Closed-Form Modeling of Layout-Dependent Mechanical Stress

Vivek Joshi, Valeriy Sukharev[±], Andres Torres[±], Kanak Agarwal*, Dennis Sylvester, David Blaauw
University of Michigan, Ann Arbor, MI {vivekj,dennis,blaauw@eecs.umich.edu}, [±]Mentor Graphics Corp., Wilsonville, OR {valeriy_sukharev,andres_torres@mentor.com}, *IBM Research Lab, Austin, TX {kba@us.ibm.com}

***Abstract* —** Modern CMOS technologies employ process-induced stress to improve carrier mobility and increase drive current. This stress has been shown to be strongly layout dependent; however there is a lack of physical models relating potential performance variation to critical layout parameters. This paper presents compact *closed-form* models that capture the layout dependence of mechanical stress induced in the device channel while considering all relevant sources of stress (STI, tensile/compressive nitride liners, and embedded SiGe). The models are calibrated using ring oscillator frequency data obtained from an experimental test chip to verify their accuracy. Results indicate that the models accurately capture the layout dependence of stress and carrier mobility for a variety of layout permutations and the root mean square error in the predicted ring oscillator frequency is less than 1% for the different layout experiments. These models can help drive layout optimization and timing/power analysis without the use of technology computer-aided design (TCAD) tools, which are slow and very limited in capacity.

Categories and Subject Descriptor: B.7.2 [IC Design Aids]
General Terms: Measurement, Performance, Theory
Keywords: Mechanical Stress, Mobility, Modeling

1. INTRODUCTION

Mechanical stress inducing layout features are used by modern CMOS processes in order to enhance carrier mobility, for higher performance. Mechanical stress breaks the crystal symmetry of Silicon, causing changes in the band scattering rates, and/or the carrier effective mass, which in turn affects carrier mobility [1, 2]. Application of the correct type of stress (tensile or compressive) results in significantly higher carrier mobility, and improves transistor performance [3]. There are three major layout dependent sources of mechanical stress: Shallow Trench Isolation (STI) generates compressive stress due to thermal mismatch with Silicon [4], embedded SiGe is epitaxially grown in the S/D regions of PMOS devices to induce high compressive stress due to lattice mismatch [5], and tensile/compressive nitride liner layers are integrated into a single, high performance process flow called the Dual Stress Liner (DSL) approach [6]. However, stress introduced in the channel, and hence carrier mobility, show a strong dependence on the device layout and its neighboring features [7]. As a result, layout properties such as active area length, number of contacts, distance of the device to the well edge, etc. become important in determining the mechanical stress induced in the channel of a device. Figure 1 shows the layout view for the three PMOS devices in a 3-input NAND gate, along with the corresponding longitudinal stress distribution under the channel, for a selected cross-section. Although the three devices have identical gate width and length, the channel stress is different in the three cases depending on other layout features such as

DAC'10, June 13-18, 2010, Anaheim, California, USA

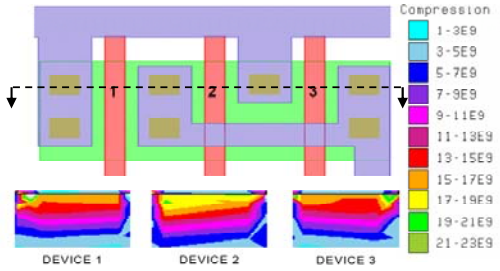

Fig 1: Channel stress distribution for PMOS devices in a 3-input NAND for a selected cross-section.

active area length, and contact placement. The device in the center (device 2) has higher stress than the two corner transistors because it is surrounded by more SiGe. This difference in stress is reflected in their performance, and simulations show that the drive currents for the center and edge devices differ by 8.2%. Such dependence can result in significant variation in the performance and leakage of devices, based on their context and layout.

Technology computer-aided design (TCAD) tools have been used to simulate device fabrication in order to capture process induced mechanical stress, and calculate its impact on device performance and leakage. However, TCAD tools based simulation frameworks involve time consuming computational steps, and have severely limited capacity in terms of the number of devices that can be accurately simulated in a single run. Hence, there is an urgent need to develop scalable, closed-form models for calculating process induced stress as a function of the device layout, and its neighboring features, to enable fast and accurate modeling and simulation of strained devices. In the past, [7, 8] have studied this layout dependence for different sources of stress, for both NMOS and PMOS devices. However, there has been very little work on comprehensive closed-form models of the layout dependence of process induced stress, and its impact on carrier mobility. Authors in [9] focused mainly on modeling mobility changes due to STI stress. [10] presented a very good method at modeling layout dependence of process induced stress through non process specific analytic models. However, while these models show a good fit for isolated device level stress simulation, they do not account for layout features such as distance of device from the well edge (tensile/compressive liner interface), presence of contacts, dummy poly, and neighboring devices. The paper also does not account for the transverse/lateral stress dependence on layout. So, while these models provide a good fit for simple device level experiments, they fail to account for key neighboring features which are critical for accurate stress simulation, when focusing on the complete circuit layout.

