

Stress-induced Effects Caused by 3D IC TSV Packaging in Advanced Semiconductor Device Performance



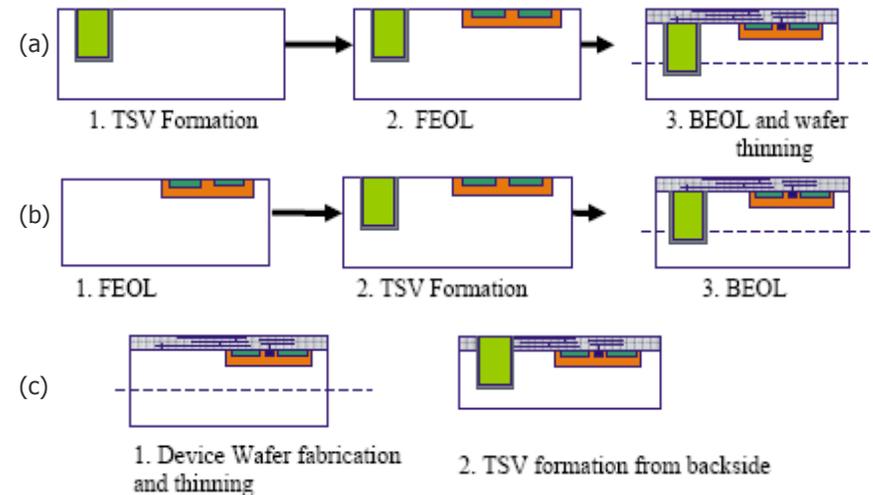
**V. Sukharev^a, A. Kteyan^a, J-H Choy^a,
H. Hovsepyan^a, A. Markosian^a, E. Zschech^b,
and R. Huebner^b**

*^aMentor Graphics Corporation, 46871 Bayside Parkway,
Fremont, CA 94538, USA*

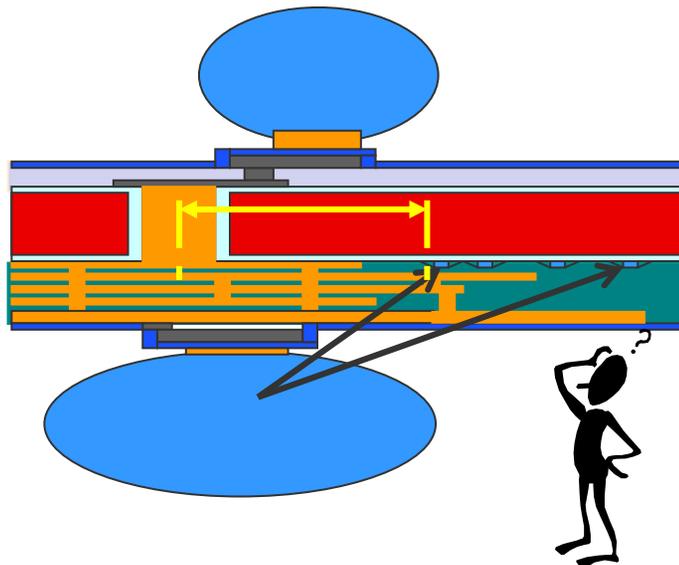
*^bFraunhofer Institute for Non-Destructive Testing, Maria-
Reiche-Strasse 2, 01109 Dresden, Germany*

New Technology Brings "New" Stress

- Number of process steps employed by 3D IC technology such as: wafer/die thinning, wafer/die mounting, chip stacking, TSV drilling and filling, solder ball solidification, etc. acts as additional stress sources that can affect the chip-stack performance.
- While wafer thinning is accompanied by a relaxation of preexisting internal stress the additional warpage-induced stress load generated by a thin die mounting on a wafer or on another die can easily propagate from the surface to the underneath device layer.



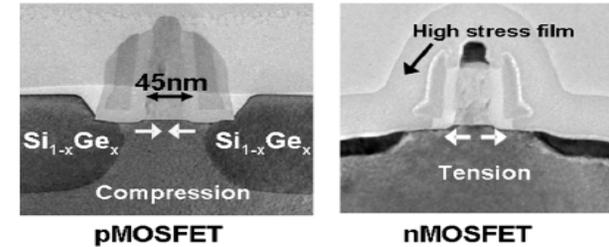
Schematics of examples of TSV construction methods, as it is shown in JEDEC Publ. 158: (a) Via-first, (b) Via middle, (c) Via-last



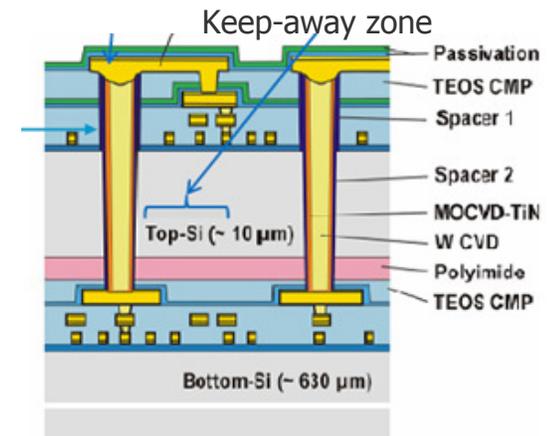
- Effects of stress generated by the difference in thermal expansions of TSV copper fill and silicon can be easily avoided by introducing the so-called "keep-out" regions around every TSV where no devices should be placed.
- The effective CTE of the silicon die is higher when it contains copper in the form of filled TSVs. As a result, on temperature cycling, the die expansion results in an enhanced loading on the solder joint.
- Stress generated in the devices right under the solder joints could modify device characteristics and should be accounted.

Effect of 3D on Pre-Engineered Stress

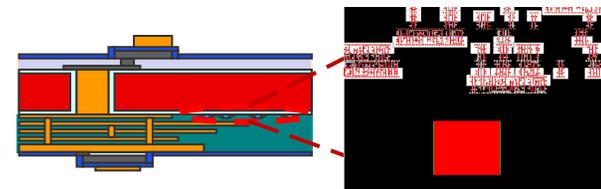
- The major device characteristics such as mobility and threshold voltage could be affected by stress
- Traditionally a number of strain engineering techniques are used for boosting the chip performance. Engineered stress sources, such as the capping stressed layer technology (CESL), epi-Si_{1-x}Ge_x structures confined to the S/D regions, a variety of stress memorization techniques, etc. should generate the needed amount of stress exactly to the targeted gate channel.
- Layout-induced stress variation makes this target too optimistic. In addition to the stress variation caused by variations in the transistor size and shape, a long-range character of the stress propagation makes a prospective gate-to-gate stress variation even more pronounced.
- In the case of 3D TSV-based technology an additional inside transistor stress variation caused by a die stacking should be accounted.
- Hence, there is a need in a simulation methodology/flow that should be physics-based and includes an interface to layout formats (GDS, OASIS, etc) which contain an entire die layout and which can be linked with package-scale simulation models (FEA).



S. E. Thompson *et al.*, *IEEE Electron Device Lett.*, vol. 25, pp. 191, 2004.

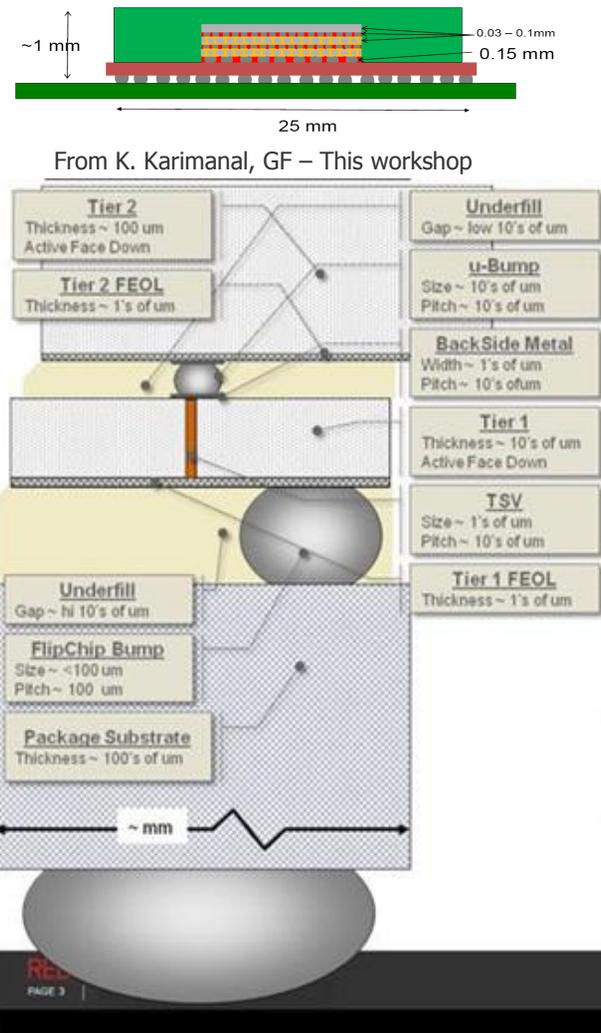


Rui Huang, Paul S. Ho, et al., 11th International Workshop on Stress-Induced Phenomena in Metallization - Dresden, 2010



DFM-type Methodology for Managing Mechanical Stresses

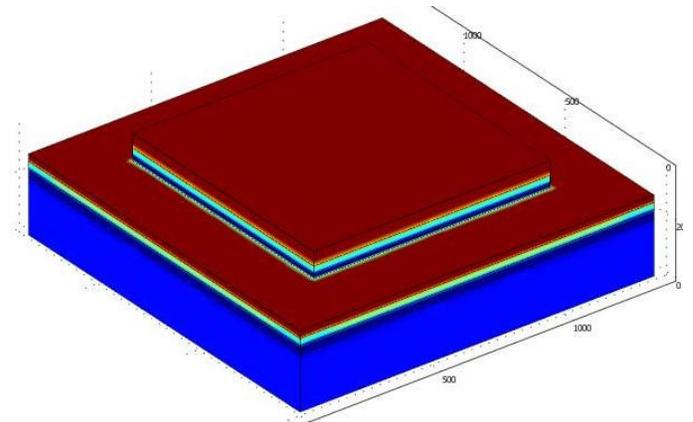
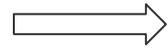
- Traditional TCAD tools based on FEA cannot be employed for a simulation of transistor channel stress distribution across a die, due to the size of a model, which can easily reach hundreds of millions degrees of freedom, and due-to the multi-scale character of the simulation problem.
- Empirical modeling cannot take into account CPI-induced variations in transistor characteristics. Because of the lack of a physical basis, this kind of modeling cannot provide a link to the physics-based package-scale simulation in order to accept a CPI-induced stress load.
- Hence, there is a need in a simulation methodology/flow that should be physics-based and includes an interface to layout formats (GDS, OASIS, etc) which contain an entire die layout and which can be linked with package-scale simulation models (FEA).
- Mentor Graphics has developed a new physics-based, DFM like methodology for calculation of the gate-to-gate variation of stress generated by chip layout and chip stacking (CPI).



