Stress Aware Layout Optimization

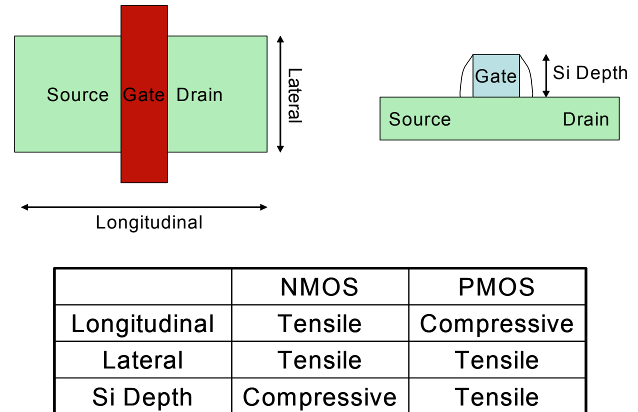
Vivek Joshi, Brian Cline, Dennis Sylvester, David Blaauw, Kanak Agarwal*
University of Michigan, Ann Arbor, MI. email: {vivekj, btcline, dennis, blaauw}@eecs.umich.edu
*IBM Research, Austin, TX. email : kba@us.ibm.com

Abstract- Process-induced mechanical stress is used to enhance carrier transport and achieve higher drive currents in current CMOS technologies. In this paper, we study how stress-induced performance enhancements are affected by layout properties and suggest guidelines for improving layouts so that performance gains are maximized. All MOS devices in this work include STI and nitride stress liners as sources of stress. Additionally, the PMOS devices incorporate the stress effects caused by the embedded SiGe S/D layer common in today's processes. First, we study how stress and drive current depend on layout parameters such as active area length and contact placement. We develop an intuition for the drive current dependency on these parameters and propose simple guidelines to improve a layout while considering mechanical stress effects. We then use these guidelines to improve the standard cell layouts in a 65nm industrial library. Experimental results show that we can enhance NMOS and PMOS drive currents by ~5% and ~12%, respectively, while only increasing NMOS leakage current by 1.48X and PMOS leakage current by 3.78X. By applying our guidelines to a 3-input NOR gate and a 3-input NAND gate, we are able to achieve a ~13.5% PMOS drive current improvement in the NOR gate and a ~7% NMOS drive current improvement in the NAND gate, without increasing cell area in either case.

1. INTRODUCTION

As industry strives to extend Moore's law through aggressive process scaling, significant challenges arise. Maintaining performance and reliability while facing fundamental scaling limitations (i.e. gate oxide thickness) is a major challenge. We can no longer scale certain device parameters such as t_{ox}, V_{th}, V_{DD} as aggressively as gate length (L) without significantly degrading reliability and exponentially increasing leakage current. Additionally, as MOSFET's continue to scale below 100nm, higher effective fields cause mobility degradation, leading to decreasing drive currents. In order to battle mobility degradation and achieve higher drive currents, modern-day fabrication processes use special means to induce mechanical stress in MOSFET's, which enhances carrier mobility. Mobility enhancement has emerged as an attractive alternative to device scaling because it can achieve similar device performance improvements with reduced affects on reliability and leakage.

Mechanical stress in Silicon leads to band splitting and alters the effective mass, which results in carrier mobility changes [1, 2]. Induced stress in the channel can be either tensile or compressive. As illustrated in Figure 1, NMOS and PMOS devices have different desired stress types (compressive or tensile) in the longitudinal, lateral and Si-depth (vertical) dimensions. By providing the correct type of stress for a device (in one or more dimensions), we can achieve higher



	NMOS	PMOS
Longitudinal	Tensile	Compressive
Lateral	Tensile	Tensile
Si Depth	Compressive	Tensile

Figure 1. Desired stress types for NMOS and PMOS [3].

drive currents. Mechanical stress can be generated by either thermal mismatch or lattice mismatch. Thermal mismatch stress is caused by differences in the thermal expansion coefficient, while lattice mismatch stress is caused by differences in lattice constants. Figure 2 shows the major sources of stress for one of the latest 65nm CMOS technologies [4]. Apart from Shallow Trench Isolation (STI), which creates compressive stress longitudinally and laterally due to thermal mismatch, other sources of stress are used to enhance transistor speed. For PMOS, an embedded SiGe process is implemented where SiGe is epitaxially grown in cavities that have been etched into the source/drain areas [5]. Lattice mismatch between Si and SiGe creates a large compressive stress in the PMOS channel, thereby resulting in significant hole mobility improvement. In this process, NMOS is protected by a capping layer to prevent Si recess and SiGe epitaxial growth.