In this paper, we propose compact closed-form models for layout dependence of process induced stress, and its impact on carrier mobility. We analyze the physics behind stress inducing process steps, and solve relevant equations describing the stress distribution, in order to develop the models. Since the derivation is based on underlying physics, the derived models are scalable. We model stress due to Shallow Trench Isolation (STI), tensile/compressive nitride liners, and embedded SiGe S/D layers (used only in PMOS devices). In order to quantify the impact of

stress on mobility, we use the piezoresistive model [12]. Since longitudinal stress varies across the device width; we propose partitioning the gate into segments, such that each segment has almost constant stress, based on measured, stress-critical, layout parameters. We calculate the stress based mobility enhancement, in terms of mobility multipliers, for each of these segments, and take a weighted average of these multipliers based on the slice widths to derive one mobility multiplier for each device. Experiments based on ring oscillator frequency data show that the model accurately captures the variation of layout dependent stress effect for a variety of layout permutations. Proposed models are calibrated using the frequency data, and then used to predict the oscillation frequency using SPICE. The root mean square error in the predicted ring oscillator frequency for the different sets of layout experiments is less than 1%, verifying the accuracy of the proposed models.

The rest of the paper is organized as follows. Section 2 discusses the derivation of stress models for the different stress inducing process steps, along with the translation of stress into impact on device mobility. Experimental results are discussed in Section 3, and Section 4 concludes the paper.

2. MODELING STRESS ENHANCED CARRIER MOBILITY

For model based simulation of strained devices, we need to calculate the mechanical stress induced in the device channel, and then translate the stress into impact on carrier mobility. This impact is quantified in terms of mobility multipliers, which can then be used in circuit simulators such as SPICE to capture the stress effect. In this section, we first present our closed-form stress models to enable fast and accurate stress modeling, and the second part of the section discusses translating these stress numbers into mobility multipliers to calculate the impact on performance and leakage by using SPICE.

2.1 Stress Models

We develop our stress models by analyzing the physics behind various stress inducing process steps, and solving relevant equations. We analyze each source of stress separately, and add up the stress due to each source, to obtain overall stress in the device channel. Since the models are based on the physics behind each process step, they are scalable for future technology generations. The sources of stress modeled are: embedded SiGe S/D layer (for PMOS devices), tensile/compressive nitride liners, and Shallow Trench Isolation (STI). The models represent a very simple combination of transverse and longitudinal direction 1D spring approximations. The physics based derivation is done under multiple simplifying assumptions and is supposed to provide a general form for the model, while the actual parameter values come from rigorous calibration-optimization. For each device, we consider all the features within a certain window of influence (of length L_W), to calculate the resulting stress.

2.1.1 Embedded SiGe source/drain

For PMOS devices, SiGe is epitaxially grown in cavities that have been etched into the source/drain areas [5]. A large compressive stress is created in the PMOS channel due to lattice mismatch between Si and SiGe, thereby resulting in significant hole mobility improvement. In this process, NMOS is protected by a capping layer to prevent Si recess, and SiGe epitaxial growth. The key to modeling the magnitude of induced stress is to identify the physics behind generation of compressive stress, and solve relevant equations by applying simple spring approximations. We assume that the widths of all structures are much bigger then their lengths (quasi 1D case).

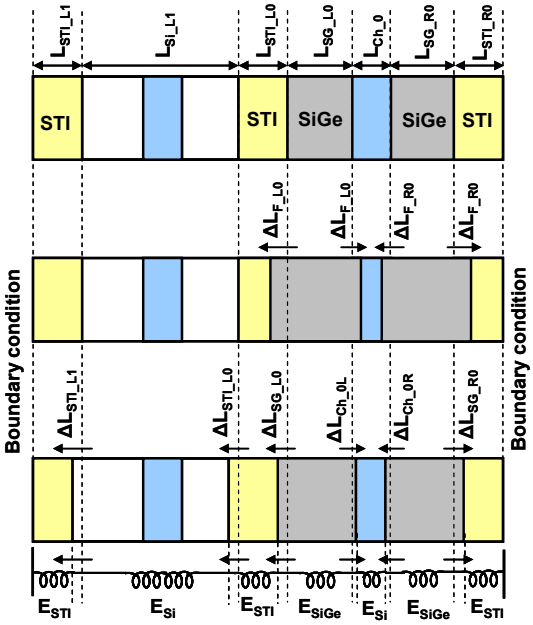

Fig 2: Before SiGe expansion (top), after non-confined SiGe expansion (middle), and after deformation of all segments due to SiGe expansion (bottom).

