

# Chip Optimization Through STI-Stress-Aware Placement Perturbations and Fill Insertion

Andrew B. Kahng, *Senior Member, IEEE*, Puneet Sharma, *Member, IEEE*, and Rasit Onur Topaloglu, *Member, IEEE*

**Abstract**—Starting at the 65-nm node, stress engineering to improve the performance of transistors has been a major industry focus. An intrinsic stress source—shallow trench isolation (STI)—has not been fully utilized up to now for circuit performance improvement. In this paper, we present a new methodology that combines detailed placement and active-layer fill insertion to exploit STI stress for performance improvement. We conduct process simulation of a 65-nm production STI technology to generate mobility and delay impact models for STI stress. We then utilize these models to perform STI-stress-aware delay analysis of critical paths using Simulation Program with Integrated Circuit Emphasis (SPICE). We present our timing-driven optimization of STI stress in standard cell designs, using detailed placement perturbation and active-layer fill insertion to improve complementary metal–oxide–semiconductor performance. We assess the proposed analysis and optimization on small designs implemented with a 65-nm production cell library and a standard synthesis place-and-route flow. Our stress-aware timing analysis improves the clock frequency by 4.68% to 6.31% over traditional worst case analysis, and our optimization improves clock frequency by 2.44% to 5.26%. The frequency improvement through exploitation of STI stress comes at practically zero cost in terms of design area and wire length.

**Index Terms**—Design for manufacturing (DFM), performance analysis and optimization, shallow trench isolation (STI), stress modeling and optimization.

## I. INTRODUCTION

AT THE 65-nm process node and beyond, it is evident that stress- and strain-based techniques for mobility improvement will dominate traditional geometric scaling to maintain Moore's law trajectories for device performance. Enabling progress has been made in the manufacturing process and technology computer-aided design (TCAD, modeling and simulation) to support stress. However, stress has not yet been exploited by layout optimizations to improve design performance. In this paper, we present a new methodology that

Manuscript received August 21, 2007; revised November 20, 2007. This paper was recommended by Associate Editor J. Hu.

A. B. Kahng is with the Department of Computer Science and Engineering and the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093-0404 USA (e-mail: abk@ucsd.edu).

P. Sharma was with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093-0409 USA. He is now with Freescale Semiconductor, Austin, TX 78727 USA (e-mail: sharma@ucsd.edu).

R. O. Topaloglu is with the Department of Computer Science and Engineering, University of California at San Diego, La Jolla, CA 92093-0404 USA, and also with Advanced Micro Devices, Sunnyvale, CA 94088 USA (e-mail: rtopalog@cs.ucsd.edu; rasit.topaloglu@amd.com).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCAD.2008.923083

combines detailed placement and active-layer fill insertion to exploit shallow trench isolation (STI) stress for IC performance improvement. Our methodology begins with process simulation of a 65-nm production STI technology, from which we generate mobility and delay impact models for STI stress. We develop STI-stress-aware Simulation Program with Integrated Circuit Emphasis (SPICE) modeling and simulation of critical paths and finally perform timing-driven optimization of STI stress in standard cell designs, using detailed placement perturbation to optimize p-channel MOS (PMOS) performance and active-layer fill insertion to optimize n-channel MOS (NMOS) performance.

### A. Stress Modulation Techniques in Process

In 65-nm processes, a number of stress modulation techniques have been introduced:

- 1) SiGe stress from underneath the channel;
- 2) embedded SiGe (e-SiGe) from the source and drain;
- 3) stress liner;
- 4) stress memorization;
- 5) hybrid orientation.

Early stress modulation methods employed a silicon-germanium (SiGe) layer underneath the channel, which improves the channel mobility. More recently, e-SiGe [10], [11], [13] has been used in the source and drain regions to exert stress along the channel and improve PMOS speed. The stress liner technique involves deposition of stressed liners over the transistors on top of the polysilicon. Single stress liners may be used to cover the entire wafer with compressive or tensile liners. Alternatively, dual stress liners [15], [16] may be used to cover the NMOS devices with a tensile liner and the PMOS devices with a compressive one. The stress memorization technique [17], [18], which is typically used to improve NMOS speed, relies on plastic deformation of certain materials due to a process step and the consequent *memorization* of the applied stress in the channel. Finally, in the hybrid orientation technique [12], [14], crystal orientations are used to separately enhance NMOS and PMOS speeds.

STI stress is the stress that is exerted by STI wells on device active regions and is generally compressive in nature. Irrespective of the use of stress modulation techniques in the process, STI stress is not negligible, and its magnitude depends on the sizes of the STI wells and the active regions for a given process. In this paper, we present a technique that modulates stress at timing critical devices to improve circuit delay by altering the STI widths (STIW) adjacent to the devices.

TABLE I  
IMPACT OF STI WIDTH ON THE PERFORMANCE OF SEVERAL STANDARD CELLS

Cell	Spacing (nm)	PMOS $STIW_L$ (nm)	PMOS $STIW_R$ (nm)	NMOS $STIW_L$ (nm)	NMOS $STIW_R$ (nm)	R-Delay (ps)	F-Delay (ps)
INVDO	0um	140	140	110	110	27.27	21.96
	5um	5140	5140	5110	5110	23.65	23.70
BUFFDO	0um	140	140	125	125	45.56e	46.11
	5um	5140	5140	5125	5125	43.84	43.53
NR2D0	0um	140	140	110	110	51.12	23.06
	5um	5140	5140	5110	5110	42.77	24.69
ND2D0	0um	140	140	110	110	29.63	35.36
	5um	5140	5140	5110	5110	25.77	38.81

### B. Stress Modeling Techniques

In the area of stress modeling and characterization, Rueda [7] has provided general models for stress. Gallon *et al.* [22] has specifically analyzed the stress-induced by STI. Bradley *et al.* [23] have characterized the piezoresistance of complementary metal-oxide-semiconductor transistors. Sheu *et al.* have modeled well-edge proximity effect on MOSFETs [28] and the effect of mechanical stress on dopant diffusion [21]. Su *et al.* [24] have proposed a scalable model for the layout dependence of stress. Miyamoto *et al.* [29] have provided a layout-dependent stress analysis of STI. Recently, Tsuno *et al.* [30] has shown 65-nm silicon data showing that STIW stress effect can impact drive current by up to 10%. However, no models or optimization methodology is presented.

With respect to the STI process, several optimizations have been developed to reduce the STI stress, but they typically fail to completely eliminate layout-dependent stress impacts. Elbel *et al.* [19] have proposed an STI process flow based on selective oxide deposition. Lee *et al.* [20] has proposed an optimization for the densification of the STI fill oxide to reduce the stress. Miyamoto *et al.* [29] have also proposed process innovations to reduce the active-area layout dependence of MOSFET characteristics. Looking forward, the introduction of e-SiGe for certain 65-nm nodes in the source and drain may reduce the stress variation due to STIW for PMOS, but NMOS performance can still be actively improved by exploiting the STIW effect. If these STI stress controlling techniques are used, we expect the improvements demonstrated in this paper to reduce but remain nonnegligible.

Stress TCAD simulations have been conducted by Moroz *et al.* [25], [27] and by Smith [26]. The work of Moroz *et al.* is significant for indicating possible ways to enhance performance using STI stress; however, no circuit-level optimizations are explained. With respect to the current body of knowledge, models are still needed to relate stress due to the STIW effect to transistor mobilities, and there is still a lack of available stress optimization methods. A fundamental research goal is to develop novel and efficient simulation, modeling, analysis, and optimization methods to support next-generation stress-aware EDA technology. Our work strives to enable this.

### C. Our Focus: Exploitation of STIW Effect

STI is an important and well-studied stress source that has not been fully exploited for design quality improvement until

now. STI usually exerts compressive stress along the channel (i.e., the current flow direction), which improves PMOS device mobility. The opposite type of stress, i.e., tensile stress, degrades the PMOS performance in this direction. NMOS is, in general, complementary to PMOS in terms of how it is affected by stress, and its mobility degrades because of STI stress.

