Journal of
Micro/Nanolithography, MEMS, and MOEMS

SPIEDigitalLibrary.org/jm3

# Stress assessment for device performance in three-dimensional IC: linked package-scale/die-scale/feature-scale simulation flow

Armen Kteyan
Gevorg Gevorgyan
Henrik Hovsepyan
Jun-Ho Choy
Valeriy Sukharev

SPIE

# Stress assessment for device performance in three-dimensional IC: linked package-scale/die-scale/feature-scale simulation flow

**Armen Kteyan,[a] Gevorg Gevorgyan,[a] Henrik Hovsepyan,[a] Jun-Ho Choy,[b] and Valeriy Sukharev[b,*]**
[a]Mentor Graphics Corporation, Yerevan 0012, Armenia
[b]Mentor Graphics Corporation, Fremont, California 94538

**Abstract.** Potential challenges with managing mechanical stress and the consequent effects on device performance for advanced three-dimensional (3-D) IC technologies are outlined. The growing need in a simulation-based design verification flow capable of analyzing a design of 3-D IC stacks and detecting across-die out-of-spec variations in MOSFET electrical characteristics caused by the die thinning and stacking-induced mechanical stress is addressed. The development of a multiscale simulation methodology for managing mechanical stresses during a sequence of designs of 3-D IC dies, stacks, and packages is focused. A set of physics-based compact models for a multiscale simulation is proposed to assess the mechanical stress across the device layers in silicon chips stacked and packaged with the 2.5D interposer-based, and true 3-D through silicon via-based technology. A simulation flow is developed for the hot-spot checking in different types of devices/circuits such as digital, analog, analog matching, memory, IO, characterized by different sensitivities to the stress-induced mobility variations. A calibration technique based on fitting to measured electrical characteristics of the test-chip devices is presented. The limited characterization or measurement capabilities for 3-D IC stacks and a strict "good die" requirement make this type of analysis critical in order to achieve an acceptable level of functional and parametric yield.  [DOI: 10.1117/1.JMM.13.1.011203]

Keywords: stress; three-dimensional IC package; bumps; through silicon via; compact model.

Paper 13128SS received Jul. 16, 2013; revised manuscript received Nov. 25, 2013; accepted for publication Dec. 11, 2013; published online Jan. 21, 2014.

## 1 Introduction

It is a common understanding that the motivation for 3-D IC integration is a mixture of economic and technical requirements, summarized within the term "More than Moore."[1–3] Three-dimensional (3-D) IC stacking technologies (including 2.5D interposer-based approaches), employing thinned wafer/through silicon via (TSV) structures, are novel solutions that result in reduced floor space, higher bandwidth, and reduced energy consumption. To enable 3-D chip stacking, new components were adopted by interconnect technology: through-silicon vias to provide connectivity between the back-end-of-line (BEoL) and back-chip redistribution layer (BRDL) metallization of some of the stacked dies, and solder or metal bumps and pillars for interconnecting the neighboring dies. The processing of high-density TSV structures through thinned dies and subsequent 3-D stacking is a promising technological alternative to the traditional two-dimensional (2-D) lithography/etch scaling. However, several issues have to be addressed to guarantee the needed product performance and reliability. Among them are management of mechanical stress and high Joule heating as well as reliability issues such as electromigration and stress migration.[4] Many of the process steps employed by 3-D IC technology act as stress sources that can affect the chip performance and reliability. These are TSV etching and filling, wafer/die thinning, wafer bumping, high-temperature solder reflow, chip stacking, etc.[3,5] Hence, it is important to have a capability to accurately assess the stress generated during 3-D IC stacking.

The traditional chip–package interaction (CPI) concerns are related to reliability issues caused by high peeling and shear stresses (hundreds of MPa), which are able to cause cracking, delamination, etc., resulting in shorts or opens. These stresses represent a serious risk for chips' durability; nevertheless, the existing measurement and screening procedures enable the failure modes detection, allowing to eliminate the detected threats. The failure risk can be minimized by using appropriate package assembly materials, which can decrease substantially the CPI stress. Analysis of the reliability issues is out of scope of this paper.

The stress simulator described here reflects the needs of designers and manufacturers of 3-D IC chip packages to control deviations of design-in MOSFET parameters that cause degradation of the chip performance. Even relatively low values of CPI stress (less than hundred MPa) can cause parametric failures of circuits due to change of charge carrier mobility in transistor's channel regions. Stress gradients across regions where circuits are located can worsen the chip performance even if the stress level is not high. Due to the lack of measurement and screening procedures for detecting mobility shift in transistors, the developed simulator becomes an additional metrology tool for 3-D IC package inspection.

It is a very challenging task to get an entire picture of the modification of the stress distribution across device layers caused by 3-D IC technology. Different scales of the stress distributions must be taken into account: die scale (i.e., several millimeters) global stress variations which are generated

*Address all correspondence to: Valeriy Sukharev, E-mail: Valeriy_Sukharev@mentor.com

due to thin die bending; bump and TSV scale (i.e., 10 to 100 $\mu$m) local stress variations, which are induced by the package component assembly; device scale (10 to 100 nm) variations due to transistor layer nonuniformity. Traditional methods such as finite-difference analysis and finite-element analysis (FEA) cannot be employed for a simulation of the transistor channel stress distribution across a die due to the size of a model, which can easily reach hundreds of millions degrees of freedom. The established FEA-based stress simulators have typically been used for addressing the traditional CPI effect where a silicon chip was modeled as a homogeneous isotropic piece of silicon. Details of chip structures (layer information, layout, etc.) have not been considered, and the problem with calculating the transistor-to-transistor intrachannel stress variation and consequent variation in transistor electrical characteristics has not been addressed yet.[6] In order to be able to consider these effects, the stress simulation methodology should be capable to resolve scales of the order of a transistor size (approximately nanometers) and, to account for all major internal (layout-induced) and external (e.g., packaging) stress sources affecting a particular device. Compact model-based approaches for the layout-induced stress effects, which typically employ the empirical modeling,[7,8] cannot take into account the package-induced variations in transistor characteristics. Because of the lack of physics-based foundation, this kind of modeling cannot provide a link to the physics-based package-scale simulation in order to include a CPI-induced stress loading. A look-up table methodology is not practical due to a large size of each device's surrounding layout area (radius of up to 5 $\mu$m) that should be accounted for a correct stress prediction. It will require the generation of an enormous amount of local layout configurations around a gate in order to obtain a proper representation. Therefore, the development of compact physics-based models is the only possible solution to achieve the ultimate goal of predicting and simulating transistor-to-transistor stress variation across a device layer. These compact models are based on analytical treatment of the elasticity problems: approximate solutions of the basic differential equations are used to generate a set of simple algebraic equations allowing fast full-chip analysis of stress values in transistor channels. Resolved stress variation should be converted further into the variation of the device electrical characteristics.

This paper describes the recently developed physics-based simulator that predicts variation in transistor's electrical characteristics, caused by CPI. The simulator provides an interface between layout formats (GDS II, OASIS) and FEA-based package-scale models.

## 2 Simulation Methodology

The proposed simulation methodology/flow, which will be discussed in detail below, results in the development of a new type of design verification tools. The tool is capable of analyzing any 3-D IC die stack design with regard to out-of-spec variations in device electrical characteristics caused by mechanical stress generated by warpages of the stacked dies, by solder bumps pressure generated in a course of die stacking, as well as by mismatch of thermomechanical properties of TSV and silicon die bulk. The target of the developed simulation flow is the calculation of across-die distributions of the stress normal components $\{\sigma_x, \sigma_y, \sigma_z\}$ (i.e., diagonal components of stress tensor) inside transistor channels and its conversion into the stress-modified electrical characteristics of the transistors.