Figure 2 shows a very simple layout used to explain the derivation of 1D models for compressive stress generated due to embedded SiGe. The layout is composed of two simple devices separated by STI, one with embedded SiGe in S/D regions (device 0), and the other without it (device 1). Ge has a lattice constant larger than Si and hence it occupies more volume than Si would occupy. The gray areas (SiGe) can be seen as trying to expand in all the directions. The scenario after epitaxial growth of SiGe is depicted in the bottom picture of Figure 2. If χ is an atomic ratio of Ge in Si and Ω_{Si} and Ω_{Ge} are the atomic volumes of Si and Ge, respectively, then it is easy to show that an initial volume V_0 (volume without introduction of SiGe in the S/D) would try to expand by $\Delta V = \left(\frac{\Omega_{Ge}}{\Omega_{Si}} - 1\right)\chi V_0$, which translates to a linear expansion of $\Delta L_F = \sqrt[3]{\left(\frac{\Omega_{Ge}}{\Omega_{Si}} - 1\right)\chi L_0}$ in all the three dimensions. As shown in the middle picture of Figure 2, in the absence of any confinement (neighboring features), SiGe would have expanded by this amount. The expansions for the left and right SiGe regions can therefore be expressed as:

$$\Delta L_{F\_L0} = \sqrt[3]{\left(\frac{\Omega_{Ge}}{\Omega_{Si}} - 1\right)\chi L_{SG\_L0}} \qquad \Delta L_{F\_R0} = \sqrt[3]{\left(\frac{\Omega_{Ge}}{\Omega_{Si}} - 1\right)\chi L_{SG\_R0}} \tag{1}$$

In reality, the presence of neighboring features opposes such an expansion, thereby creating compressive stress in the device channel. The deformation of the SiGe sub-segment as compared to the non-confined case can be expressed as the difference between the non-confined and the actual confined case expansion. The bottom picture in Figure 2 shows the deformation of different segments after SiGe expansion. We consider each layout segment as being represented by a spring (or an elastic beam) characterized by different elasticity. It is assumed that displacements at ends of the considered segment (leftmost and rightmost edges) are equal to zero. This might be treated as the symmetry boundary conditions. At equilibrium, the forces acting from one sub-segment on another at the points of contact are equal. It provides us, in the frame of the accepted approximation, with the condition

of equal stress along the entire line of cross-section. This stress value will depend on the layout composition in the region of interest. So, we can express the generated longitudinal stress in different segments with following equations.

$\sigma_{STI\_L1} = \frac{-\Delta L_{STI\_L1}}{L_{STI\_L1}} E_{STI}$	$\sigma_{Si\_L1} = \frac{\Delta L_{STI\_L1} - \Delta L_{STI\_L0}}{L_{Si\_L1}} E_{Si}$
$\sigma_{STI\_L0} = \frac{\Delta L_{STI\_L0} - \Delta L_{SG\_L0}}{L_{STI\_L0}} E_{STI}$	$\sigma_{SG\_L0} = -\frac{\Delta L_{F\_L0} - (\Delta L_{SG\_L0} + \Delta L_{Ch\_0L})}{L_{SG\_L0}} E_{SiGe}$
$\sigma_{Ch\_0L} = -\frac{\Delta L_{Ch\_0R} + \Delta L_{Ch\_0L}}{L_{Ch\_0}} E_{Si}$	$\sigma_{SG\_R0} = -\frac{\Delta L_{F\_R0} - (\Delta L_{SG\_R0} + \Delta L_{Ch\_0R})}{L_{SG\_R0}} E_{SiGe}$
$\sigma_{STI\_R0} = \frac{\Delta L_{STI\_R0} - \Delta L_{SG\_R0}}{L_{STI\_R0}} E_{STI}$	

(2)

Here E_{SG}, E_{Si} and E_{STI} are the elasticity constants of the $Si_{1-x}Ge_x$, silicon, and STI, various ΔL and L are the deformations and nominal dimensions as shown in the figure, and σ_a is the stress generated in segment a due to SiGe expansion.

Using the condition of equal stress, we can set up the system of equations for determination of unknown deformations of the segments. The deformation numbers for each segment can then be used to determine the value of stress generated. Upon solving these equations, we obtain the longitudinal stress in the channel:

$$\sigma^{L}_{Ch\_0} = -\frac{\beta(L_{SG\_L0} + L_{SG\_R0})}{d_{STI} + d_{Si} + (1+\beta)d_{SiGe}} \tag{3}$$

$$d_{STI} = \frac{L_{STI\_L1} + L_{STI\_L0} + L_{STI\_R0}}{E_{STI}} \qquad d_{Si} = \frac{L_{Ch\_0} + L_{Si\_L1}}{E_{Si}}$$

$$d_{SiGe} = \frac{(L_{SG\_L0} + L_{SG\_R0})}{E_{SiGe}} \qquad \beta = \sqrt[3]{\left(\frac{\Omega_{Ge}}{\Omega_{Si}} - 1\right)\chi}$$

