# The Impact of Shallow Trench Isolation Effects on Circuit Performance

Sravan K. Marella and Sachin S. Sapatnekar
Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455

***Abstract*— In nanometer technologies, shallow trench isolation (STI) induces thermal residual stress in active silicon due to post-manufacturing thermal mismatch. The amount of STI around an active region depends on the layout of the design, and the biaxial stress due to STI results in placement-dependent variations in the the transistor mobilities and threshold voltages of the active devices. An analytical model based on inclusion theory in micromechanics is employed to accurately estimate the stresses and the strains induced in the active region by the surrounding STI in the layout. The induced changes in mobility and threshold voltage changes are computed at the transistor level, and then propagated to the gate and circuit levels to predict circuit-level delay and leakage power for a given placement.**

***Key Terms* : Shallow Trench Isolation, Static Timing Analysis, Analytical Model, Inclusion Theory**

## I. Introduction

In nanometer technologies, shallow trench isolation (STI) is used to isolate active transistor regions in the layout. In typical fabrication technologies, *shallow* blocks of STI, made of $SiO_2$, are inserted into a much larger three-dimensional silicon structure. Figure 1(a) shows a representative layout of a standard cell showing a 2D view of STI.

During manufacturing, the STI oxide is grown from Si around an active region at a temperature of 1000°C using oxidation. When the chip returns to room temperature, the unequal coefficients of thermal expansion (CTEs) of $SiO_2$ and Si result in an unintentional residual thermal stress in the active Si. This stress can affect the mobility and threshold voltage of the transistors, and hence the circuit performance. The work in [1] documents the impact of STI stress and shows that the PMOS (NMOS) delay of a CMOS inverter improves (degrades) by about 17% (8%) when moved from a denser layout region with many surrounding gates to a sparser region with no neighbours.

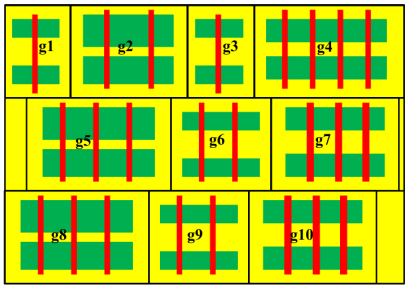



Fig. 1. A segment of a circuit layout showing how the STI in adjacent cells, or in gaps between cells, imply that the shape of an STI region depends on the layout of neighboring cells.

This STI-induced stress, and hence its performance impact, is highly layout-dependent since STI surrounds and abuts the active region in the physical layout in nonuniform ways. Therefore, the amount of STI around a transistor is determined by the relative locations and layouts of its neighbouring cells. For instance, to evaluate the stress affecting gate g6 in the middle row in the Figure 1, we must consider STI contributions from its eight neighbours g2 through g10, and also the STI within g6.

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

Therefore, STI stress can only be correctly evaluated after layout. In theory, it may be possible to precharacterize the stress by parameterizing the layout of the neighbors of a cell, but the number of cases to be characterized for all possible neighbors can be large. In the published literature [1], [2], the only known accurate method involves computationally expensive finite element simulation for each transistor, which is impractical for layouts of realistic-sized circuits.

An alternative to finite element simulations involves the use of analytical models, which can be evaluated fast enough to permit the analysis of large layouts. Much of the literature in this area [3]–[6] is based entirely on the use of one-dimensional models that account for stress components only along the longitudinal direction (i.e., along the channel direction). However, finite element simulations in [1], [2] show that STI stress in the transverse direction, perpendicular to the channel direction, also impacts the circuit performance. Furthermore, [3]–[6] use only a single component of the stress tensor for performance evaluation, while the entire stress tensor must be evaluated to accurately analyze STI-induced circuit performance variation. The work in [7] uses both longitudinal and transverse direction STI contributions, but is based on an empirically fitted model that is not scalable for nonrectangular shaped active/STI regions.

In this work, we present an analytical method to accurately capture the effects of STI on circuit performance for a given layout, taking into account the three-dimensional geometry of the STI together with its nonrectangular shape around an active region. Specifically, we

- model the effects of STI using a three-dimensional stress model based on inclusion theory in micromechanics,
- translate STI-induced stress effects into corresponding transistor mobility and threshold voltage variations.
- capture the dependencies of gate delay and leakage variations on placement for single and multifingered standard cells, and
- analyze the impact of STI on circuit timing and leakage power.

The paper is organized as follows. In Section II, we describe the electrical effects of STI stress, and determine the precise stress and strain components that must be modeled. Next, a stress modeling approach based on results in inclusion theory is described in Section III. In Section IV, we see how all of this information is drawn together to evaluate performance. The results of our method are presented in Section V, followed by concluding remarks.

## II. Electrical Effects of STI-induced Stress

Applied mechanical stress causes changes in transistor electrical properties, specifically in the *mobility* and the *threshold voltage*. Mobility variations are caused by the piezoresistive behavior of silicon, while threshold voltage variations occur due to changes in electronic band potentials due to applied stress. The induced changes in the mobility and threshold voltage can be expressed in terms of the stress and strain tensor, which characterize the mechanical stress and are described in greater detail in Section III.

### A. Variation of Mobility with Stress

According to piezoresistivity, an applied mechanical stress causes changes in resistivity and hence the mobility of the transistors. Most

integrated circuits are manufactured on wafers with their channels parallel or perpendicular to [110] silicon crystal orientation, which also corresponds to the wafer flat direction [8]. The axis perpendicular to the wafer surface usually corresponds to (001) Si crystal orientation. Thus a natural coordinate system would be along [110], $[\bar{1}10]$ and [001] [8], which corresponds to a 45° rotation of the Cartesian coordinate system. A complete mathematical model for piezoresistivity has been presented and demonstrated in silicon in [8]. The relative change in mobility for transistors oriented along [110] crystallographic direction is given as:

$$\frac{\Delta\mu'}{\mu'} = \pi'_{11}\sigma^{total}_{x'x'} + \pi'_{12}\sigma^{total}_{y'y'} \tag{1}$$

Here, $\pi'_{11}$ and $\pi'_{12}$ are the piezoresistive coefficients in $[110]-[\bar{1}10]$ coordinate system. The values of the piezoresistive coefficients are given in Table I. Here, $\sigma_{x'x'}$ and $\sigma_{y'y'}$ are two primary components of the stress tensor that significantly affect the transistor mobilities.