This simulation flow, shown schematically in Fig. 1, can be described as a sequence of the simulation steps performed with different simulation tools:

1. FEA-based package-scale simulation of the stack deformation, which is originated by the package assembly. Output: fields of displacements on the surfaces of all stacked dies.
2. Compact model for bump-induced displacements. Output: fields of total displacements, caused by die warpage and bump-die contact mechanics, on the die surfaces.
3. Compact models for nonuniform, coordinate-dependent mechanical properties of BEoL and BRDL interconnects and silicon bulk with TSVs.
4. FEA-based die-scale simulation of the strains generated in the die by the surface displacements calculated at the step 2; material properties are described by matrices created at the step 3. Output: CPI-induced strain and stress distributions across the device layer.
5. Compact model for TSV-induced stresses.
6. Compact model-based simulation of the transistor-to-transistor stress variation across the device layer, and the conversion of the final stress into the mobility multiplier (transistor instant parameter MULU0). This stress variation is caused by a layout-induced relaxation of the stress simulated at the previous step. Output: components of CPI and TSV induced stresses, and instant parameters for each analysed transistor; the annotated SPICE netlist.

Both FEA simulations and compact modelling are performed assuming an elastic behavior of all materials involved in package assembly. The possibilities of plastic deformation of solder bumps and viscoelastic behaviour of interlayer dielectric and underfill are not considered in the presented simulation flow. Viscoelastic properties of underfill layer above glass transition temperature $T_g$ are not modeled assuming that the major input into stress caused

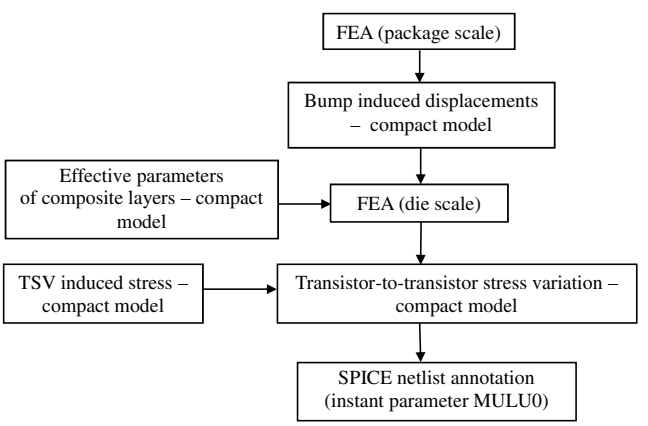



**Fig. 1** Structure of the developed multiscale simulation flow.

by the bumps contact interactions with the dies (thin multilayered plates) was developed at temperatures below $T_g$. Within the accepted linear theory of elasticity, the superposition of stresses induced by various assembly components can be used to obtain the total stress distribution. According to the flow schematics (Fig. 1), the stress created by TSVs is considered separately and is assumed to be independent of wafer deformation. Obviously, TSVs cannot be resolved in package scale FEA simulations (due to their small sizes), which implies that the possible effect of TSVs on assembly-induced wafer deformation is neglected. It should be mentioned that this effect is captured at the FEA die-scale simulation step by introducing the nonuniform, coordinate-dependent mechanical properties of the silicon die bulk layer governed by TSV locations.

The described simulations methodology has resulted in the development of a flow for assessment of stress-induced variations in the carrier (electrons and holes) mobility in MOSFET channels, which, when it is plugged into the standard design flow, can be used for the design hot-spots screening and for accurate simulation of a variety of chip characteristics such as performance, leakage, power, etc.

### 2.1 *Package-Scale FEA-Based Simulation*

Assembly of 3-D IC stack involves the wafer-to-wafer, or die-to-wafer or die-to-die mounting. Mechanical and electrical integrities of the stacked dies are achieved by employment of solder balls/metal pillars and TSVs (Fig. 2). Large flip-chip (FC) bumps connect the tier1 die to the substrate, while the set of TSVs and micro-bumps provide the die-to-die connection. The assembly of 2.5D packages is similar but involves only a die-to-interposer mounting. The encapsulation of the dies is performed by employment of an underfill resin in the gaps between dies, and finally molding the whole stack.

Curing the package at high temperatures (~150°C) and subsequent cooling down to room temperature create substantial thermal load, resulting in the bending (warpage) of the dies, due to mismatch of the thermomechanical parameters of different layers. This process generates large stresses in the dies, mainly due to large values of the coefficient of thermal expansion (CTE) of the underfill, which results in "shrinking" of silicon die in lateral directions. FEA-based simulation, which is a traditional method for analyzing a mechanical behavior of the package, is included in the developed flow. Another possible approach is an approximate analytical analysis of stresses in multilayered structures.[9,10] A review of different types of analytical techniques available for resolving the stress-related problems in silicon technology was presented in Ref. 11. Although the analytical modeling can enable a fast computing of some of the stress components in 3-D IC stack, it is expected to be less accurate than the FEA analysis in the case when tiers of different lateral sizes are stacked in the package as shown in Fig. 2. Obviously, the FEA-based package simulation cannot resolve geometrically the effect of hundreds of FC-bumps and microbumps. Therefore, all layers in the stack are considered as the homogeneous thin plates. At this step, the thermomechanical characteristics of underfill layers are defined using the so-called rule of mixtures.[12] This means that the bumps are assumed to be "smeared" throughout the entire layer. The mechanical characteristics of this "smeared" layer are obtained by the volume averaging of the corresponding properties of bumps and underfill in accordance with their volume fractions.

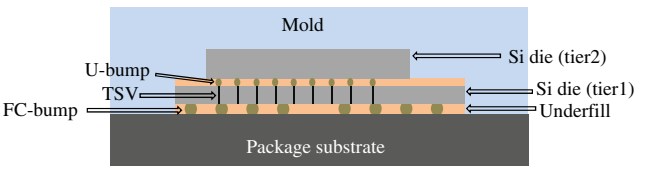
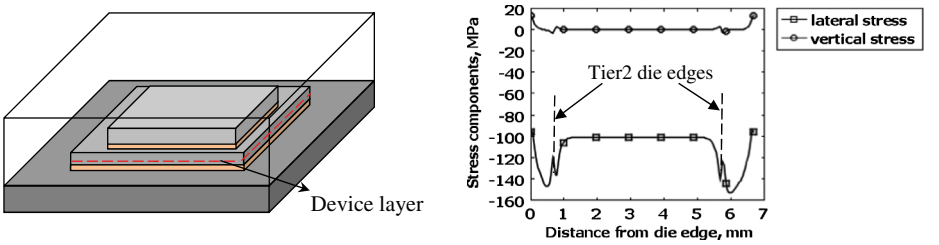



**Fig. 2** Schematics of 3-D chip stack.

An example of the simulated distribution of the stress components in two-tier stack across tier1 is shown in Fig. 3. This stress, which is generated due to mismatch in thermomechanical properties of the constituent layers, is accompanied by the warpage of the dies. With the thicknesses of the layers in the package and values of material properties collected in Table 1, the values of lateral components of the stress exceed 100 MPa. Larger stress can be generated with decreasing die thickness. Evidently, stress distribution is nonuniform near the edges of the analysed (tier1) die. In addition, a strong nonuniformity is observed in the region under the periphery of upper (tier2) die, which is caused by vertical pressure created by tier2 die edges.

### 2.2 *Compact Model for Bump-Induced Displacements*

Although the FEA simulation allows us to obtain the detailed variation of the warpage-induced stresses (Fig. 3), it does not take into account the local deformation/stress




(a) (b)

**Fig. 3** (a) The architecture of the analyzed die; (b) distribution of global stress components in tier1 across the subsurface region indicated in (a) by the dashed line.