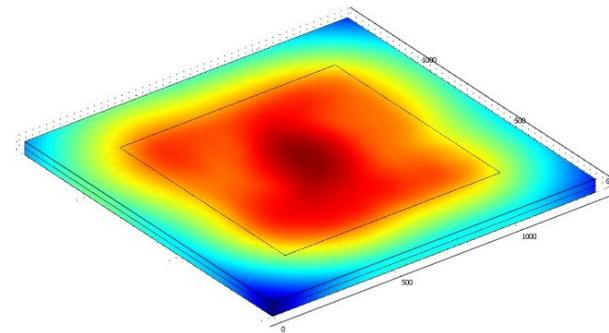
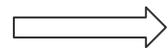
Qualcomm's example of the 3D ICTSV based technology stack

DFM-like Methodology for Calculation of the Gate-to-Gate Variation of Stress

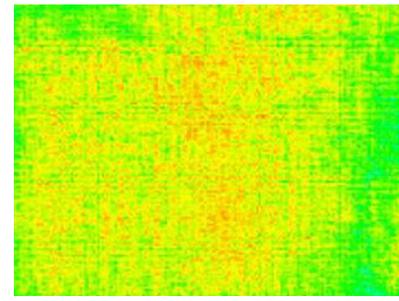
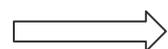
Package-scale simulation (FEA)
Input: geometry; material properties; smeared mechanical properties for RDLs, Silicon/TSV bulk, interconnect.
Output: field of displacement components on the die faces.



Die-scale simulation (FEA)
Input: geometry; field of displacements on the die faces; coordinate-dependent mechanical properties for RDLs, Silicon/TSV bulk, interconnect.
Output: Distribution of the strain components across device layer.

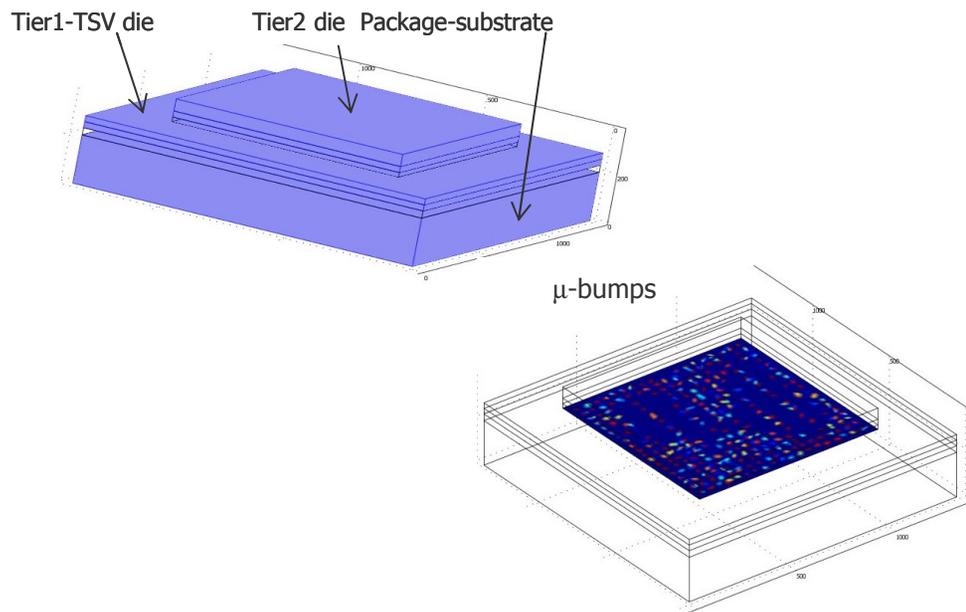


Layout-scale with feature-scale resolution (compact model):
Input: GDS; distribution of the strain components across device layer.
Output: Transistor-to-transistor variation in stress components.

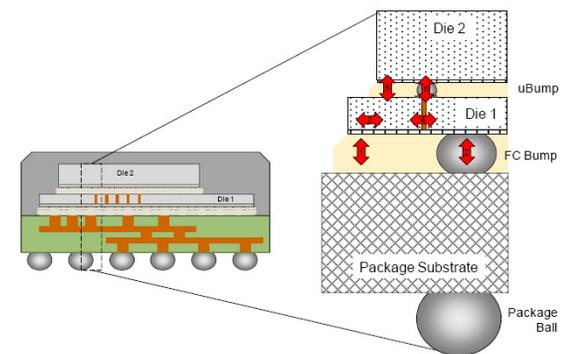


Linked Package/Die-Scale Simulations

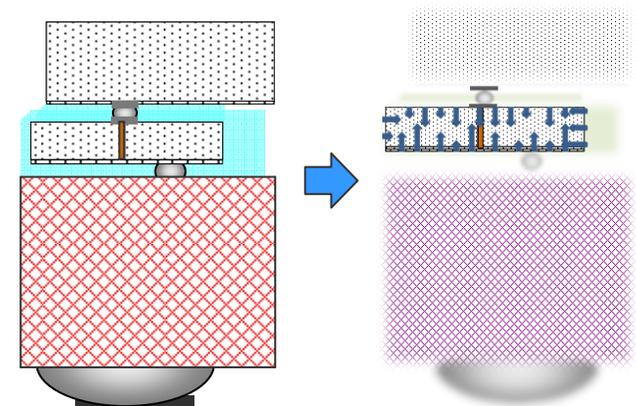
- The FEA-based package scale simulation tool is used for generating a set of BC describing packaging-induced load (displacement field) at the faces of already thinned die.
- At the package-scale simulation the interconnect layer, silicon substrate, interposer, etc. are approximated by smear layers with the effective mechanical properties representing the averaged characteristics of the composite material (for example, copper/low-k composite for interconnect).



“Package-scale simulation”

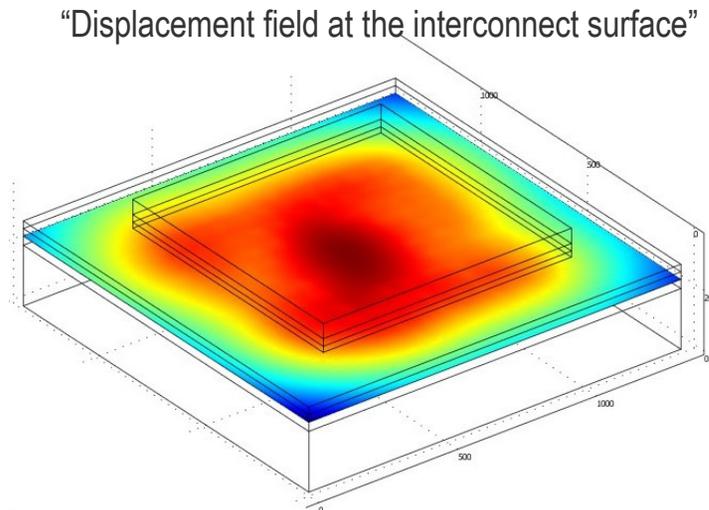


“Generation of the die-scale BC”



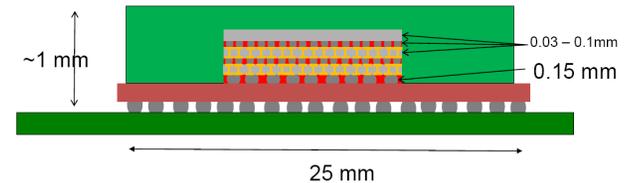
Package-Scale Simulation

- In order to evaluate the impact of packaging the entire packaging process should be considered. Both the die attach process as well as reliability cycling might be critical to generating stress in the device layer. Materials parameters have to be determined as input data for these simulations. Employing a different level of representation of the material properties (for example elasticity vs. plasto-elasticity or visco-elasticity or temperature dependent vs. constant Young's modulus and coefficient of thermal expansion (CTE)) can generate different accuracy of predicted strain/stress distributions.
- Simulation input:
 - Packaging geometry
 - Stress-free temperature
 - Thermal load and mechanical load
 - Mechanical properties of relevant materials.
- Simulation output:
 - Distributions of displacement components across the faces of the considered thin die.



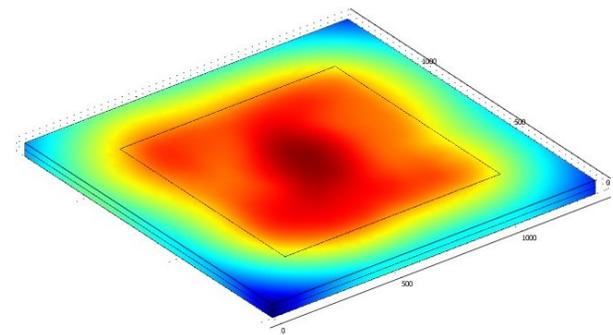
Die-Scale Simulation with Layout-Dependent Mechanical Properties

- Strain distribution across a device layer caused by packaging is calculated with the FEA tool by implementing the displacement BC.
- Interconnect, BRDL and thin silicon/TSV layers should be approximated as layers with a spatial distribution of elastic properties determined by their layouts.
- A calculation methodology of interconnect, BRDL and silicon/TSV effective Young modulus, CTE and Poisson factor as functions of metal density in all metal levels was developed based on the model of mechanical properties of unisotropic composite materials (rules of mixtures).
- Position dependent mechanical characteristics of interconnect are represented in a format readable by FEA tool.
- FEA tool calculates distributions of strain components across device layer.