TABLE I
PIEZORESISTIVITY COEFFICIENTS (X$10^{-12}$ Pa$^{-1}$) IN [100] SI [9]

| | $\pi_{11}$ | $\pi_{12}$ | $\pi_{44}$ | $\pi'_{11}$ | $\pi'_{12}$ | $\pi'_{44}$ |
|---|---|---|---|---|---|---|
| NMOS | 1022.0 | −537.0 | 136.0 | 310.5 | 174.5 | 1559.0 |
| PMOS | −66.0 | 11.0 | −1381.0 | −717.5 | 662.5 | −77.0 |

### B. Variation of Threshold Voltage with Stress

According to deformation potential theory [10], [11], mechanical strain in the channel causes shifts and splits in conduction and valence band potentials. This results in corresponding shifts in the threshold voltage of the transistors and can be attributed to changes in silicon electron affinity, band gap, and valence band density of states. The changes in conduction and valence band potentials are given by [10]:

$$\begin{aligned}
&\Delta E_C^{(i)}(\epsilon) = \Xi_d(\epsilon_{xx}+\epsilon_{yy}+\epsilon_{zz}) + \Xi_u\epsilon_{ii}, i \in \{x,y,z\} \\
&\Delta E_V^{(hh,lh)}(\epsilon) = a(\epsilon_1+\epsilon_2+\epsilon_3) \\
&\pm\sqrt{\frac{b^2}{4}(\epsilon_{xx}+\epsilon_{yy}-2\epsilon_{zz})^2 + \frac{3b^2}{4}(\epsilon_{xx}-\epsilon_{yy})^2 + d^2\epsilon_{xy}^2}
\end{aligned} \tag{2}$$

Here, $\Delta E_C^{(i)}$ is the change in the conduction band potential energy in the carrier band number $i$. The term $E_V^{hh}$ [$E_V^{lh}$] denotes the heavy-hole [light-hole] valence band potential. The positive [negative] sign is used for $E_V^{hh}$ [$E_V^{lh}$]. The terms $\Xi_d$ and $a$ are the hydrostatic deformation potential constants and the terms $\Xi_u$, $b$, and $d$ are the shear splitting deformation potential constants whose values are given in Table II. The terms $\epsilon_i, i \in \{1,\cdots,t\}$ correspond to the six strain components in the Cartesian coordinate system, and correspond to $\epsilon_{xx}$, $\epsilon_{yy}$, $\epsilon_{zz}$, $2\epsilon_{yz}$, $2\epsilon_{zx}$, and $2\epsilon_{xy}$, respectively. The expressions for the strains are given in the Section III-C.

TABLE II
BAND EDGE DEFORMATION POTENTIAL CONSTANTS [10]

| $\Xi_d$ (eV) | $\Xi_u$ (eV) | $a$ (eV) | $b$ (eV) | $d$ (eV) |
|---|---|---|---|---|
| 1.13 | 9.16 | 2.46 | −2.35 | −5.08 |

The threshold voltage is a function of band-gap potential and thus can be expressed as a function of the changes in conduction band and valence band potentials. In this work, the changes in the electronic band potentials are dependent on the STI-induced stains. Ignoring negligible contributions from density of states changes [11],

$$\begin{aligned}
q\Delta V_{thp} &= m\Delta E_C - (m-1)\Delta E_V \\
q\Delta V_{thn} &= m\Delta E_V + (m-1)\Delta E_C
\end{aligned} \tag{3}$$

where $\Delta V_{thp}$ and $\Delta V_{thn}$ are the changes in PMOS and NMOS threshold voltages, respectively. The term $q = 1.6\times10^{-19}$C represents the electron charge and the term $m$ is the body-effect coefficient with a typical value of 1.3 to 1.4. The term $\Delta E_C$ is the minimum of the changes in conduction band potentials, $\Delta E_C^i$ and the term $\Delta E_V$ denotes the maximum of the changes in valence band potentials, $\Delta E_V^{hh}$ and $\Delta E_V^{lh}$. It should be noted that $\Delta E_C$ [$\Delta E_V$] is always negative [positive] valued and denotes conduction [valence] band lowering [raising] with STI stress. This reduces the band-gap and improves th threshold voltages of both NMOS and PMOS.

## III. STRESS MODELING

As seen in Section II, the changes in electrical properties require the computation of specific components of the STI-induced stress in Si: specifically, we must determine the two components $\sigma_{x'x'}$ and $\sigma_{y'y'}$ of the stress tensor, as well as the six strain tensor components.

In the Manhattan geometries employed in chip design, STI shapes are rectilinear. In this work, we work directly with three-dimensional cuboidal shapes by employing *inclusion theory* from micromechanics [12] to estimate the stresses and strains in the active silicon arising due to thermal mismatches with cuboidal STI shapes that have finite sizes in three dimensions. In micromechanics, an inclusion is a subdomain with an initial strain embedded in a larger domain, either having similar or dissimilar mechanical properties.

We will first present a solution to the basic problem of finding the stress due to a cuboidal STI structure, with finite dimensions along all three coordinate axes, embedded in silicon. However, general STI geometries may be have arbitrary three-dimensional rectilinear shapes, as observed in Figure 1. It is common practice [13] in micromechanics to divide an arbitrary shaped inclusion into smaller substructures and use linear superposition to find the total stress. Here, a general STI geometry is as a union of smaller cuboidal shapes, whose stress and strain contributions are superposed.