As shown in Figure 2, mechanical stress can also be transferred to the channel through the active area and polysilicon gate by depositing a permanent stressed liner over the device [6]. Tensile liners improve electron mobility in NMOS devices, while compressive liners improve hole mobility in PMOS devices. The latest high performance process nodes have simultaneously incorporated both tensile and compressively stressed liners into a single high performance CMOS

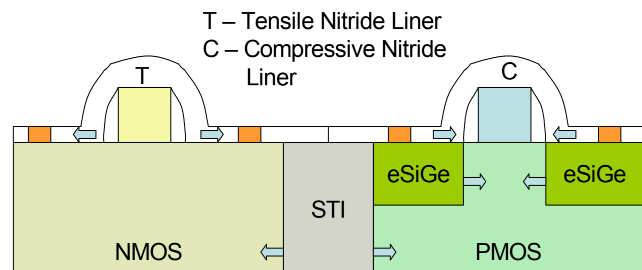


Figure 2. Sources of stress for NMOS and PMOS.

DRAFT

flow, called the Dual Stress Liner approach. In this process, a highly tensile Si_3N_4 liner is uniformly deposited over the entire wafer. The film is then patterned and etched from the PMOS regions. Next, a highly compressive Si_3N_4 liner is deposited, patterned and etched from the NMOS regions. In addition to the permanent tensile liner shown in Figure 2, a Stress Memorization Technique (SMT) is also used to increase the stress in n-Type MOSFET's [7]. In this technique, a stressed dielectric layer is deposited over all of the NMOS regions, thermally annealed and then completely removed. The stress effect is transferred from the dielectric layer to the channel during annealing and is "memorized" during the re-crystallization of the active area and gate polysilicon.

A closer examination of these stress sources shows that the amount of stress transferred to the channel, and, consequently, the drive current enhancement, has a strong dependency on the layout. The amount of SiGe (and hence the stress), for example, depends upon the length of the active area. Longer active area also means that the STI will be pushed further away from the channel, which will lower its effect on the total channel stress. Therefore, the drive current of a transistor depends not only upon the gate length L and width W, but also the exact layout of the individual transistor and its neighboring transistors. This means that the performance of two transistors with identical gate lengths and widths can actually differ significantly, depending on their layouts. The goal of this work is to study the layout dependence of stress-based performance enhancement for different device configurations and develop simple guidelines to improve the layout so that the performance gains are maximized. The idea is to identify the key layout parameters that a layout designer can change to affect the transistor performance. Since we are interested in optimizing the layout, uniform techniques such as SMT can be ignored while modeling the layout dependence of stress because SMT involves a uniform film deposition, anneal and removal over all of the NMOS regions, which leads to a uniform shift in NMOS drive current that is relatively independent of layout.

To date, there has been limited research on the layout dependence of stress-based current improvement. Most of the published work has focused on the effects of STI [8]. However, the papers that have analyzed other sources of mechanical stress do not include all of the sources (such as epitaxial SiGe) and only study the PMOS stressors [9, 10]. More importantly, none of these works address the issue of modifying the layout to maximize the mechanical stress-based performance enhancement, using the intuition developed. This is where the key contribution of this paper lies. We performed a comprehensive study in order to determine how various layout parameters affected device stress, and then analyzed their impact on device performance. From the study we developed general layout rules that serve as guidelines for optimizing transistor performance. We then show how standard cell layouts from an industrial 65nm CMOS technology can be improved by following these simple rules. Experimental results show that we can obtain a 12% performance enhancement for PMOS devices (up to about 20%), while only increasing the leakage current by ~3.78X. For NMOS devices we can achieve a drive current improvement of about 5% while increasing the leakage current by only 1.4X. Furthermore, we discovered that there is ample scope to improve the drive current for standard cells by altering the layout (with zero area penalty) in accordance with the guidelines proposed. We increased PMOS drive current in a 3-input NOR gate by ~13.5%, and increased NMOS drive current in a 3-input NAND gate by ~7% by applying our guidelines to the corresponding layouts. Since delay is inversely proportional to drive current, an increase in drive current results directly in an improvement in delay; for example, a 13.5% increase in PMOS current translates to a ~12% decrease in pin-delay.

