analytical solution in Equation (1), 3D FEA simulations were performed using the ABAQUS tool with realistic TSV structures using the methodology described in [8]. All materials (TSV, liner, silicon) are assumed to be linear, elastic, and isotropic. The annealing process is modeled in FEA by applying a temperature load with an initial temperature of 250°C and final temperature of 25°C . For the 3D FEA simulations, the copper TSV diameter is $5\mu\text{m}$, height is $30\mu\text{m}$ and the liner thickness is 125nm [6]. In addition, we define the KOZ to be $1\mu\text{m}$ from the edge of the TSV or $3.5\mu\text{m}$ from the center of the TSV. The KOZ is chosen to ensure that there is no more than 33% mobility variation in any transistor around an isolated TSV. In practice, the KOZ constraint is driven by the mobility degradation of PMOS transistors, which exceeds that of NMOS devices.

In 3D processes, a thin layer of a liner material is deposited at the sidewall between copper and silicon. This reduces the stress in silicon and also improves the mechanical reliability of the TSV structure. Two popular choices of the liner material are silicon dioxide (SiO_2) and benzocyclobutene (BCB) whose material properties are given in Table I. Our analytical model, as explained above, ignores the liner. The primary effect of the liner is in altering material properties: the TSV+liner combination has a new effective CTE, effective Young's modulus, and effective Poisson's ratio, each of which differs from the no-liner case. We find that, a simple practical way is to use an empirical scaling factor to multiply the components of the stress tensor capturing this change. We precharacterize this scaling factor (this is done once for a given technology) against the simulations and maintain the nature of the analytical model, while needing none of the additional storage and computational overhead associated with FEM-based approaches.

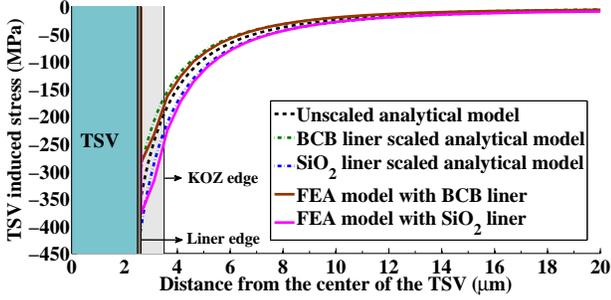
The effect of the copper landing pad is ignored in this analysis, since the landing pad size is always within the KOZ boundary and its main influence is felt only at the edge of the TSV.

The 2D analytical formulation is compared against the following three structural configurations of a copper TSV surrounded by silicon: (i) with no liner, (ii) with an SiO_2 liner, and (iii) with a BCB liner.

The unscaled analytical solution and its scaled versions are compared against actual FEA stress with BCB and SiO_2 liners, respectively. The corresponding empirically determined multiplicative factors for the BCB and SiO_2 liner cases are 0.85 and 1.16. Fig. 4 shows the comparison of the different analytical and FEA models against σ_{rr} and $\sigma_{\theta\theta}$ components. It can be observed that the unscaled



(a)



(b)

Fig. 4. Comparison of (a) σ_{rr} and (b) $\sigma_{\theta\theta}$ between the analytical and the FEA models. The TSV edge, liner edge and KOZ edge are at $2.5\mu\text{m}$, $2.625\mu\text{m}$ and $3.5\mu\text{m}$ respectively.

analytical model is inaccurate in representing the true FEA data even outside the KOZ. However, the corresponding scaled versions for BCB and SiO_2 liner cases closely follow their FEA counterparts outside the KOZ.

It will be shown in Section V that the error in stress between the scaled analytical models and the actual FEA data are bounded by 15MPa and 20MPa for the TSV with BCB and SiO_2 liners, respectively, a small fraction of the total stress. The errors in delay are even smaller. For instance, it was found that, for a two input NAND gate the worst case delay error outside the KOZ is bounded between 1.5ps and 4ps under the analytical models B and C, respectively.

III. PIEZORESISTIVITY

Based on the stress formulations developed in Section II we obtain the analytical models for evaluating mobility and threshold voltage variations in both NMOS and PMOS transistors.

From the basic axiom of the theory of conduction of electrical charge, the current density vector is a function of electric field vector. Alternatively, the electric field vector is related to the current density vector by the resistivity tensor, which can be related to mobility. According to piezoresistive theory, the resistivity tensor components vary with applied mechanical stress in piezoresistive materials such as silicon [11]. A complete mathematical model for piezoresistivity has been presented and demonstrated in silicon in [10].