**Table 1** Thicknesses and thermomechanical properties of materials in the 3-D package used in FEA simulations.

| | Thickness ($\mu$m) | Young's modulus (GPa) | Poisson's ratio | CTE (ppm/°C) |
|---|---|---|---|---|
| Substrate | 200 | 30 | 0.2 | 12 |
| Underfill (tier1) | 80 | 12 | 0.3 | 30 |
| Silicon die (tier1) | 50 | 170 | 0.27 | 2.6 |
| Underfill (tier2) | 30 | 10 | 0.3 | 25 |
| Silicon die (tier2) | 100 | 170 | 0.27 | 2.6 |
| Mold | 700 | 25 | 0.2 | 8 |

generated by bumps. To capture this deformation, an analytical modelling linked with the package-scale FEA simulation was developed. Fields of warpage-related displacements $\boldsymbol{u}^W(x,y) = \{\boldsymbol{u}^W_{\text{top}}(x,y), \boldsymbol{u}^W_{\text{bot}}(x,y)\}$ of the top and bottom surfaces of all layers calculated by FEA tool can be used for linking the package-scale simulations with the bump-scale model. These displacements allow calculating analytically the approximate local deformations generated around each bump. Here and below the displacement $u$ implies 3-D vector $\boldsymbol{u} = \{u_x, u_y, u_z\}$, which defines the stress components in the employed approximation of elastic materials.

Figure 4 shows the local deformation that is generated around the FC bump due to chip cooling from the processing temperature down to operating temperature. Bumps, consisting of intermetallic compounds (in the case of FC bumps) or copper (micro-bumps), are characterized by the substantially larger Young's modulus ($E_b$) than the underfill material ($E_u$), i.e., $E_b > E_u$, but by smaller CTE (i.e., $\alpha_b < \alpha_u$). As a result, the placement of the bump into the underfill layer should modify the warpage-related displacement fields both in lateral $(x, y)$ and vertical $(z)$ directions[13] (Fig. 4). Due to a mismatch between the bump height shrinkage and the vertical deformation of the underfill $\Delta u^W_z$ at the bump location, both caused by the cooling, the bump acts as an indenter deforming both the substrate and die subsurface regions in the case of FC bumps. A vertical load generated by the bump can be written as

$$P_z \sim E_b\left(\frac{\Delta u^W_z}{H} + (\alpha_b - \alpha_u)\Delta T\right). \tag{1}$$

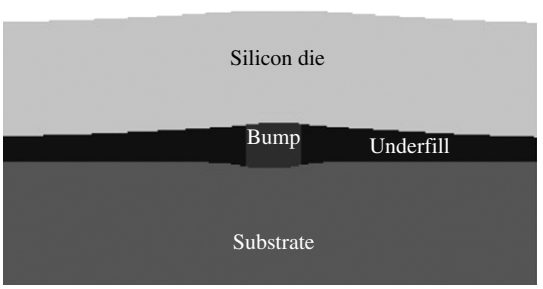



**Fig. 4** Bump induced deformation of silicon.

Here, $H$ is the bump height, $\Delta u^W_z/H$ represents warpage-related vertical strain at the bump location, and $\Delta T$ describes the thermal load. Similarly, due to warpage-induced lateral displacements $u^W_{x(y)}$, each bump generates the following tangential load

$$P_{x(y)} \sim (E_b - E_u)\frac{\Delta u^W_{x(y)}}{D}, \tag{2}$$

where $D$ is the bump diameter, and $u^W_{x(y)}/D$ represents warpage-related lateral strain in the bump region. It should be mentioned that the thermal mismatch of bump and underfill is able to reduce the tangential load provided by Eq. (2). Besides, an extra load is generated by the shear deformation of the bumps, which originates due to different CTE of the silicon die and of the substrate. This shear component of bump deformation is more essential for the bumps located near the die edges. Formulas (1) and (2) establish the link between the bump compact model-induced displacements and the warpage-related displacements calculated with the FEA model.

The additional deformation of the die surface $u^{\text{bump}}$ due to loading Eqs. (1) and (2) can be calculated by solving the corresponding problems of contact mechanics.[14,15] Additional field of $z$-displacements at the die surface that is generated by a single bump is

$$u^{\text{bump}}_z = \begin{cases} \frac{8(1-\nu_\gamma)DP_z}{3\pi E_\gamma}, & \text{if } r \le \frac{D}{2} \\ \frac{8(1-\nu_\gamma^2)DP_z}{\pi E_\gamma}\left(E\left(\frac{D}{2r}\right) - \left(1 - \frac{D^2}{4r^2}\right)K\left(\frac{D}{2r}\right)\right)r, & \text{if } r > \frac{D}{2} \end{cases}. \tag{3}$$

Here, $r$ is the distance to the center of the bump, $K(D/2r)$ and $E(D/2r)$ are the complete elliptic integrals of the first and second kinds correspondingly, with modulus $D/2r$, and $E_\gamma$ and $\nu_\gamma$ are the Young's modulus and Poisson's factor of the material which is in contact with the bump. Taking into account a fast decay of the bump contribution to the total vertical displacements at large distance $r \gg D$, we have introduced an effective "radius of bump interaction" $r_{\text{eff}}$, which restricts the size of the area affected by the bump. This $r_{\text{eff}}$ estimates the bump model accuracy. Similarly to Eq. (3), the lateral displacements caused by loads (1) and (2) can be obtained by using the force-balance equation based analysis.

Finally, the total displacements at the die surfaces can be obtained by adding the bump-induced displacements $\boldsymbol{u}^{\text{bump}}$ to warpage-induced one $\boldsymbol{u}^W$:

$$\boldsymbol{u}^{\text{surf}} = \boldsymbol{u}^W + \sum_{|\boldsymbol{r}-\boldsymbol{r}_i| \le r_{\text{eff}}} \boldsymbol{u}^{\text{bump}}(\boldsymbol{r} - \boldsymbol{r}_i). \tag{4}$$

Equation (4) shows that all bumps at $\boldsymbol{r} = \boldsymbol{r}_i$ within the region $|\boldsymbol{r} - \boldsymbol{r}_i| \le r_{\text{eff}}$ affect the displacement generated at this point.

### 2.3 *Effect of Nonuniform Interconnect, BRDL and Silicon Bulk Mechanical Characteristics on the CPI-Induced Stress Simulated with the Die-Scale FEA Model*

Strain and stress distribution across a device layer, which is a layer located in the silicon interior just nanometers below

the interface with the BEoL interconnect, should be calculated with a FEA tool by implementing the field of displacements (4) as the boundary conditions (BCs) for the die faces. At this step, each die is considered as a multilayer stack consisting of a silicon layer, BEoL interconnect, and BRDL layers. These layers are characterized by the nonuniform spatial distributions of the elastic and thermal properties determined by their layouts. An assessment of additional stress variation caused by the nonuniformity of the mechanical properties of these layers is implemented in the developed simulation flow.

A calculation methodology of the effective Young's modulus, Poisson's ratio, and CTE as functions of metal density $\rho_M$ in each metal layer has been developed based on the theory of mechanical properties of anisotropic composite materials.[12] This methodology requires a division of all considered composite layers into a number of bins. Average values of thermomechanical properties of the composite material are calculated for each bin on the basis of the average metal density $\rho_M$ inside the bin. The size of the bin should be chosen based on required simulation accuracy: the finer partitioning provides more accurate results at the expense of the run time. Depending on routing direction of the metal wires, the Young's modulus of $i$'th bin at $j$'th layer should be calculated using one of the following formulas:[12]

$$E_{\parallel}^{i,j} = E_M \rho_M^{i,j} + E_D(1-\rho_M^{i,j}),$$
$$E_{\perp}^{i,j} = \frac{E_M E_D}{E_D \rho_M^{i,j} + E_M(1-\rho_M^{i,j})}, \tag{5}$$

where the symbols $\parallel$ and $\perp$ denote directions parallel and perpendicular to routing direction correspondingly; $E_M$ and $E_D$ are Young's modulus of metal and interlayer dielectric. The similar approach is used for calculating Poisson's ratio and CTE of all composite layers.