In general, for any given layout we can write the longitudinal stress in a channel as:

$$\sigma^{L}_{SiGe} = -\frac{\beta\left(\sum_j L_{SG\_j} + \sum_i L_{neigh\_L(R)i}\right) + (\Delta L_{BC\_L} + \Delta L_{BC\_R})}{\frac{\sum_k L_{STI\_k}}{E_{STI}} + \frac{\sum_n L_{Si\_n}}{E_{Si}} + \frac{(1+\beta)\left(\sum_j L_{SG\_j} + \sum_i L_{neigh\_L(R)i}\right)}{E_{SG}}} \tag{4}$$

Here L_{SG_j} is the length of the j-th $Si_{1-x}Ge_x$ S/D segment on the same active area as the device while L_{neigh_Lj} and L_{neigh_Rj} are the length of the neighboring SiGe active areas. L_{STI_k} is the k-th STI width, L_{Si_n} is the length of the n-th gate or non-SiGe source/drain (NMOS active) area in the longitudinal direction. ΔL_{BC_L} and ΔL_{BC_R} are the boundary conditions at the left and right window edges representing stress-induced edge displacements.

In addition to the generation of compressive longitudinal stress due to SiGe growth, transverse strain/stress is also generated because of traction between channel segment and adjacent SiGe structures. The expansion of these SiGe drain structures in transverse direction causes the adjacent silicon (channel area) to expand as well. This is illustrated in Figure 3. Hence, in order to estimate SiGe induced transverse stress in the device channel, we need to account for stress caused by the traction with adjacent SiGe areas due to SiGe expansion.

The transversal stress can be calculated as

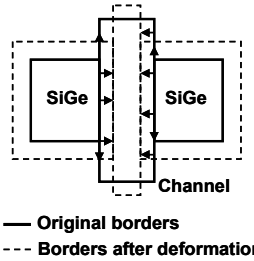

Fig 3: Sample device layout showing generation of transverse stress

$$\sigma^{T}_{Ch\_0} = \frac{E_{Si}}{W_{Ch}}\left(\frac{\sigma^{T}_{SG\_L0} + \sigma^{T}_{SG\_R0}}{2E_{SG}} + \frac{\beta}{1+\beta}\right) \tag{5}$$

Here W_{Ch} is the width of the channel. $\sigma^{T}_{SG\_L0}$ and $\sigma^{T}_{SG\_L0}$ are the stress in adjacent left and right S/D $Si_{1-x}Ge_x$ structures, which can be calculated in a manner similar to Equation 4 by replacing all L (horizontal distances) by W (vertical distances). Indexes T and B are for top and bottom, respectively.

$$\sigma^{T}_{SiGe} = -\frac{\beta\left(\sum_j W_{SG\_j} + \sum_i W_{neigh\_T(B)i}\right) + (\Delta L_{BC\_T} + \Delta L_{BC\_B})}{\frac{\sum_k W_{STI\_k}}{E_{STI}} + \frac{\sum_n W_{Si\_n}}{E_{Si}} + \frac{(1+\beta)\left(\sum_j W^{j}_{SG} + \sum_i W_{neigh\_T(B)i}\right)}{E_{SG}}} \tag{6}$$

2.1.2 Nitride Liner

Capping stressed layer technology is one of the most important techniques employed to generate a desirable stress in device channel. Traditionally, a silicon nitride based contact etch stop layer (CESL) is used as the source of the tensile stress. In this technology, a $Si_xN_yH_z$ layer is deposited followed by a special type of anneal to release hydrogen. This results in volume shrinking, which generates strong tensile stress in the surrounding confinement that gets transferred into the channel region of NMOS devices. In order to avoid tensile stress generation in PMOS devices, different technological steps were introduced. The most effective way was to dope the CESL in the PMOS regions with a *Ge* implant that results in volume expansion, and compressive stress generation in the confinement [11]. Latest high performance process nodes have simultaneously incorporated both tensile and compressive nitride liners into a single high performance CMOS flow, called the Dual Stress Liner approach [6]. Nwell mask is generally used while defining the compressive and tensile regions and nwell edges can be seen as the interface of compressive and tensile nitride.