In the rotated (x', y') coordinate system described earlier, the relative change in mobility is given by the expression:

$$\frac{\Delta\mu'}{\mu'} = [\pi'_{11}\sigma_{x'x'} + \pi'_{12}\sigma_{y'y'}] \cos^2 \phi' + [\pi'_{11}\sigma_{y'y'} + \pi'_{12}\sigma_{x'x'}] \sin^2 \phi' + [\pi'_{44}\tau_{x'y'}] \sin 2\phi' \quad (6)$$

Here, π'_{11} , π'_{12} and π'_{44} are the three unique piezoresistivity coefficients defined along the primed coordinate axes, and ϕ' is the angle made by the transistor channel with the x' -axis, i.e., the [110] axis.

This implies that $\phi' = 0$ for the transistor channels that are oriented along this direction, and $\phi' = \pi/2$ when they are orthogonal to this axis. As we will see, the piezoresistivity coefficients and the stress tensor components vary with the channel orientation, implying that the mobility variation depends on the transistor channel orientation.

In practice, the piezoresistivity coefficients for silicon are typically listed in databooks along the crystallographic axes. The transformation to the primed axes is straightforward. Using standard techniques for coordinate rotation, it can be shown that [12]:

$$\begin{aligned} \pi'_{11} &= \frac{\pi_{11} + \pi_{12} + \pi_{44}}{2} \\ \pi'_{12} &= \frac{\pi_{11} + \pi_{12} - \pi_{44}}{2} \\ \pi'_{44} &= \pi_{11} - \pi_{22} \end{aligned} \quad (7)$$

Here, the terms π_{11} , π_{22} , and π_{44} are the primary piezoresistive coefficients along the crystallographic axes. Table II shows the values for the primary piezoresistivity coefficients [2] in both coordinates.

TABLE II
PIEZORESISTIVITY COEFFICIENTS ($\times 10^{-12} Pa^{-1}$) IN (100) Si [2]

	π_{11}	π_{12}	π_{44}	π'_{11}	π'_{12}	π'_{44}
NMOS	1022.0	-537.0	136.0	310.5	174.5	1559.0
PMOS	-66.0	11.0	-1381.0	-717.5	662.5	-77.0

Biaxial case: For a transistor oriented along the [110] axis, $\phi' = 0$. From Equations (4), (6), and (7),

$$\frac{\Delta\mu'}{\mu'} = \pi'_{11}\sigma_{x'x'} + \pi'_{12}\sigma_{y'y'} = \pi_{44}\sigma_{x'x'} = \pi_{44}\sigma_{rr} \sin 2\theta. \quad (8)$$

Recall that θ is the angle made by the vector from the origin to the center of the transistor with the unprimed x -axis. Similarly, for a transistor in the orthogonal direction, $\phi' = \pi/2$, and

$$\frac{\Delta\mu'}{\mu'} = \pi'_{11}\sigma_{y'y'} + \pi'_{12}\sigma_{x'x'} = -\pi_{44}\sigma_{x'x'} = -\pi_{44}\sigma_{rr} \sin 2\theta. \quad (9)$$

Based on the above analysis, we can observe that:

- For the same stress and orientation, PMOS and NMOS devices experience opposite mobility variation effects: both depend on π_{44} , which has a different sign for PMOS and NMOS (Table II). Under tensile stress, PMOS mobility degrades while NMOS mobility improves; the opposite is true under compressive stress.
- For the same stress, PMOS devices experience greater mobility variation as compared to NMOS devices, since the π_{44} value of PMOS is an order of magnitude greater than that of the NMOS.
- The relative mobility variation depends on the operating temperature since stress varies linearly with temperature (Section II).

Uniaxial case: For $\phi' = 0$, from Equations (3), (5), and (6), the corresponding mobility variation can be expressed as:

$$\frac{\Delta\mu'}{\mu'} = \pi'_{11}\sigma_{x'x'} + \pi'_{12}\sigma_{y'y'} = \pi'_{11}\sigma_{rr} \cos^2 \theta + \pi'_{12}\sigma_{rr} \sin^2 \theta. \quad (10)$$

For the orthogonal transistor orientation, $\phi' = \pi/2$, and therefore

$$\frac{\Delta\mu'}{\mu'} = \pi'_{11}\sigma_{y'y'} + \pi'_{12}\sigma_{x'x'} = \pi'_{11}\sigma_{rr} \sin^2 \theta + \pi'_{12}\sigma_{rr} \cos^2 \theta. \quad (11)$$

Comparison: From Equation (2), TSV stress is biaxial, and we now examine the error from the uniaxial assumption. We consider transistors oriented along the [110] axis ($\phi' = 0$). From equations (8) and (10), the relative mobility variation depends only upon π_{44} in the biaxial formulation, while in the uniaxial formulation it depends

on π'_{11} and π'_{12} . The inaccuracies in using the uniaxial formulation can be identified by observing two cases:

- $\theta = \frac{\pi}{2}$ (Figure 5 (a)): For the NMOS transistor, the biaxial analysis correctly predicts a mobility degradation while the uniaxial case mispredicts an improvement. For the PMOS transistor, both formulations predict a mobility improvement, but the uniaxial formulation underestimates the variation.
- $\phi' = 0$ (Fig. 5 (b)): For the same stress, the uniaxial case shows the same trends with T as the biaxial case, but overestimates the NMOS mobility variation and underestimates PMOS variation. The percentage inaccuracies are significant.