FEA model is used for the die-scale simulation of the stress components $\sigma = \{\sigma_x, \sigma_y, \sigma_z\}$. In this simulation, the BCs are represented by previously calculated die face displacements, and the material properties are represented by the calculated position-dependent mechanical properties of the composite layers:

$$\sigma(x,y;z \in \text{devicelayer}) = f[\boldsymbol{u}^{\text{surf}}(x,y), E(x,y), \nu(x,y), \alpha(x,y)]. \tag{6}$$

Figure 5 demonstrates the distributions of stress components across the device layer of tier1 of the stack shown in Fig. 3(a). The studied region is located far from the die edges; therefore, the obtained periodical distribution of the stress components reflects a periodical pattern in bump locations. Across die variation of the BEoL mechanical properties results in the stress distribution blurring. The colormaps, showing stress values in Fig. 5, were obtained using the material properties presented in Table 1 and the properties of the bumps, interlayer dielectric (ILD), and copper presented in Table 2.

### 2.4 *Compact Model for TSV-Induced Stress*

TSV fabrication generates strain in the surrounding silicon: a thermal load $\Delta T$ generated by cooling the chip down from copper anneal temperature to room temperature creates a thermal mismatch strain $\varepsilon_{\text{th}} = (\alpha_{\text{Cu}} - \alpha_{\text{Si}})\Delta T$ due to large CTE difference between copper and silicon. The distribution of radial and circumferential stress components around the TSV, at a distance $r > D_{\text{TSV}}/2$, where $D_{\text{TSV}}$ is the TSV diameter, can be approximated by Lame formula:[14]

$$\sigma_r = -\sigma_\theta \approx \frac{E_{\text{Si}}\varepsilon_{\text{th}}}{1-2\nu_{\text{Si}}}\frac{D_{\text{TSV}}^2}{4r^2}. \tag{7}$$

Due to axial symmetry, the stress components of Eq. (7) in polar coordinates are independent of azimuthal angle $\theta$. Equation (7) is valid when the TSV is placed far enough from the die periphery (at a distance much larger than $D_{\text{TSV}}$), where the boundary effects can be neglected. It should be mentioned that for more accurate description of TSV-induced stress distribution, the deformation of silicon surface due to prevailing vertical shrinking of the TSV in comparison with silicon should be considered.[16] However, it can be shown that for the device layer, which is located very close to the silicon/BEoL interface, this surface effect results in just

**Table 2** Thermomechanical properties of bumps and back-end-of-line (BEoL) interconnect materials.

| | Young's modulus (GPa) | Poisson's ratio | CTE (ppm/°C) |
|---|---|---|---|
| Bump | 40 | 0.35 | 21 |
| ILD | 10 | 0.18 | 5.5 |
| Copper | 120 | 0.34 | 16.5 |

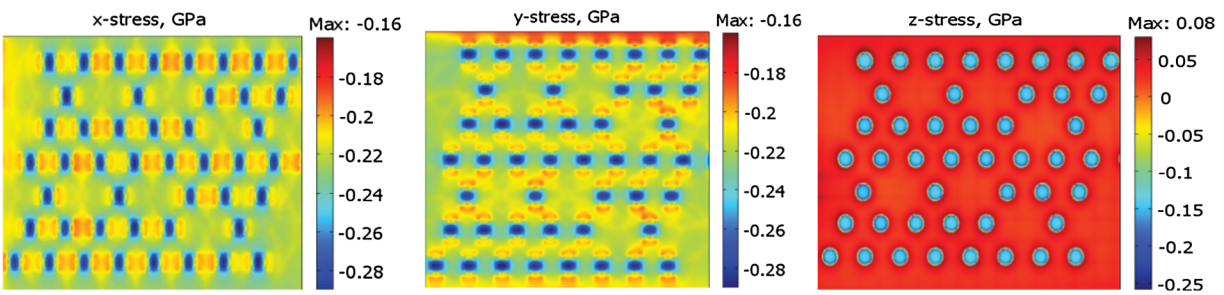



**Fig. 5** An example of the die-scale simulation results: distributions of stress components across the device layer of tier1 die. The obtained patterns are due to the effect of FC bumps of 80-$\mu$m diameter. The size of the demonstrated region is $1900 \times 1300$ $\mu$m$^2$.

a renormalization of the parameter $\varepsilon_{\text{th}}$. In the discussed simulation flow, the value of this parameter should be determined by calibration. This should validate the employment of the Eq. (7) as an appropriate approximation for the effect of TSV on device characteristics. Vertical component of stress $\sigma_z$ is essential only in the immediate vicinity of TSV and can be neglected since the devices are not allowed to be placed inside the so-called "keep-out-zones" around TSVs, where zone size is prescribed by the design rules.

The radial and circumferential components [Eq. (7)] of TSV-induced stress can be transformed to the Cartesian components $\sigma_x, \sigma_y$ by

$$\sigma_x = \frac{\sigma_r + \sigma_\theta}{2} + \frac{\sigma_r - \sigma_\theta}{2}\cos\,2\theta$$
$$\sigma_y = \frac{\sigma_r + \sigma_\theta}{2} - \frac{\sigma_r - \sigma_\theta}{2}\cos\,2\,\theta.$$

This transformation provides

$$\sigma_x = -\sigma_y = \frac{E_{\text{Si}}\varepsilon_{\text{th}}}{1-2\nu_{Si}}\frac{D_{\text{TSV}}^2}{4r^2}\cos 2\,\theta. \tag{8}$$

These stress components should be determined in the same grid points, which were used for calculation of CPI stress components [Eq. (5)]. Finally, assuming that CPI and TSV stresses are independent and contribute additively, the total stress distribution is determined as