We define α as the coefficient of proportionality between the as-drawn length (L_{CESL}) of a CESL segment (stress effect is not accounted), and the confinement-free length (L_{CESL}^{*}) of the same segment if the nitride layer was allowed to expand/contract without any confinement imposed by neighboring features: $L_{CESL}^{*} = \alpha.L_{CESL}$. Having defined that, we can then proceed to calculate the stress generated due to nitride in a manner similar to embedded SiGe. The quasi 1D approximation yields the following expression for capping layer induced longitudinal stress as a function of layout geometry.

$$\sigma^{L}_{CESL} = \frac{(1-\alpha)\sum_i L_{CESL\_i} + (\Delta L_{BC\_L} + \Delta L_{BC\_R})}{\frac{\alpha\sum_i L_{CESL\_i}}{E_{CESL}} + \frac{\sum_j L_{Poly\_j}}{E_{Si}} + \frac{\sum_k L_{Contact\_k}}{E_{Contact}}} \tag{7}$$

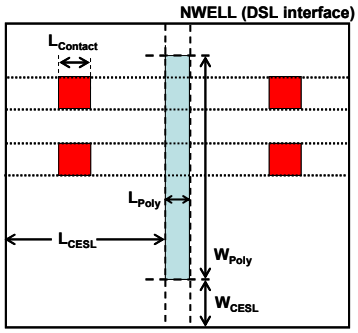

Fig 4: Sample layout parameters for CESL stress calculation

Here, L_{CESL_i} is the length of *i*-th stress layer segment either between two neighboring poly, or between poly and contact, or poly and border of the chosen window, L_{Poly_j} is the length of the *j*-th gate (channel length), and $L_{Contact\_k}$ is the contact size, all in the longitudinal direction. Similar to the SiGe case, ΔL_{BC_L} and ΔL_{BC_R} are the boundary conditions at the left and right window edges representing stress-induced edge displacements and E_{CESL}, E_{Si} and $E_{Contact}$ are the elasticity constants of the capping layer, silicon, and the contact material, respectively. In the absence of contacts, $L_{Contact}$ is taken as 0.

Stress in the transverse direction can be obtained by replacing all the longitudinal measurements with transverse measurements and left and right boundary conditions with the corresponding top and bottom limits. The traverse stress can then be expressed as:

$$\sigma_{CESL}^{T} = \frac{(1-\alpha)\sum_i W_{CESL\_i} + \left(\Delta L_{BC\_T} + \Delta L_{BC\_B}\right)}{\frac{\alpha\sum_i W_{CESL\_i}}{E_{CESL}} + \frac{\sum_j W_{Poly\_j}}{E_{Si}} + \frac{\sum_k L_{Contact\_k}}{E_{Contact}}} \quad (8)$$

Figure 4 shows a set of relevant layout parameters for CESL stress calculation. As predicted by the proposed model, the presence of polysilicon gates and contacts decreases the stress due to nitride liner by breaking the continuity of the deposited nitride liner layer. Contacts create holes in the liner layer, while polysilicon gates cause a bump in the deposited liner layer to bring down the stress. As a result, an isolated device with no contacts will have the highest stress due to nitride. These effects are included in the models expressed in Equations 7 and 8.

2.1.3 Shallow Trench Isolation

Shallow Trench Isolation (STI) creates stress due to thermal mismatch between silicon and STI. The difference in the thermal expansion coefficients causes compressive stresses to develop in the device once the wafer is cooled down post annealing. We can quantify the magnitude of generated stress using the expression for linear contraction that causes the stress to develop. For a given silicon segment, contraction upon cooling can be quantified as:

$$\Delta L_{x,y} = \left(\alpha_{Si} \times \Delta T \times L_{x,y}\right) \quad (9)$$

Here $\Delta L_{x,y}$ is change in length upon cooling, α_{Si} is the thermal expansion coefficient of Silicon, ΔT is the difference between the anneal temperature and the final temperature, and $L_{x,y}$ is the as-drawn length of the considered segment. This is the contraction that would occur in the absence of any confinement. We can then proceed to calculate STI stress for a given layout segment, by following an approach similar to that used for calculating stress due to SiGe, and nitride.

The longitudinal stress can be expressed as:

$$\sigma_{STI}^{L} = -\frac{(\alpha_{Si} - \alpha_{STI})\Delta T\sum_i L_{STI\_i} + \left(\Delta L_{BC\_L} + \Delta L_{BC\_R}\right)}{\frac{(1-\alpha_{STI}\Delta T)}{E_{STI}}\sum_i L_{STI\_i} + \frac{(1-\alpha_{Si}\Delta T)}{E_{Si}}\sum_j L_{Si\_j}} \quad (10)$$

Here L_{STI_i} is the length of the *i*-th STI segment, L_{Si_j} is the length of the *j*-th silicon segment, α_{Si} and α_{STI} are the coefficients of thermal expansion for silicon, and STI, respectively. Replacing longitudinal measurements by lateral (transverse) measurements and left and right boundary conditions by top and bottom edges, we get the following expression for transverse stress:

$$\sigma_{STI}^{T} = -\frac{(\alpha_{Si} - \alpha_{STI})\Delta T\sum_i W_{STI\_i} + \left(\Delta L_{BC\_T} + \Delta L_{BC\_B}\right)}{\frac{(1-\alpha_{STI}\Delta T)}{E_{STI}}\sum_i W_{STI\_i} + \frac{(1-\alpha_{Si}\Delta T)}{E_{Si}}\sum_j W_{Si\_j}} \quad (11)$$