### A. Notation and Fundamental Equations of Elasticity

Before we develop the stress model, we describe the notation and the fundamental equations used in describing a stress state. In this paper, all materials are assumed to be isotropic and homogeneous. We employ the standard concise Einstein notation, where repeated indices imply summation, and we represent the three coordinate axes as $(x_1, x_2, x_3)$, respectively. In general, to obtain the stress state of a mechanical system, we need 15 components:

- six unique stress components $\sigma_{ij}$ (stress tensor),
- six unique strain components $\epsilon_{ij}$ (strain tensor), and
- three displacements $u_i$ (displacement tensor)

where $i,j \in \{x_1,x_2,x_3\}$ for any orthogonal coordinate system. The 15 unknowns are determined by solving the following 15 equations:

- 6 stress-strain equations (Hooke's Law):

$$\sigma_{ij} = C_{ijkl}(\epsilon_{kl} - \delta_{kl}\alpha\Delta T) \tag{4}$$

- 6 strain-displacement equations:

$$\epsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial j} + \frac{\partial u_j}{\partial i}\right) + \delta_{ij}\alpha\Delta T \tag{5}$$

- 3 force-balance equations:

$$\frac{\partial\sigma_{ix_1}}{\partial x_1} + \frac{\partial\sigma_{ix_2}}{\partial x_2} + \frac{\partial\sigma_{ix_3}}{\partial x_3} + B_i = 0 \tag{6}$$

Here, $i,j,k,l \in \{x_1,x_2,x_3\}$, $\delta_{ij}$ is Kronecker's delta function, $\alpha$ denotes the coefficient of thermal expansion, $\Delta T$ refers to the change in temperature, and $B_i$ is the external body force. The values of the physical constants used in this work are given in Table III.

TABLE III
PHYSICAL CONSTANTS FOR STRESS COMPUTATION

| | E (GPa) | CTE (ppm/°C) | $\nu$ |
|---|---|---|---|
| Silicon | 162.0 | 3.05 | 0.28 |
| $SiO_2$ | 71.7 | 0.51 | 0.16 |

The $C_{ijkl}$ elements here represent the components of the stiffness tensor and is a function of Young's modulus $E$ and Poisson's ratio $\nu$ of the material. The nonzero components are given below:

$$C_{1111} = C_{2222} = C_{3333} = \frac{E(1-\nu)}{(1+\nu)(1-2\nu)}$$
$$C_{1122} = C_{2233} = C_{1133} = \frac{E\nu}{(1+\nu)(1-2\nu)}$$
$$C_{2211} = C_{3322} = C_{3311} = \frac{E\nu}{(1+\nu)(1-2\nu)}$$
$$C_{1212} = C_{3131} = C_{2323} = \frac{E}{2(1+\nu)} \quad (7)$$

The solution of Equations (5) and (6) depends upon the geometry and boundary conditions of the mechanical system. The Equation (4) purely depends upon the material under consideration.

When the body forces $B_i, i \in \{x_1, x_2, x_3\}$ are zero, it can be shown that the displacements or stresses can be represented in terms of a function $\Phi$ that satisfies the relation:

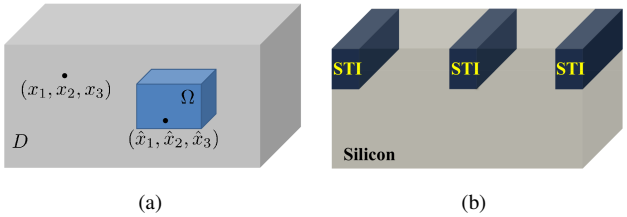
$$\nabla^4 \Phi = 0 \quad (8)$$

The solution to the system of elasticity equations can be found in terms of a *biharmonic function*, $\Phi$, that satisfies the specified boundary conditions of the system. A biharmonic [harmonic] function is a function whose fourth [second] order partial derivative is zero. This useful result has been exploited in micromechanical stress modeling to deduce the stress state for complex geometries. In a displacement formulation [stress formulation] the displacement [stress] is equated to the second partial derivative of a biharmonic function that satisfies the boundary conditions [14]. Once the displacement [stress] is known, the other unknowns of the stress state can be determined from Equations (4), (5), and (6). For the rest of this section, the terms qualified by a superscript $M \in \{\text{Si, SiO}_2\}$ that refers to the terms corresponding to the material $M$.

### B. The Inclusion Problem in Micromechanics

In continuum mechanics, inelastic strains are those that occur even in the absence of external body forces and thus can never be removed. Residual strains such as thermal mismatch strains, initial strains, and misfit strains (due to crystal defects) are examples of inelastic strains. In micromechanics such strains as termed as eigenstrains [12]. The six possible eigenstrains in any coordinate system $(x_1, x_2, x_3)$ are denoted by $e_{ij}$ for $i, j \in \{x_1, x_2, x_3\}$.

Furthermore, any subdomain $\Omega$ having an initial nonzero eigenstrain, embedded in a domain $D$ with zero initial eigenstrains, and either having similar or dissimilar mechanical properties, is known as a mechanical inclusion. Figure 2(a) shows an example of a cuboidal inclusion embedded in a semi-infinite space. A homogeneous [inhomogeneous] inclusion is one with domain $D$ and subdomain $\Omega$ having similar [dissimilar] mechanical properties. The domain has typically much larger dimensions as compared to the subdomain. The inclusion problem in micromechanics finds the stress state of such a system. There is a rich body of work on this class of problems in micromechanics [13], [15]–[17].




(a) (b)

Fig. 2. (a) A general inclusion in half-space. (b) STI as a cuboidal inclusion.

Shallow trench isolation (STI) is made up of $SiO_2$ and is embedded in silicon at a high temperature of 1000°C. The thickness of STI is of the order of few hundreds of nanometers, while the thickness of silicon substrate is typically of the order of several tens or hundreds of micrometers. Figure 2(b) shows three STI inclusions in silicon.

After manufacturing, owing to the CTE mismatch, seen in Table III, between Si and $SiO_2$, there is a residual thermal stress induced in active silicon. Compared to when it was manufactured, STI is comparatively smaller in volume to the silicon substrate and causes inelastic thermal strains, and it can be considered as an inhomogeneous inclusion within Si. In general, an STI structure is in the form of an arbitrary rectilinear shape, and we decompose this shape STI into elementary cuboidal shapes and superpose known solutions for cuboidal inclusion problems. Thus, we can treat STI as a cuboidal inclusion and obtain the effective eigenstrains in silicon by following a series of fictitious mechanical operations, as is the case with most inhomogeneous inclusion problems [12].