$$\sigma = \sigma^{\text{CPI}} + \sigma^{\text{TSV}}. \tag{9}$$

### 2.5 *Simulation of the Transistor Intrachannel Stress Components*

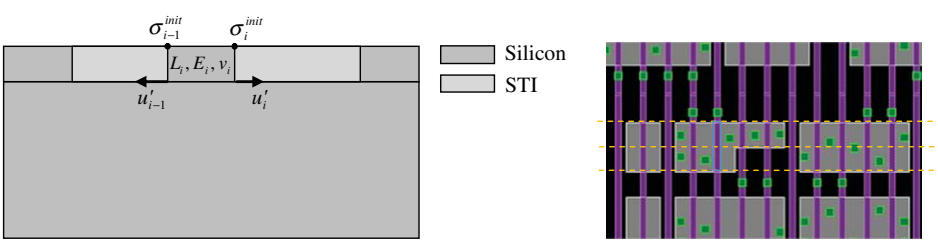
The generated across-die distributions of stress components make it possible to calculate the averaged stress components inside transistor channels by using a simple interpolation procedure. However, at a device scale, the composite nature of the device layer becomes important. In fact, this layer represents a sequence of silicon islands separated by the shallow trench isolation (STI) regions filled by deposited silicon oxide. Keeping in mind a drastic difference in the mechanical characteristics of silicon and silicon oxide, as well as the fact that both the Si islands (i.e., the regions where the transistors are located) and STI regions are characterized by wide dispersion in sizes and shapes, we can conclude that strain distribution should depend on the local layout configurations of the device layer. FEA-based die-scale simulations of the CPI stress described in Sec. 2.3 cannot resolve the billions of shapes existing in devices layout. Therefore, in the next step of the multiscale simulation flow, a compact model-based calculation of the device-scale stress variations should be performed. This compact model considers the strain/stress distribution, calculated with Eq. (9), as an intial distribution that should be changed (relaxed) in accordance with the layout geometries. To calculate this stress redistribution, each transistor channel and the neighboring layout are portitioned on a set of "cut-layers" (both in $x$- and $y$-directions). Each cut-layer consists of the device layer and the silicon bulk, as shown in Fig. 6(a). The device layer consists of the sequence of segments representing slices of silicon islands and STIs separating silicon islands. Each $i$'th segment in the device "composite" layer is characterized by its length $L_i$, Young's modulus $E_i$ and Poisson's ratio $\nu_i$. The stresses given by Eq. (9) are considered as the "initial" stresses in each segment of the cutline. The difference in elastic properties of the neighbouring segments results in redistribution of this stress: each segment edge experiences additional lateral displacement $u_i'$ due to the action of the forces $F_i \sim (1 - E_{\text{STI}}/E_{\text{Si}})\sigma_i^{\text{init}}$. These displacements can be obtained from the solution of the force balance equation that takes into account the interaction between adjacent segments and the traction with the silicon bulk. Initial stress redistribution can be described as generation of an additional stress $\sigma'$, which can be obtained from solution of the corresponding force balance equation. For example, for each cut-layer directed along $x$-axis, the force balance equation for calculating the stress component $\sigma_x'$ is the following

$$\frac{\partial\sigma_x'}{\partial x} + \frac{\partial\tau_{xz}'}{\partial z} = 0, \tag{10}$$

where $\tau_{xz}$ is the shear stress component. Assuming that the vertical displacements everywhere in the plane are small in comparison with the lateral displacements, we can employ the following representation of stresses as functions of displacements

$$\sigma_x' = \frac{E}{1-\nu^2}\frac{\partial u_x'}{\partial x} \quad \tau_{xz}' = \frac{E}{2(1+\nu)}\frac{\partial u_x'}{\partial z}. \tag{11}$$

It reduces the problem to the solution of the following equation for the lateral displacement $\mu_x'$:




(a) (b)

**Fig. 6** Schematics of the cut-line: (a) vertical slice, (b) top view demonstrating device partitioning into cut-lines. Here, light polygons are silicon islands, narrow lines are the poly gates, small squares are contacts, and dark background is STI.

$$\frac{\partial^2 u_x'}{\partial x^2} + \frac{1-\nu_{Si}}{2}\frac{\partial^2 u_x'}{\partial z^2} = 0. \tag{12}$$

As it was shown in Refs. 17 and 18, this equation can be further reduced to the system of linear equations for the lateral displacements at each node of the considered cut-line:

$$E_i'' u_{i-1} - (E_i' - E_{i+1}')u_i + E_{i+1}'' u_{i+1} = (E_i - E_{i+1})\varepsilon_i^0. \tag{13}$$

Here, $E_i$ and $E_{i+1}$ are the Young's modulus of $i$'th and $(i+1)$'th segments; $E_i''$ and $E_i'$ are the known functions of the materials properties and segment geometries; $\varepsilon_i^0$ is the initial strain in $i$'th node. These equations clearly demonstrate that the coupling between the global stress load (TSVs, packaging, bumps, etc.) and the resulted layout-induced stress distribution is introduced through the initial strain distribution calculated with the combined FEA-compact model. Solution of the system of Eq. (13) provides the distribution of displacements along all considered cut-lines, which after standard transformation [Eqs. (14) and (15)] generate the distribution of stress components everywhere in the cut-lines, and particularly inside the transistor channels, Fig. 6(b).

$$\varepsilon_x^j = \frac{u_i - u_{i-1}}{L_i}, \qquad \varepsilon_x = \frac{\sum_j W^j \varepsilon_x^j}{W_{gate}}, \tag{14}$$

$$\begin{aligned}
\sigma_x &= \frac{E_{Si}}{(1+\nu_{Si})(1-2\nu_{Si})}[(1-\nu_{Si})\varepsilon_x + \nu_{Si}(\varepsilon_y + \varepsilon_z)] \\
\sigma_y &= \frac{E_{Si}}{(1+\nu_{Si})(1-2\nu_{Si})}[(1-\nu_{Si})\varepsilon_y + \nu_{Si}(\varepsilon_x + \varepsilon_z)] \\
\sigma_z &= \frac{E_{Si}}{(1+\nu_{Si})(1-2\nu_{Si})}[(1-\nu_{Si})\varepsilon_z + \nu_{Si}(\varepsilon_x + \varepsilon_y)].
\end{aligned} \tag{15}$$

### 2.6 *Stress-Induced Variation in Device Electrical Characteristics*

The piezoresistive effect describes change in the electrical resistivity of a semiconductor or metal when mechanical strain is applied. In the case of semiconductors, this resistivity change is caused by the strain-induced modification of their band structure and results in mobility changes of the conductive electrons and holes. Stress-induced modification of the charge carrier mobility is calculated based on well-determined piezoresistance coefficients:[19]

$$\frac{\Delta U}{U0} = -(\pi_x\sigma_x + \pi_y\sigma_y + \pi_z\sigma_z). \tag{16}$$

Here, $\Delta U = U(\sigma) - U0$, $U(\sigma)$ is the low-field mobility in the stressed silicon, $U0$ is the mobility at the zero stress condition;[20] the signs and values of piezoresistance coefficients $\pi_x, \pi_y, \pi_z$ are known to be dependent on crystallographic orientation of silicon surface and on the transistor channel orientation.[19,21] Correspondingly, for the SPICE instance parameter MULU0, we can write

$$\mathrm{MULU0} = \frac{U(\sigma)}{U0} \equiv 1 + \frac{\Delta U}{U0} = 1 - (\pi_x\sigma_x + \pi_y\sigma_y + \pi_z\sigma_z). \tag{17}$$

Figure 7 demonstrates the simulated mobility distributions in NMOS and PMOS transistors caused by an FC bump array for the case of ⟨110⟩ channel direction. The magnitude of the mobility change depends on the pitch of bump pattern and on device location relative to bumps.

SPICE netlist annotation with the calculated instant parameters MULU0 makes it possible to calculate stress-modified electrical characteristics of any device in the analyzed design by running SPICE simulator. It should be mentioned that the target of the developed simulation flow is the assessment of the effect of stresses associated with the 3-D IC technology on chip performance. These stresses are the unintentional and unwanted addition to the stresses already existing in the device layer generated by intentional stress sources, such as stressed layers (CESLs), source-drain epi-$Si_{1-x}Ge_x$, stress memorization, etc., and unintentional ones (STI, residual, etc.). The total stress-induced variation of the transistor characteristics is caused by a combination of the layout-induced and package (CPI/TSV)-induced stresses. An instant parameter MULU0_layout describing the effect of the layout-induced stresses on the mobility can be calculated with the foundry calibrated SPICE models. A combined effect of these two types of stresses on the device characteristics can be obtained from SPICE simulations by employing the netlist annotated with MULU0 that represents the product of MULU0_layout and MULU0_3D.