It should be noted that all the derived formulas which describe the stress generated by different stress sources ((4), (6), (7), (8), (10), (11) contain the window edge displacements terms ΔL_{BC}. These displacements generally should be equal to zero, in accordance with the assumption of symmetry boundary condition. However, in some specific cases, when the effect of global load, such as packaging, chip mounting or 3D integration, on the variation of transistor-to-transistor characteristics is of interest, these terms should come from the *global* finite element based simulation. Also note that these models provide a general form for functions to estimate stress, the values for parameters such as *E*, *α*, etc. are obtained by calibration optimization and might be different from the actual physical values.

2.2 Converting Stress to Mobility

The layout dependence of process induced stress leads to gates with non uniform stress, and, hence, non uniform mobility, in the device channel across the width of the device. Based on the closed-form models, we know the layout parameters that affect the stress induced in the channel (such as number of contacts, distance of device from well-edge, active area length, etc.). This knowledge can used to partition the device gate into segments, such that these stress-critical geometrical parameters for a given segment are constant throughout the segment width. We can then calculate stress, and its impact on mobility, for each of these segments independently, and take a weighted average of mobility multipliers for different segments (based on segment width), to determine one single value of mobility multiplier for the strained device. Figure 5 shows a sample device layout (selected from a larger MUX circuit layout) partitioned into segments. For each segment, we can then proceed to calculate stress due to different sources, and sum it up to obtain the overall stress in each direction. However, in accordance with Poisson's Effect, layout generated longitudinal strain also produces a transverse strain which is given by $\varepsilon^T = -\nu\varepsilon^L$; where ν is the Poisson's factor, ε^T is the transverse strain, and ε^L is the longitudinal strain. Similar relationship exists for longitudinal stress caused by layout generated transverse stress/strain. A complete stress distribution in the *j*-th segment can then be expressed as following:

$$\begin{aligned} \sigma_j^{L(tot)} &= \sigma_j^{L} - \nu\sigma_{Ch}^{T} \\ \sigma_j^{T(tot)} &= \sigma_{Ch}^{T} - \nu\sigma_j^{L} \end{aligned} \quad (12)$$

where, $\sigma_j^{L(tot)}$ and $\sigma_j^{T(tot)}$ are the total longitudinal and the total transverse stress in the segment; σ_j^{L} and σ_{Ch}^{T} are the longitudinal and transverse stress values calculated based on the model, and *v* is the Poisson factor. As shown in Figure 5, the longitudinal stress is different for different segments of the device based on the

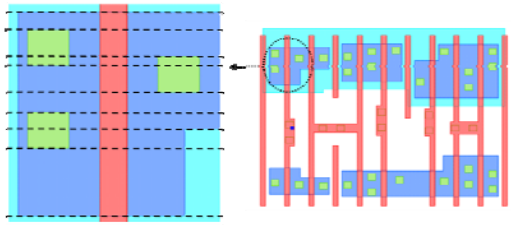

Fig 5: MUX layout showing stress based partitioning of a random PMOS device.

longitudinal layout parameters while the traverse stress is same for the entire channel. Finally, we use piezoresistive coefficients to convert from stress to mobility [12]. Mobility multiplier (μ_{mult}) for a given segment is expressed as:

$$\mu_{mult} = 1 + \frac{\Delta\mu}{\mu} = 1 + \pi_L \sigma_j^{L(tot)} + \pi_T \sigma_j^{T(tot)} \quad (13)$$

Here, π_L and π_T are the longitudinal and transverse piezoresistive coefficients, respectively. Since the piezoresistive coefficients have a strong dependence on the doping concentrations [12], we assume that these coefficients come from calibration optimization as well. Finally, we can take a width based weighted average of these multipliers to obtain an overall device mobility multiplier, which can then be used in a circuit simulator such as SPICE for accurate simulation of strained devices.

3. EXPERIMENTAL RESULTS

In order to verify the accuracy of proposed stress models, we used TCAD simulation based stress and on-current (I_{on}) data for NMOS and PMOS devices in various configurations. We also validated our models against ring oscillator frequency data from an experimental test chip fabricated in a process that contains both nitride liner and SiGe stress enhancement techniques. The models were separately calibrated for each case by setting up a system of equations in terms of the unknown model coefficients (π_L, π_T, α, etc.) using measured layout parameters. We wrote a simple layout editor script to measure layout distances, and segment the device gate into regions with equal stress. As discussed in the previous section, this segmentation is done such that the stress-critical layout parameters such as active area length, etc. are constant for each segment. SPICE based simulations were used to generate tables for dependence of I_{on} on mobility multipliers. Finally, MATLAB code based on least squares fitting is used to solve for model coefficients using these equations.