Device mobility increase corresponds to speed increase. Hence, it is possible to utilize STI, which is used to separate NMOS and PMOS regions, to improve performance. Table I shows the impact of STIW on the rise and fall delays (averaged over all timing arcs) of several 65-nm standard cells using the models developed in this paper. Fig. 1 illustrates the change in STIW for the INVDO when the intercell spacing is increased from 0  $\mu\text{m}$  (i.e., abutting neighbors) to 5  $\mu\text{m}$ . The impact of placement on STIW and, consequently, on rise (*R-Delay*) and fall (*F-Delay*) delays for a few cells is provided. For each cell in the table, three instances of it are placed with different spacings between them, and the delay of the center instance is reported. In Table I, *Spacing* is the spacing between cells, and PMOS  $STIW_L$  (NMOS  $STIW_L$ ) and PMOS  $STIW_R$  (PMOS  $STIW_R$ ) are the STIWs next to the left and right sides of the positive active (negative active) regions of the center cell. It is possible to both speed up and slow down cells by controlling the STIW and, thereby, the stress that is applied to a cell. In particular, larger STIW will generate more stress in neighboring transistors. In this paper, we propose placement perturbation and the insertion of active-layer fills to control the STIW in a performance-driven manner.

The proposed active-layer fill insertion and placement perturbation do not require additional process steps or add complications to resolution enhancement techniques. Active-layer fill insertion is a standard process step that is performed in all designs to control active-layer density. Placement perturbation yields a new valid placement. We ensure that the design is design-rule correct after we perform these two steps.

### D. Organization of the Paper

The remainder of this paper is organized as follows: In Section II, we describe our STIW-induced stress models that we have developed. After a brief introduction to stress, we review the process steps we have simulated using TCAD tools and our proposed stress models. In Section III, we present our STI-stress-aware timing analysis approach. Section IV describes our timing optimization methodology. In Section V, we present our circuit-level optimization results. Section VI concludes this paper.

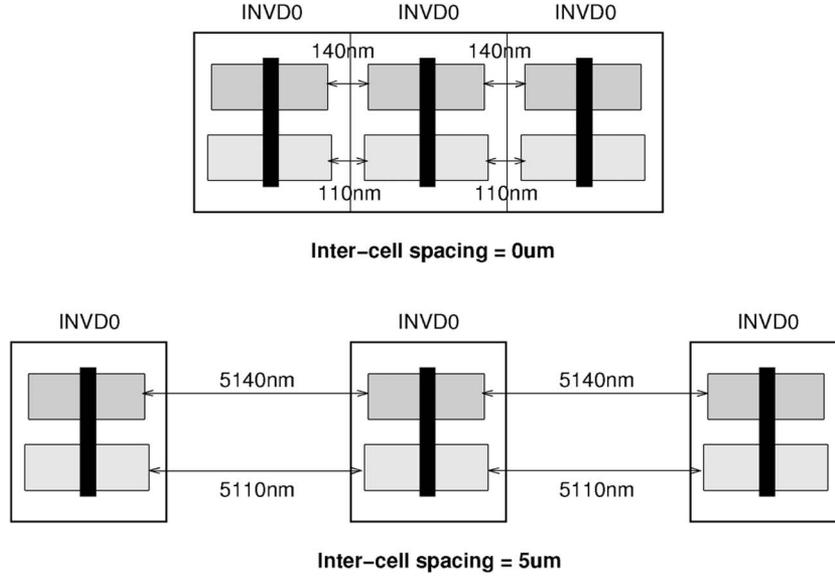


Fig. 1. Change in STIW on increasing the intercell spacing from 0 to 5  $\mu\text{m}$ .

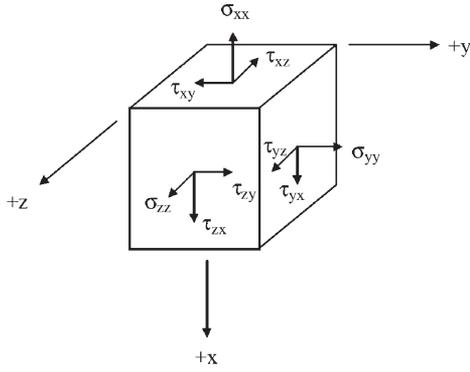


Fig. 2. Stress components on a unit volume.

## II. STIW STRESS MODELING

Based on the understanding in [7], the stress components on a unit cell are as shown in Fig. 2. The stress vector  $T_x$  acting normal to  $x$  is given as  $T_x = \sigma_{xx} \cdot \hat{x} + \sigma_{xy} \cdot \hat{y} + \sigma_{xz} \cdot \hat{z}$ . The stress tensor is defined by the three stress vectors, i.e.,

$$\sigma_{ij} = \begin{bmatrix} \sigma_{xx} & \tau_{xy} & \tau_{xz} \\ \tau_{yz} & \sigma_{yy} & \tau_{yz} \\ \tau_{zx} & \tau_{zy} & \sigma_{zz} \end{bmatrix}.$$

In this equation, the  $\sigma_{ii}$ 's are stress components normal to the unit cube faces, whereas the  $\tau_{ij}$ 's are shear components directed toward  $j$  on the orthogonal face to  $i$ . The  $\sigma_{ii}$ 's are used for analyzing the impact of stress. Using the individual stress components, we may convert the stress values to mobility [27].

The remainder of this section describes a generic STI process flow, along with the STIW models we propose. The STIW parameter is as shown in Fig. 3.

### A. Process Steps

The process recipe that we use for the simulation of STI stress is summarized in Table II. We have simulated the struc-

ture up to the gate deposition step using the Synopsys Sentaurus 2005.12 process simulator.<sup>1</sup> We make three observations.

- 1) We use a high mesh density, particularly between the STI and underneath the channel, to obtain accurate finite-element calculations close to the channel.
- 2) Temperature cycling (steps 7 and 19) and densification steps (steps 10–12) are responsible for the stress buildup. Due to viscoelastic material behavior, materials cannot recover to their original state after stress is withheld. Thermal cycles result in stress due to thermal mismatch between different materials, which have different thermal expansion coefficients. As a result, stress builds up in the STI oxide, and this stress remains there even at room temperature at the end of the process. Final stress shows its effects all the way into the channel of neighboring transistors, in a space-dependent trend, during the lifetime of a chip.
- 3) In step 14, STI chemical mechanical polishing (CMP) is applied. At the end of this step, the top of the STI is left above the active region on purpose. The basic reason is to avoid defectivity, such as delamination of the STI oxide. At the edges of the channel, this step height difference would introduce threshold voltage variations and so-called “width effects.”

### B. STI Stress Modeling

The popularly used Berkeley Short-channel IGFET Model (BSIM) SPICE model (revision 4.3 and higher) contains an explicit STI model. However, only the impact of the distance from the transistor channel to the STI boundary is modeled. Hence, the dependence on the STIW is not present in the BSIM4 model. Our simulations, as well as the simulations

<sup>1</sup>When foundry models are used, the exact process steps are not known, except for hints provided in the literature or various collateral documentation. Foundries should provide STIW impact models in such a scenario.

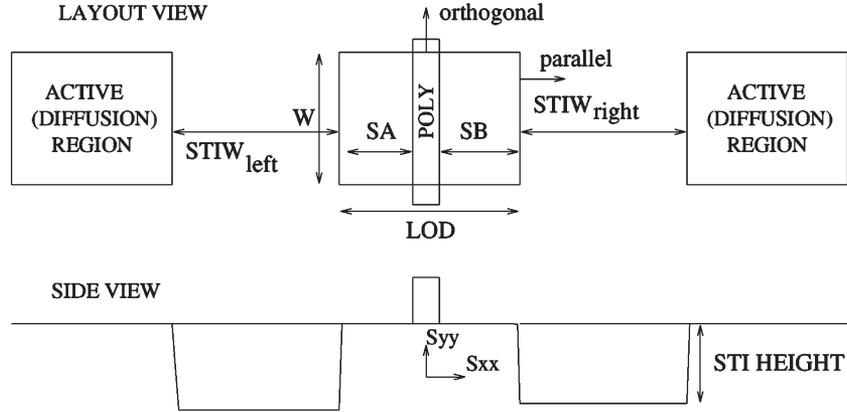


Fig. 3. STIW parameter. LOD is accounted for in BSIM models. The STIW impact is not modeled. Parallel and orthogonal distances with respect to a transistor is also indicated in the figure.

