TSV Stress Aware Timing Analysis with Applications to 3D-IC Layout Optimization

Jae-Seok Yang, Krit Athikulwongse∗, Young-Joon Lee∗, Sung Kyu Lim∗, and David Z. Pan
Department of ECE, University of Texas at Austin, Austin, TX USA
∗School of ECE, Georgia Institute of Technology, Atlanta, Georgia, USA
jsyang@cerc.utexas.edu, {krit, yjlee, limsk}@ece.gatech.edu, dpan@ece.utexas.edu

ABSTRACT
As the geometry shrinking faces severe limitations, 3D wafer stacking with through-silicon via (TSV) has gained interest for future SOC integration. Since TSV fill material and silicon have different coefficients of thermal expansion (CTE), TSV causes silicon deformation due to different temperatures at chip manufacturing and operating. The widely used TSV fill material is copper which causes tensile stress on silicon near TSV. In this paper, we propose systematic TSV stress aware timing analysis and show how to optimize layout for better performance. First, we generate a stress contour map with an analytical radial stress model. Then, the tensile stress is converted to hole and electron mobility variations depending on geometrical relation between TSVs and transistors. Mobility variation aware cell library and netlist are generated and incorporated in an industrial timing engine for 3D-IC timing analysis. It is interesting to observe that rise and fall time react differently to stress and relative locations with respect to TSVs. Overall, TSV stress induced timing variations can be as much as ±10% for an individual cell. Thus as an application for layout optimization, we can exploit the stress-induced mobility enhancement to improve timing on critical cells. We show that stress-aware perturbation could reduce cell delay by up to 14.0% and critical path delay by 6.5% in our test case.

Categories and Subject Descriptors
B.7.2 [Hardware, Integrated Circuit]: Design Aids

General Terms
Design

Keywords
3DIC, TSV, stress, mobility variation, timing analysis

1. INTRODUCTION
3D-IC stacking has gained tremendous interests for IC integration in order to reduce wire length and increase density [1–4]. TSVs are inserted for wafer-to-wafer connection in 3D-ICs. Tungsten (W), poly-silicon, and copper (Cu) have all been considered as fill materials of TSVs. Since copper has low resistivity, it is widely used material for TSV fill. However, copper CTE differs from silicon CTE which can be a source of silicon strain. CTE of copper is 17 × 10⁻⁶ K⁻¹ at 20°C, while CTE of silicon is 3 × 10⁻⁶ K⁻¹ at 20°C [5]. The CTE mismatch between copper and silicon causes tensile stress on silicon near TSVs. Because copper electroplating and annealing temperature is higher than operating temperature, tensile stress appears on silicon after cooling down to room temperature.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

DAC 2010, June 13-18, 2010, Anaheim, California, USA.
Copyright 2010 ACM ACM 978-1-4503-0002-5 ...$10.00.

The tensile stress on silicon causes reliability problem such as cracking. In addition, the stress can change mobility of carriers. Therefore, TSV stress induced by CTE mismatch may cause timing violation if cells on a critical path are placed near TSVs. Tensile stress enhances electron mobility. However, hole mobility is either enhanced or degraded depending on TSV and transistor channel direction. Longitudinal tensile stress reduces hole mobility while transverse tensile stress increases the mobility [8]. When TSV induced tensile stress is 100MPa and the stress works for longitudinal direction, hole mobility degradation can be up to 7.2%, which makes PMOS transition slow. If PMOS is on a critical path, it can cause unexpected setup time violation which is not detected with the current timing analysis flow.

Even though several papers have been published regarding TSV stress for reliability, this is the first paper addressing TSV stress from the circuit design perspective, to our best knowledge. In this paper, we propose a design flow to analyze timing variation by TSV induced stress, and show its implications for layout optimizations during 3D-IC design. The first step of our framework is to generate stress map which is used to estimate hole and electron mobility variation. Stress calculation is based on analytical model and linear super-position. Since every cell near TSVs has a different mobility depending on stress and orientation between channel and TSV, we substitute a cell near TSVs to another cell having the same topology but having different timing characteristics according to the estimated hole and electron mobility change.

To show the benefit of our framework, we present that TSV stress aware design plays an important role to optimize timing by adjusting cell locations to take advantage of enhanced mobility property due to TSV stress. Since hole mobility contour differs from electron mobility contour, PMOS and NMOS should be optimized separately. If a PMOS transistor in a cell is on a critical path, the cell becomes a critical cell for hole mobility optimization. An NMOS critical cell can be optimally placed in a similar manner.

The main contributions of this paper include the following:

- We show that TSV stress can change hole mobility quite significantly, e.g., from -22% to +10%, which means more than 20% variation for single cell delay. Thus it can deteriorate the overall chip performance thus must be considered during timing analysis and optimization.

- We propose several layout optimization techniques including small perturbation and optimal cell rotation, and show that the optimization can improve single cell delay by 14% and critical path delay up to 6.5%.