3.1 TCAD Experiments

In this set of experiments we use a setup comprising of Tsuprem4 (for simulating fabrication process to generate stress data), and Davinci (for simulating the on current values using Tsuprem4 generated stress data) to generate on current values for different layout configurations of 65nm NMOS and PMOS devices. The TCAD setup is accurately calibrated to the SPICE models for the 65nm technology. Once calibrated, the models are used to generate mobility multipliers which are then used in SPICE based simulation of the devices, and the result is then compared to the TCAD simulation data for each configuration.

We first look at the impact of active area length on device stress. Active area length is one of the most important layout parameters that impacts channel stress quite significantly by increasing the SiGe region around the channel(for PMOS). Figure 6 shows the variation of longitudinal channel stress with source/drain length ($L_{s/d}$) (normalized to minimum value of $L_{s/d}$) of an isolated 65nm PMOS device as simulated in Tsuprem4 and Davinci TCAD tool. Also shown in the figure is the stress predicted by the proposed model. Stress values are normalized to the value of stress at minimum $L_{s/d}$ for the technology. The figure shows that increasing $L_{s/d}$ increases stress in the channel and this dependence is captured quite accurately by the proposed stress model.

Next we focus on the CESL stress and predict the TCAD results with the proposed model. The most critical layout parameter for CESL is the distance to well edge which serves as an interface between the compressive and the tensile nitride liners. Figure 7 shows TCAD based simulation for dependence of PMOS channel stress due to nitride liner as a function of distance from the well edge in the longitudinal direction. As the distance from the well edge increases, so does the compressive stress [7]. The stress values are normalized to the value at minimum allowed distance from well edge for the 65nm technology.

We then analyze on-current predictions for NMOS and PMOS obtained from TCAD simulations and the proposed models. For this we generate a set of layout experiments by varying critical layout parameters. Different combinations of various layout parameters, as shown in Figure 8, are varied to generate several different experiments. The first few experiments try to increase the stress based mobility by a combination of increasing the active area length, moving the device away from well edge (in the longitudinal direction), sharing the active area with other devices, etc., while the last few experiments try to decrease the stress based mobility by moving the devices closer to the well edge, introducing more contacts, and decreasing active area length.

Figure 9 shows the predicted and simulated on current values (normalized to the on current for isolated NMOS and PMOS devices with one contact) for various TCAD experiments. The proposed model accurately predicts the current values, and the

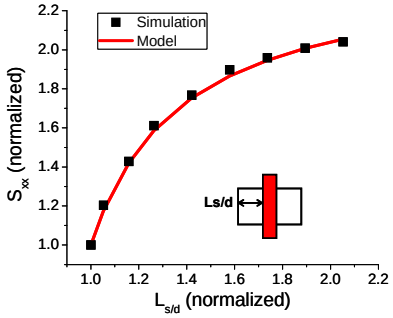

Fig 6: Longitudinal channel stress as a function of active area length as obtained by TCAD simulations and after proposed model fitting

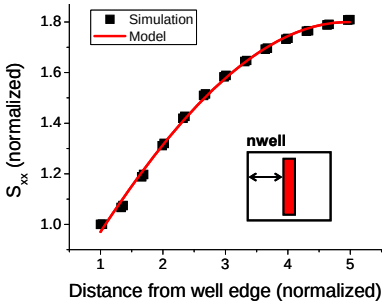

Fig 7: Longitudinal channel stress as a function of distance from well edge as obtained by TCAD simulations and after proposed model fitting

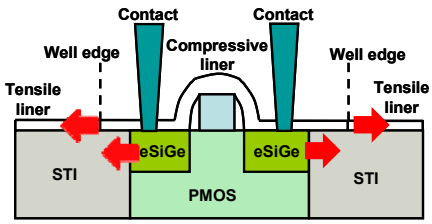

Fig 8: Layout permutations in TCAD experiments for model verification

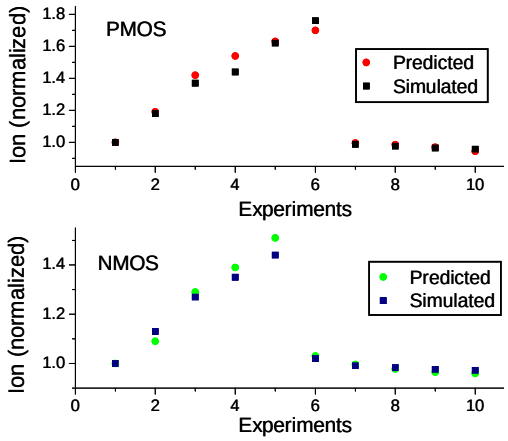

Fig 9: Experimental (TCAD) and predicted on current values for NMOS and PMOS devices

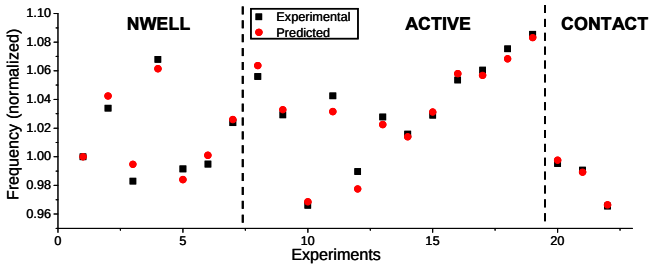

Fig 10: Experimental (hardware) and predicted ring oscillator frequencies for different layout configurations

root mean square error in predicted on current value is less than 0.8% for both PMOS and NMOS experiments.