TABLE II  
STI PROCESS STEPS

1. Deposit pad oxide
2. Deposit pad nitride
3. Deposit photoresist for STI lithography
4. Anisotropically etch nitride and oxide
5. Strip photoresist
6. Directional etch at a rate of $0.01\mu\text{m/s}$ for 40 seconds at $86^\circ$ angle
7. Ramp up temperature to $600^\circ\text{C}$
8. Deposit TEOS oxide at a rate of $0.01\mu\text{m/s}$ for 20 seconds
9. Deposit trench fill oxide
10. Temperature ramp up from $600^\circ\text{C}$ to $1000^\circ\text{C}$ at a rate of $50^\circ\text{C/min}$
11. Hold temperature for 1 minute at $1000^\circ\text{C}$
12. Temperature ramp down from $1000^\circ\text{C}$ to $600^\circ\text{C}$ at a rate of $50^\circ\text{C/min}$
13. Diffuse oxide
14. STI CMP
15. Etch nitride isotropically at a rate of $0.015\mu\text{m/s}$ for 15 minutes
16. Etch oxide isotropically at a rate of $0.02\mu\text{m/s}$ for 1 minute
17. Temperature ramp up from $600^\circ\text{C}$ to $800^\circ\text{C}$ at a rate of $40^\circ\text{C/min}$
18. Diffuse for 5 minutes using dry $O_2$
19. Temperature ramp down from $800^\circ\text{C}$ to $600^\circ\text{C}$ at a rate of $40^\circ\text{C/min}$
20. Deposit polysilicon gate
21. Ramp down to room temperature

and data in the literature, show that STIW impact cannot be neglected. Thus, as previously noted, this paper not only models STIW impact but also builds upon this modeling to improve circuit performance at no area cost.

The STI impact as a function of length of diffusion (LOD) (refer to Fig. 3) is already incorporated into the BSIM4 model. Our objective is to isolate and correct for the impact of STIW, in a manner that can be applied on top of existing BSIM4 stress modeling. Using 2-D simulations, we have developed the model given in (1)–(4) to capture the STIW effect in the parallel direction (shown in Fig. 3). The LOD parameter still appears in the equation, as the STIW impact differs according to the choice of LOD. In addition, for purposes of this paper, we do not require or discuss STIW impact modeling in the orthogonal direction (shown in Fig. 3), as the STIW effects are blocked in the orthogonal directions by active regions for the

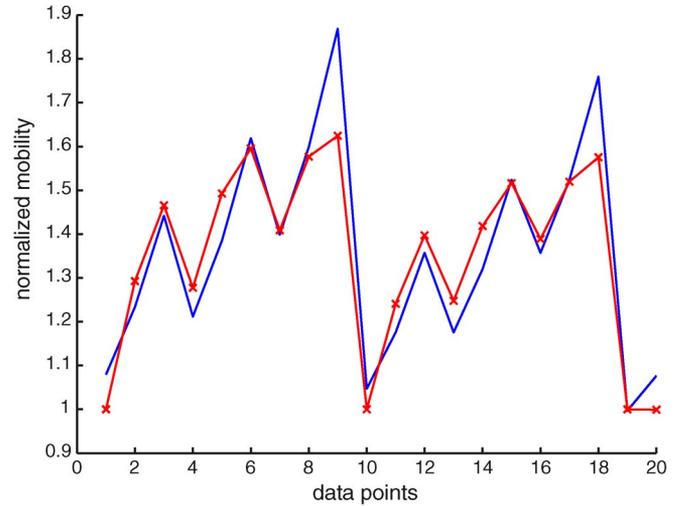


Fig. 4. Model versus data. Plot with crosses corresponding to the data. Each data point corresponds to a DOE instance.

type of standard cells we have used. At the end of the TCAD simulations, we obtain stress values in pascals. We then convert the stress values to mobilities using the methodology in [8] and normalize the mobilities. The NMOS equation is given as

$$MOB_{L,R} = \zeta + (1 - (STIW_{L,R}/2)^\alpha) / S\{A, B\}^\beta \quad (1)$$

$$MOB = [MOB_L * MOB_R]^{0.26}. \quad (2)$$

In (2),  $MOB$  is the mobility multiplier. Parameters L and R indicate left and right directions with respect to the channel. The equation states that the final mobility multiplier (i.e.,  $MOB$ ) is the product of the mobility multipliers from the left and right directions (i.e.,  $MOB_L$  and  $MOB_R$ ). The PMOS equation is given as

$$MOB_{L,R} = \zeta + ((STIW_{L,R}/2)^\alpha) / S\{A, B\}^\beta \quad (3)$$

$$MOB = [MOB_L * MOB_R]^{0.14}. \quad (4)$$

The model and data comparison is shown in Fig. 4. In the figure, the  $x$ -axis is a given data point in the design of experiments (DOE), i.e., a given  $SA$ ,  $SB$ ,  $STIW_L$ , and  $STIW_R$  combination, and the  $y$  axis is the mobility multiplier. The

TABLE III  
MODEL PARAMETER TABLE

	$\zeta$	$\alpha$	$\beta$
NMOS	1.03	0.076	0.48
PMOS	0.49	0.48	0.57

models provide an average of 7.5% accuracy with respect to the data, along with preserving physical intuition.

The physical intuition is simply based on a relational understanding of how distance to a stress source impacts the stress, depending on the type of transistor. As  $S\{A, B\}$  increases in (3), the stress source becomes farther away from the channel. Hence, mobility should decrease for PMOS as STI has compressive stress in this technology. This is captured in the equation. Increasing STIW in (3) should improve the mobility. This can also be seen from the equation. On the contrary, NMOS equation (1) has these physical terms to yield the opposite behavior. The mathematical constants are used to enable a fitting for the particular process technology.

When BSIM models are being generated from test structures, assuming that device modeling test structures with worst case corners for STIW are present, the slow-N–slow-P corner would correspond to data from test structures with large STIW near NMOS modeling devices, and/or test structures with minimum STIW near PMOS devices. Our normalization step should consider the slow corner conditions as described herein, as well as the presence of STI models in BSIM models. The latter is necessary to eliminate any possibility of the double counting for the channel to STI edge stress, which is already covered by the BSIM models.

In order to enable an accurate normalization, we have set the slow N or P devices to a mobility of 1. Other STIW values result in an increase in the mobility. As multiple  $SA$  and  $SB$  values, which can share the same STIW, are present in the DOE, there are multiple devices that have a mobility of 1 after normalization. The models are then fit to the normalized results of the process simulation DOE. The corresponding parameters of the model are given in Table III.

*Consideration of Nonrectangular Active Regions:* For non-rectangular active regions, an average distance to the channel from the STI boundary and for STIW can be computed as follows:

$$S\{A, B\} = \frac{\sum_i w_i * S\{A, B\}_i}{W} \quad (5)$$

$$STIW_{L,R} = \frac{\sum_i t_i * STIW_{L,Ri}}{W}. \quad (6)$$

In (5) and (6),  $i$  is an enumeration over active region edges, and  $w_i$  and  $t_i$  are the widths of each such edge parallel to the channel for the distance from the channel to the active edge and STIW, respectively. Parameter  $W$  is the width of the channel. Similar width-weighted averaging is used by BSIM 4.3.0 for the calculation of stress effects without explicitly using parameters to capture irregular active region shapes.

A related discussion pertains to the choice of active diffusion fill widths. Fringing effects due to STI next to PMOS may degrade the NMOS speed. However, we consider the fringing effects to be small, because NMOS and PMOS are separated by

TABLE IV  
LINER STRESS ANALYSIS DOE PARAMETERS

$SA$	$SB$	$STIW_L$
0.2,4 $\mu\text{m}$	0.2,4 $\mu\text{m}$	0.1,2 $\mu\text{m}$
$STIW_R$	LinerHeight	LinerStress
0.1,2 $\mu\text{m}$	0.1,0.2 $\mu\text{m}$	1,2GPa

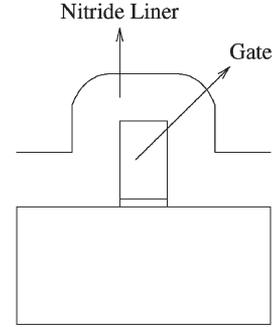


Fig. 5. Stressed nitride liner as mobility enhancer.

hundreds of micrometers. To further limit this issue, the width of the STI near PMOS can be decreased by a few nanometers. The exact value to be used would depend on particular technology and can be obtained through the foundry or by conducting 3-D stress simulations.

*STI CMP:* Traditionally, active region fills are usually inserted at the tape-out stage to minimize active region density variations. When active region fills need to be utilized for performance, a number of observations need to be made. A buffer distance can be determined, such that, above this distance, the stress width has no further effect on channel. This distance can be selected as, for example, 10  $\mu\text{m}$  for the process we have studied. It is advisable that such postfill insertion algorithms use such buffer distances and do not insert fills inside buffer regions near critical gates. Furthermore, these fills are inserted using a window-based scheme to minimize density variations across windows. With optimization, fills will be inserted next to NMOS but not next to PMOS. Hence, there will be approximately 50% active region fill in an optimized window. If the density is lower and needs to be increased, then users can start to insert fills near noncritical gates as well. This approach will have negligible impact on timing closure.