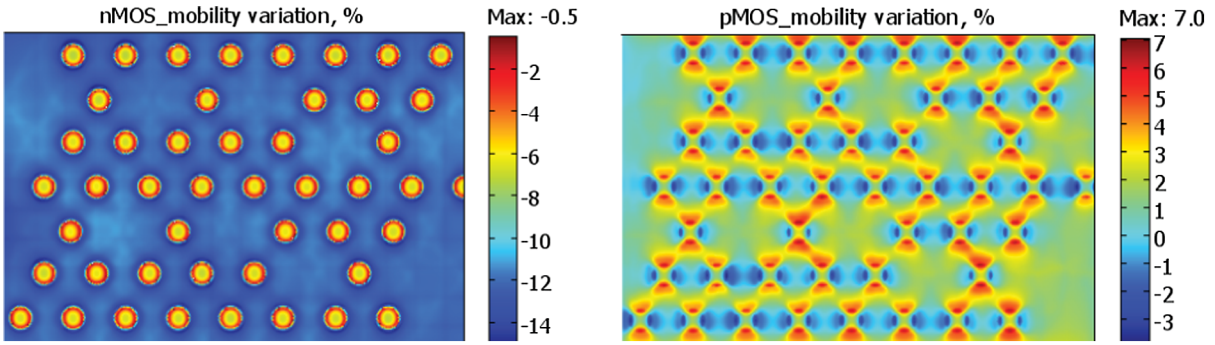



**Fig. 7** Mobility variation in nMOS and pMOS with ⟨110⟩ channel orientation caused by bump pattern.

## 3 Outline of the Calibration and Validation Procedures

For calibration of the parameters involved into the developed compact models, the measured MOSFETs drain current should be used. Stress-induced variations of drain current are assumed to be caused by mobility variation, which is related to stress by means of relation.[16] A representative set of transistors should be used for calibration, which means that the chosen transistors should be located on different distances from the die edges, corners, TSVs, and bumps. The drain current $I_d$ can be measured and simulated either in linear or in saturation regimes. As it was mentioned in the previous section, the 3-D stacking induced stress is not the only cause of $I_d$ variations. Layout-proximity effects (e.g., stress induced by built-in stressors, process/litho variations, wall-proximity effect, etc.) are able to introduce substantial variations of the device characteristics and should be treated properly. Therefore, the difference of the measured current $I_d^{\text{meas}}$, and the current calculated with the foundry calibrated SPICE model $I_d^{\text{SPICE}}$ should be used for calibration: $\Delta I_d^{\text{meas}} = I_d^{\text{meas}} - I_d^{\text{SPICE}}$.

The predicted values of the current variations $\Delta I_d^{\text{simul}}$ caused by the 3-D stress in CMOS channels are related to the simulated stress components by the expression similar to Eq. (16):

$$\frac{\Delta I_d^{\text{simul}}}{I_d^{\text{SPICE}}} = -(\pi_x^I \sigma_x + \pi_y^I \sigma_y + \pi_z^I \sigma_z), \tag{18}$$

where the coefficients $\pi_i^I$ are related to piezoresistance coefficients $\pi_i$ by means of special transformation matrix that can be obtained by SPICE simulations.[7] Then, the calibration engine tunes all the model parameters by minimizing the target function:

$$\text{TF} = \sqrt{\sum (\Delta I_d^{\text{meas}} - \Delta I_d^{\text{sim}})^2}.$$

The calibration results demonstrated below were obtained with the measurements done on a single-chip package where a 100-$\mu$m-thin silicon chip was mounted on the package substrate using FC bumps. Since this package did not contain TSVs, only the die warpage- and bump pressure-related effects were considered. The thicknesses and properties of the involved materials are listed in Tables 1 and 2.

Figure 8 demonstrates the schematics of the structures used for the model calibration. They contain FC bump arrays located: Fig. 8(a) shows near the die edge and Fig. 8(b) shows near the die corner. MOSFET devices, located at different distances from FC bumps, were used for calibration. All these devices represent identical p-type transistors, having $1 \times 1$ $\mu$m channel sizes. The identity of transistor sizes and the identical neighboring layouts have allowed us to exclude layout-induced variations in mobility, and thus to avoid the issues related to the accuracy of SPICE models of layout-proximity effects. The calibration results are demonstrated in Fig. 9. The relative changes of the drain current [i.e., $\%\Delta I_d = (\Delta I_d/I_d) \cdot 100\%$] for the selected p-type transistors are in the range of ~(3 to 7)% for the die-edge region, and ~(−3 to 3)% for the die corner. These changes are caused by the demonstrated lateral stresses (having average values ~200 MPa). The effect of vertical stress component is much smaller due to small value of the corresponding piezoresistance coefficient: for the considered ⟨110⟩ orientation of the channels the values of the piezoresistance coefficients are $\pi_x = 0.718$, $\pi_y = -0.663$, $\pi_z = -0.011$ [in GPa$^{-1}$)]. The smaller values of the current variation in die corner region are due to higher values of $y$-component of stress tensor, which reduces the current and for some transistors can surpass the positive effect of the $x$-component. The obtained correlation between the measured and simulated values clearly demonstrates the calibration capability of the developed compact-model-based simulation flow. Fit of the model predictions to the measurements for all transistors, demonstrated in Fig. 9(c), is characterized by mean-square deviation of ~0.5% and maximal deviation of ~1% (in terms of relative change of the drain current). This corresponds to the stress prediction accuracy of ~15 MPa.

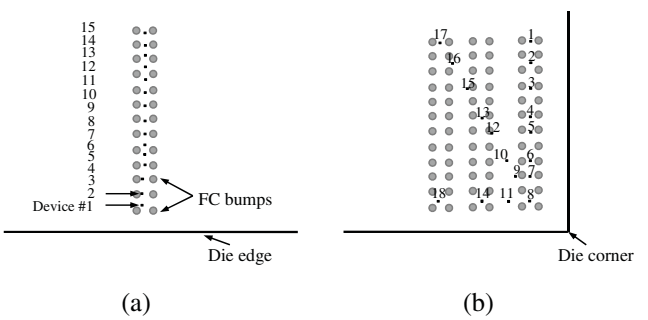



**Fig. 8** Schematics of the test structures used for model calibration: die edge (a), and die corner (b) regions.

## 4 Hot-Spot Analysis

Developed model, once calibrated with the test-chip data from a particular foundry and a technology node, can be employed for analyzing any design of the chips as long as the foundry manufacturing process and technology node are unchanged. It addresses the transistor-to-transistor variation of electrical characteristics caused by 3-D IC technology-induced stresses. However, full-chip detailed analysis of each transistor performance can be expensive in terms of time and CPU resources. So, it seems reasonable to have a capability of fast hot-spot analysis, or in other words, a capability for fast detecting the devices or circuits for which the CPI/TSV-induced stress impact can destroy the nets functionalities. Potential impact of the most stressed regions can be realized by means of the hot-spot analysis, consisting of the following steps:

1. For each circuit in the design (IO, digital, analog, etc.), the corresponding failing criteria are specified, i.e., the maximal acceptable current variation $(\Delta I_d/I_d)_{\text{threshold}}$ is provided as an input parameter.
2. A coarse grid covering the die is constructed. The bin size of the grid is comparable with the characteristic size of the analyzed circuits.
3. For each bin, the maximal value of the CPI/TSV stress is obtained using the calculated stress distribution [Eq. (9)].
4. Maximal current variation $(\Delta I_d/I_d)_{\max}$ is determined for each bin. If this variation exceeds the threshold, i.e., $(\Delta I_d/I_d)_{\max} > (\Delta I_d/I_d)_{\text{threshold}}$, for any type of circuits presented in this bin, then the bin is recognized