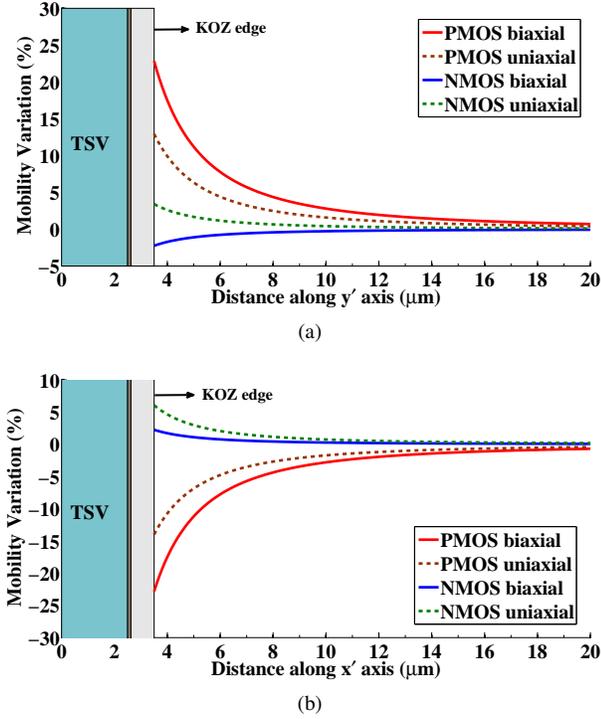


Fig. 5. Mobility variation comparison in uniaxial and biaxial formulations with distance along (a) y' -axis (b) x' -axis. Here edge of the TSV = $2.5\mu\text{m}$.

IV. TIMING ANALYSIS UNDER MOBILITY VARIATIONS

Our circuit-level input is a characterized cell library and a placed netlist, based on which the stresses may be computed using the techniques in Section II; this stress can be converted to determine the transistor mobility variations using the methods in Section III.

A. Delay dependence on temperature

We consider the effect of temperature on delay for the case without TSV stress; TSV stress effects are added to these effects.

The traditional assumption that has guided timing analysis is that the delays of library cells increase monotonically with temperature, corresponding to the NTD case. However, with technology scaling and the increased use of lower V_{dd} and V_t values, PTD and MTD are also often seen. Gate delays change with T in two ways:

- (1) The *mobility change* for charge carriers, $\Delta\mu_T$, is given by:

$$\Delta\mu_T = \mu(T_0) \left(\frac{T}{T_0} \right)^{-m} \quad (12)$$

Here T_0 is the room temperature, and $m > 0$ is the mobility temperature exponent, with a typical value of 1.7 in highly doped silicon, and

1.4 in nanometer silicon layers, where boundary scattering becomes important [13]. This reduction in μ increases the delay.

- (2) The *threshold voltage change*, ΔV_t , for a transistor is given by:

$$\Delta V_t = -\kappa(T - T_0) \quad (13)$$

where $\kappa > 0$ has a typical value of 2.5mV/K [14]. Thus, the delay decreases with T due to this effect.

The two phenomena above have opposite effects on gate delays, and the trend of delay with temperature depends on which of the two is more dominant, resulting in PTD, NTD, or MTD.

B. Gate characterization

The variation in the mobility translates into variations in the gate delay of a gate. The delay, D_{str} , of a gate under stress is given by:

$$D_{str} = D_{nom} + \left(\frac{\partial D}{\partial \mu} \right) (\Delta\mu_{TSV} + \Delta\mu_T) + \left(\frac{\partial D}{\partial V_t} \right) \Delta V_t \quad (14)$$

where D_{nom} is the delay without temperature or TSV effects, $\partial D / \partial \mu$ [$\partial D / \partial V_t$] is the sensitivity of the delay to mobility [V_t] variations, and $\Delta\mu_{TSV}$ is the mobility change due to TSV stress.

The sensitivity is a nonlinear function of the nominal point, and is stored as a look-up table (LUT) rather than a constant sensitivity value. During delay calculation, linear interpolation is used between the stored points. This results in improved accuracy, e.g., for a NAND2 gate in the library, the delay error using our approach is 3%, vs. 10% for a constant sensitivity model.