The rest of the paper is organized as follows. First, the overall stress-aware timing analysis and design flow is presented in section 2. We propose compact mobility modeling in section 3. In section 4, we will explain how to analyze timing with TSV stress. Experimental results are shown in section 5, followed by conclusion in section 6.

2. TSV STRESS AWARE DESIGN FLOW

Overall flow for 3D-IC design methodology is proposed in Fig. 1. Our timing analysis consists of two steps. The first step is to calculate TSV induced stress and mobility change.
Figure 1: Overall flow for TSV stress aware design. Since FEA simulation which provides an accurate solution takes several hours to simulate stress for one TSV, we use the analytical model proposed in [6]. Mobility(μ) change as a function of applied stress(σ) has been proposed by the following formula [9].

\[ \frac{\Delta \mu}{\mu} = -\alpha \times \sigma \]  

(1)

where \( \alpha \) is the tensor of piezo-resistive coefficients, and \( \sigma \) is the applied stress in silicon. Positive \( \sigma \) means tensile stress while compressive stress is represented by negative \( \sigma \). We will explain the stress and mobility modeling in section 3. The second step is 3D timing analysis with TSV stress. We use PrimeTime as a STA (static timing analysis) engine. In section 4, we explain how to deal with verilog netlist and timing library to consider mobility variation. The timing result can be used for layout optimization. Intuitively, if a PMOS in a cell is on a critical path, the cell should be moved to hole mobility enhanced zone. Then, we can run timing analysis iteratively to verify perturbation effects.

3. MOBILITY VARIATION MODELING

In this paper, we assume that TSV shape is cylindrical, which is widely used for better manufacturability. FEA based TSV simulation has been proposed [5,6]. The simulation approaches provide an accurate solution with long runtime which is not acceptable for our design flow that should calculate stress for several thousands of TSVs iteratively after each optimization. Assuming 2-D planar stress, we use the following analytical solution which is known as Lame’ stress solution in [6].

\[ \sigma_{rr} = \frac{B\Delta \alpha \Delta T}{2} \left( \frac{R}{r} \right)^2 \]  

(2)

The analytical stress model provides a relatively accurate solution [6]. In the formula (2), \( B \) is biaxial modulus, \( \Delta \alpha \) is CTE difference between copper and silicon, \( \Delta T \) is the temperature difference between copper annealing and operating temperature. \( R \) is TSV radius, and \( r \) is a distance from the center of TSV. We assume that \( \Delta T \) is 175°C which is the case of 25°C for the room temperature and 200°C for the copper annealing temperature [10]. The formula shows that the thermal stress near TSV depends on the ratio of TSV radius and a distance from a TSV edge.

The formula (1) provides an efficient way to calculate mobility variation due to \( \sigma_{rr} \). Mobility change depends on not only \( \sigma_{rr} \) but also orientation between applied force and a transistor channel. The empirical value for showing the relation of mobility change and a channel direction has been proposed in [11]. We extend the formula (1) to consider stress and channel direction in (3).

\[ \frac{\Delta \mu}{\mu} (\theta) = -\alpha \times \sigma_{rr} \times \alpha (\theta) \]  

(3)

Figure 2: Optimal orientation of MOSFET to maximize mobility for (001) surface, (110) channel.

\[ \theta = \tan^{-1} \frac{Y_{TSV} - Y_{poly}}{X_{TSV} - X_{poly}} - \frac{\pi}{2} \]  

(3)

where \( \alpha (\theta) \) is an orientation factor as a function of \( \theta \) which is defined as degree between the center of TSV and the center of a transistor channel when a transistor is placed vertically as shown in Fig. 2(a),(b). \( \alpha \) is the piezo-resistive coefficient at \( \theta = 0 \) which works as longitudinal stress.

In Fig. 2(a), if NMOS is in right side of TSV, \( \theta \) becomes zero, and \( \alpha(0) \) becomes one, which enhances NMOS mobility at its maximum. However, if NMOS is in upper side of TSV, \( \alpha(\pi/2) \) is 0.5, which means that NMOS mobility increase is half of the enhancement at \( \theta=0 \). PMOS shows opposite trends, which has the best mobility enhancement at \( \theta=\pi/2 \). Fig. 2(c) and (d) show the transistor direction for the best performance. Even though the mixed channel direction is not allowed due to the patterning difficulty, the observation provides a way to optimize layout for 3D-ICs.

Figure 3: Stress contour map with 0.5um KOZ. We generate stress contour map based on (2). Fig. 3 shows contour for a TSV having radius of 1.5um. Since the region near TSV may have a crack or extremely high stress, we define that 0.5um from TSV edge is Keep-Out-Zone(KOZ), in which no cell is allowed to be placed. We can see stress of more than 200MPa out of KOZ. Approximately, 100MPa stress appears on the region of 1um from a KOZ edge.

Fig. 4(a) shows a contour map for hole mobility variation. From the contour, we can see that hole mobility decreases in a horizontal direction, while it increases in a vertical region. 45° direction has no hole mobility change. Contour map for electron mobility variation is presented in Fig. 4(b). As we see in Fig. 2(a), horizontal direction has more mobility.
renamed to INVX1_N8.P8 which means -8% hole mobility, +8% electron mobility in Fig. 6.