*Impact of Stress Liner:* To evaluate the impact of stress liners on STI stress, we have used a similar simulation setup with the parameters given in Table IV. A nitride liner is shown in Fig. 5. The nitride liner height and intrinsic stress are varied to observe the influence on STI stress. The parameters are shown in Fig. 3. Combinations of all parameters are individually simulated. We have observed that, in the given parameter range, the addition of a 1-GPa 0.1- $\mu\text{m}$ -thick stressed nitride layer increases the impact of STI by 9.9% in terms of an average stress under the channel. A 0.2- $\mu\text{m}$  stressed nitride layer, on the other hand, increases the STIW impact by 12.9%. A 2-GPa liner can increase impact by 10.3%. These changes seem negligible and also indicate that STI stress width effect will still be important across such process variants. The silicon data reported in [30] support our findings, i.e., STI stress width effect is still important even in the presence of stressed nitride liners. The nominal stress values

TABLE V  
STI HEIGHT ANALYSIS DOE PARAMETERS

$SA$	$SB$	$STIW_L$	$STIW_R$	$STIHeight$
0.2,4 $\mu$ m	0.2,4 $\mu$ m	0.1,2 $\mu$ m	0.1,2 $\mu$ m	0.2,0.4 $\mu$ m

usually increase for the  $S_{yy}$ , which is shown in Fig. 3, as this component is in the direction of the nitride liner. The  $S_{xx}$  stress component typically reduces with these changes.

*Impact of STI Height:* The parameters used to analyze the impact of STI height on stress are shown in Table V. The height of the STI trench is changed to see the impact of STI height on stress. We have observed that increasing the STI height can result in a reversal of the STI stress, i.e., changing it from compressive to tensile for the  $S_{yy}$  component. This observation is in line with what has been presented in [31]. Comparison of average stress values shows a reduction of up to 6% for the STIW impact due to variation of STI height. As with stress liner impact, we believe that our conclusions remain valid across STI heights used by different processes.

Analyses and optimizations proposed in the rest of this paper are not tethered to the models developed in this section and can be used with other models after appropriate modifications. For example, there are known STI processes that may induce tensile, instead of compressive, stress. This may be due to STI trench height, or material and thermal processing differences, such as high-density plasma chemical vapor deposition, as used in [9]. The optimization procedure presented here can be adopted to such a scenario with minor changes. Furthermore, the proposed models show monotonic response with respect to the STI proximity and widths. This results in an optimization scheme where the maximum or minimum allowed dimensions will improve performance. With models that are nonmonotonic, the optimization algorithm would need to be altered to provide an optimal solution.

Even as new mobility enhancement techniques show substantial variation in mobility, STI has been a mainstream methodology for over a decade. STI processes are much better controlled than the new stress engineering methods. Hence, there is much smaller observed variability in the STI mobility impact.

Lithography effects will show negligible impact on the mobility change due to STI, as long as the layout is designed considering design for manufacturability rules, such as using regular active regions without unnecessary corners.

Finally, since active-layer fill insertion is a knob that we propose to exploit, we comment on additional process considerations for active-layer filling.

- 1) There exist design rules that restrict the maximum active-layer density, with this constraint arising for reasons of STI CMP uniformity. Such rules must be observed.
- 2) Insertion of active-layer fills can potentially increase the total capacitance of intercell M1 routing and may induce additional resistance, capacitance extraction modeling and characterization for a given process technology. However, our methodology should not affect the extraction of intracell M1 routing; as our active-layer fills are floating, their impact is smaller.

- 3) Reduced STIW may slightly increase the leakage between NMOS and PMOS transistors as well as that between devices of the same type. However, the active-to-active design rules are typically set such that this leakage is minimized to a negligible level.

### III. STRESS-AWARE TIMING ANALYSIS

In this section, we describe our STI-stress-aware timing analysis methodology. We adapt the traditional SPICE-based timing analysis flow to consider stress induced by STIW.

#### A. Traditional SPICE-Based Timing Analysis

Cell-level static timing analysis (STA) tools such as PrimeTime offer a good tradeoff between accuracy and analysis speed. Full designs or their blocks are typically analyzed and signed off with circuit-level STA. However, if greater accuracy is desired, SPICE-based analysis, which has better accuracy but substantially slower analysis speed, is employed. Since running full-chip SPICE analysis is not feasible, critical paths are first identified with STA and then simulated with SPICE.

A typical netlist input to SPICE is layered into three tiers.

- 1) *Device-level models*, which contain transistor parameters in the form of coefficients of functions defined in BSIM or equivalent formats. Device-level models allow output waveforms for PMOS and NMOS devices to be simulated.
- 2) *Cell-level netlists*, which describe the connectivity of the devices that comprise individual cells. Cell-level netlists instantiate device-level models and allow SPICE to simulate waveforms at the outputs of cells in the library when subjected to a stimuli.
- 3) *Critical path netlists*, which describe the connectivity between the cells for each critical path. Critical path netlists instantiate cell-level netlists and can be simulated to calculate the delay of the critical paths.

As previously noted, stress-induced device mobility change is determined by: 1) the separation between the gate and the active edges and 2) by the size of the STI region that surrounds the active region of the device. Fortunately, the separation between gate and active edges is fixed when the cells are designed, and the contribution of this separation to stress and mobility can be modeled at the cell level. Specifically, in the BSIM 4.3.0 device-level models, stress parameters  $SA$ ,  $SB$ , and  $SC$  have been introduced to model the stress effect as a function of gate and active edge separation. In cell-level netlists, these parameters are passed with the instantiation of the device-level models. Cell-level netlists are used in library characterization to generate gate-level timing models for use in STA. An example of device-level instantiation with stress parameters is shown in Fig. 6.

The stress effect due to STIW is not modeled primarily for two reasons.

- 1) STIW is determined by the placement of the cells, so that stress effect due to STI cannot be captured in library characterization. A new methodology that analyzes a placed design and annotates STIW information for use in timing analysis is required.

```

.subckt INVX1 A Z
MM1 D G S B NCH SA=0.2u SB=0.2
MM2 D G S B PCH SA=0.19u SB=0.19u
...
.ends

.model NCH NMOS (
*Other stress parameters defined
)

```

Fig. 6. Instantiation of device-level models in a standard cell SPICE netlist. The parameters added in BSIM 4.3.0 to partially model stress are shown in bold.

```

*Critical path 00001
X01 N1 N2 INVX1 PL=0.08u PR=4.08u NL=0.06u NR=4.06u
X02 N2 1 N2 NAND2X1 PL=5.0u PR=5.0u NL=5.0u NR=5.0u
X03 N3 N4 BUFFX1 PL=2.1u PR=5.0u NL=2.04u NR=5.0u
...

.subckt INVX1 A Z
.param PMOB = Our_PMOS_Model (PL, PR, NL, NR)
.param NMOB = Our_NMOS_Model (PL, PR, NL, NR)
MM1 D G S B NCH SA=0.2u SB=0.2 MOB=NMOB
MM2 D G S B PCH SA=0.19u SB=0.19u MOB=PMOB
...
.ends

.model NCH NMOS (
*Other stress parameters defined
)

```

Fig. 7. Critical paths instantiate cell-level netlists, which instantiate device-level models. Our modifications to the traditional flow to model STIW-dependent stress are shown in bold.

- 2) Stress effect due to STI is of smaller magnitude than gate and active edge separation.

### B. STI-Stress-Aware Timing Analysis

Our approach analyzes the placement of a design and the standard cell layouts to calculate the STIWs for all critical cells in the design. The STIWs are then passed as parameters, which are used in the models developed in the previous section.

We modify the cell-level netlists such that parameters  $PL$ ,  $PR$ ,  $NL$ , and  $NR$ , which capture the STIW, are passed to them. Parameter  $PL$  is the spacing between the boundary of a cell and the neighboring active region to the left of its positive active region. Similarly, parameter  $PR$  is the spacing between the boundary of a cell and the neighboring active region to the right of its positive active region (PRX). Parameters  $NL$ , and  $NR$  are similarly defined for *negative* active regions (NRX). The parameters are set in the critical path netlists when cells are instantiated, as shown in the example in Fig. 7.