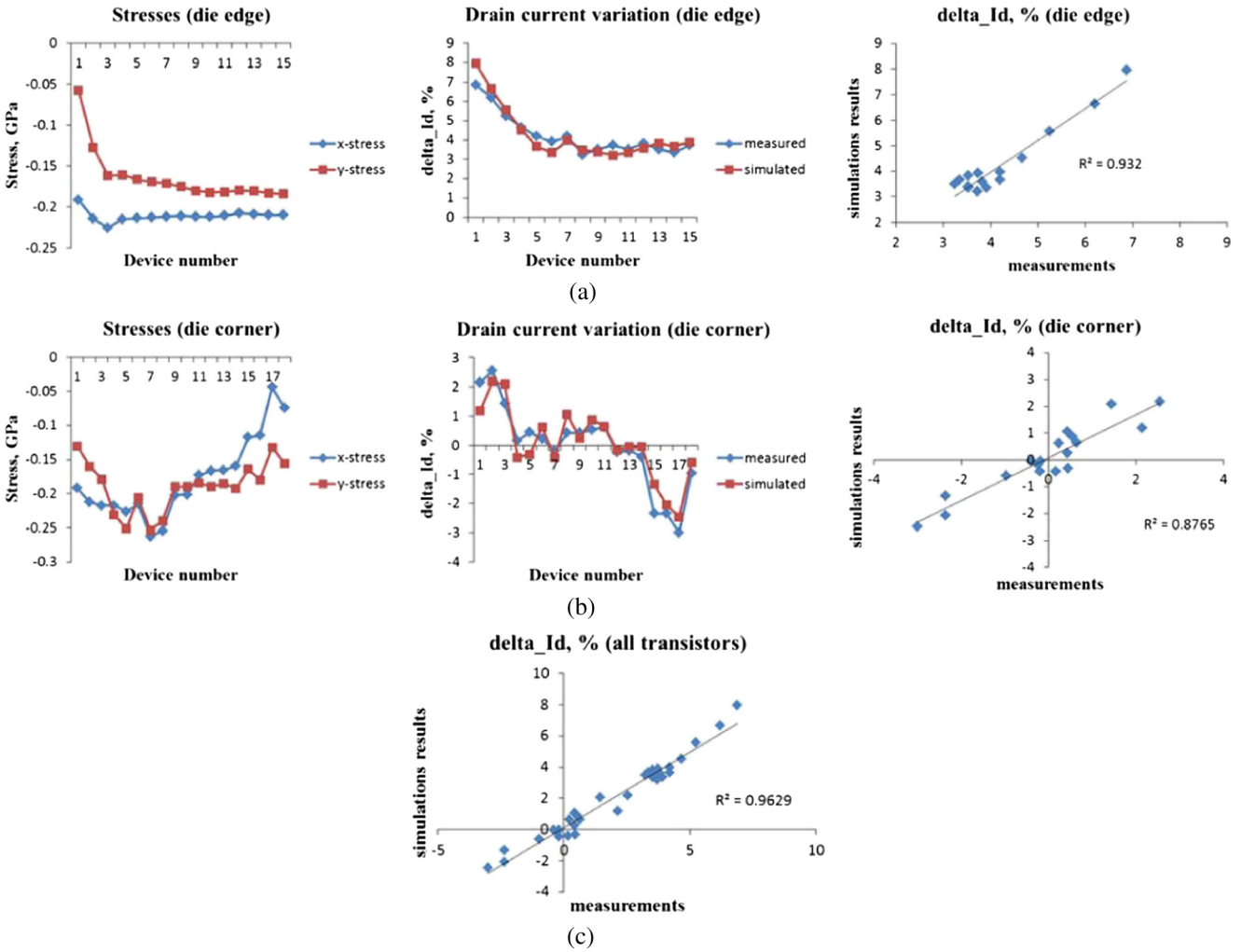



**Fig. 9** Model calibration results: 3-D stress components and the induced drain current variations for p-type transistors, located: (a) in the die edge region; (b) in the die corner region; (c) overall fit of the model prediction to measurements.

as a hot spot and can be visualized by means of color maps as displayed in Fig. 10.

5. A detailed analysis of stress impact on the device/circuit characteristics should be performed for this bin as it was described in Secs. 2.5 and 2.6. Figure 11(a) depicts visualization of detailed analysis results. For the highlighted transistor, the calculated mobility variation MULU0 is shown, as well as the stress components induced by TSV, CPI (denoted as "packaging"), and the total stress (sigma). Figure 11(b) demonstrates the obtained results when the methodology of Secs. 2.5 and 2.6 was employed for the analysis of the layout-induced stress effect on the device performance: the developed capabilities have allowed to

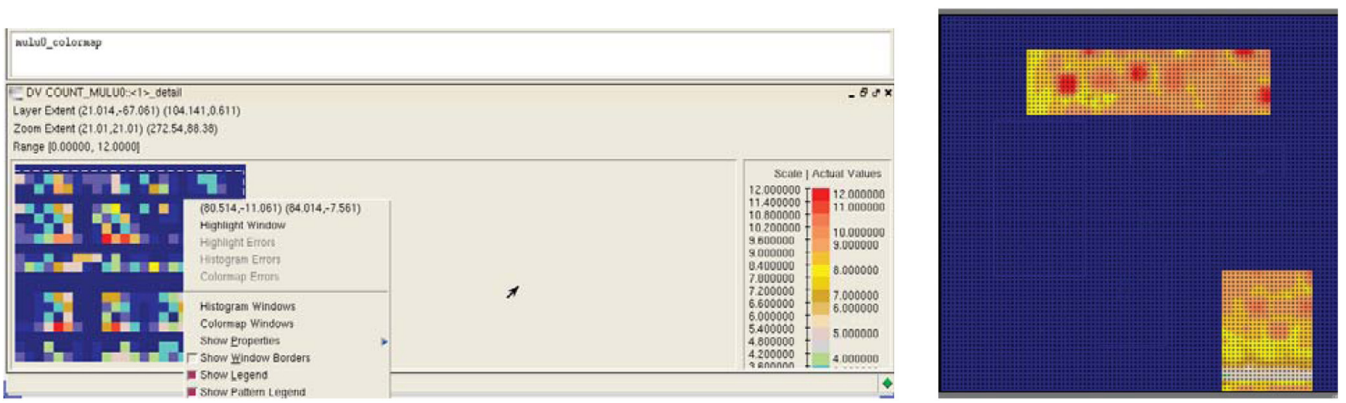
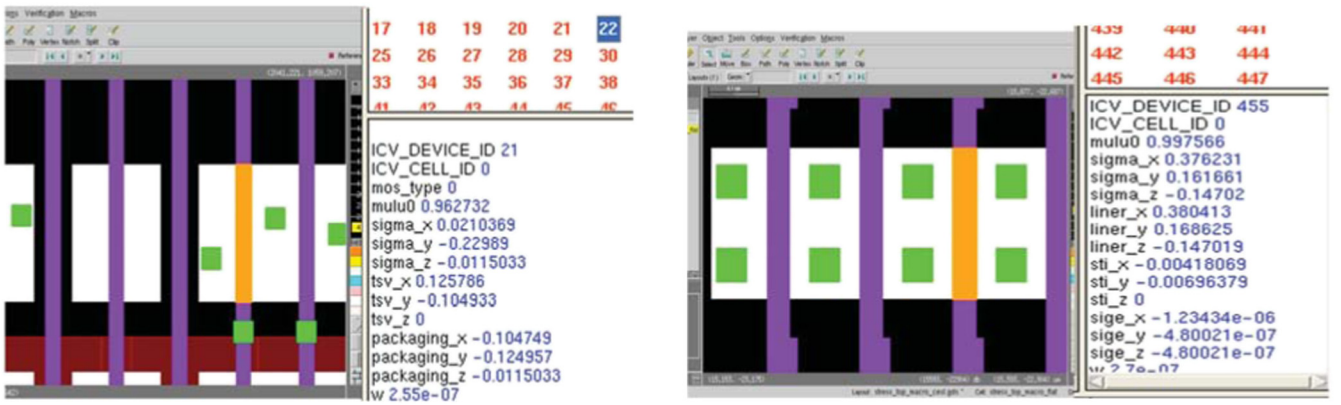



**Fig. 10** Visualization of the hot-spot bins locations at the die layout.




(a) (b)

**Fig. 11** Results of detailed analysis for the highlighted transistors: variations in mobility and stresses caused by: 3-D IC packaging, (a), and built-in layout sources, (b).

calculate Id variations caused by different stress sources like STI, $Si_{(1-x)}Ge_x$, CESL, etc.