LUT characterization is a one-time exercise for a library. The range of the LUT reflects the observed range of variations. For example, for mobility sensitivity, using HSPICE, we characterize a 45nm gate library for five delay values with corresponding PMOS mobility variations ranging from $\pm 50\%$, and using a linear approximation for the NMOS mobility variations, considering a range of $\pm 5\%$, respectively. The standard cell library is characterized from -25°C to 125°C , along with different corners of supply voltage, load capacitance, and input slope.

C. Timing analysis framework

For the placed netlist that is provided as an input to the procedure, the left bottom coordinates and width and height of each cell in the layout can be determined. The computation then proceeds as follows:

- 1) From the above placement information, the centers of the TSV and the standard cells are computed.
- 2) The equations in (2) and (4) are used to calculate the stress tensor from to every TSV present in the circuit, capturing the transistor channel orientation with respect to the wafer flat. The stress tensor with respect to different TSVs are added up.
- 3) The mobility variations are calculated according to Equation (8) for transistor channels oriented along the [110] axis.
- 4) The computed mobility variation is employed to obtain accurate cell delays using linear interpolation with the characterized five delay values in conjunction with Equation (14) during static timing analysis.
- 5) Finally, the delay of the circuit is computed at different temperature points ranging from -25°C to 125°C in steps of 20°C .

V. RESULTS

A. Gate delay comparison: Analytical solution vs. FEA

In this section, we compare the errors in predicting stress and gate delays based on the scaled analytical models of Section II-D, as compared to the results from true FEA simulation.

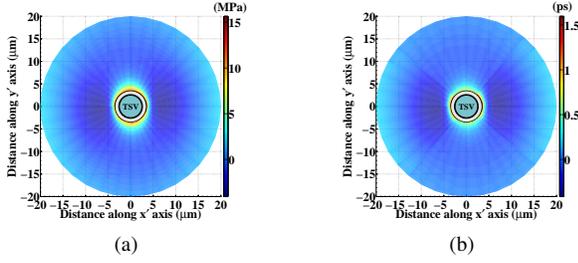


Fig. 6. Contours of (a) $\sigma_{x'x'}$ difference and (b) rise time difference of NAND2 gate around a TSV with BCB liner.

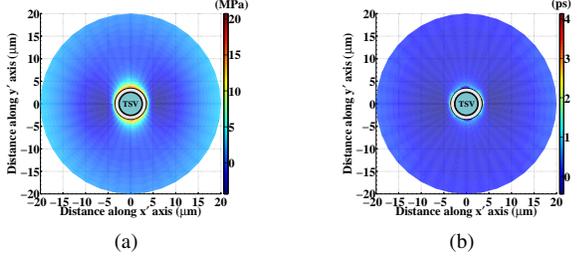


Fig. 7. Contours of (a) $\sigma_{x'x'}$ difference and (b) rise time difference of NAND2 gate around a TSV with SiO₂ liner.

Fig. 6(a) and Fig. 7(a) shows the errors in the dominant $\sigma_{x'x'}$ stress component in a two-dimensional region around the TSV for the BCB and SiO₂ liners, respectively. From the legend, it can be observed that the stress difference in using the corresponding scaled analytical formulations for the BCB and SiO₂ liners is bounded by 15MPa and 20 MPa, respectively, which is a small fraction of the peak stress in the KOZ of ≈ 200 MPa (see the 1D plot in Figure 4).

The impact of these errors in stress prediction on the rise delay of a NAND2 gate in the standard cell library are shown in Fig. 6(b) and Fig. 7(b). The errors here are significantly attenuated, and are no more than 1.5ps and 4ps, respectively. Further, it can be observed that these errors rapidly vanish even slightly away from the KOZ.

Therefore, the use of a single simple scaling factor is extremely effective in ensuring excellent timing accuracy, and empirically scaled analytical models will be adequate to determine the circuit behavior. The overhead of determining the scaling factor is a one-time (for a given technology) comparison with the FEA stress. This removes the need for the storage overhead of store FEA models, or the computational overhead of on-the-fly FEA.

B. Effect of TSV-induced stress on circuit path delays.

We apply our techniques on a set of IWLS 2005 benchmarks [15] whose attributes are as shown in Table III, where #PO denotes the number of primary outputs in the design. The parameters chosen in our experiments are listed below:

- The scaled analytical solution with scaling factors 0.85 and 1.16 for TSV models with BCB and SiO₂ liners, respectively.
- A cell library characterized under 45nm PTM models [16]
- All transistor orientations parallel to the [110] axis, i.e., $\phi' = 0$
- A TSV diameter of $5\mu\text{m}$. TSV is surrounded by either BCB or SiO₂ liner with a liner thickness of 125nm.
- Our KOZ defined as the point where the mobility variations are below 33%; this corresponds to a KOZ size of $1\mu\text{m}$ from the TSV edge.