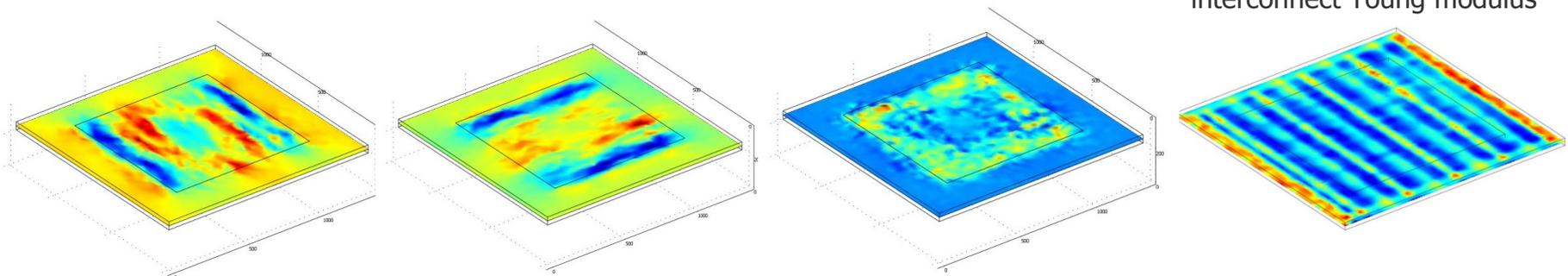


From K. Karimanal, GF – 3D Stress workshop, Santa Clara, 2011

Z-displacement load distributions at the die surfaces



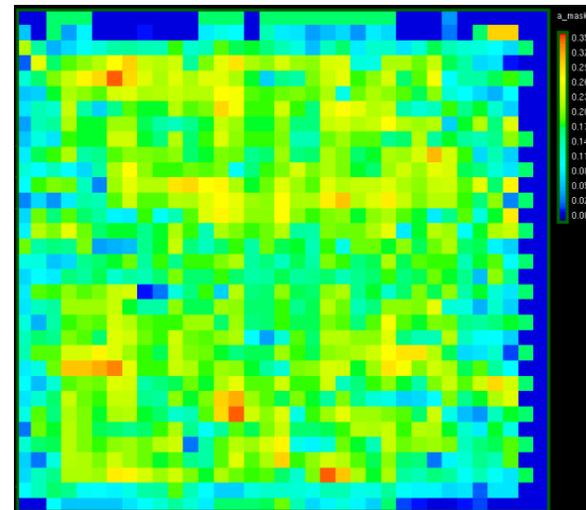
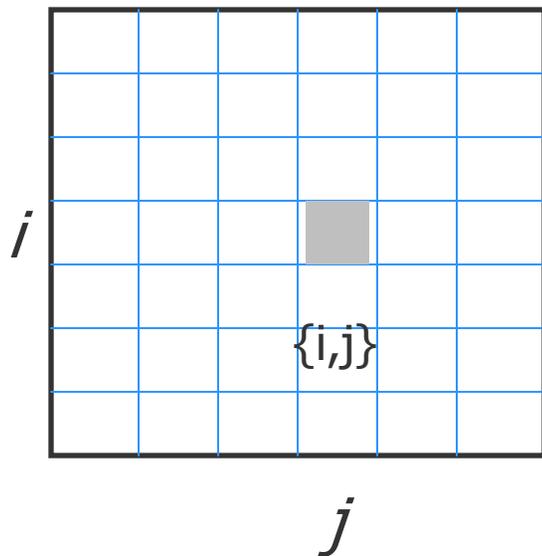
Across die distribution of interconnect Young modulus



Distributions of strain x, y and z-component across device layer

Layout-dependent mechanical properties - Calculation Method

- Bin-base approach is applied to define average metal density $\rho_M^{\{i,j\}}$
- Our tool extracts metal densities from layout file (GDSII, Oasis, etc)

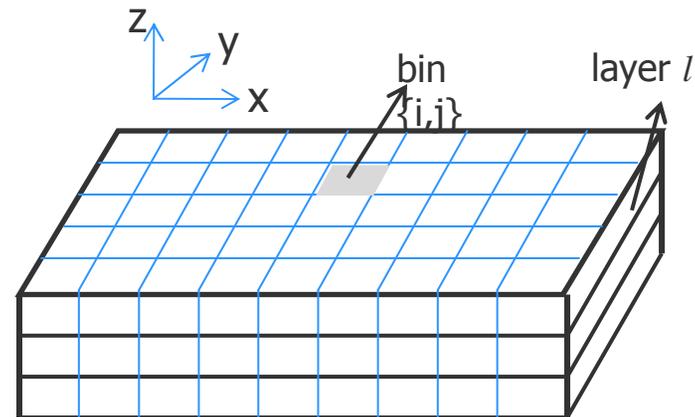


Young's Modulus: Employed Formalism (In-plane Components)

- For each bin of each layer of interconnect, depending on routing direction:

$$E_{||}^{\{i,j\}l} = E_M \rho_M^{\{i,j\}l} + E_D (1 - \rho_M^{\{i,j\}l})$$

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

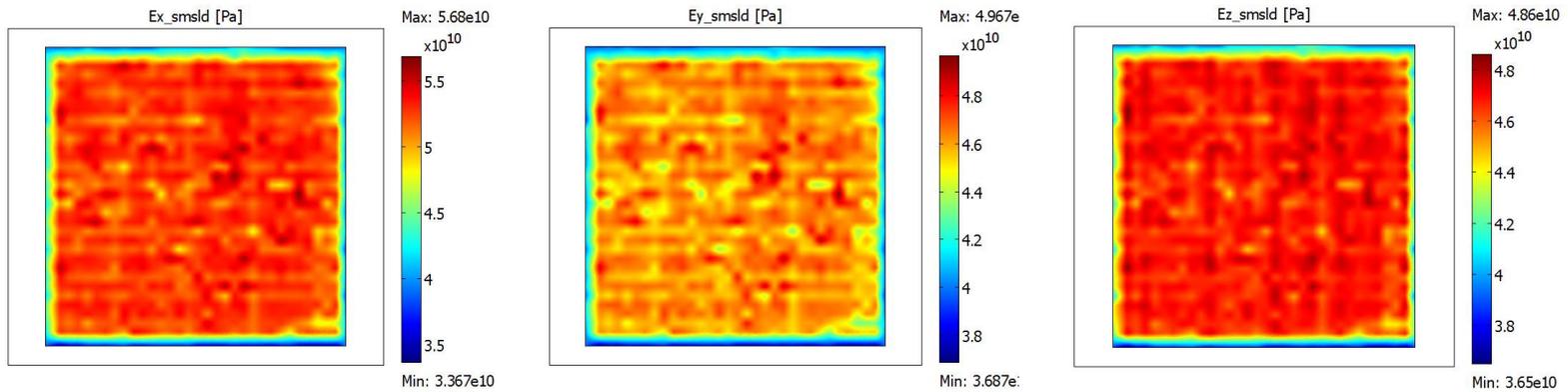


- E_{\perp} expression is used for via layers
- For the bin $\{i,j\}$ - for whole interconnect:

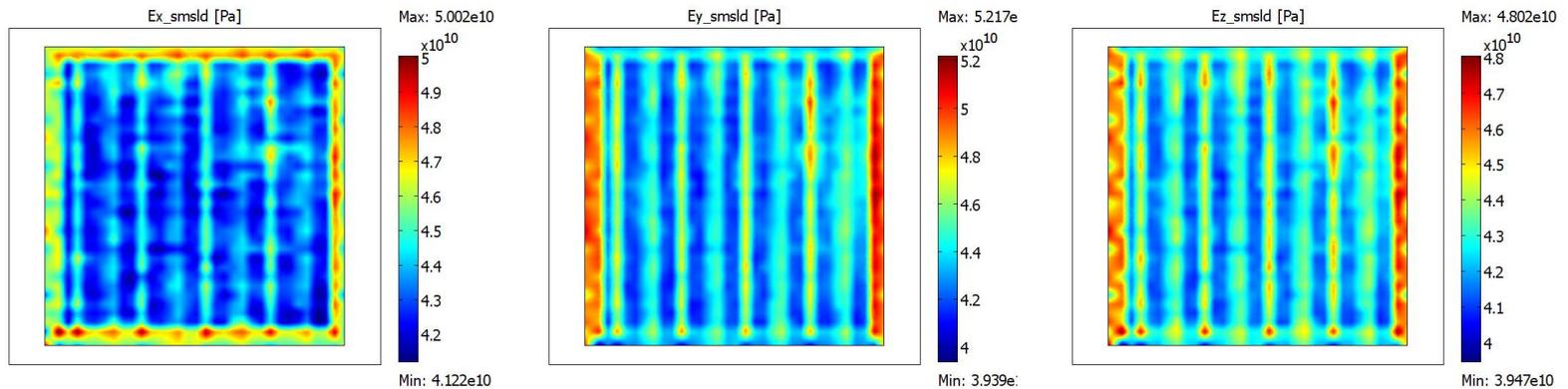
$$E_x^{\{i,j\}} = \frac{\sum_l E_x^{\{i,j\}l} h^l}{\sum_l h^l}, \quad E_y^{\{i,j\}} = \frac{\sum_l E_y^{\{i,j\}l} h^l}{\sum_l h^l}$$

Results: Visualization of E Components

Young's modulus, layer M1



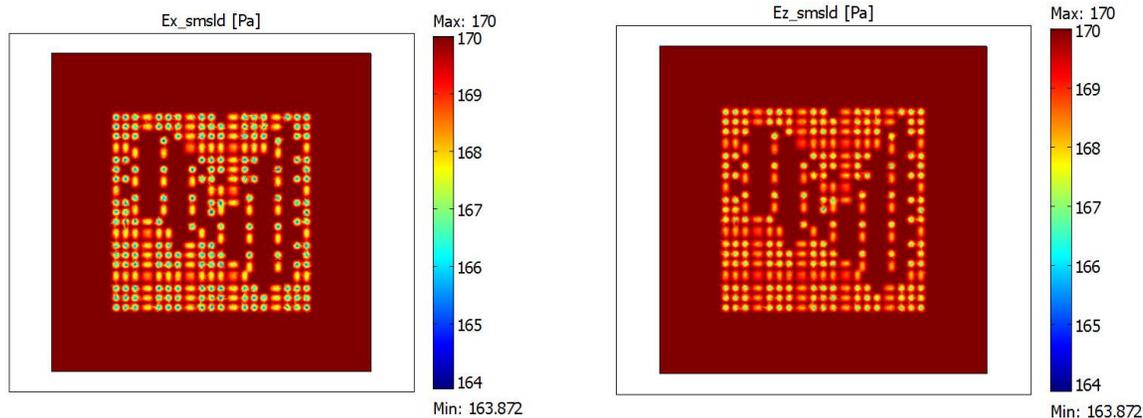
Young's modulus, averaged over all layers:



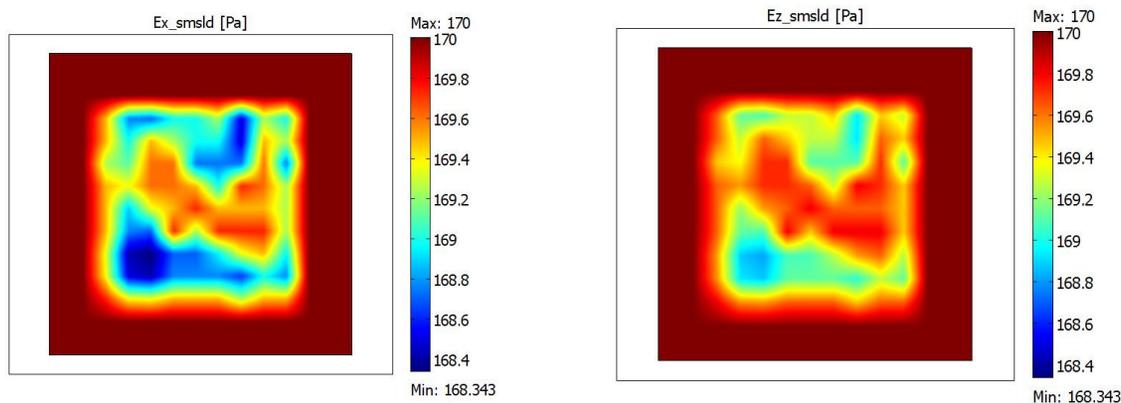
Si Bulk Properties: Impact of TSV Density

Visualizations for Different Bin Sizes

- Young's modulus: TSV diameter – 6 μm , min. spacing – 40 μm



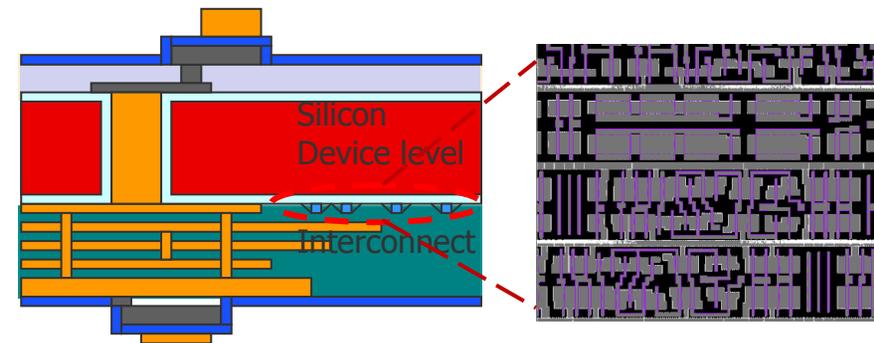
Bin size – 20 μm



Bin size – 100 μm

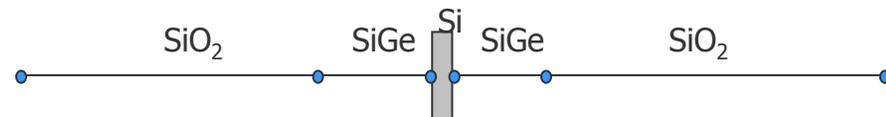
Package-Induced Stress Inside Transistor Channel

- Across-die distribution of strain components calculated at the previous step didn't take into account a composite nature of the device layer (silicon islands-diffusions, STI, contacts, poly, etc).
- Traditional methods such as FEM, FED etc. cannot be employed for simulation of intra-channel stress distribution across the die.
- A complexity of the layout (billions of transistors) cannot be captured by FEM-based simulation when redistribution of stress caused by composite nature of a device layer should be accounted.
- New approximate methodology for calculation of gate-to-gate variation of stress generated by packaging should be developed/adopted

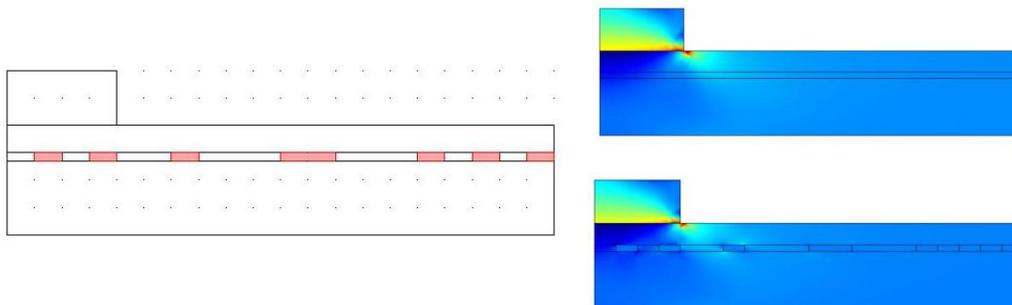
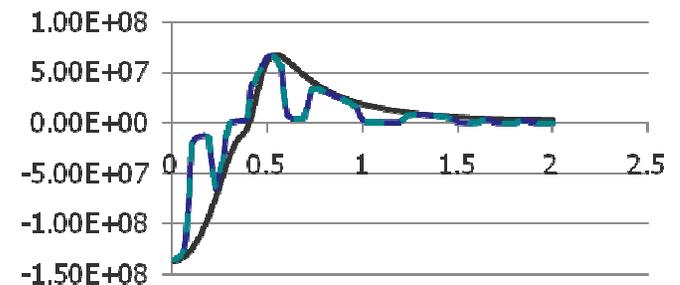


Example of cutline

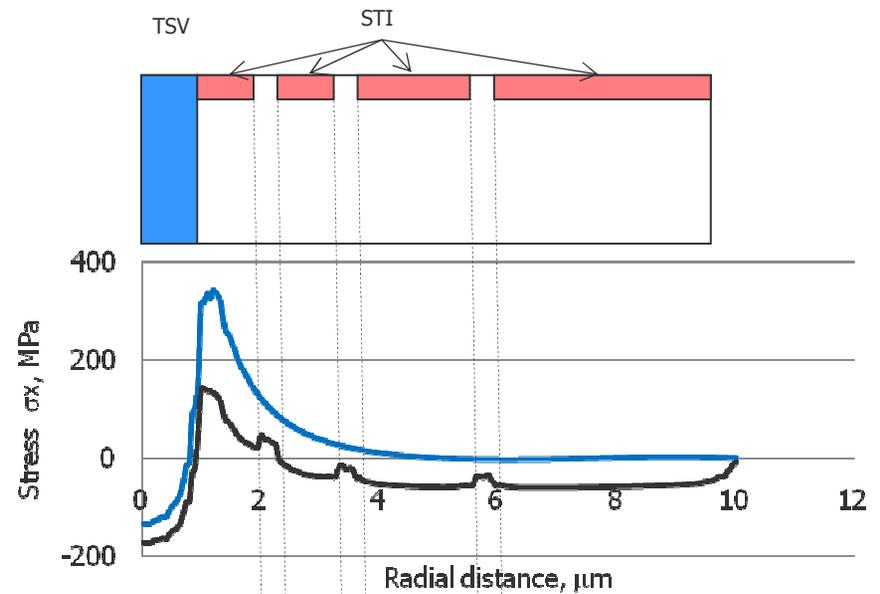
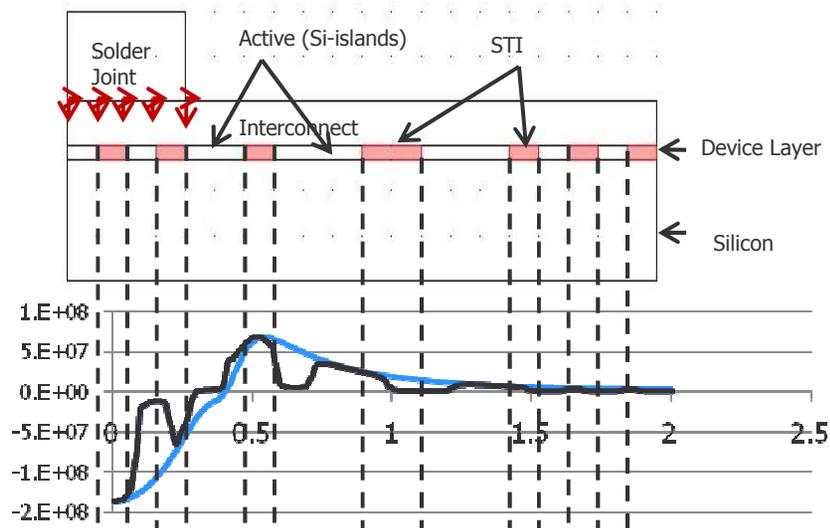
```
{{-1000,"SiO2"}, {-132,"SiGe"}, {-24,"Si"}, {24,"SiGe"}, {132,"SiO2"}, {1000,"end"}}
```



Relaxation of the initial strain ϵ_0 in each segment creates additional displacements. These displacements depend on mechanical properties of materials in cutline as well as on substrate traction.



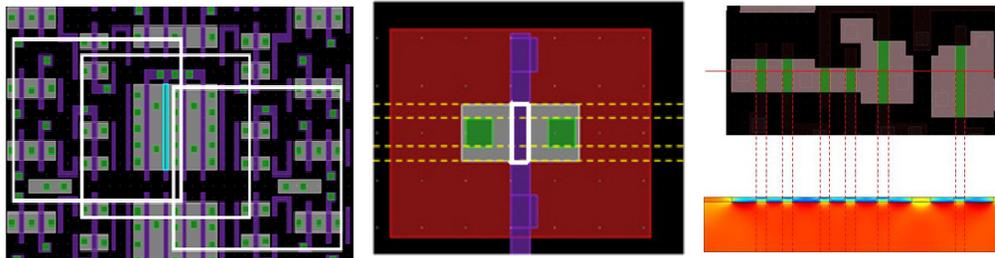
Effect of Composite Nature of Device Layer



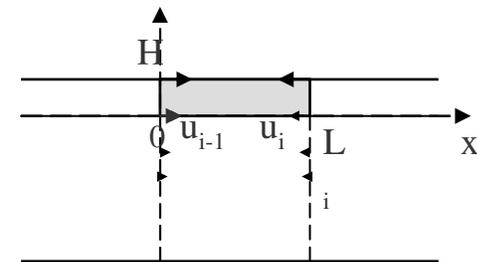
Example of the distribution of longitudinal stress along a cut-line in the device layer when its composite structure (-STI-active-STI-) was accounted (black line) and was not (blue line).