The  $PL$ ,  $PR$ ,  $NL$ , and  $NR$  parameters can be calculated from the placement and the cell's layouts, specifically, the cell bound-

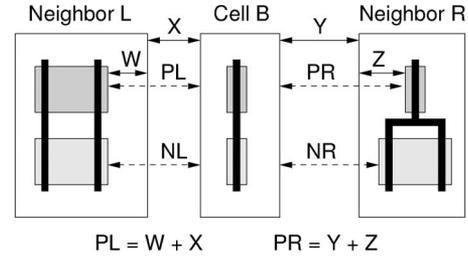


Fig. 8. Calculation of parameters  $PL$ ,  $PR$ ,  $NL$ , and  $NR$  from intercell spacings and active-to-cell-boundary spacings.

ary to active spacings. Computation of  $PL$  for a cell, which is the spacing between the cell's boundary and the positive active region of the cell to its left, is given as follows: The spacing between the cell and its left neighboring cell is found from the placement. The spacing between the positive active region of the neighbor and its cell boundary is found from layout analysis of the neighbor. The two spacings are then added, with correct consideration of the orientations of the cell and its neighbor. Other parameters  $PR$ ,  $NL$ , and  $NR$  are similarly calculated. Fig. 8 illustrates the calculation.

We note that our flow needs modifications to work for cells with complex active shapes, such as flip-flops and multiplexors. Active shape complexities include nonrectangular shapes and noncontinuous shapes. To model STI stress impact for nonrectangular active shapes, modifications such as those employed by BSIM to handle nonrectangular active may be used. For cells with noncontinuous active shapes, devices can be completely shielded from STIW outside the cell, and our flow should not alter their mobility. In our analysis and optimization, we focus on the cells with simple active shapes and do not change the mobilities for cells with complex active shapes (i.e., use traditional analysis and no optimization for them). Fortunately, the most frequent cells such as inverters, buffers, NANDs, NORs, ANDs, and ORs have simple active shapes, so we consider and optimize most cells in our designs.

### C. Alternative Flow

STI-stress-aware timing analysis can also be performed by cell-level STA. Toward this standard, cells in the library can be characterized for different STIW configurations around them. Since stress dependence on STIW is relatively gradual, STIWs can be binned into a small number of bins to reduce the total number of STIW configurations. For each standard cell, variants may be created corresponding to each STIW configuration. Given the STIW, models presented in the previous section are used in library characterization. The STIW of a cell in a design can be computed from the placement and standard cells layouts, and can be used to find the variant that has the closest STIW configuration. The cell can then be bound to the variant in the library and cell-level STA run to perform STI-stress-aware timing analysis.

## IV. TIMING OPTIMIZATION

In this section, we present our timing optimization methodology. The basic idea exploited in our optimization is that

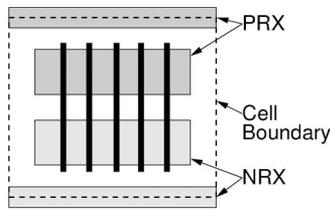


Fig. 9. Generic standard cell, with polysilicon positive active regions (PRX), negative active regions (NRX), and cell boundary shown.

the STIW of devices can be altered to change their mobility and improve performance. Specifically, the alteration involves increasing the STIW for PMOS devices and decreasing them for NMOS devices. We identify the timing critical cells and alter their STIW to improve the circuit performance. In our approach, we use two knobs to alter the STIW.

- 1) Placement perturbation. The placement of a layout can be changed to increase or decrease the spacing between neighboring cells, which directly increases or decreases the STIW. Additionally, spacing cells apart can allow fills, for which, initially, there was insufficient space, to be inserted.
- 2) Active-layer fill (*RX fill*) insertion. Active-layer fills are rectangular dummy geometries inserted on the active (RX) layer primarily to improve planarity after CMP. However, such geometries also reduce the STIW of the devices next to which they are inserted. The STIW after insertion of an RX fill next to a device is the spacing between the active region of the device and the fill.

We now present the details of the two aforementioned knobs.

#### A. Active-Layer (RX) Fill Insertion

Even though RX fills are nonfunctional geometries, their effect on the stress is identical to that of active regions of devices. When inserted next to the active region of an NMOS device, fills substantially reduce the STIW and stress of the device and consequently improve the performance of the NMOS device. On the other hand, fills inserted next to a PMOS device reduce STIW and stress but consequently degrade the performance. Hence, inserting fill next to the NMOS devices but not next to the PMOS devices of a cell improves performance.

Circuit delay improves when the delay of setup-critical cells is reduced. Thus, we insert rectangular RX fills next to the NMOS devices, to the left and right of the cell. No RX fills are inserted next to the PMOS devices; thus, the PMOS remains exposed to a large STIW and stress. The devices closer to the active boundary experience the maximum benefit of this optimization. Since the most frequently used cells in the designs are small, a large fraction of devices in the design benefit from fill insertion. Our technique can also be employed for hold-time-critical cells in the reverse manner, i.e., insert fills next to the PMOS devices but not next to NMOS devices to slow down the cell.

Fig. 9 shows a sample standard cell with PRX (active regions for PMOS devices) and NRX (active regions for NMOS devices). As can be seen, active regions exist under the top and bottom cell boundaries that completely shield the cell from STI

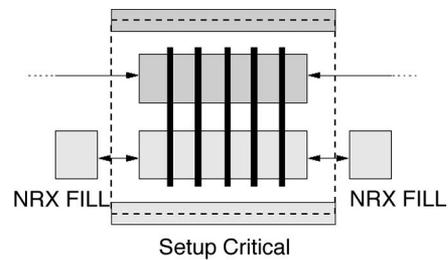


Fig. 10. Generic cell of Fig. 9 optimized with fill insertion for setup-time criticality.

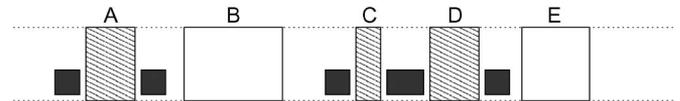


Fig. 11. Row of standard cells after active-layer fill insertion for setup-time improvement. The cells patterned with diagonal lines are the setup-critical cells, and the filled rectangles are the inserted active-layer fills.

stress effects in the direction orthogonal to the carrier (current) flow direction. Hence, we only apply our optimization in the parallel direction by inserting fill to the right and left of a cell. Fig. 10 illustrates fill insertion for a setup-critical cell; NRX fills are inserted next to the NRX region to reduce stress and *fasten* the NMOS devices. Fig. 11 illustrates the approach for a few setup-critical cells in a standard cell row.

All fills are inserted subject to the design rule constraints (DRCs) and introduce no DRC violations. We have already noted that no additional mask step is required and that M1 capacitance impact is likely negligible. Since the fill insertion knob can only decrease STIW, NMOS performance can be improved, but PMOS performance can, at best, be kept constant. However, neighboring cells, which have very small spacing and between which fills cannot be inserted, can be spaced apart by placement perturbation to allow fills to be inserted.

#### B. Intrarow Placement Optimization

We now present the placement perturbation knob, which can increase (decrease) the STIW and improve PMOS (NMOS) performance. Placement of a cell determines its location (consequently, its neighbors and spacings with them) and its orientation. In our optimization, we change the locations of the cells such that spacings are altered, but the ordering of cells in a standard cell row is not affected. Increased spacing next to a cell increases the STIW and improves the delay of the PMOS devices. However, the delay of the NMOS devices increases with increased spacing. Fortunately, we can utilize our first knob, i.e., RX fill insertion, to reduce the NMOS STIW and improve its delay as well. In fact, if the spacing between cells is too small for fill insertion, placement can facilitate fill to be inserted by creating additional space for it. The placement perturbation just reorganizes the white space in the standard cell row of the cell without requiring any additional space.

1) *Minimizing Delay Increase Due to Wire Length Increase:* The perturbation of detailed placement from the original placement results in small wire length change, which can impact wire parasitics and, consequently, timing. Even though our

**Input:** Placed design; set of timing critical cells,  $T$ ; set of fixed cells,  $F$ ; maximum spacing to create,  $S$   
**Output:** New placement with altered inter-cell spacings

```

1  forall cells  $t \in T$ 
1.1 createRightSpace( $t$ )
1.2 createLeftSpace( $t$ )
createRightSpace( $t$ )
1   $n = \text{cellsToMoveRight}(t)$ ;
2  moveCellsRight( $t, n$ );
cellsToMoveRight( $t$ )
1   $i \leftarrow t$ ;
2   $j \leftarrow \text{cellToRightOf}(t)$ ;
3  accumulatedSpacing = 0;
4  cellsToMove = 0;
5  while accumulatedSpacing  $\leq S$  and  $j \notin F$ 
   and cellToRightOf( $j$ )  $\notin T$ 
5.1  accumulatedSpacing +=
interCellSpacing( $i, j$ );
5.2  cellsToMove++;
5.3   $i \leftarrow j$ 
5.4   $j \leftarrow \text{cellToRightOf}(i)$ ;
6  return cellsToMove;
moveCellsRight( $t, n$ )
1  // flush  $n$  cells to the right of Cell  $t$ 
   towards the right to create space  $S$ 
2   $F \leftarrow F \cup \{n \text{ cells to the right of Cell } t\}$ 
createLeftSpace( $t$ )
1  // similar to createRightSpace( $t$ )

```

Fig. 12. Pseudocode for intrarow placement optimization.

localized placement perturbations do not significantly affect timing, small changes in the timing of critical paths can affect the minimum clock cycle time. To minimize the timing change of critical paths, we fix the cells and nets in the critical paths. Fixed cells are not moved during optimization, and fixed nets are not changed during the engineering change order (ECO) routing that is performed after optimization. Since the nets in the critical path are fixed, all cells connected to these nets should also be marked as fixed and not moved during optimization. We note that the delay of such nets can marginally change due to the coupling capacitance with neighboring nets, the routing for which may change. We also fix all flip-flops, clock buffers, and clock nets to avoid any impact on the clock tree. Thus, our list of *fixed cells* comprises timing critical cells, their fan-out cells, flip-flops, and clock buffers.