Summarizing the procedure for analyzing an STI inclusion in Si,

1) We first conceptually "remove" the STI from substrate at $T = 1000°\text{C}$ and allow both STI and the silicon substrate to undergo thermal contraction to room temperature, i.e., 25°C. This implies that $\Delta T$ = 975°C can be used in the stress formulation. The thermal strains in STI and silicon are $\epsilon_{ij}^{T(SiO_2)} = \delta_{ij}\alpha^{SiO_2}\Delta T$ and $\epsilon_{ij}^{T(Si)} = \delta_{ij}\alpha^{Si}\Delta T$, respectively. Since the inclusion (STI) as well as the domain (silicon) undergo free thermal contractions, the stresses in both materials are zero.
2) Next, we apply a fictitious tensile force of $F_{ij}^{SiO_2} = C_{ijkl}^{SiO_2}\epsilon_{ij}^{T(SiO_2)}$ on the STI inclusion and a fictitious compressive force of $-F_{ij}^{Si} = -C_{ijkl}^{Si}\epsilon_{ij}^{T(Si)}$ on silicon to bring them to original shapes.
3) The $SiO_2$ is now considered to be welded back into the silicon and the fictitious forces are removed and are replaced by an effective force applied on the insides of the silicon domain of $\Delta F_{ij} = F_{ij}^{Si} - F_{ij}^{SiO_2}$. $\Delta F_{ij}$ is the equivalent force applied by a homogeneous inclusion with a initial strain.
4) The equivalent eigenstrain due to this equivalent force in silicon is given by $e_{ij}^{Si} = {C_{ijkl}^{Si}}^{-1}\Delta F_{ij}$.

### C. Galerkin Vector Function Based Stress Formulation

From Section III-A, in the absence of body forces, the system of elasticity equations are reduced to a biharmonic equation. Using displacement potential theory, the elastic displacement can be expressed as a second partial derivative of a single vector function, the Galerkin vector function [13]. Elastic strains and stresses can be deduced from Equations (4) and (5). The form of these potentials depends on the geometry of the exterior domain and the inclusion subdomain.

In a general coordinate system, any point can be represented by a tuple $(x_1, x_2, x_3)$ and the corresponding position vector is denoted by $\mathbf{x}$. The points in an inclusion are known as source points and the points in the domain are known as observation points. We are interested in computing the stress state at the observation points. Let $(\hat{x}_1, \hat{x}_2, \hat{x}_3)$ denote a point in the source subdomain; the corresponding position vector is denoted by $\hat{\mathbf{x}}$. The elastic displacements $u_i$ and stress components $\sigma_{ij}$ due to eigenstrains $e_{ij}, i, j \in \{x_1, x_2, x_3\}$ in terms of a Galerkin vector function $\mathbf{\Phi}(\mathbf{x})$ are given by [13]:

$$\begin{aligned} 2\mu u_i(\mathbf{x}) &= 2(1-\nu)\Phi_{i,jj} - \Phi_{k,ki} \\ \sigma_{ij}(\mathbf{x}) &= \nu\Phi_{k,kmm}\delta_{ij} - \Phi_{k,kij} + \nu(\Phi_{i,kkj} + \Phi_{j,kkj}), \mathbf{x} \notin \Omega \\ \sigma_{ij}(\mathbf{x}) &= \nu\Phi_{k,kmm}\delta_{ij} - \Phi_{k,kij} + \nu(\Phi_{i,kkj} + \Phi_{j,kkj}) \\ &\quad -2\mu e_{ij} - \lambda e_{kk}\delta_{ij}, \mathbf{x} \in \Omega \end{aligned} \quad (9)$$

Here, $\mu$ and $\lambda$ are the elastic Lamé constants given in Table IV. The Galerkin vector function $\mathbf{\Phi}(\mathbf{x})$ is biharmonic and satisfies $\nabla^4\mathbf{\Phi}(\mathbf{x}) = 0$, and is in turn a function of elementary Galerkin vectors composed

of biharmonic and harmonic potential functions. It is chosen so that it satisfies two primary boundary conditions of the inclusion problem:

- all components of stress should vanish at infinite distance from the inclusion, $\sigma_{ij}^{D}(\infty) = 0$ for $i, j \in \{x_1, x_2, x_3\}$.
- there should be a displacement continuity across the inclusion and domain boundary. $u_i^{\Omega} = u_i^{D}$ for every $i \in \{x_1, x_2, x_3\}$.

A general solution for a cuboidal inclusion has been presented in [13]. The work presents a detailed mathematical framework based on Galerkin vector formulation. The general solution in [13] can predict the stress state at every point in the domain for an any given eigenstrain tensor. For the STI-induced thermal stress problem, further simplifications are possible based on two observations:

- For a thermal stress problem, only the normal components of the eigenstrain tensor are present, $e_{ij}^{Si} \neq 0$ for $i = j$; zero otherwise.
- Since STI is near the surface of silicon and electrical current flows near the device surface, $z_1 = 0$ for the observation points.

Making use of these ensuing simplifications, we obtain closed-form expressions for the major stress and strain components used in computing electrical variations as seen in Section II. As pointed out in Section II-A, since integrated circuits are manufactured in the primed coordinate system, $(x_1, x_2, x_3)$ can be replaced by $(x', y', z')$ to represent the stress and strain tensor components in this primed system. The strain components in Cartesian coordinate system can be obtained by Hooke's Law and by appropriate coordinate transformations. For an cuboidal inclusion whose coordinates are described by the closed intervals, $\hat{x}' \in [a_1, a_2]$, $\hat{y}' \in [b_1, b_2]$, and $\hat{z}' \in [c_1, c_2]$, the final closed-form expressions are given in Table IV in terms of elementary functions and constants. The constant $C^{\sigma}$ denotes the multiplicative constant for the stress components.