Cut-Layer Based Model for Package-Induced Strain

- By interpolating the FEA-generated strain, the average “initial” package-induced strain ε_i^{CPI} is determined inside each segment of the transistor layer.
- For the calculation of the stress components inside a transistor channel the tool generates the cut-lines which cross the transistor channel and define “2D” structures along these lines within the corresponding window
- Due to difference in mechanical properties of segments in the transistor layer, the boundaries between segments experience displacements – stress relaxation. These displacements must be determined from the set of equations similar to one used for calculation of layout-induced stress: $E_i''u_{i-1} - (E_i' + E_{i+1}')u_i + E_{i+1}''u_{i+1} = E_i\varepsilon_i^{CPI} - E_{i+1}\varepsilon_{i+1}^{CPI}$, here, E_i'' and E_{i+1}' are the known functions of the materials properties; ε_i^{CPI} and ε_{i+1}^{CPI} are the initial strain in i -th and $(i+1)$ -th segments.



“Floating” windows - white square s(left). Each transistor channel (white frame) is divided into a number of segments by cut-lines (dashed lines) crossing the whole window (center). Cut-line view in vertical direction; dot lines indicate the channels (right).



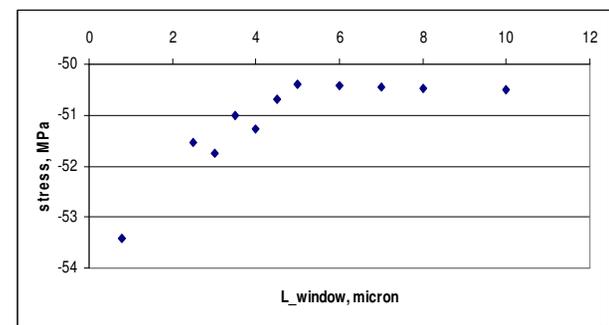
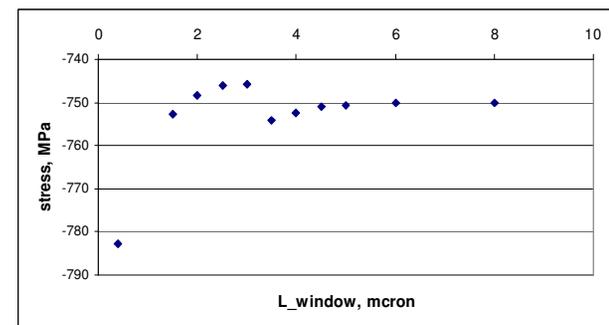
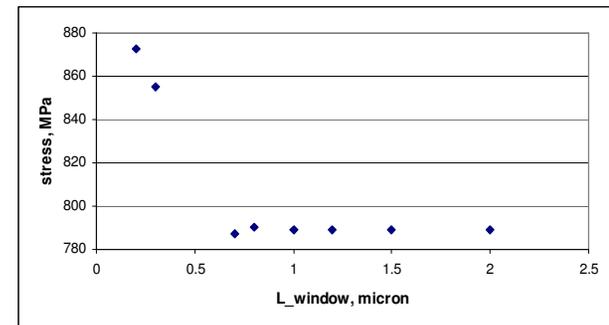
Schematics of the traction between cutline segments and substrate

Model-Based “Stress” Prototype

- Mentor full-chip EDA tool is capable of predicting stress everywhere in the layout caused by a variety of sources:
 - - stressed liners,
 - - epi-Si_{1-x}Ge_x structures confined to the S/D regions,
 - - tensile contacts,
 - - STI , TSV
 - - Packaging
- These sources are located inside a floating window surrounding each gate. The minimal size of the window is determined by a saturation in the dependency of stress in a channel vs. window size.
- Prototype generates the stress distribution using the approximate analytical solution (compact models) of the corresponding elasticity problems.

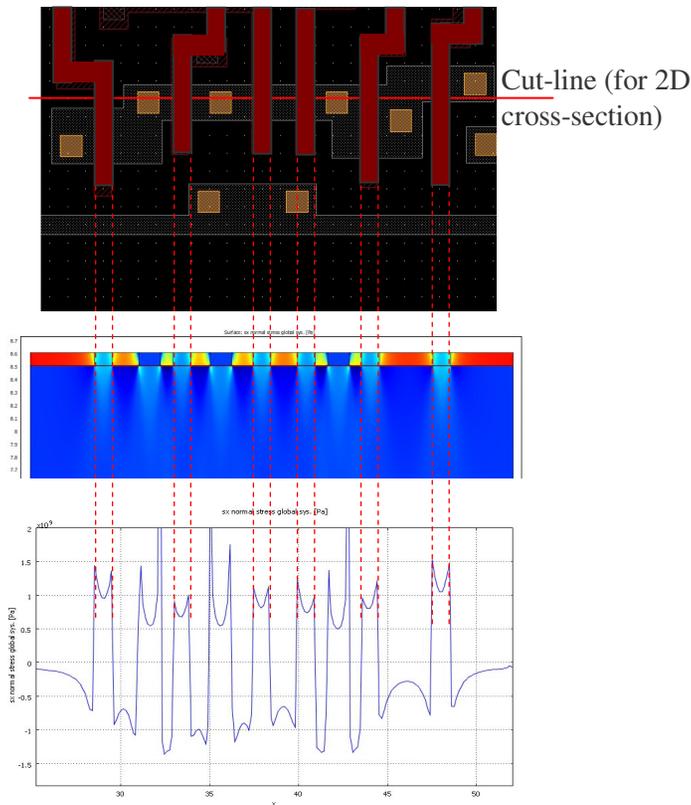
Based on the preliminary estimations for the 45 nm technology node the window size should be:

- ~1 μm for silicon nitride liner as a stress source
- ~4 μm for epi-Si_{1-x}Ge_x confined in S/D as a stress source
- ~5 μm for STI as a stress source

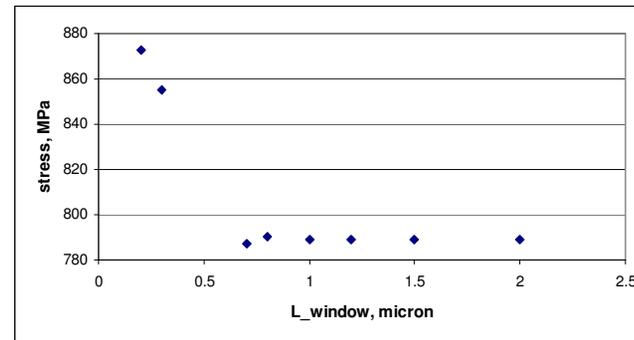


Liner

In a floating window around the n-gate the sequence of the following materials must be defined: Poly, contacts, and liner. For i -th material (with length L_i) we define displacements of the right edge as u_i . Every material is characterized by the parameters E_i and m_i . The “inherent” strain of the liner is e_0 , liner thickness is H_{liner}



Liner-induced stress vs. windows size

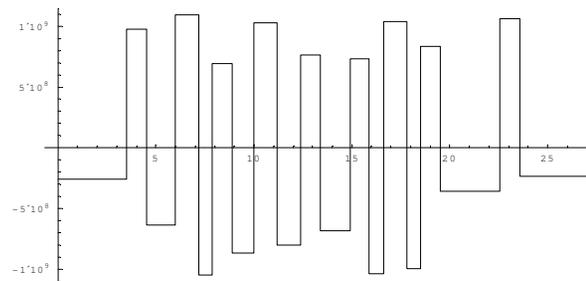


Based on the preliminary estimations for the 45 nm technology node the window size should be ~1 micron for silicon nitride liner as a stress source.

Comparison of results provided by the model and FEM simulations (liner)

Stress values in channels, 10^8 Pa

FEM simulations	10.5	7.8	9.3	8.4	9	11.6
Model	11	7.7	8.5	8.2	9.3	12

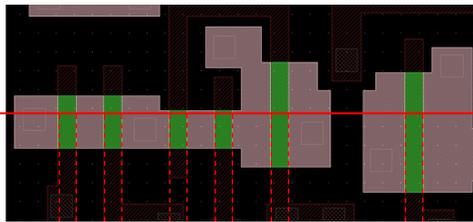


2D Analytical Model

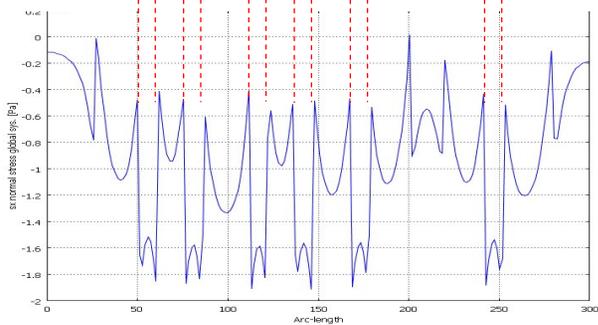
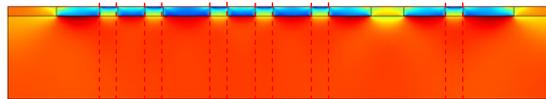
Epi-Si_{1-x}Ge_x

In a floating window around the p-gate the sequence of the following materials must be defined:

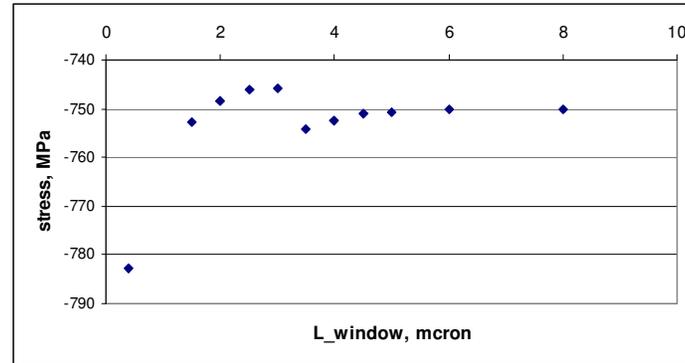
- In n-wells: gate is Si, active (without gate) is Si everything else is SiO₂
- In p-wells: gate is Si, active is Si_{1-x}Ge_x, everything else is SiO₂



Cut-line (for 2D cross-section)



epi-Si_{1-x}Ge_x - induced stress vs. windows size



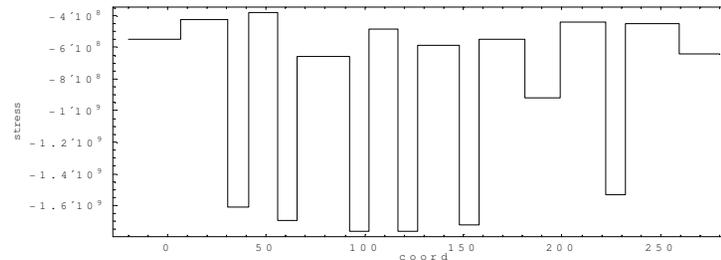
Based on the preliminary estimations for the 45 nm technology node the window size should ~4 micron for Si_{1-x}Ge_x as a stress source.