The rest of the paper is organized as follows. Motivation for this work is discussed in Section 2. Section 3 presents a study on the layout dependence of stress-based performance enhancement, while Section 4 develops simple guidelines for improving the layout. Experimental results from applying these guidelines to 65nm industrial CMOS standard cells is discussed in Section 5, and Section 6 concludes the paper.

2. MOTIVATION

Mechanical stress in Silicon breaks crystal symmetry and removes the 2-fold and 6-fold degeneracy of the valence and conduction bands, respectively. This leads to changes in the band scattering rates and/or the carrier effective mass, which in turn affects carrier mobility. Since changes in mobility directly influence the drive current, higher carrier mobility improves transistor performance. However, increased mobility not only improves the drain current in the saturation regime of MOSFET operation, but it also increases the subthreshold leakage current. In order to study the trade-offs involved, we need to examine the saturation and subthreshold current equations in order to determine their dependency on carrier mobility. This also allows us to compare mobility enhancement to other performance enhancement techniques, such as V_{th} reduction. Equations 1 and 2 below give the expressions for drain current [11, 12] when the transistor is operating in the saturation and subthreshold regimes, respectively.

$$I_D = \frac{\mu_0}{[1 + U_0(V_{GS} - V_T)]} \cdot \frac{C_{ox}}{2aV} \cdot \frac{W}{L_{eff}} \cdot (V_{GS} - V_T)^2$$
$$V = \frac{1 + v_c + \sqrt{1 + 2v_c}}{2} \qquad v_c = U_1((V_{GS} - V_T)/a) \tag{1}$$

$$I_{sub} = A \cdot e^{\frac{1}{nv_T} \cdot (V_G - V_S - V_{th0} - \gamma' V_S + \eta V_{DS})} \cdot (1 - e^{(-V_{DS})/v_T})$$
$$A = \mu_0 C'_{ox} \frac{W}{L_{eff}} v_T^2 e^{1.8} e^{-\frac{\Delta V_{th}}{\eta v_T}} \tag{2}$$

As seen in (1), the saturation drain current, I_D, has a sublinear dependence on mobility, μ_0. On the other hand, as shown in (2), the subthreshold drain current dependence on mobility is linear. The dependence on V_{th}, however, is almost linear in saturation, but is exponential in the subthreshold regime. Therefore, if we obtain identical saturation current improvement using two separate enhancement techniques: 1) stress-based mobility enhancement, and 2) V_{th} reduction, then the corresponding increase in leakage current for the reduced V_{th} case will be much higher (due to the exponential dependence of I_{sub} on V_{th}). Consequently, the reduced

DRAFT

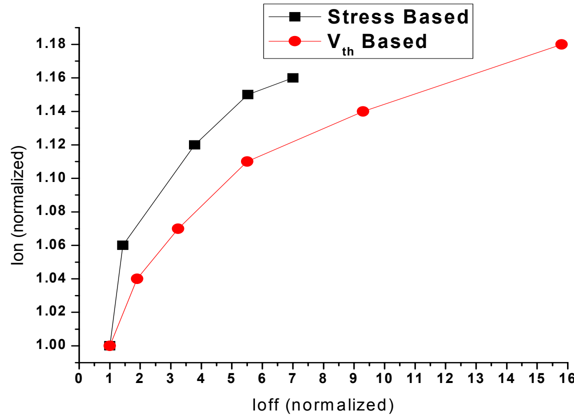


Figure 3. Ion versus Ioff curves for Vth assignment and stress-based performance enhancement for a 65nm PMOS device.

increase in leakage current makes mobility enhancement a more attractive option than its V_{th} counterpart.

The benefits of using mobility enhancement over V_{th} reduction is illustrated in Figure 3, which shows the normalized Ion versus Ioff curves for stress-based and V_{th}-based performance enhancements for an isolated 65nm PMOS device. The device has three sources of stress: STI, a compressive nitride liner, and embedded SiGe source/drain regions. Stress is varied by changing the active area length, while the n-channel doping is changed to vary V_{th}. The curves clearly show that the trade-off is better for stress variation. For a 12% improvement in Ion, the leakage for the V_{th} case is nearly twice as large as that for the stress-based improvement, and the difference is only amplified for higher values of improvement. Also, stress-based improvement allows for more fine-grain improvement control than V_{th} assignment, where only 2-3 V_{th} values are typically allowed. For these reasons, a designer would prefer to achieve performance enhancement through increasing stress whenever possible. The superiority of the stress-based performance improvement technique makes it an appealing option for further investigation. Thus, the next two sections study the layout dependence of stress, and develop guidelines for optimizing layouts so that stress-induced enhancements are maximized.