TSVs are placed in the layout with equal horizontal and vertical spacing. The number of TSVs inserted in a circuit depends upon the size of the benchmark and the TSV spacing used. Four separate layouts are generated using the Capo placer [17]:

TABLE III

IWLS 2005 [15] CIRCUITS

Circuit	# Gates	Dimension	# POs	#V1	#V2	#V3
		H×W ($\mu\text{m}\times\mu\text{m}$)				
ac97_ctrl	11308	130×80	4204	70	54	35
aes_core	12223	87×85	12313	49	36	25
des	4647	68×85	332	35	24	15
ethernet	29739	104×170	32149	170	84	60
i2c	1221	16×74	204	6	5	4
mem_ctrl	10094	94×84	2522	49	36	25
pci_bridge32	11148	127×85	9025	70	48	35
spi	3632	48×87	564	21	18	10
systemcdes	2694	50×71	549	18	15	8
usb_funct	12987	76×113	3930	54	40	28

- TSVless contains no TSVs.
- TSV_{*i*}, $i \in \{3, 7, 10\}$ correspond to regularly-spaced horizontal and vertical TSVs with a spacing of 3, 7, and 10 μm , respectively, between the edges of the KOZs for the TSVs.

The number of the TSV's in the TSV₃, TSV₇, and TSV₁₀ circuits are represented by #V1, #V2, and #V3 in Table III.

Tables IV and Table V show how the critical path changes, when TSV with corresponding BCB and SiO₂ liners are taken into account. We found that, the interconnect lengths were short in the critical paths of the circuits considered here. Hence the gate delay component dominates the interconnect delay component. Thus, addition of interconnect delays will not significantly alter the timing results presented here. In Table IV, D0 represents the critical path delay for the TSVless case, and the temperature at which this delay is seen. The columns designated by D1, D2, and D3 represent the critical path delays of TSV₃, TSV₇, and TSV₁₀ layouts with the BCB liner effects taken into account. The corresponding delays in the three layouts with SiO₂ liner are denoted by D4, D5, and D6 in Table V. The temperatures at which the maximum occurs is shown alongside each delay. Each circuit is seen to exhibit MTD as its worst case delay occurs in the interior of the temperature range of [-25C, 125C].

TABLE IV

COMPARISON OF CRITICAL PATH DELAY OF CIRCUITS WITHOUT AND WITH {TSV + BCB LINER} EFFECTS

Circuit	TSVless		TSV ₃			TSV ₇			TSV ₁₀		
	D0 (ps)	T (°C)	D1 (ps)	T (°C)	$\Delta D1$ (%)	D2 (ps)	T (°C)	$\Delta D2$ (%)	D3 (ps)	T (°C)	$\Delta D3$ (%)
ac97_ctrl	505	55	504	35	-0.2%	502	35	-0.6%	510	35	1.0%
aes_core	516	35	523	15	1.4%	549	15	6.4%	512	35	-0.8%
des	1024	35	1040	-5	1.6%	1036	15	1.2%	1025	35	0.1%
ethernet	914	15	939	-5	2.7%	916	15	0.2%	911	15	-0.3%
i2c	444	35	449	15	1.1%	450	15	1.4%	448	15	0.9%
mem_ctrl	979	35	991	15	1.2%	995	-5	1.6%	989	15	1.0%
pci_bridge32	738	35	741	15	0.4%	744	35	0.8%	735	15	-0.4%
spi	954	15	965	-5	1.2%	985	-5	3.2%	953	15	-0.1%
systemcdes	855	15	876	-5	2.5%	876	-5	2.5%	859	15	0.5%
usb_funct	702	15	724	-5	3.1%	715	15	1.9%	701	15	-0.1%

TABLE V

CRITICAL PATH DELAY OF CIRCUITS WITH {TSV + SiO₂ LINER} EFFECTS

Circuit	TSV ₃			TSV ₇			TSV ₁₀		
	D4 (ps)	T (°C)	$\Delta D4$ (%)	D5 (ps)	T (°C)	$\Delta D5$ (%)	D6 (ps)	T (°C)	$\Delta D6$ (%)
ac97_ctrl	505	35	0.0%	499	15	-1.2%	513	35	1.6%
aes_core	526	15	1.9%	562	-5	8.9%	511	35	-1.0%
des	1066	-5	4.1%	1045	15	2.1%	1025	35	0.1%
ethernet	956	-5	4.6%	916	-5	0.2%	910	15	-0.4%
i2c	456	-5	2.7%	452	15	1.8%	449	15	1.1%
mem_ctrl	998	15	1.9%	1003	15	2.5%	992	15	1.3%
pci_bridge32	744	-5	0.8%	746	35	1.1%	734	35	-0.5%
spi	976	-5	2.3%	1000	-5	4.8%	953	15	-0.1%
systemcdes	906	15	6.0%	885	-5	3.5%	860	15	0.6%
usb_funct	741	-5	5.6%	730	-5	4.0%	700	15	-0.3%