To obtain the overall STI impact, we divide the STI in the transverse and longitudinal directions around an active region into nonintersecting cuboidal shapes and use the solution presented in Table IV. We apply linear superposition and add all contributions from the adjoining STI to find the total stress and strains:

$$\sigma_{ij}^{total} = \sum_{STI} \sigma_{ij}^{Si}; \qquad \epsilon_{ij}^{total} = \sum_{STI} \epsilon_{ij}^{Si} \tag{10}$$

### D. Comparison with the Finite Element Method

To verify the accuracy of the analytical stress model and the validity of linear superposition we perform finite element (FE) simulations using ABAQUS [18] on representative active silicon regions surrounded by STI ($SiO_2$) on all sides. To demonstrate the effectiveness of the superposition we use an irregular shaped active region as shown in Figure 3. We consider four diffusion connected transistors T1, T2, T3, and T4. This represents the series pull-down NMOS transistors of a NAND4 gate with T1 being closest to the output. Each active region (green) is about 250 nm wide. The electrical widths or the physical heights of the transistors are: W(T1) = 100nm, W(T2) = 200nm, W(T3) = 300nm, and W(T4) = 400nm. The channel length is 50nm. The boundary of the STI is 1600nm $\times$ 1200nm. We decompose these STI regions into smaller cuboids as shown in the top view in Figure 3. We then apply our model described in Section III-C and use linear superposition to add contributions from each STI cuboid. The resultant stress components probed under the channel region below the poly (red) and are shown in Figure 4. Our analytical model provides a good match even for nonrectangular active or STI regions. Table V compares the NAND4 FO4 fall-time delays in a 45nm technology for low-to-high transitions on inputs of each of the transistors, obtained using our analytical stress model and the FEM model. The delays are computed using HSPICE. It can be seen that although the FEM stress can be different from the analytical models, the delay error in using our analytical model compared to the FEM model is well under 1%.

TABLE IV

STRESS AND STRAIN TENSOR COMPONENTS

Stress components used in mobility computations

$$\begin{aligned}
\sigma_{x'x'} &= C^{\sigma}\Big[(2+4\nu^{Si})\phi_1 + (6-4\nu^{Si}-8(\nu^{Si})^2)\bar{\phi}_1 + 2\nu^{Si}\phi_2 \\
&\quad -2\nu^{Si}\bar{\phi}_2 + 2\nu^{Si}\phi_3 - 2\nu^{Si}(5+4\nu^{Si})\bar{\phi}_3\Big]_{x_1-a_2, x_2-b_2, x_3\pm c_2}^{x_1-a_1, x_2-b_1, x_3\pm c_1} \\
\sigma_{y'y'} &= C^{\sigma}\Big[(2+4\nu^{Si})\phi_2 + (6-4\nu^{Si}-8(\nu^{Si})^2)\bar{\phi}_2 + 2\nu^{Si}\phi_1 \\
&\quad -2\nu^{Si}\bar{\phi}_1 + 2\nu^{Si}\phi_3 - 2\nu^{Si}(5+4\nu^{Si})\bar{\phi}_3\Big]_{x_1-a_2, x_2-b_2, x_3\pm c_2}^{x_1-a_1, x_2-b_1, x_3\pm c_1} \\
\sigma_{x'y'} &= C^{\sigma}\Big[(2+2\nu^{Si})\chi + (6-2\nu^{Si}-8(\nu^{Si})^2)\bar{\chi} - \psi - (3-4\nu^{Si})\bar{\psi} \\
&\quad +4(1-2\nu^{Si})(1-\nu^{Si})\bar{\eta}\Big]_{x_1-a_2, x_2-b_2, x_3\pm c_2}^{x_1-a_1, x_2-b_1, x_3\pm c_1}
\end{aligned}$$

Strain components used in threshold voltage computations

$$\begin{aligned}
\epsilon_{xx} &= \frac{1}{2E^{Si}}[(1-\nu^{Si})(\sigma_{x'x'}+\sigma_{y'y'}) + 2(1+\nu^{Si})\sigma_{x'y'}] \\
\epsilon_{yy} &= \frac{1}{2E^{Si}}[(1-\nu^{Si})(\sigma_{x'x'}+\sigma_{y'y'}) - 2(1+\nu^{Si})\sigma_{x'y'}] \\
\epsilon_{xy} &= \frac{(1+\nu^{Si})}{2E^{Si}}[\sigma_{y'y'} - \sigma_{x'x'}] \\
\epsilon_{zz} &= \epsilon_{zx} = \epsilon_{zy} = 0
\end{aligned}$$

Elementary functions and constants

$$\phi_1 = -\tan^{-1}\left(\frac{\xi_2\xi_3}{\xi_1 r}\right);\ \phi_2 = -\tan^{-1}\left(\frac{\xi_1\xi_3}{\xi_1 r}\right);\ \phi_3 = -\tan^{-1}\left(\frac{\xi_1\xi_2}{\xi_3 r}\right)$$

$$\bar{\phi}_1 = -\tan^{-1}\left(\frac{\xi_2\bar{\xi}_3}{\xi_1 \bar{r}}\right);\ \bar{\phi}_2 = -\tan^{-1}\left(\frac{\xi_1\bar{\xi}_3}{\xi_1 \bar{r}}\right);\ \bar{\phi}_3 = -\tan^{-1}\left(\frac{\xi_1\xi_2}{\bar{\xi}_3 \bar{r}}\right)$$

$$\chi = \log(r+\xi_3);\ \bar{\chi} = \log(\bar{r}+\bar{\xi}_3);$$

$$\psi = \frac{\xi_1^2+\xi_2^2}{r(r+\xi_3)} + \frac{\xi_3}{r};\ \bar{\psi} = \frac{\xi_1^2+\xi_2^2}{\bar{r}(\bar{r}+\bar{\xi}_3)} + \frac{\bar{\xi}_3}{\bar{r}};\ \bar{\eta} = \frac{\xi_1^2+\xi_2^2}{2(\bar{r}+\bar{\xi}_3)^2} + \frac{\bar{\xi}_3}{\bar{r}+\bar{\xi}_3}$$

$$r = \sqrt{\xi_1^2+\xi_2^2+\xi_3^2};\ \bar{r} = \sqrt{\xi_1^2+\xi_2^2+\bar{\xi}_3^2};$$

$$\xi_1 = x' - \hat{x}';\ \xi_2 = y' - \hat{y}';\ \xi_3 = z' - \hat{z}';\ \bar{\xi}_3 = z' + \hat{z}'$$

$$C^{\sigma} = \frac{\mu e^{Si}}{8\pi(1-\nu^{Si})};\ e^{Si} = \frac{1-2\nu^{Si}}{E^{Si}}\left(\frac{E^{Si}\alpha^{Si}\Delta T}{1-2\nu^{Si}} - \frac{E^{SiO_2}\alpha^{SiO_2}\Delta T}{1-2\nu^{SiO_2}}\right)$$

$$\mu^{M} = \frac{E^{M}}{2(1+\nu^{M})};\ \lambda^{M} = \frac{E^{M}\nu^{M}}{(1+\nu^{M})(1-2\nu^{M})}, \text{for} M \in \{Si, SiO_2\}$$

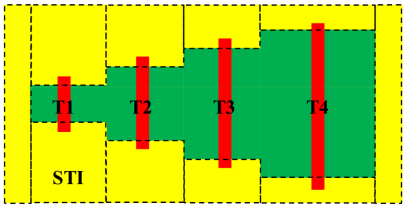



Fig. 3. An irregular shaped active region in STI. The STI is fragmented into smaller cuboids (rectangles in 2D) around the active regions.