3. LAYOUT DEPENDENCE OF STRESS-BASED PERFORMANCE ENHANCEMENT

In order to study the layout dependence of stress-based performance enhancement, we used the Davinci 3D TCAD tool [13], which has an extensive set of stress related features. Additionally, we followed the layout rules from an industrial 65nm CMOS technology and the device fabrication was simulated in Tsuprem4 [14] (in order to capture the process-induced stress). The stress values were then imported into Davinci, which simulated the device and solved for the stress-based mobility enhancement equations. The resulting values for drive current and stress were found to be consistent with previously published 65nm technology data. Our consistency with these fabricated measurements can be attributed to the fact that we model all of the layout dependent sources of stress in the industrial 65nm technology. For a PMOS device, the sources of stress that are layout dependent include the compressive nitride liner, embedded SiGe source/drain, and STI. The NMOS sources, on the other hand, only include the tensile nitride liner and STI. We have ignored the Stress Memorization Technique (SMT) in our simulations, since it involves a uniform deposition and eventual removal of a dielectric layer over all NMOS devices (as discussed previously in Section 1). SMT, therefore, does not depend on layout properties and can be accurately treated as a uniform increase in NMOS drive current, independent of layout.

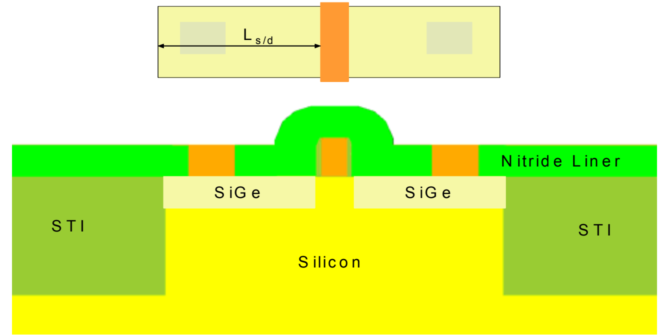


Figure 4. Isolated PMOS device.

Figure 4 shows the 2D cross-section of an isolated PMOS device, surrounded by STI, and the corresponding layout view. For the device shown, we increase the active area length ($L_{s/d}$) and examine the corresponding changes in drive current. Increasing active area length has a number of effects: 1) it increases the amount of SiGe, causing more stress to be transferred to the channel; 2) it increases the distance between the channel and the STI, decreasing the effect STI has on channel stress; 3) it allows more nitride over the active area. The nitride layer actually transfers stress in two ways – vertically through the gate and longitudinally through the active area. Since active contacts create openings in the nitride layer, the longitudinal component of nitride stress can be increased by moving the contacts away from the channel. Similarly, a source/drain region that does not have any contacts (or has a smaller number of contacts) will have higher channel stress than one that has a high contact density.

Figure 5a shows the longitudinal stress (S_{xx}) in the same isolated PMOS device for two normalized $L_{s/d}$ values of 1 and 1.58. The $L_{s/d}$ values have been normalized to the minimum possible S/D length for a region that contains a contact, in accordance with the layout design rules of the industrial library. Figure 6 shows the drive current, Ion, and leakage

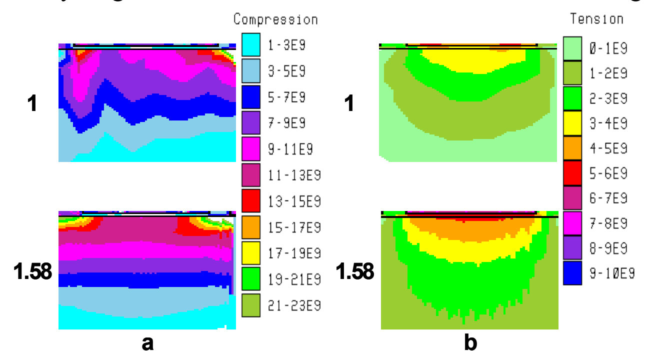


Figure 5. Longitudinal stress component S_{xx} (in Pascals) for normalized $L_{s/d}$ of 1 and 1.58 for a) PMOS b) NMOS.