## 5 Conclusions

The paper describes the developed multiscale simulation methodology and flow for the assessment of the stress-induced performance variation in 3-D IC chips. The proposed approach allows the linkage between the package-scale FEA formats and the chip layout formats by means of the developed physics-based compact models. The accuracy of the stress prediction is enhanced by developing and implementing the compact models for the effective thermo-mechanical properties of the composite BEoL interconnect and BRDL layers. Finally, the compact model of the device-scale stress relaxation allows taking into account every individual transistor and its neighboring layout content. Depending on the design stage where the analysis is made, the simulation results can be used for hot-spot analysis, or for back annotating the SPICE netlist with instant parameter MULU0 allowing more accurate circuit simulations can be done.

### *Acknowledgments*

The authors would like to thank Drs. Riko Radojcic and Mark Nakamoto of Qualcomm for numerous target-oriented discussions.

### *References*

1. R. Radojcic, M. Nowak, and M. Nakamoto, "TechTuning: stress management for 3D through-Si-Via stacking technologies," in *AIP Conf. Proc.*, Albany, New York, Vol. 1378, pp. 5–20 (2011).
2. G. Van der Plas et al., "Design issues and considerations for low-cost 3D TSV IC technology," *IEEE J. Solid State Circuits* **46**(1), 293–307 (2011).
3. S. Borkar, "3D Integration for energy efficient system design," in *Proc. IEEE Symp. VLSI Technology*, Digest of Technical Papers, pp. 58–59 (2009).
4. K. N. Tu, "Reliability challenges in 3D IC packaging technology," *Microelectron. Reliability* **51**(3), 517–523 (2011).
5. L. Mercado et al., "Impact of flip-chip packaging on copper/low-k structures," *IEEE Trans. Adv. Packag.* **26**(4), 433–440 (2003).
6. S. Flachowsky et al., "Understanding strain-induced drive-current enhancement in strained-silicon n-MOSFET and p-MOSFET," *IEEE Trans. Electron Devices* **57**(6), 1343–1354 (2010).
7. K. Yamada et al., "Layout-aware compact model of MOSFET characteristics variations induced by STI stress," *IEICE Trans. Electron.* **E91-C** (7), 1142–1150 (2008).
8. C.-Y. Cheng et al., "New observations in LOD effect of 45-nm P-MOSFETs with strained SiGe source/drain and dummy gate," *Trans. Electron Devices* **56**(8), 1618–1623 (2009).
9. E. Suhir, "Predicted thermally induced stresses in, and the bow of, a circular substrate/thin-film structure," *J. of Appl. Phys.* **88**(5), 2363–2370 (2000).
10. E. Suhir, "An approximate analysis of stresses in multilayer elastic thin films," *ASME J. Appl. Mech.* **55**(1), 143–148 (1988).
11. S. M. Hu, "Stress-related problems in silicon technology," *J. Appl. Phys.* **70**(6), R53–R80 (1991).
12. R. M. Jones, *Mechanics of Composite Materials*, Hemisphere Publishing Corporation, New York (1975).
13. A. Ivankovic et al., "Thermo-mechanical impact of the Underfill-micro-bump interaction in 3D stacked integrated circuits," in *IEEE 13th Electronic Packaging Technology Conf.*, pp. 34–38 (2011).
14. S. Timoshenko and J. N. Goodier, *Theory of Elasticity*, McGraw-Hill, New York (1952).
15. K. L. Johnson, *Contact Mechanics*, Cambridge University Press, UK (1985).
16. S. K. Ryu et al., "Impact of near-surface thermal stresses on interfacial reliability of through-silicon-vias for 3-D interconnects," *IEEE Trans. Device Mater. Reliab.* **11**(1), 35–43 (2011).
17. V. Sukharev et al., "3D IC TSV-based technology: stress assessment for chip performance," in *AIP Conf. Proc.*, Vol. 1300, pp. 202–213 (2010).
18. V. Sukharev et al., "Multi-scale simulation methodology for stress assessment in 3D IC: effect of die stacking on device performance," *J. Electron. Test.* **28**(1), 63–72 (2012).
19. C. Smith, "Piezoresistance effect in germanium and silicon," *Phys. Rev.* **94**(1), 42–49 (1954).
20. M. Dunga et al., *BSIM4.6.0 MOSFET Model*, University of California, Berkeley (2006).
21. S. Thompson et al., "Uniaxial-process induced strained-Si: extending the CMOS roadmap," *IEEE Trans. Electron Devices* **53**(5), 1010–1020 (2006).

**Armen Kteyan** received his MS degree in semiconductor physics and technology from Yerevan State University, Yerevan, Armenia, in 1994, and a PhD degree in solid state physics from the Institute of Radiophysics and Electronics (IRPhE) of Armenian National Academy of Sciences in 1998. The research activity at IRPhE was in the field of theory of dislocations in metals and semiconductors, and theory of superconductivity. Currently, he is a member of the technical research staff at Mentor Graphics and involved in development of physics-based models for DFM applications.

**Gevorg Gevorgyan** received his MS degree in computer science in 2012 from the Yerevan State Engineering University, Yerevan, Armenia. Currently, he is an engineer in Mentor Graphics Corporation and is involved in software development for EDA applications.

**Henrik Hovsepyan** received his MS degree in computer science from Yerevan State University, Yerevan, Armenia. He held senior technical positions at Ponte Solutions, Mountain View, California, working on physical design/verification, critical area analysis. Since 2008, he has been with Mentor Graphics Corporation, Fremont, California, as a part of the Ponte Solutions acquisition, where he was a lead engineer. His major research activity is related to the development of new

full-chip modeling and simulation capabilities for the semiconductor processing and DFM applications.

**Jun-Ho Choy** is an active staff engineer of Mentor Graphics Corp., Fremont, California. He received his BS degree at Seoul National University, his MS at Sung Kyun Kwan University, and his PhD at Michigan Technological University, Houghton, Houghton, MI, all in the field of metallurgical and materials engineering. He held the positions of device engineer in the Memory R&D Division at Hynix Semiconductor, Inc., and a principal engineer at LSI Logic Corp, Milpitas, California. Subsequently, he joined Ponte Solutions, Inc. (later acquired by Mentor Graphics), and participated in the development of full-chip DFM/DFR EDA tools.

**Valeriy Sukharev** is a technical lead at the Design to Silicon Division of Mentor Graphics Corporation, Fremont, California. He leads research and development of new full-chip modeling and simulation capabilities for the semiconductor processing and DFM/ DFR applications. Prior to Mentor Graphics, he was a visiting professor with Brown University, Providence, Rhode Island, and a guest researcher with NIST, Gaithersburg, MD. He held senior technical positions at LSI Logic Advanced Development Lab, Milpitas, California. He holds a PhD in physical chemistry from the Russian Academy of Sciences.