3.2 Hardware Experiments

In this set of experiments, the proposed stress models are calibrated and verified using ring oscillator frequency data from an experimental test chip. The ring oscillator data is measured and averaged over several dies to reduce the impact of random and die-to-die systematic variations. For the purpose of calibration, we assume that the frequency of oscillation is directly proportional to average drive current for the ring oscillator, which was confirmed to be a valid assumption using SPICE based simulations of the ring oscillator circuit. Once calibrated, the models are used to calculate impact of stress in terms of mobility multipliers for different ring oscillator layout configurations. Figure 10 shows the comparison between the measured frequency data and the predicted frequency (normalized) for various layout experiments. The plot is divided into three distinct regions corresponding to three different set of layout configurations constituting the hardware experiments. In the nwell experiments, we vary the distance between nwell edge and device in both lateral and longitudinal directions. Since nwell mask is used to define the interface between compressive and tensile nitride liners, such changes have an impact on longitudinal and traverse stress due to the nitride layer. In the second set of experiments, active area layout and length was varied to change the amount of embedded SiGe next to the channel (only for PMOS devices), and the distance between STI edge and gate. In the contact experiments, we varied the number of contacts in the devices constituting the ring oscillator. The plot shows that the models exhibit a very good fit to the hardware data with the root mean square error between simulated and measured data to be only 0.9%.

4. CONCLUSION

In this work, we propose compact, closed-form models for layout dependence of process induced stress. We partition each device channel into segments with equal stress in order to calculate the impact on mobility in terms of mobility multipliers. We extensively verify our models against hardware and TCAD simulation data for a large number of layout permutations. The models enable fast and accurate stress prediction for a device in a given layout environment. The root mean square error in the predicted behavior is observed to be less than 1% for the different experiments, thereby, verifying the accuracy of the models.

References

[1] F. Andrieu *et al.*, "Experimental and Comparative Investigation of Low and High Field Transport in Substrate- and Process-Induced Strained Nanoscale MOSFETs," *Proc. VLSI Tech. Symp. Tech. Dig.*, pp. 176-177, 2005.

[2] K. Mistry *et al.*, "Delaying Forever: Uniaxial Strained Silicon Transistors in a 90nm CMOS Technology," *Proc. VLSI Tech. Symp. Tech. Dig.*, pp. 50-51, 2005.

[3] V. Chan *et al.*, "Strain for PMOS performance Improvement," *Proc. CICC*, pp. 667-674, 2005.

[4] R. A. Bianchi *et al.*, "Accurate modeling of trench isolation induced mechanical stress effects on MOSFET electrical performance," *Proc. IEDM*, pp. 117-120, 2002.

[5] Z. Luo *et al.*, "Design of high performance PFETs with strained Si channel and laser anneal," *Proc. IEDM*, pp. 489-492, 2005.

[6] H. S. Yang *et al.*, "Dual stress liner for high performance sub -45nm gate length SOI CMOS manufacturing," *Proc. IEDM*, pp. 1075-1077, 2004.

[7] V. Joshi *et al.*, "Stress Aware Layout Optimization," *Proc. International Symposium on Physical Design (ISPD),* pp. 168-174, 2008.

[8] V. Moroz *et al.*, "The Impact of Layout on Stress-Enhanced Transistor Performance," *Proc. SISPAD*, pp. 143-146, 2005.

[9] R. A. Bianchi *et al.*, "Accurate modeling of trench isolation induced mechanical stress effects on MOSFET electrical performance," *Proc. IEDM*, pp. 117-120, 2002.

[10] M. V. Dunga *et al.*, "A Holistic Model for Mobility Enhancement through Process-Induced Stress," *Proc. IEEE Conf. on Electron Devices and Solid-State Circuits*, pp. 43-46, 2005.

[11] Shimizu *et al.*, "Local mechanical-stress control (LMC): a new technique for CMOS-performance enhancement," *Proc. IEDM*, 2001, pp. 433-436.

[12] Y. Kanda, "A Graphical Representation of the Piezoresistance Coefficients in Silicon," IEEE Transactions on Electron Devices, Ed. 29, No. 1, 1982, pp. 64-70.