Note that since the four layouts are different, these delays should not be directly compared. However, it is instructive to observe the portion of the delays, ΔD_i , $i \in \{1, 2, 3, 4, 5, 6\}$, that can explicitly be attributed to the TSV+liner effects (clearly, ΔD_0 is zero since the TSVless layout has no TSVs). To compute each ΔD_i , we first find the critical path delay for the corresponding layout while ignoring TSV stress effects, then the critical path delay when TSV stresses are added in, and we show the percentage change. The liner effects are always considered when the TSV is present. Note that the critical path can (and often does) change when TSV stress is accounted for.

The improvements (negative changes) in critical path delays indicate that even with the smaller, more aggressive KOZ used here, we can mitigate the TSV effects on the critical path delays to some extent by careful design choices during initial circuit placement. Additionally, temperature dependence of the circuits is also altered when TSV effects are taken into account. From Tables IV and V, it can be observed that there is a wider range of delay variation in the TSV inserted layouts with SiO₂ liner as compared to the corresponding layouts with BCB liner. For instance, in the TSV_7 layouts, the critical path variations with SiO₂ ranges from -1.2% to 8.9%. However, the variation within the same layout with the BCB liner taken into account ranges from -0.6 to 6.4%. Similar trends can be observed in the TSV_3 and TSV_10 layouts. This indicates that the BCB liner is preferable over SiO₂ liner from a circuit timing perspective. The improvement in mechanical reliability in using BCB liner over SiO₂ is already shown in [6]. For these reasons, we shall focus on the layouts with TSV surrounded by the BCB liner for the rest of the discussion.

TABLE VI

DELAY CHANGES IN THE TSV_7 CIRCUITS WITH {TSV + BCB LINER}

Circuit	D_{P1} (ps)	ΔD_{P1} (%)	D_{P2} (ps)	ΔD_{P2} (%)	D_{P3} (ps)	ΔD_{P3} (%)	ΔTPS (ps)	ΔTNS (ps)
ac97_ctrl	505	-0.6%	372	8.33%	351	-4.3%	-5790	0
aes_core	516	6.4%	548	6.20%	426	-4.0%	-2228405	-597
des	1012	2.4%	802	6.36%	834	-2.5%	-71540	-86
ethernet	914	0.2%	690	6.67%	632	-4.4%	-2187618	-2
i2c	444	1.4%	353	7.08%	295	-4.4%	-28238	-21
mem_ctrl	977	1.8%	610	6.89%	616	-3.2%	-614491	-260
pci_bridge32	738	0.8%	580	8.10%	648	-3.9%	-1664605	-19
spi	950	3.7%	795	6.16%	463	-3.2%	-115770	-264
systemcdes	855	2.5%	756	5.56%	491	-4.1%	-144792	-100
usb_funcnt	683	4.7%	632	6.33%	362	-4.1%	-796113	-118

In order to gain more insights into the circuit timing behavior we further examine one set of circuits, the TSV_7 circuits. Let P1 denote the critical path in the circuit with TSV effects. Let P2 and P3 represent the paths that show maximum delay degradation, and delay improvement, respectively, when TSV effects are considered. For each circuit, Table VI describes the extent of delay changes in these paths due TSV-induced mobility variations. Here D_{P1} , D_{P2} and D_{P3} denote the nominal path delays of paths P1, P2, and P3, respectively, and ΔD_{P1} , ΔD_{P2} , and ΔD_{P3} , respectively, are the changes in the delay of each of these paths due to TSV-stress-induced variations. Note that D_{P1} and ΔD_{P1} together evaluate to the actual critical path delay of the circuit show in column D2 of Table IV. This table also shows the amount of change in the circuit total positive slack (TPS) and the total negative slack (TNS) when TSV effects are considered, are denoted by ΔTPS and ΔTNS , respectively. While computing slacks, we consider the worst case path delay of the circuit without TSV effects as the required time specification to be met. From the table we can observe that:

- The actual change on the critical path denoted by ΔD_{P1} can be more than the change in the worst case path delay observed