Our intrarow placement optimization attempts to create space on the right and left sides of each timing critical cell. In the process, the minimum number of cells is displaced to minimize the wire length impact. Fig. 12 presents the pseudocode for our intrarow placement optimization. For each timing critical cell, right and left spacings are increased by functions *createRightSpace* and *createLeftSpace*, respectively, to attain a spacing of up to  $S$ . The spacing,  $S$ , may not always be attainable because of the presence of fixed cells and availability of limited space in the row. For the right side, function *cellsToMoveRight* finds the minimum number of cells to move. Then, function *moveCellsRight* flushes the computed number of cells to the right as much as possible.

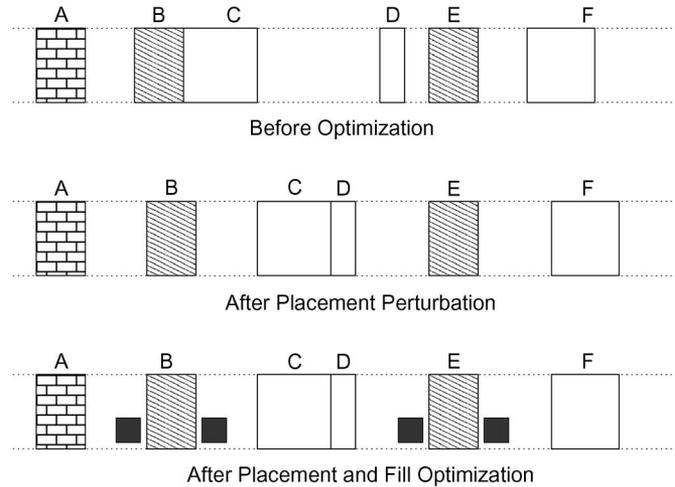


Fig. 13. Placement change and fill insertion for setup-time optimization. A standard cell row is shown before optimization, after placement perturbation, and after fill insertion. The cells patterned by diagonal lines are the setup-critical cells for which timing is optimized. Fixed cells are patterned with the brick pattern, and their placement cannot be changed.

Our algorithm sequentially processes critical cells in decreasing order of their criticality. Cells displaced in an iteration to create space are added to the list of fixed cells to lock them for successive iterations. This can limit the optimization of critical cells processed later in the algorithm. Therefore, we run the algorithm multiple times with increasing value of  $S$ . This enhancement allows a fair distribution of white space among all critical cells. We increase the value of  $S$  from 0.6 to 1.8  $\mu\text{m}$ , in steps of 0.2  $\mu\text{m}$ . Starting with a smaller value of  $S$  leads to a more equitable distribution of white space at the expense of runtime. For designs with high utilization ratios, starting  $S$  of less than 0.6  $\mu\text{m}$  may be desirable. We have found that the STIW effect saturates at 1.8  $\mu\text{m}$  and there is negligible change in stress beyond that.

Our second enhancement is perturbing the critical cells to balance the space on the right and left sides of them. Since the stress effect rapidly decays with space, nearly equal spacings on both sides are desirable. We limit the perturbation to 0.6  $\mu\text{m}$  to minimize wire length and the associated delay increase. The space required to insert RX fill is typically very small and in the 0.2- $\mu\text{m}$  range. Therefore, if the optimization creates any space for PMOS optimization, fill can always be inserted to improve the deteriorated NMOS performance. Fig. 13 illustrates placement perturbation and fill insertion for setup-time optimization on a standard cell row.

While it is possible to perform fill insertion without placement perturbation, we have found the associated performance benefits to be very small. Both knobs complement each other: Placement creates space for fill insertion, and fill insertion improves the performance of the NMOS devices that are slowed down by placement perturbation.

Our overall STI-stress-aware placement and fill optimization flow has five steps.

- 1) Identify critical paths and critical cells.
- 2) Perform intrarow placement optimization.
- 3) Perform ECO routing followed by parasitic extraction.

TABLE VI  
TEST CASES USED IN EXPERIMENTAL VALIDATION.  
MCT IS THE MINIMUM CYCLE TIME

Circuit	Source	#cells	Utilization	MCT (ns)
C5315	ISCAS'85	1,408	82%	0.912
ALU	opencores.org	11,106	78%	4.333
S38417	ISCAS'85	8,514	79%	3.086
AES	opencores.org	21,000	78%	4.738

- 4) Perform RX fill insertion.
- 5) Evaluate the optimized layout with STI-stress-aware timing analysis.

## V. EXPERIMENTAL STUDY

We now present our experiments to evaluate the proposed optimization methodology. Our experiments assess the impact of our optimization on the minimum clock cycle time, delay of top critical paths, and final routed wire length.

### A. Experimental Setup

The details of the test cases used in our experiments are presented in Table VI. We use *Synopsys Design Compiler vW-2004.12.SP3* [2] for synthesis; *Cadence SOC Encounter (v5.2)* [1] for placement, clock tree synthesis, routing, and parasitic extraction; *Synopsys PrimeTime vW-2004.12.SP2* [5] for cell-level timing analysis; and *Synopsys HSPICE vY-2006.03* [3] for SPICE simulations. For our experiments, we use the 50 most frequently used cells from high- $V_{th}$  and nominal- $V_{th}$  65-nm high-speed libraries. SPICE device models and cell netlists were supplied by a foundry. We built our optimizer on top of *OpenAccess API v2.2.4* [4].

### B. Experimental Results

We first compare the proposed stress-aware timing analysis with traditional analysis. Since traditional analysis does not account for STI stress and must correctly analyze for all STI configurations, it is conservative. Traditional analysis is corner-based and uses the worst case cell delay, which reflect worst case STI stress effects in addition to worst case process variations. Worst case analysis, while correct, leaves valuable performance on the table. Stress-aware timing analysis reduces pessimism in analysis by explicitly accounting for STI stress. We therefore expect stress-aware timing analysis to report circuit delays that are smaller than those from traditional analysis.

Table VII presents the comparison between traditional timing analysis and stress-aware timing analysis on our four test cases. We study two delay metrics: 1) minimum cycle time (MCT) and 2) *top paths delay* (TPD), which is the sum of the delays of top 100 critical paths. While MCT determines the maximum speed at which the circuits can be run, TPD determines the robustness to variations. We observe that stress-aware analysis reduces MCT by 5.75% and TPD by 5.28%, on average. We use stress-aware analysis to evaluate our optimization in the remainder of this section.

In Section IV, we presented two optimization knobs: 1) fill insertion and 2) placement perturbation. Although the two tech-

niques complement each other, we separately evaluate the fill insertion knob. Placement perturbation, without fill insertion, is not interesting, because it slows down the NMOS devices while speeding the PMOS. Table VIII presents the improvements in MCT and TPD due to fill insertion. Since we optimize several critical paths, TPD reduces. However, we observe that reductions in MCT and TPD are typically under 1%.

We now evaluate the simultaneous use of the proposed placement perturbation and fill insertion knobs. In addition to comparing MCT and TPD results, we also compare the wire length, which changes because of placement perturbation. After placement perturbation, several nets are left dangling; we perform ECO routing to route them and follow by resistance, capacitance extraction and stress-aware timing analysis to accurately report the MCT and TPD results for the optimized case. The runtime of our placement and fill optimization is generally small; it depends on the circuit size and the number of critical paths to be optimized. In our experiments, the runtime was under 1 min for all test cases on a 2.2-GHz AMD Opteron/8-GB random-access memory machine running Linux 2.6.