Comparison of results provided by the model and FEM simulations (Si_{1-x}Ge_x)

Stress values in channels, 10⁹ Pa

FEM simulations	-1.68	-1.74	-1.76	-1.75	-1.74	-1.67
Model	-1.61	-1.69	-1.76	-1.76	-1.72	-1.54

2D Analytical Model



Methodology for Stress Calculation

- Using the calculated displacements, strain in every segment can be obtained as:

$$\epsilon_i = \frac{u_i - u_{i-1}}{L_i} + \epsilon_i^{CPI}$$

$$\sigma_x = \frac{E_{Si}}{(1 + \nu_{Si})(1 - 2\nu_{Si})} [(1 - \nu_{Si})\epsilon_x + \nu_{Si}(\epsilon_y + \epsilon_z)]$$

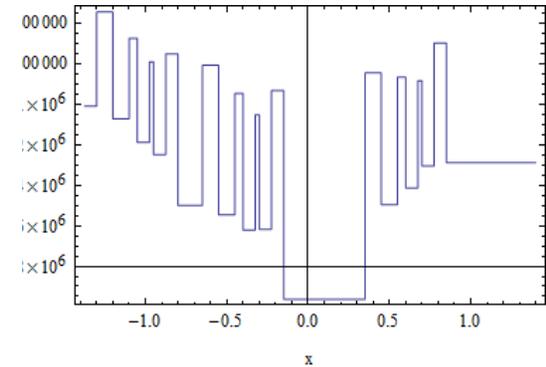
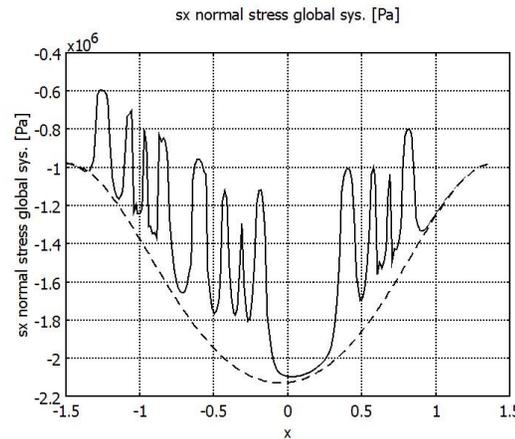
Stress is determined as:

$$\sigma_y = \frac{E_{Si}}{(1 + \nu_{Si})(1 - 2\nu_{Si})} [(1 - \nu_{Si})\epsilon_y + \nu_{Si}(\epsilon_x + \epsilon_z)]$$

CPI-INDUCED STRESS

$$\sigma_z = \frac{E_{Si}}{(1 + \nu_{Si})(1 - 2\nu_{Si})} [(1 - \nu_{Si})\epsilon_z + \nu_{Si}(\epsilon_x + \epsilon_y)]$$

A correspondence between the longitudinal stresses, caused by CPI, calculated with FEM tool (left) and with the developed compact model (right).



Simulation input:

- GDSII – layout for the device layer
- Across-layout distribution of the package induced strain components (output of the die-scale simulation)
- Mechanical properties of relevant materials .

Simulation output:

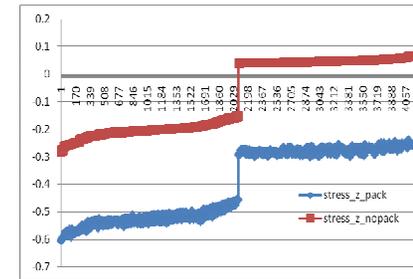
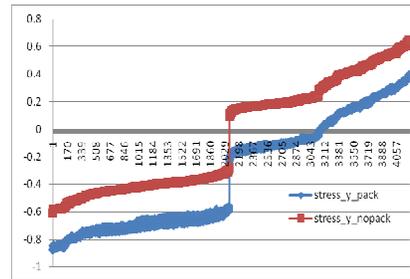
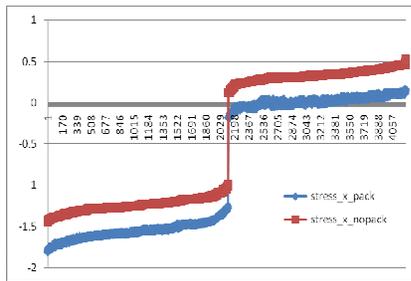
- Intra-channel stress components for all transistors in the analyzed design.

Effect of TSV and CPI on I_{dlin} Variation

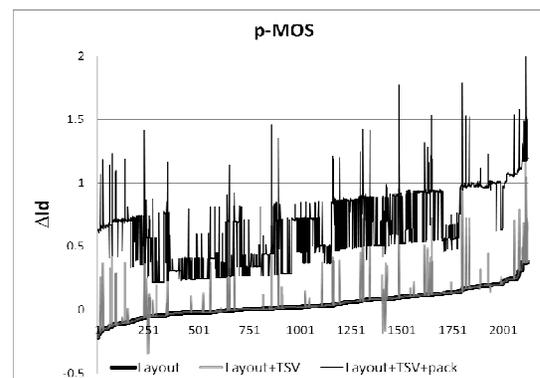
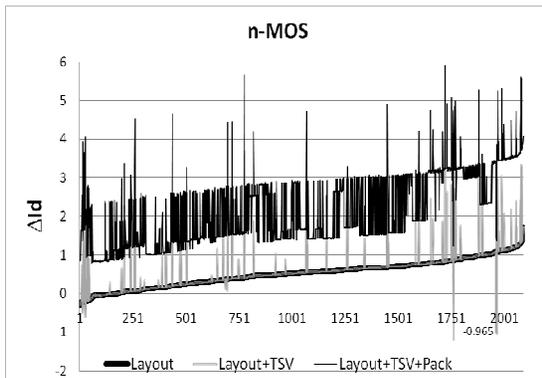
Across-die calculation of stress components inside transistor channels (compact model)

Compact model-based conversion of the stress values into corrections to the U_0 (low-field mobility) and V_{TH0} (zero biased threshold voltage for long channel transistor) for each considered transistor: MULU0 and DELVTO

• Transistor characteristics are calculated with a circuit simulator by annotating the corresponding netlist with instant parameters: MULU0 and DELVTO.

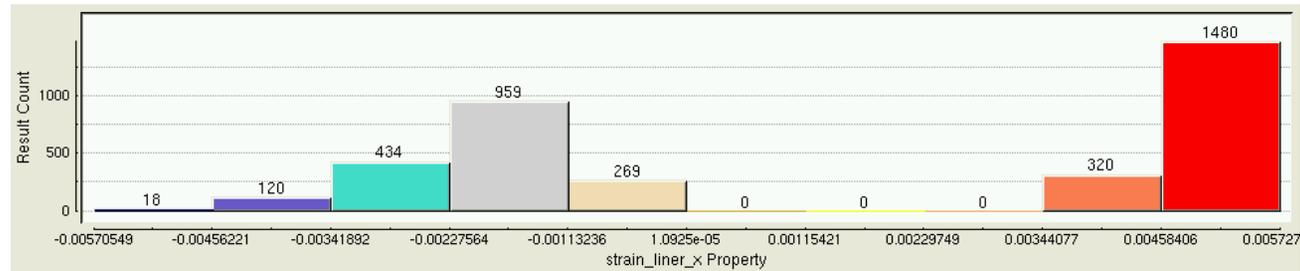


An example of the "compact model"-based simulation: distributions of x (left), y (middle) and z (right) stress components through all simulated transistors. Red line shows a sorted distribution of these characteristics when package-induced stress was not accounted, blue line shows these characteristics for the same transistors when package-induced stress was combined into the total stress.

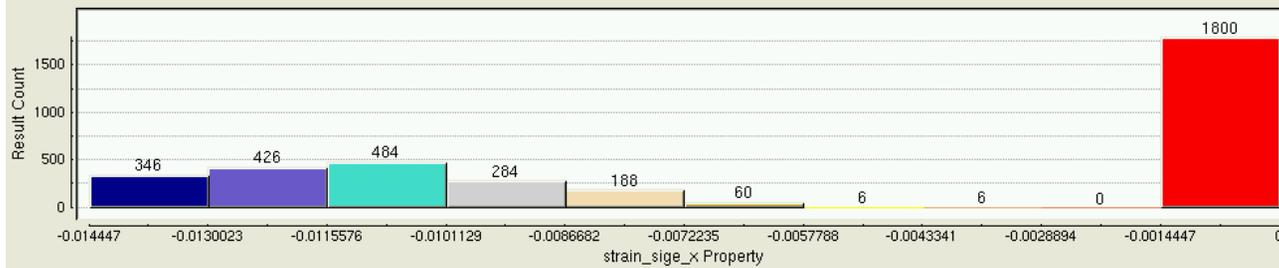


Sorted distributions of I_{dlin} for NMOS and PMOS when layout-induced stress sources were accounted (thick black lines); TSV-induced stress was added (grey lines); TSV and package-induced stresses were added (thin black lines).