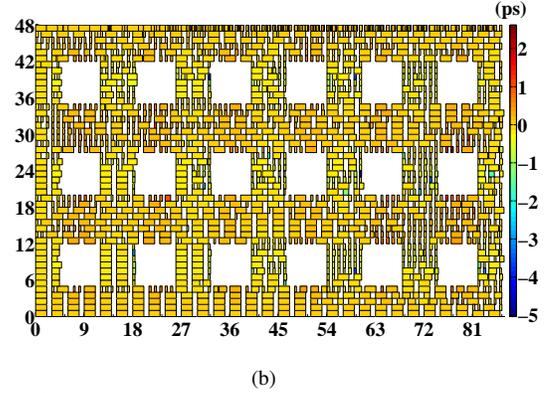
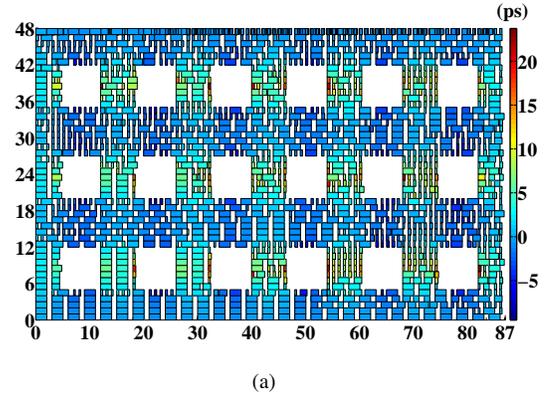


Fig. 8. spi (a) PMOS Δ Delay map (b) NMOS Δ Delay map.

at the circuit level shown in ΔD_2 in Table IV.

- A noncritical path can become timing-critical when TSV effects are considered. This is observed by comparing the delays in D_{P1} and its percentage change, ΔD_{P1} in Table VI with the circuit critical path delay D_2 and the circuit level change, ΔD_2 in Table IV.
- The maximum delay degradation or improvement, given by ΔD_{P2} and ΔD_{P3} , respectively, among all paths is significantly greater than the worst case path delay changes observed at the circuit level.
- The negative changes in ΔTPS of the circuits reveal that a majority of paths experience delay degradation and there is lower positive slack available in the circuit under TSV effects.
- The wide distribution in the ΔTNS indicates that many non-critical paths in the circuit can violate timing constraints when TSV effects are taken into account,

Figure 8 shows the color maps of the delay changes in PMOS and NMOS transistors in the gates for the spi circuit. The square white portions represent the TSV locations. Consistent with Figure 3, we can see that the maximum delay changes are observed in the regions horizontal and vertical to the TSVs where the stresses are concentrated. By comparing the scales we can conclude that the changes in the PMOS delays are dominant in the circuit as compared to the NMOS delays. Moreover, the amount of path delay variation depends on the relative location of the gates with the TSVs. Thus we can conclude that path delay degradations are due to the gates placed in the horizontal region between TSVs and the delay improvements are due to the gates placed in the vertical regions between TSVs. The effects are opposite when all the transistor channels are perpendicular to the [110] axis.

TABLE VII

MINIMUM PATH DELAY OF TSV₇ CIRCUITS WITH {TSV + BCB LINER}

Circuit	w/o TSV effects		with TSV effects	
	$D_{min}(ps)$	# Violations	$D_{min}(ps)$	# Violations
ac97_ctrl	22	998	18	1112
aes_core	22	3802	19	3684
des	29	28	24	50
ethernet	22	2480	19	2304
i2c	22	80	22	77
mem_ctrl	22	500	21	500
pci_bridge32	22	4140	21	3822
spi	22	48	22	91
systemcdes	29	238	23	255
usb_funct	22	908	21	1046

Short path variations: Finally, we examine the effects of TSV stress on short paths and hold time constraints, since it is possible for path delays to decrease under TSV-induced mobility variations, depending on their placement relative to the TSVs. Table VII shows the minimum path delays and the number of violations observed in the circuits without and with TSV effects. The minimum path delay in each case is denoted by D_{min} and we consider a minimum path delay requirement of 50 ps to report the number of path violations with and without TSV effects. We can see that the minimum path delay D_{min} differs in the two cases. Moreover, when we incorporate TSV stress effects, the number of paths violating the minimum delay constraint either increase or decrease for the circuits except for mem_ctrl, where they are equal. Thus, during sequential circuit design in the presence of TSVs, the impact on minimum path delays should also be accounted for.

Layout guidelines: Based on this analysis, it has been demonstrated that the delay changes within the circuit are very significant, but their effects are attenuated at the outputs due to the effect of the max operation in timing analysis, which changes the critical path. This suggests that this freedom can be exploited by layout tools to “hide” the delay increases. Based on our analysis of stress patterns, we can draw the following general layout strategies that optimize delay:

- In general, to minimize the variations in gate-delays, the regions diagonal to the TSVs should be preferred.
- For paths that are timing-critical or near-critical, the gates should be placed in the vertical (horizontal) regions between TSVs when transistors are parallel (perpendicular) to the wafer flat.
- On paths with low minimum delay margins, the gates should be placed in the horizontal (vertical) regions between TSVs when transistors are parallel (perpendicular) to the wafer flat direction.