Table IX presents our results for our four test cases. For negligible increase in wire length, we observe 4.37% and 5.15% reductions in (stress-aware) MCT and TPD averaged over the test cases *C5315*, *ALU*, and *AES*. The test case *S38417*, however, demonstrates smaller improvements. We attribute this to the fact that *S38417* is an artificial test case, with over 50% of its cells being flip-flops. We do not allow our optimization to change the locations of flip-flops, to avoid changes to the clock tree; hence, in the *S38417* test case, we can perturb the placement of fewer cells. Fig. 14 shows the histograms for the delays of top 200 critical paths of our test case *AES* before and after optimization. As can be seen, the delay distribution has substantially shifted to the left (lower delay).

We also tried the technique to optimize hold-critical paths but found negligible improvement in hold slack for our test cases. This is because stress optimization can only change cell delays by 10%–20%, and for hold-critical paths, the cell delays are very small. Thus, the change in the delay of hold-critical paths is insignificant with our approach, and traditional delay introduction methods such as insertion of delay elements or wire snaking must be used.

## VI. CONCLUSION

We have conducted TCAD process simulations to generate models that relate the dependence of transistor mobilities to stress induced by STIW. We have proposed an STIW-aware design methodology for standard cell place-and-route designs. The proposed stress-aware timing analysis technique reduces pessimism in delay analysis. Over traditional corner-based analysis, delays reported by stress-aware analysis were, on average, 5.75% lower. We have also devised an optimization methodology based on cell placement perturbation to create extra space around critical cells; this is followed by dummy diffusion insertion. The proposed optimization flow, while demonstrated with our models, can be generalized to other STI stress models. We have applied the proposed optimization flow on a number of test cases implemented with industry

TABLE VII  
TRADITIONAL VERSUS STRESS-AWARE TIMING ANALYSIS

Circuit	Traditional		Stress-Aware			
	MCT (ns)	TPD (ns)	MCT (ns)	MCT Improv. (%)	TPD (ns)	TPD Improv. (%)
C5315	0.977	87.43	0.915	6.31	81.93	6.29
ALU	1.885	185.50	1.778	5.68	175.24	5.53
S38417	1.068	104.95	1.018	4.68	99.58	5.11
AES	1.739	165.82	1.655	4.83	158.88	4.19

TABLE VIII  
TIMING OPTIMIZATION RESULTS WITH FILL INSERTION. MCT IS THE MINIMUM CYCLE TIME. WL IS THE WIRE LENGTH. TPD STANDS FOR TOP PATHS DELAY AND IS THE SUM OF THE DELAYS OF THE TOP 100 CRITICAL PATHS

Circuit	Original		Fill Opt			
	MCT (ns)	TPD (ns)	MCT (ns)	MCT Improv. (%)	TPD (ns)	TPD Improv. (%)
C5315	0.915	81.83	0.903	1.32	81.35	0.71
ALU	1.778	175.24	1.771	0.39	174.53	0.40
S38417	1.018	99.58	1.010	0.79	99.92	0.39
AES	1.655	158.88	1.651	0.24	158.55	0.21

TABLE IX  
TIMING OPTIMIZATION RESULTS WITH PLACEMENT AND FILL INSERTION. MCT IS THE MINIMUM CYCLE TIME. WL IS THE WIRE LENGTH. TPD STANDS FOR TOP PATHS DELAY AND IS THE SUM OF THE DELAYS OF THE TOP 100 CRITICAL PATHS

Circuit	Original			Placement & Fill Opt					
	MCT (ns)	TPD (ns)	WL (mm)	MCT (ns)	MCT Improv. (%)	TPD (ns)	TPD Improv. (%)	WL (mm)	$\Delta$ WL (%)
C5315	0.915	81.93	17.8	0.879	3.97	75.50	7.85	17.9	+0.67
ALU	1.778	175.24	196.1	1.709	3.88	168.14	4.05	196.8	+0.36
S38417	1.018	99.58	96.4	0.993	2.44	97.94	1.65	96.64	+0.23
AES	1.655	158.88	374.7	1.568	5.26	153.21	3.56	3.75	+0.08

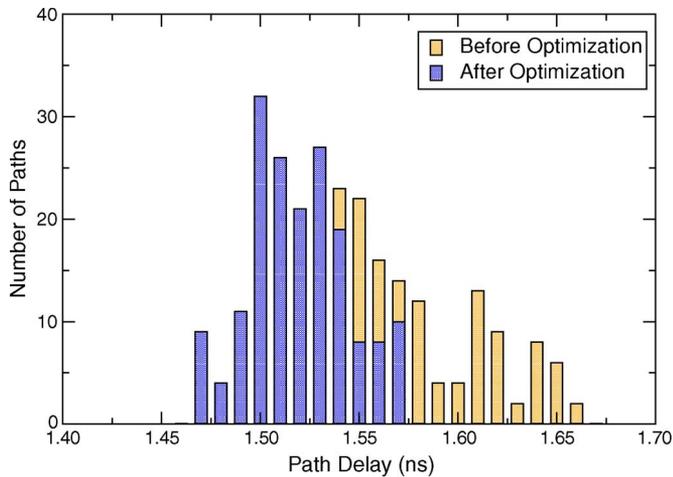


Fig. 14. Path delay histograms for the top 200 critical paths of test case AES before and after optimization.

65-nm libraries. Our data show that STIW optimization can increase performance by 2.44% to 5.26% with no area penalty. The proposed optimization can form the basis of circuit optimization that exploits upcoming stress-engineered transistor technologies in 65-nm and below processes.

Beyond our preliminary work reported in [6], in this paper, we provide detailed explanations of our simulations and report recent improvements to our models to handle asymmetric left and right STI width conditions. Implementation of these asymmetries required additional TCAD simulations as well as new modeling work. We investigate how the STI stress

impact is affected when stress liners are used for mobility improvement. Through TCAD simulations, we observe the impact of stress liner intrinsic stress and stress liner height on STI stress. Furthermore, STI height is a process parameter that can change the stress levels transferred to the channel. Hence, we also provide additional simulation results for STI height impact on stress. On the circuit-level optimization side, we have enhanced our optimization and repeated the experimentation with the improved models and realistically implemented test cases. Additionally, we compare the proposed stress-aware timing analysis with traditional corner-based timing analysis by turning on and off our stress modeling on top of BSIM models. This analysis quantifies the accuracy lost on not considering STI stress impact in the timing analysis. The improvements described in this paper thus provide a more complete stress optimization framework based on an extensive set of analyses.

#### ACKNOWLEDGMENT

The authors would like to thank Prof. H.-J. Lee of Sangmyung University, Korea, and F. Geelhaar of AMD for the useful discussions, and S. Muddu for help with the implementation and the test cases.

#### REFERENCES

- [1] *Cadence SoC Encounter*. [Online]. Available: [http://www.cadence.com/products/digital\\_ic/soc\\_encounter/index.aspx](http://www.cadence.com/products/digital_ic/soc_encounter/index.aspx)
- [2] *Synopsys Design Compiler*. [Online]. Available: [http://www.synopsys.com/products/logic/design\\_compiler.html](http://www.synopsys.com/products/logic/design_compiler.html)
- [3] *Synopsys HSPICE*. [Online]. Available: <http://www.synopsys.com/products/mixedsignal/hspice/hspice.html>