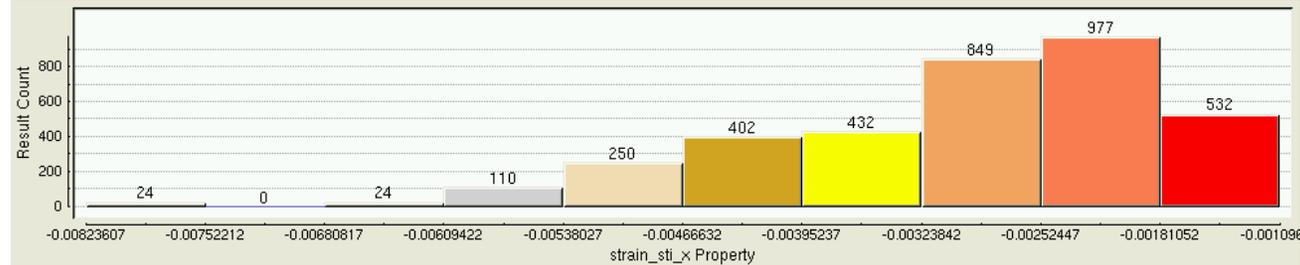
Histograms for Strain Populations



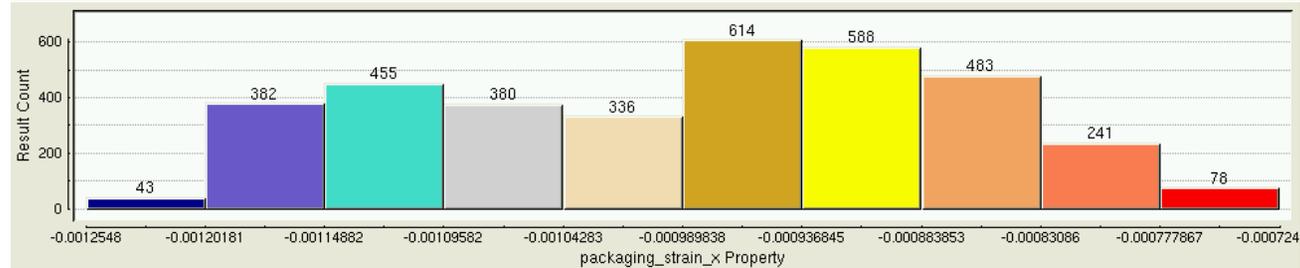
CESL



Si_{1-x}Ge_x



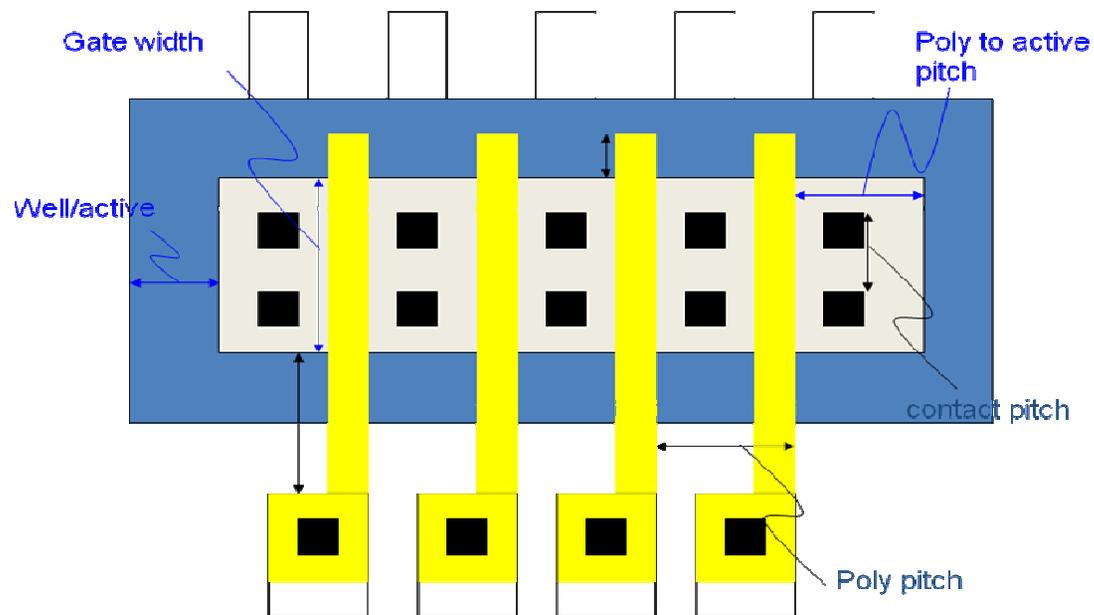
STI



CPI

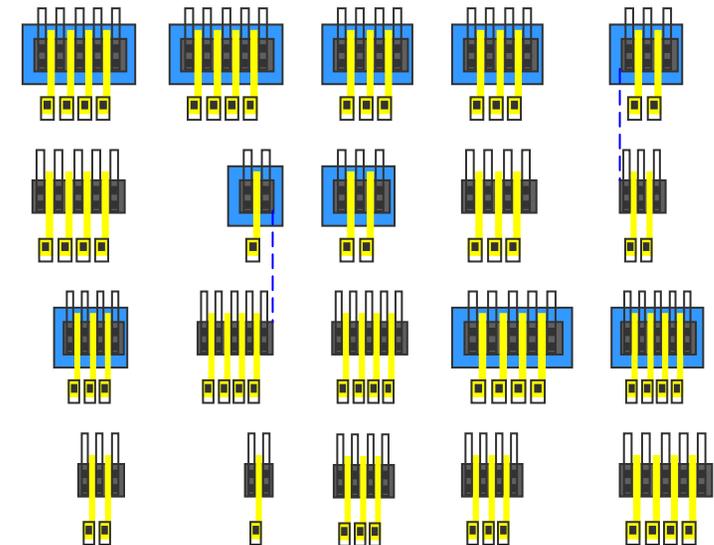
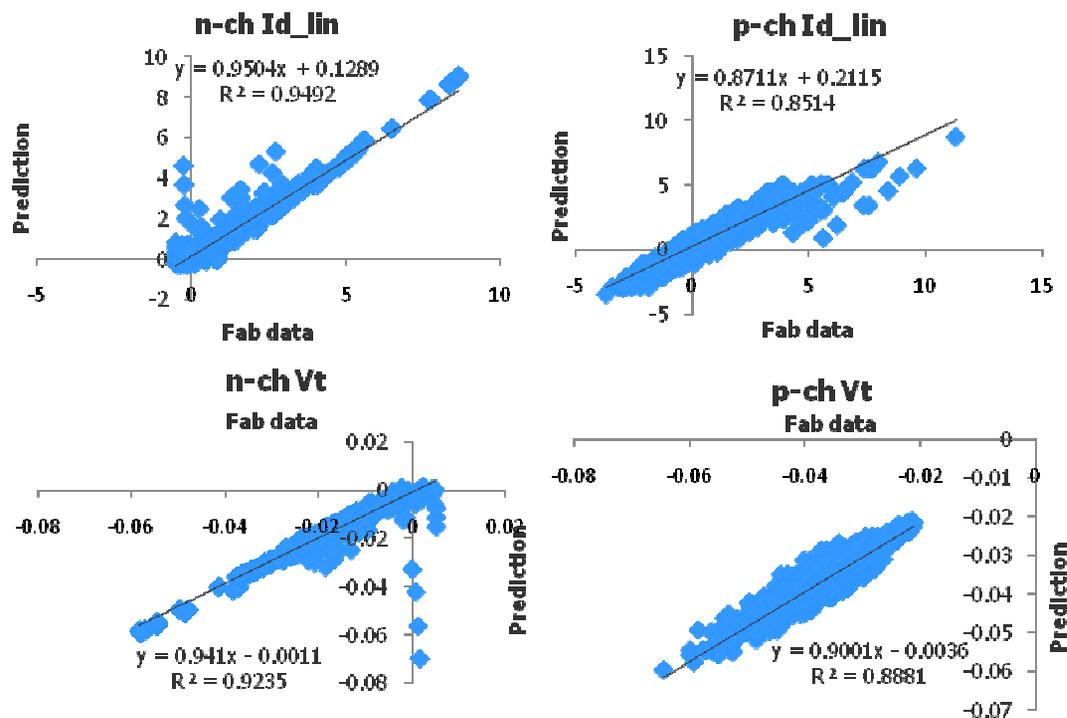
Calibration/Validation of the Compact Models for Device Characteristics vs. Stress

For calibration of the model which is interrelated the intra-channel stress with the device characteristics a special test-chip should be designed in such a way that all devices could be independently loaded electrically (decoupled) in order to avoid an electrical connectivity. In order to generate a reasonable amount of stress variation the variations in gate width, N-well to active distance, as well as variation in active size and in number of contacts and fingers should be introduced into the test-chip layout.



Calibration with the Foundry Model

Calibration was performed on ~100 gates
Prediction was made for all (~4000) gates



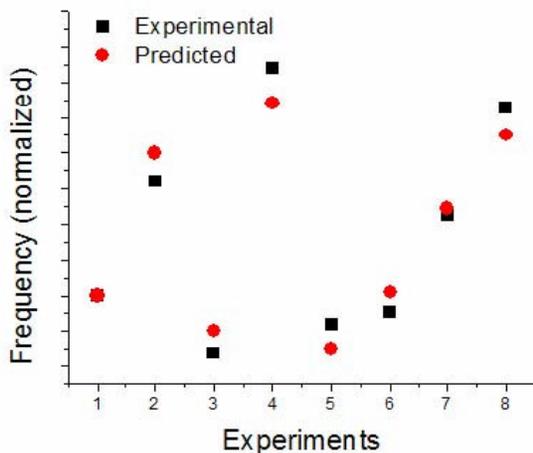
Test-chip segment

Fit between the predicted electrical characteristics of 3600 transistors in the developed “electrical” test-chip and the predictions obtained with the calibrated foundry model.

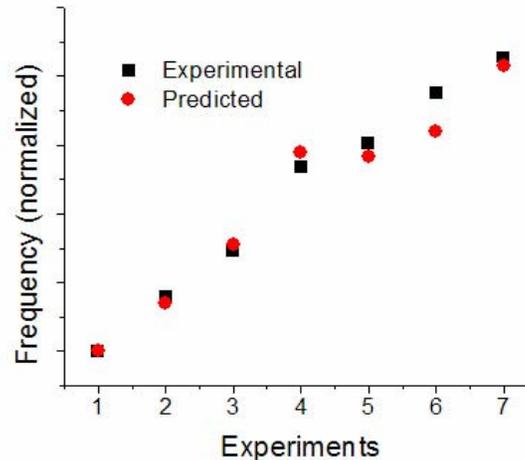
Model Predictions vs. Measurements

In order to verify the proposed stress models, we calibrate them using ring oscillator frequency data from an experimental test chip.