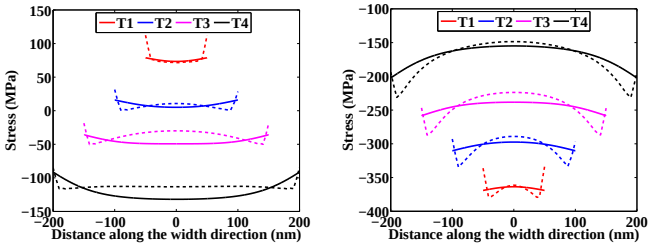



Fig. 4. Solid [dashed] lines showing our [FEM] model. (a) $\sigma_{x'x'}$ (b) $\sigma_{y'y'}$

TABLE V

DELAY COMPARISON BETWEEN FEM AND ANALYTICAL MODELS

| Transistor | **T1** | **T2** | **T3** | **T4** |
|---|---|---|---|---|
| FEM (ps) | 44.45 | 46.11 | 47.61 | 48.04 |
| Analytical (ps) | 44.62 | 46.23 | 47.7 | 48.06 |
| Error (%) | 0.38% | 0.26% | 0.19% | 0.04% |

## IV. CIRCUIT PERFORMANCE EVALUATION

Using the methods described in Sections II and III, for a given layout, the changes in the device mobility and threshold voltage can been computed for each transistor along the width of the transistor. We use the average of the electrical variations along the transistor width. We then evaluate the variations in circuit performance by conducting static timing analysis and leakage power analysis.

For a gate with $n$ transistors, the delay under variations in the threshold voltage $V_{th,i}^{str}$ and mobility $\mu_i^{str}$ for the $i^{\text{th}}$ transistor, $1 \leq i \leq n$, can be computed using a first-order Taylor expansion:

$$D^{str} = D^0 + \sum_{i=1}^{n} \left( \frac{\partial D}{\partial \mu_i}\bigg|_0 \Delta\mu_i^{str} + \frac{\partial D}{\partial V_{th,i}}\bigg|_0 \Delta V_{th,i}^{str} \right) \quad (11)$$

where $D^{str}$ is the total gate delay due to STI stress, $D^0$ is the nominal delay of the gate without any electrical variations, and the partial derivatives of delay with $\mu_i$ and $V_{th,i}$ denote the delay sensitivity of the gate to the mobility and threshold voltage, respectively, of transistor $i$, computed at the nominal point.

The leakage power of a transistor exponentially increases (decreases) with its decreasing (increasing) threshold voltage. However, for small changes in threshold voltage of a transistor, the gate-level leakage power varies almost linearly. STI-induced threshold voltage variations in transistors are typically few tens of millivolts, while the nominal threshold voltage of a transistor is about 400 mV in this work. Thus the leakage power of a gate under unequal changes in threshold voltages of $n$ transistors of a gate can also be computed using a first order Taylor series expansion as:

$$L_{gate}^{str} = L_{gate}^0 + \sum_{i=1}^{n} \frac{\partial L_{gate}}{\partial V_{thi}}\bigg|_0 \Delta V_{thi}^{str} \quad (12)$$

where $L_{gate}^{str}$ is the leakage power of a gate under STI-induced stress and $L_{gate}^0$ is the nominal leakage power of the gate under no stress. The partial derivative of $L_{gate}$ with $V_{thi}$ represents the sensitivity of the leakage current of the gate to changes in the threshold voltage of transistor $i$, evaluated at the nominal point. Our relative error in computing leakage power of standard cells in this work is under 1%.

For a given placement, we use the analytical framework developed so far to compute the circuit performance as follows:

- From the layout information for a circuit, we recover the STI configuration affecting the transistors within each standard cell. We then compute the stress using the models in Section III.
- Based on the stress computations, we then proceed to compute the changes in mobility and threshold voltage of each transistor using Equations (1) and (3), respectively.
- Knowing the changes in electrical parameters of individual transistors in a logic gate, we compute the delay and leakage power using Equations (11) and (12), respectively.
- We then perform static timing analysis and leakage computation.

## V. RESULTS

Shallow trench isolation effects are highly layout-dependent. The magnitude of electrical variation in a standard cell depends on its layout, and its relative position to its neighbours and their layouts. We apply our methods on a set of IWLS benchmarks [19], listed in Table VI, where H [W] represents the height [width] of the layouts, #PO denotes the number of primary outputs, and $D_0$ [$L_0$] denote the critical path delay [leakage power] without STI effects.

TABLE VI

IWLS 2005 [19] CIRCUITS

| Ckt. | Index | # Gates | H×W ($\mu$m×$\mu$m) | # POs | $D_0$ (ps) | $L_0$ ($\mu$W) |
|---|---|---|---|---|---|---|
| ac97_ctrl | C1 | 9047 | 92 × 171 | 4204 | 429 | 298 |
| aes_core | C2 | 11346 | 64 × 259 | 12313 | 418 | 226 |
| des | C3 | 4443 | 50 × 178 | 332 | 870 | 177 |
| ethernet | C4 | 27060 | 184 × 242 | 32149 | 644 | 562 |
| i2c | C5 | 1110 | 23 × 76 | 204 | 389 | 35 |
| mem_ctrl | C6 | 8860 | 78 × 201 | 2522 | 842 | 251 |
| pci_bridge32 | C7 | 9988 | 92 × 200 | 9025 | 636 | 325 |
| spi | C8 | 3216 | 60 × 117 | 564 | 693 | 117 |
| systemcdes | C9 | 2600 | 48 × 119 | 549 | 694 | 118 |
| usb_funct | C10 | 10667 | 79 × 201 | 3930 | 624 | 248 |

Our standard cell layouts are based on the 45nm Nangate standard cell library [20]. The cells consist of gates with single-, two-, and four-fingered layouts. The standard cells are characterised for different load capacitances and input slopes at a supply voltage of 1.0V and a temperature of about 25$^o$C. Since STI is manufactured at 1000$^o$C, it can be noted that the $\Delta T$ is almost the same over the operating range of temperatures. We employ Capo [21] to obtain legalized placements of the IWLS circuits. From the circuit placement information and active layer information of the standard cell layouts, STI information is extracted as a set of nonintersecting cuboids around the active region. We then employ our analytical stress model from Section III to compute the stress in the active transistor regions.