- [4] *OpenAccess API*. [Online]. Available: <http://openeda.si2.org>
- [5] *Synopsys PrimeTime*. [Online]. Available: [http://www.synopsys.com/products/analysis/primetime\\_ds.html](http://www.synopsys.com/products/analysis/primetime_ds.html)
- [6] A. B. Kahng, P. Sharma, and R. O. Topaloglu, "Exploiting STI stress for performance," in *Proc. ICCAD*, 2007, pp. 83–90.
- [7] H. A. Rueda, "Modeling of mechanical stress in silicon isolation technology and its influence on device characteristics," Ph.D. dissertation, Univ. Florida, Gainesville, FL, 1999.
- [8] C. S. Smith, "Piezoresistance effect in germanium and silicon," *Phys. Rev.*, vol. 94, no. 1, pp. 42–49, Apr. 1954.
- [9] S. W. Chung *et al.*, "Novel shallow trench isolation process using flowable oxide CVD for sub-100 nm DRAM," in *Proc. IEDM*, 2002, pp. 233–236.
- [10] J.-P. Han *et al.*, "Novel enhanced stressor with graded embedded SiGe source/drain for high performance CMOS devices," in *Proc. IEDM*, 2006, pp. 1–4.
- [11] Q. Ouyang *et al.*, "Characteristics of high performance PFETs with embedded SiGe source/drain and (100) channels on 45° rotated wafers," in *Proc. Int. Symp. VLSI Technol.*, 2005, pp. 27–28.
- [12] Y. Tateshita *et al.*, "High-performance and low-power CMOS device technologies featuring metal/high-k gate stacks with uniaxial strained silicon channels on (100) and (110) substrates," in *Proc. IEDM*, 2006, pp. 1–4.
- [13] Q. Ouyang *et al.*, "Investigation of CMOS devices with embedded SiGe source/drain on hybrid orientation substrates," in *Proc. Symp. VLSI Technol.*, 2005, pp. 28–29.
- [14] M. Yang *et al.*, "Hybrid-orientation technology (HOT): Opportunities and challenges," *IEEE Trans. Electron Devices*, vol. 53, no. 5, pp. 965–978, May 2006.
- [15] W.-H. Lee *et al.*, "High performance 65 nm SOI technology with enhanced transistor strain and advanced-low-K BEOL," in *Proc. IEDM*, 2005, pp. 61–64.
- [16] H. S. Yang *et al.*, "Dual stress liner for high performance sub-45 nm gate length SOI CMOS manufacturing," in *Proc. IEDM*, 2004, pp. 1075–1077.
- [17] C. Ortolland, "Stress memorization technique (SMT) optimization for 45 nm CMOS," in *Proc. Symp. VLSI Technol.*, 2006, pp. 78–79.
- [18] C.-H. Chen *et al.*, "Stress memorization technique (SMT) by selectively strained-nitride capping for sub-65 nm high-performance strained-Si device application," in *Proc. Symp. VLSI Technol.*, 2004, pp. 56–57.
- [19] N. Elbel, Z. Gabric, W. Langheinrich, and B. Neureither, "A new STI process based on selective oxide deposition," in *Proc. Symp. VLSI Technol. Dig. Tech. Papers*, 1998, pp. 208–209.
- [20] H. S. Lee *et al.*, "An optimized densification of the filled oxide for quarter micron shallow trench isolation (STI)," in *Proc. Symp. VLSI Technol. Dig. Tech. Papers*, 1996, pp. 158–159.
- [21] Y.-M. Sheu *et al.*, "Modeling mechanical stress effect on dopant diffusion in scaled MOSFETs," *IEEE Trans. Electron Devices*, vol. 52, no. 1, pp. 30–38, Jan. 2005.
- [22] C. Gallon *et al.*, "Electrical analysis of mechanical stress induced by STI in short MOSFETs using externally applied stress," *IEEE Trans. Electron Devices*, vol. 51, no. 8, pp. 1254–1261, Aug. 2004.
- [23] A. T. Bradley, R. C. Jaeger, J. C. Suhling, and K. J. O'Connor, "Piezoresistive characteristics of short-channel MOSFETs on (100) silicon," *IEEE Trans. Electron Devices*, vol. 48, no. 9, pp. 2009–2015, Sep. 2001.
- [24] K.-W. Su *et al.*, "A scaleable model for STI mechanical stress effect on layout dependence of MOS electrical characteristics," in *Proc. IEEE Custom Integr. Circuits Conf.*, 2003, pp. 245–248.
- [25] V. Moroz *et al.*, "The impact of layout on stress-enhanced transistor performance," in *Proc. Int. Conf. Simul. Semicond. Processes Devices*, 2005, pp. 143–146.
- [26] L. Smith, "TCAD modeling of strain-engineered MOSFETs," in *Proc. Mater. Res. Soc. Symp.*, 2006, vol. 913, Paper ID 0913-D05-05.
- [27] V. Moroz, L. Smith, X.-W. Lin, D. Pramanik, and G. Rollins, "Stress-aware design methodology," in *Proc. Int. Symp. Quality Electron. Des.*, 2006, pp. 807–812.
- [28] Y.-M. Sheu *et al.*, "Modeling well edge proximity effect on highly-scaled MOSFETs," in *Proc. IEEE Custom Integr. Circuits Conf.*, 2005, pp. 831–834.
- [29] M. Miyamoto, H. Ohta, Y. Kumagai, Y. Sonobe, K. Ishibashi, and Y. Tainaka, "Impact of reducing STI-induced stress on layout dependence of MOSFET characteristics," *IEEE Trans. Electron Devices*, vol. 51, no. 3, pp. 440–443, Mar. 2004.
- [30] H. Tsuno *et al.*, "Advanced analysis and modeling of MOSFET characteristic fluctuation caused by layout variation," in *Proc. Symp. VLSI Technol.*, 2007, pp. 204–205.
- [31] R. Arghavani *et al.*, "Stress management in sub-90-nm transistor architecture," *IEEE Trans. Electron Devices*, vol. 51, no. 10, pp. 1740–1744, Oct. 2004.



**Andrew B. Kahng** (SM'07) received the A.B. degree in applied mathematics (physics) from Harvard College, Cambridge, MA, and the M.S. and Ph.D. degrees in computer science from the University of California at San Diego (UCSD), La Jolla.

He joined the Department of Computer Science, University of California, Los Angeles (UCLA), as an Assistant Professor, in July 1989, and became an Associate Professor, in July 1994, and a Full Professor, in July 1998. In January 2001, he joined UCSD as a Professor in the Departments of Computer Science and Engineering (CSE) and Electrical and Computer Engineering. From 2003 to 2004, he served as Associate Chair of the Department of CSE, UCSD. He was the Chair of both the U.S. Design Technology Working Group and the Design International Technology Working Group (ITWG), from 2000 to 2003, and continues to serve as Cochair of the Design ITWG. He has also served as a member of the EDA Council's EDA 200X Task Force. He has been an Executive Committee Member of the MARCO Gigascale Systems Research Center since its inception in 1998. In October 2004, he cofounded Blaze DFM, Inc., Sunnyvale, CA (a company that provides cost and yield optimization tools at the VLSI design-to-manufacturing interface) and served as its Chief Technical Officer until resuming his duties at UCSD in September 2006. He has published well over 300 journal and conference papers. Since 1997, his research on IC design for manufacturability has pioneered methods for automated phase-shift mask layout, variability-aware analyses and optimizations, CMP fill synthesis, and parametric yield-driven cost-driven methodologies for chip implementation.

Dr. Kahng was the founding General Chair of the 1997 ACM/IEEE International Symposium on Physical Design and cofounder of the ACM Workshop on System-Level Interconnect Prediction. He also defined the physical design roadmap as a Member of the Design Tools and Test Technology Working Group for the 1997, 1998, and 1999 renewals of the International Technology Roadmap for Semiconductors. He was the recipient of the NSF Research Initiation and Young Investigator awards, 11 Best Paper nominations, and six Best Paper awards [DAC, ISQED (2), ICCD, ASP-DAC/VLSI Design, and BACUS].



**Puneet Sharma** (S'02–M'05) received the B.Tech. degree in computer science and engineering from Indian Institute of Technology, Delhi, India, in 2002, and the M.S. and Ph.D. degrees in electrical engineering from the University of California, San Diego, in 2005 and 2007, respectively.

He is currently with Freescale Semiconductor, Austin, TX, in the Design for Manufacturing Department. His duties include the development of physical and electrical DFM methodologies. He has published more than 20 articles. His research interests include design for manufacturing, power analysis, and power optimization.



**Rasit Onur Topaloglu** (S'02–M'07) received the B.S. degree in electrical and electronic engineering from Bogazici University, Istanbul, Turkey, in 2002, and the M.S. degree in computer science from the University of California, San Diego (UCSD), in 2005. He is currently working toward the Ph.D. degree in computer engineering at UCSD.

He has worked on SRC and MARCO projects. He has worked part-time for Qualcomm, San Diego, CA, in 2003 and 2005; National Semiconductor, Santa Clara, CA, in 2004; and AMD, Sunnyvale, CA, in 2005. Since 2006, he has been with Advanced Micro Devices, Sunnyvale, CA, where he is with the Compact Modeling and Characterization Group. His duties include interconnect (BEOL) and device (FEOL) modeling, in particular, stress, reliability, leakage, CMP and variation characterization, modeling and design methodology development, transistor and interconnect roadmapping, and silicon test structure design, silicon correlation, TCAD simulations, and RLC characterization. He is also an SRC liaison for four projects. He has published more than 20 articles, including three invited papers. He is the holder of four pending patents. His research interests include modeling, design for manufacturability, TCAD, semiconductor manufacturing modeling, design (analog, digital, and RF), and design for test, in particular, automated fill pattern optimization, RLC extraction, and circuit optimization.

Mr. Topaloglu was a recipient of a Best Paper Award at ISQED 2007.