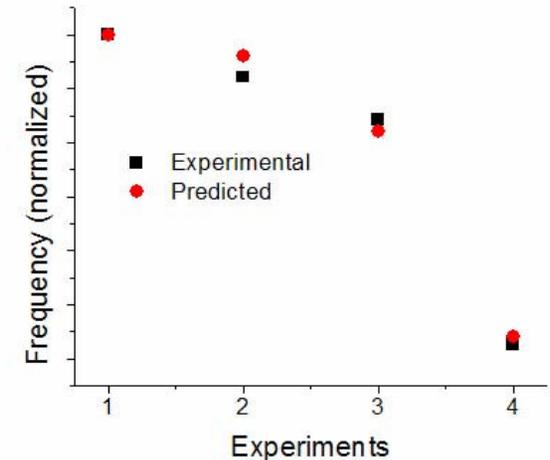
EXPERIMENTAL VS. PREDICTED FREQUENCY FOR:



various distances between n-well edge and device in longitudinal and transverse directions



various active area lengths

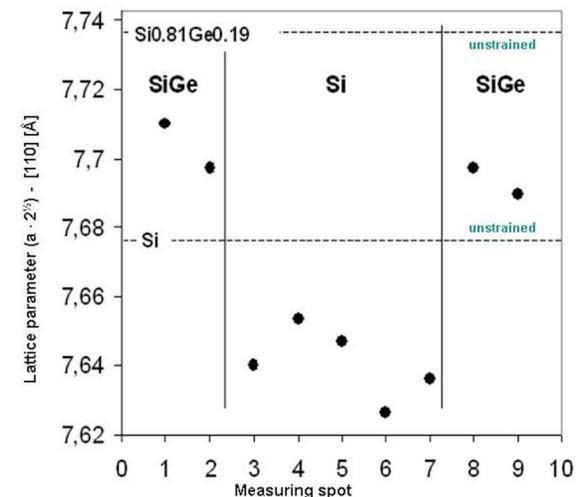
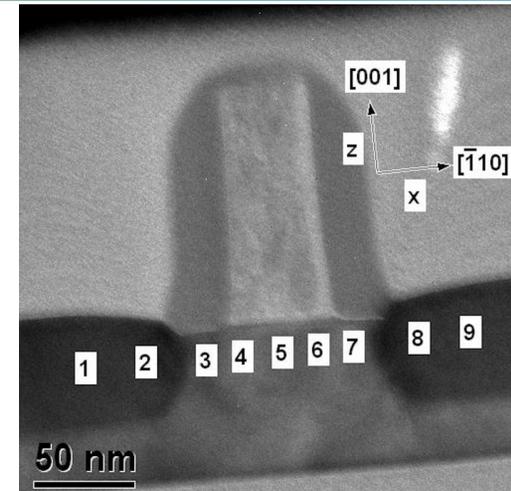


various number of contacts in the device

V. Joshi, V. Sukharev, A. Torres, K. Agarwal, D. Sylvester, and D. Blaauw, "Closed-form modeling of layout-dependent mechanical stress", in *Proceedings of Design Automation Conference (DAC), 2010*, pp. 673-678.

Two-step Calibration

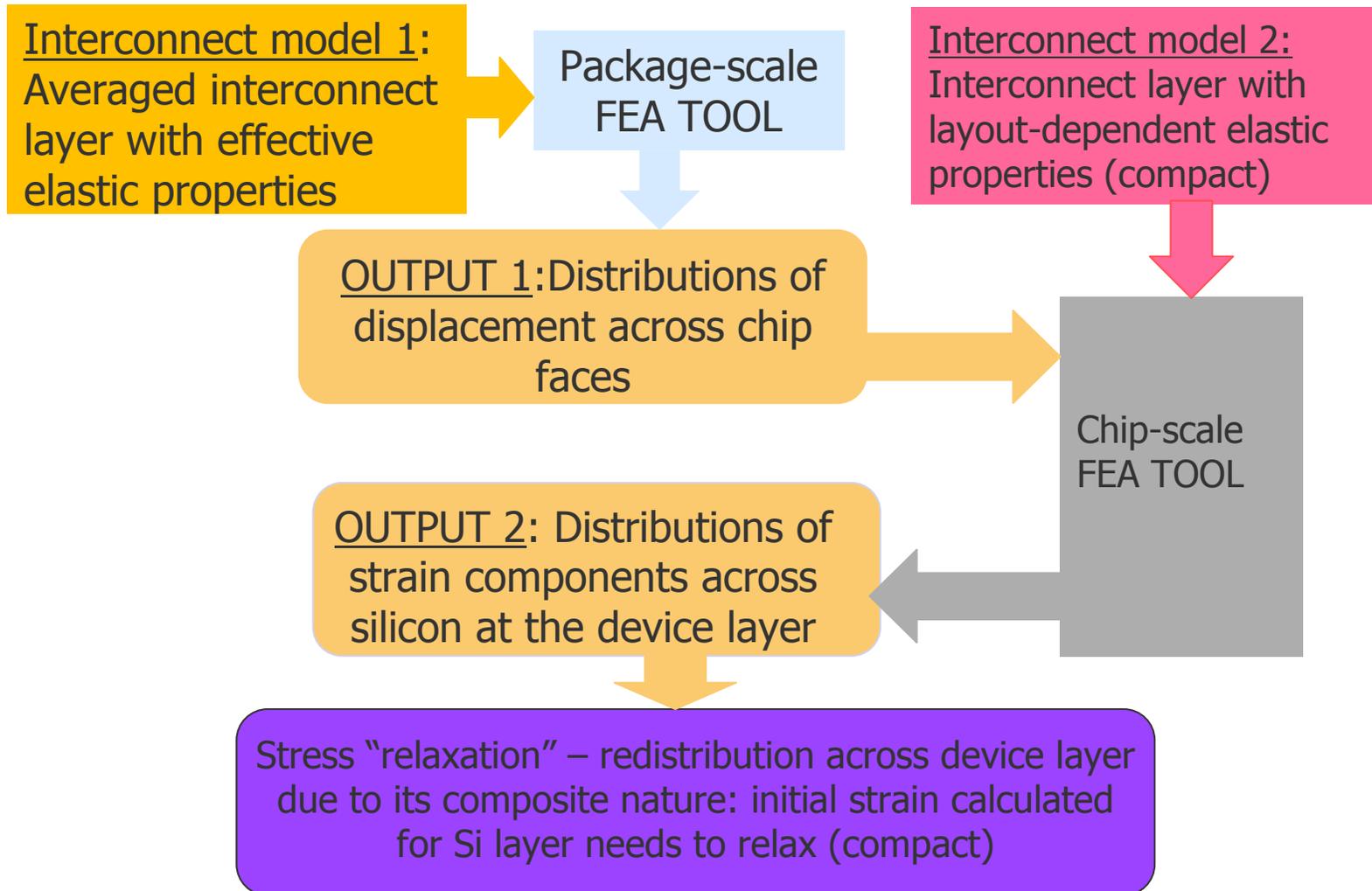
- Since the mechanical stress is not the only cause of variations in the transistor characteristics, the calibration procedure should be modified.
- The first step should be designed for the stress model calibration.
- The second one for calibrating the model transferring the intra-channel stress into U_0 and V_{TH0} .
- For these purposes, the high-resolution strain measurements in Si channels of the test-chip devices are needed.
- The diffraction-based techniques CBED and NBED are probe-based with a spatial resolution of 5 nm and 10 nm, respectively. They are performed at selected points on the specimen as shown in the Figure.



Engelmann HJ, Geisler H, Huebner R, et al (2008),

Proceedings of 4th EMC 2: 13-14

3D IC Technology: Simulation Flow for Stress-Assessment



PDK Requirements

(<http://wiki.sematech.org/MultiScaleSimulationCharacterization>)

TABLE 1. A preliminary list of materials parameters required for the package-scale simulation

Material	Meas. as function of temp. [-50...200°C]	Underfill (epoxy resin)	Cu pillars (copper)	Molding compound (ceramic-filled epoxy polymer)	Flip-Chip ball (approx. 100µm)	Micro ball (10s of µm)	BGA bump (>200µm)	Effective interconnect (Low-k/copper composite)	Effective bulk of silicon die (silicon/copper TSV)	Effective package substrate (organic PCB/copper)
CTE (ppm/K)	Yes	X	X	X	X	X	X	X	X	X
Young's modulus (GPa)	Yes	X	X	X	X	X	X	X	X	X
Poisson's ratio	No	X	X	X	X	X	X	X	X	X
Plasticity (MPa)	Yes		X							
Visco-plasticity (MPa)	Yes				X	X	X			
Visco-elasticity (MPa)	Yes	X		X						X
Glass transition temp. (°C)	No									X
Die attach T (°C)	No		X		X	X	X			

TABLE 2. Measurement techniques to determine the properties of interconnects (BEOL) layer

Property	Technique	Resolution	Accuracy	Specimen	Notes
E (T)	Nanoindentation	20 nm	± 10nN	Real specimen, cross section or blanket films	-
Poisson's ratio	Nanoindentation and FE-parameter extraction	20 nm	± 10nN	Real specimen, cross section or blanket films	* Calculation procedure to be validated
CTE (T)	Digital image correlation/SEM X-ray reflectometry	tbd tbd	tbd tbd	blanket films blanked films	* customized samples * customized samples

Conclusions

- This presentation describes the developed multi-scale simulation methodology and flow for stress assessment in 3D TSV-based chip stacks.
- The core of the proposed approach is the physics-based compact model which allows for the making of a link between the package-scale FEA tools and chip layout formats (GDSII, OASIS, etc.).
- The described compact model represents an extension of the previously developed model for the assessment of the layout-induced stress.
- The major difference between these two models is in the way of introduction of the initial stressor strains. In the case of layout-induced stress they were introduced as parameters that were extracted at the calibration stage. In the current model they are the results of the FEA-based simulations performed at the package-scale and die-scale steps.

Thank you!