We prepare a verilog netlist and a parasitic extraction file (SPEF) for each die. In addition, we make a top level verilog netlist that instantiates the dies and connects them using wires which corresponds to TSV connections. Then we make a top level SPEF file for the TSV connections. With a proper timing constraints file, we can run PrimeTime and get the 3D STA results.

To consider the systematic variation during timing analysis, we characterize a cell with different mobility corners as shown in Fig. 7. Hole mobility variation is from -14% to +8%, and electron mobility variation is up to +8% to cover stress caused by TSVs in Fig. 6. I1 in Fig. 6 is matched the corner near FF corner, while I3 is in FS corner. With mobility variation aware library and verilog netlist having renamed cells, we run PrimeTime to do timing analysis with TSV stress.

To cover mobility variation caused by multiple TSVs, the mobility variation range needs to be extended to PMOS: from -22% to +10%, NMOS: from 0% to +24%. If mobility step is 2%, we need to characterize 221 library with mobility variation range needs to be extended to PMOS: from -22% to +10%, NMOS: from 0% to +24%. If mobility step is 2%, we need to characterize 221 library with different mobility values which is not available. However, we can observe that rising delay only depends on \((\Delta \mu/\mu)_h\), falling delay variation depends on \((\Delta \mu/\mu)_e\) from Fig. 8. When we simulate inverter rising delay with mobility variation, electron mobility variation does not work for the delay. Similarly, we can see that falling delay only depends on electron mobility variation. In addition, from Fig. 8, we can see that hole mobility variation can cause more than 20% PMOS performance variation depending on device technology, and electron mobility variation can enhance NMOS performance up to 7.5%. We use inverter in NCSU library and PTM spice model [12] to obtain Fig. 8. Therefore, we can fix \((\Delta \mu/\mu)\), when we sweep \((\Delta \mu/\mu)_h\), 30 (=17+13) library characterization will be enough to cover the entire mobility set. If mobility step is 4%, 16 (=9+7) library characterization will be enough to cover the entire mobility set. If mobility step is 4%, 16 (=9+7) library set is required. Since delay variation has semi-linear dependency on mobility variation, we can use interpolation for the mobility value between two libraries.

5. EXPERIMENTAL RESULTS

We implement TSV stress aware 3-D timing analysis flow in C++ and generate the mobility aware library based on NCSU 45nm cell library. TSV used in this experimentation is in Table 1.
First, we show the efficiency of our compact stress and mobility modeling. Even though we generate mobility contour for a block having die size: 1.75\times1.75mm^2, #TSVs: 462, it takes only 14.9s. The proposed timing analysis with compact process/device model is fast enough to be used for iterative optimization purpose.

Second, we compare stress aware timing result with no stress case. Ten benchmark circuits are used to show the timing variation in Table 2. The benchmark circuits are placed for wire length minimization [1] without TSV stress consideration. We assume that there are four dies stacking, and the number of inserted TSVs are 10% of #cells in each circuit. When we consider TSV stress effect, the longest path delay of the benchmarks has variation from -1.32% to 3.58%. Some benchmarks have timing gain while some benchmarks have timing penalty. If we consider TSV stress effect during cells and TSVs placement, we can expect performance improvement for every benchmark. TNS (total negative slack) has more variation from -12.43% to 22.68% which is bigger than delay variation. That motivates the need of TSV stress aware layout optimization.

Last, we manually optimize a critical path in 8051 to present the potential benefit of TSV stress aware layout optimization. Before optimization, the path delay is 4.64ns with stress aware timing analysis. However, we could reduce the delay to 4.62ns with small layout perturbation which is 6.5% improvement. Table 3 shows the gates on the path. We can see that cell relocation according to the mobility variation is critical. We adjust each cell location with small perturbation so that each cell has timing gain. The maximum timing gain in a cell is 14%. Fig. 9 shows how cell relocation works for timing optimization. We capture the placement result on die2 with mobility variation in Fig. 9. As the cell in logic depth 2, 4, 6, and 9 are mobility critical cells because the timing arc is rising on the path. Therefore, we perturb the cells to be placed close to green area in hole mobility contour. However, the cells in logic depth 3 and 5 are electron mobility critical. Therefore, we push the cells to electron mobility enhancement zone in Fig. 9(c) (d).

6. CONCLUSIONS

The 3D IC stacking requires TSVs for interconnection between wafers. Cu TSV causes thermal stress which can lead to significant timing variations. Stress, though commonly believed to have negative impact on timing can actually be taken advantage of for timing optimization, since it is a strongly layout dependent, systematic effect. In this paper, we present the potential benefit of TSV stress aware timing analysis framework for 3D-IC also opens the opportunity for stress-aware layout optimizations, such as placement and TSV optimizations.

ACKNOWLEDGMENTS

This work is supported in part by NSF and SRC Interconnect Focus Center (IFC).

7. REFERENCES