VI. CONCLUSION

We have developed a holistic framework to analyze the effects of temperature in 3D ICs. A 2D analytical plane stress model has been developed for TSVs embedded in silicon and translated into mobility variations using piezoresistivity theory. The crystal orientation and the transistor channel orientation effects have also been taken into account. A thorough comparison between our biaxial model and the uniaxial models employed in prior art have been shown. The biaxial analytical model has been compared with FEA simulations using realistic TSV structures and liner configurations. Empirically scaled analytical models were employed to improve the accuracy in predicting stress. A sensitivity based delay model has been employed to compute the delay variations in circuits due to both temperature and the TSV-effects. A detailed analysis of circuit delays has been presented, and layout guidelines are suggested for delay optimization in 3D ICs.

REFERENCES

- [1] A. Dasdan and I. Hom, “Handling inverted temperature dependence in static timing analysis,” *ACM Transactions on Design Automation of Electronic Systems*, vol. 11, pp. 306–324, April 2006.
- [2] K.-H. Lu, S.-K. Ryu, J.-H. Im, R. Huang, and P. S. Ho, “Thermomechanical reliability of through-silicon vias in 3D interconnects,” in *IEEE International Reliability Physics Symposium*, pp. 3D.1.1–3D.1.7, 2011.
- [3] A. Karmarkar, X. Xu, and V. Moroz, “Performanace and reliability analysis of 3D-integration structures employing through silicon via (TSV),” in *IEEE International Reliability Physics Symposium*, pp. 682–687, 2009.
- [4] J.-S. Yang, K. Athikulwongse, Y.-J. Lee, S. K. Lim, and D. Pan, “TSV stress aware timing analysis with applications to 3D-IC layout optimization,” in *Proceedings of the ACM/IEEE Design Automation Conference*, pp. 803–806, 2010.
- [5] K. H. Lu, X. Zhang, S.-K. Ryu, J.-H. Im, R. Huang, and P. S. Ho, “Thermo-mechanical reliability of 3D ICs containing through silicon vias,” in *Electronics Components and Technology Conference*, pp. 630–634, 2009.
- [6] J. Jung, J. Mitra, D. Z. Pan, and S. K. Lim, “TSV stress-aware full-chip mechanical reliability analysis and optimization for 3D IC,” in *Proceedings of the ACM/IEEE Design Automation Conference*, pp. 188–193, June 2011.
- [7] D. H. Kim, K. Athikulwongse, and S. K. Lim, “A study of through-silicon-via impact on the 3D stacked IC layout,” in *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pp. 674–680, 2009.
- [8] S.-K. Ryu, K.-H. Lu, X. Zhang, J.-H. Im, P. S. Ho, and R. Huang, “Impact of near-surface thermal stresses on interfacial reliability of through-silicon vias for 3-D interconnects,” *IEEE Transactions on Device and Materials Reliability*, pp. 35–43, 2011.
- [9] D. A. Bittle, J. C. Suhling, R. E. Beaty, R. C. Jaeger, and R. W. Johnson, “Piezoresistive stress sensors for structural analysis of electronic packages,” *Journal of Electronic Packaging*, vol. 113, no. 3, pp. 203–215, 1991.
- [10] R. C. Jaeger, J. C. Suhling, R. Ramani, A. T. Bradley, and J. Xu, “CMOS stress sensors on (100) silicon,” *IEEE Journal of Solid-State Circuits*, vol. 35, pp. 85–95, 2000.
- [11] Y. Kanda, “A graphical representation of the piezoresistance coefficients in silicon,” *IEEE Transactions on Electron Devices*, pp. 64–70, 1982.
- [12] W. G. Pfann and R. N. Thurston, “Semiconducting stress transducers utilizing the transverse and shear piezoresistance effects,” *Journal of Applied Physics*, vol. 32, pp. 2008–2019, 1961.
- [13] E. Pop, S. Sinha, and K. E. Goodson, “Heat generation and transport in nanometer-scale transistors,” *Proceedings of the IEEE*, vol. 94, pp. 1587–1601, August 2006.
- [14] K. Kanda, K. Nose, H. Kawaguchi, and T. Sakurai, “Design impact of positive temperature dependence on drain current in sub-1-V CMOS VLSIs,” *IEEE Journal of Solid-State Circuits*, vol. 36, pp. 1559–1564, October 2001.
- [15] “IWLS 2005 Benchmarks.” available at <http://www.iwls.org/iwls2005/benchmarks.html>.
- [16] “Predictive Technology Model.” available at [http://www.eas.asu.edu/\\$\sim\\$sim\\$ptm](http://www.eas.asu.edu/\simsim$ptm).
- [17] A. E. Caldwell, A. B. Kahng., and L. L. Markov, “Can recursive bisection alone produce routable, placements?,” in *Proceedings of the ACM/IEEE Design Automation Conference*, pp. 477–482, 2000.