In the rest of the section, the STI along the active width [height] direction is termed as longitudinal [transverse] STI. Tensile [compressive] stress indicates stress is positive [negative] valued. Using the techniques in Section IV, we perform static timing analysis and leakage power analysis on the circuits under three conditions:

- Nominal: STI effects in the layout are ignored.
- 3D STI: Our 3D stress model, superposing effects from STI rectangles in transverse and longitudinal directions, is used.
- 1D STI: Only the effects of STI rectangles in the longitudinal direction are considered and transverse effects are ignored.

Note that our 1D approach is more accurate than conventional 1D models which assume uniformity in the $z$ direction, since it also considers finite depth effects along the $z$ axis.

TABLE VII

COMPARISON OF DELAY AND LEAKAGE POWER UNDER STI

| Ckt. | 3D STI | | | | | | 1D STI | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\Delta D_{3D}$ (%) | $\Delta L_{3D}$ (%) | $D_+$ (ps) | $\Delta D_+$ (%) | $D_-$ (ps) | $\Delta D_-$ (%) | $\Delta D_{1D}$ (%) | $\Delta L_{1D}$ (%) | $D_-$ (ps) | $\Delta D_-$ (%) |
| C1 | -5.3% | 24.7% | 108 | 15.7% | 381 | -8.7% | -8.3% | 16.9% | 370 | -11.9% |
| C2 | -3.9% | 32.6% | 173 | 2.9% | 335 | -9.6% | -6.1% | 26.5% | 327 | -12.5% |
| C3 | -4.1% | 23.2% | 354 | 2.0% | 568 | -8.1% | -5.7% | 15.8% | 541 | -11.3% |
| C4 | -2.2% | 33.2% | 434 | 1.6% | 496 | -8.9% | -5.7% | 26.6% | 530 | -12.3% |
| C5 | -6.6% | 26.5% | 192 | 10.4% | 356 | -9.0% | -9.1% | 18.7% | 345 | -12.5% |
| C6 | -5.2% | 27.8% | 473 | 1.3% | 731 | -8.1% | -7.1% | 20.6% | 345 | -12.5% |
| C7 | -3.7% | 26.8% | 350 | 1.1% | 538 | -11.5% | -6.1% | 18.5% | 521 | -15.2% |
| C8 | -2.2% | 24.1% | 476 | 2.7% | 540 | -8.1% | -3.3% | 17.0% | 520 | -12.5% |
| C9 | -2.1% | 21.4% | 458 | 2.6% | 622 | -5.0% | -4.1% | 14.2% | 607 | -7.6% |
| C10 | -4.3% | 30.6% | 289 | 1.7% | 460 | -8.3% | -6.3% | 23.4% | 511 | -10.4% |

Table VII shows the results under the 3D and 1D STI cases. The columns $\Delta D_{3D}$ and $\Delta L_{3D}$ [$\Delta D_{1D}$ and $\Delta L_{1D}$] provide the changes in critical path delay and leakage power, respectively, in the 3D STI [1D STI] case with respect to the nominal values, $D_0$ and $L_0$, from Table VI. Note that the critical path may not be identical in the nominal circuit and the stressed circuit. Here, positive [negative] changes denote increases [reductions] in the delay or leakage.

The above numbers only capture the delay changes in the worst-case path, where in all cases, the delay happens to reduce for our benchmark set: between $-2.1\%$ and $-6.6\%$ for 3D and between $-3.3\%$ and $-9.1\%$ for 1D. However, it is instructive to observe what happens on noncritical paths by examining the largest delay shifts, over all paths in a circuit, from the nominal to the stressed cases. Let $D_+$ and $D_-$, respectively, represent the delays (under stress)

of the paths in each circuit that show the largest delay increase and reduction. The corresponding maximum delay increases and reductions observed on these paths are denoted by $\Delta D_+$ and $\Delta D_-$. Note that for the 1D STI case, we only show $D_-$ and $\Delta D_-$ since only path delay reductions are observed, and no increases are seen, i.e., $\Delta D_+$ is uniformly zero in 1D. On the other hand, the value of $\Delta D_+$ varies from $1.1\%$ to $15.7\%$ for the 3D case. The values of $\Delta D_-$ range from $-5.0\%$ to $-11.5\%$ for 3D, and are overestimated in 1D where they lie in the range $-7.6\%$ to $-15.2\%$.

To understand these results, we further analyze the 1D and 3D stress cases to explain the observed trends in the data:

- When longitudinal STI is alone taken into account, as in the 1D case, the $\sigma_{x'x'}$ stress component is provably always compressive, while $\sigma_{y'y'}$ is tensile. Furthermore, the magnitude of $\sigma_{y'y'}$ is typically smaller than $\sigma_{x'x'}$. Consequently in the 1D STI case, from Equation (1) and the signs of $\pi'_{11}$ and $\pi'_{12}$ in Table I, PMOS [NMOS] mobility always improves [degrades].
- When transverse STI effects are also considered, as in the 3D case, in the $\sigma_{x'x'}$ component could be tensile or compressive, depending on the dimensions of the active region and the STI, while $\sigma_{y'y'}$ is seen to be compressive in practice, as observed in Fig. 4. Thus, for 3D STI, the PMOS mobilities may improve or degrade, while NMOS mobilities always degrade. Furthermore, the magnitudes of PMOS [NMOS] mobility variations in the 3D STI case are smaller [greater] than the 1D STI case.
- In determining the impact of stress on circuit delay, STI-induced threshold voltage reductions attenuate [fortify] increases [reductions] in the mobility. While PMOS and NMOS devices show similar levels of mobility shifts, the threshold voltage reductions for PMOS are much lower than for NMOS. Therefore, PMOS devices are mostly mobility-dominated, while NMOS device performance is determined by the balance between the shifts in mobility and threshold voltage. This is reflected at the circuit level in terms of the increase or reduction in path delays.
- Under STI effects, threshold voltages of both PMOS and NMOS transistors are lowered, and the reduction depends on the amount of surrounding STI (which is higher in the 3D case than the 1D case). It is observed that the magnitude of $\sigma_{y'y'}$ in the 3D STI case is significantly higher than the 1D STI for this reason. Therefore, the leakage power is seen to increase from the nominal case to either the 1D or 3D case. The shift the 3D STI [1D STI] formulation, $\Delta L_{3D}$ [$\Delta L_{1D}$] can vary from $21.4\%$ to $33.2\%$ [$14.2\%$ to $26.6\%$]. Thus, when STI effects are neglected, the leakage power can be significantly underestimated.

**Layout guidelines:** Based on the above analysis, for a given row-based placement, the following guidelines are obtained:

- To optimize delay, gates on critical/near-critical paths should have higher [smaller] longitudinal [transverse] spacing with respect to their neighbours in the same row [adjacent rows].
- To optimize leakage, noncritical gates should have minimum spacing with neighbours in the row (longitudinal STI). Spaces in the rows above/below (transverse STI) should be avoided.

## VI. Conclusion

We have developed an analytical framework to analyze the circuit performance under both longitudinal and transverse STI-induced stress variations. An accurate analytical stress model based on inclusion theory has been employed to find the stress state in silicon by modeling STI as a cuboidal inclusion, and closed-form expressions for stress are presented. Using the stress and strain tensor components thus generated, layout-dependent electrical variations in individual transistors are then computed. The gate delay and leakage power are subsequently evaluated for unequal variations in the constituent transistors, based on first-order Taylor series expansions. The circuit level timing and leakage power analysis is performed on ten IWLS layouts using our analytical models and is shown to be more accurate than existing approaches. Finally, layout guidelines for delay and leakage power optimization are provided.

## References


[1] V. Moroz *et al.*, "Stress-aware design methodology," in *Proceedings of the IEEE International Symposium on Quality Electronic Design*, pp. 806–812, March 2006.

[2] J. Xue *et al.*, "A framework for layout-dependent STI stress analysis and stress-aware circuit optimization," *IEEE Transactions on VLSI Systems*, vol. 20, no. 3, pp. 498–511, 2012.

[3] V. Joshi *et al.*, "Closed-form modeling of layout-dependent mechanical stress," in *Proceedings of the ACM/IEEE Design Automation Conference*, pp. 673–678, June 2010.

[4] A. Kahng, P. Sharma, and R. Topaloglu, "Chip optimization through STI-stress-aware placement perturbations and fill insertion," *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pp. 1241–1252, July 2008.

[5] R. Bianchi, G. Bouche, and O. Roux-dit Buisson, "Accurate modeling of trench isolation induced mechanical stress effects on mosfet electrical performance," in *IEEE International Electronic Devices Meeting*, pp. 117–120, 2002.

[6] B. T. Cline *et al.*, "STEEL: A technique for stress-enhanced standard cell library design," in *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pp. 691–697, 2008.

[7] X. Li *et al.*, "A two-dimensional analysis method on STI-aware layout-dependent stress effect," *IEEE Transactions on Electronic Devices*, vol. 59, no. 11, pp. 2964–2972, 2012.

[8] R. C. Jaeger *et al.*, "CMOS stress sensors on (100) silicon," *IEEE Journal of Solid-State Circuits*, vol. 35, pp. 85–95, 2000.

[9] K.-H. Lu *et al.*, "Thermomechanical reliability of through-silicon vias in 3D interconnects," in *IEEE International Reliability Physics Symposium*, pp. 3D.1.1–3D.1.7, 2011.

[10] J.-S. Lim, S. E. Thompson, and J. G. Fossum, "Comparison of threshold-voltage shifts for uniaxial and biaxial tensile-stressed n-MOSFETs," *IEEE Electron Device Letters*, vol. 25, pp. 731–733, November 2004.

[11] W. Zhang and J. G. Fossum, "On the threshold voltage of strained-Si-$Si_{1-x}Ge_x$ MOSFETs," *IEEE Transactions on Electron Devices*, vol. 52, pp. 263–268, February 2005.

[12] T. Mura, *Micromechanics of defects in solids*. The Hauge, The Netherlands: Martinus Nijhoff, 1987.

[13] S. Liu *et al.*, "Analytical solution for elastic fields caused by eigenstrains in a half-space and numerical implementation based on fft," *International Journal of Plasticity*, vol. 35, pp. 135–154, 2012.

[14] M. Saad, *Elasticity: Theory, Applications and Numerics*. Oxford, U.K.: Elsevier Academic Press, 2004.

[15] J. Eshelby, "The elastic field outside an ellipsoidal inclusion," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 252, no. 1271, pp. 561–569, 1959.

[16] R. Mindlin and D. Cheng, "Nuclei of strain in a semi-infinte solid," *Journal of Applied Physics*, vol. 21, pp. 926–930, 1950.

[17] Y. P. Chiu, "On the stress field and surface deformation in a half space with a cuboidal zone in which initial strains are uniform," *ASME Journal of Applied Mechanics*, vol. 45, pp. 302–306, 1978.

[18] "ABAQUS CAE Online Documentation." available at `http://www.sharcnet.ca/Software/Abaqus/6.11.2/index.html`.

[19] "IWLS 2005 Benchmarks." available at `http://www.iwls.org/iwls2005/benchmarks.html`.

[20] "Nangate Open Cell Library." available at `http://www.si2.org/openeda.si2.org/projects/nangatelib`.

[21] A. E. Caldwell, A. B.Kahng., and I. L. Markov, "Can recursive bisection alone produce routable, placements?," in *Proceedings of the ACM/IEEE Design Automation Conference*, pp. 477–482, 2000.