DRAFT

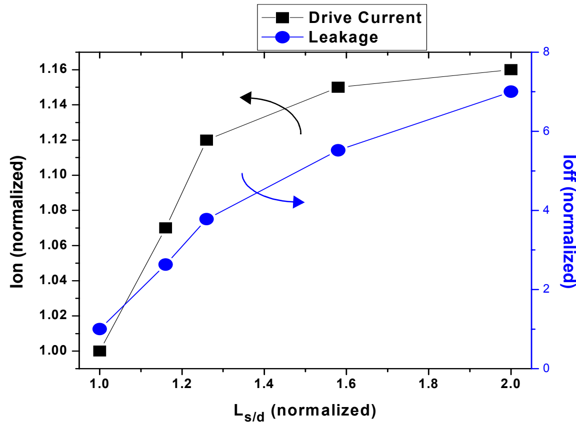


Figure 6. Ioff and Ion versus $L_{s/d}$ curves for stress-based performance enhancement in an isolated PMOS device.

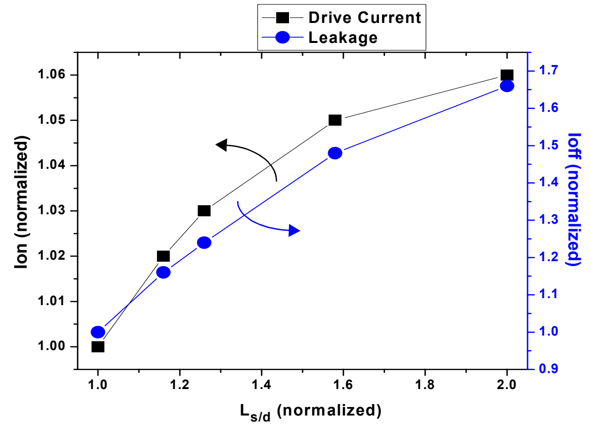


Figure 7. Ioff and Ion versus $L_{s/d}$ curves for stress-based performance enhancement in an isolated NMOS device.

current, Ioff, plotted against the S/D length, $L_{s/d}$. Results show that for a 12% performance increase, leakage current only increases by 3.78X. This Ion versus Ioff trade-off, as we have already shown in Section 2, is much better when compared to the enhancement technique where V_{th} is reduced. In addition to illustrating the layout-dependent trade-offs, Figure 6 also shows the limit for extending the S/D length. Increasing the source/drain length beyond 1.58 (normalized value) yields minimal performance gains, even when active area length and leakage current are increased substantially. As mentioned previously, the performance enhancement is also sensitive to contact placement. The experimental results show that about 65% of stress is transferred through the gate and the rest is transferred through the active area. Moving the contacts away from the channel accounts for nearly 2.6% of the drive current improvement. Also, a device with no contacts on the drain side (typically seen for devices in series) has about 4% higher performance.

Unlike its PMOS counterpart, NMOS device performance is actually degraded by STI since it induces compressive stress in the channel. The other stressor present in NMOS devices is the tensile nitride layer, which again transfers stress through the gate and active area (influenced by the contacts). However, in NMOS devices, the contact placement becomes much more important. Increasing the active area pushes away the compressive STI and allows more room for the contacts to be separated from the channel. Figure 5b shows the longitudinal stress in an isolated NMOS device for normalized $L_{s/d}$ values of 1 and 1.58. Figure 7 shows the corresponding Ion and Ioff plotted against the active area length ($L_{s/d}$). We can achieve a 5% performance gain for a leakage current increase of only 1.48X. Analogous to the PMOS S/D extension limits discussed previously, NMOS S/D extensions also have an upperbound – 1.58 (normalized value). Beyond this value, the area and leakage current penalties do not warrant the minimal gains in Ion. The increase in performance here is limited by the fact that we are increasing only the nitride's longitudinal stress through the active area (about 35% of the total stress due to nitride), and pushing away the STI (which has a relatively smaller contribution to the overall channel stress). Unlike the case of the PMOS, a major portion of the overall performance enhancement can be attributed to moving the contacts away from the channel. Experimental results show that almost 80% of the total improvement is due to moving the contacts. Also, a device with no contacts on the drain side has about 2% higher performance.

Next, we studied transistor performance in denser layouts. Figure 8 shows the channel stress and the corresponding layout view for three PMOS transistors in a 3-input NAND gate. The device in the center (device 2) has higher stress than the two corner transistors because it is surrounded by more SiGe. This difference in stress is reflected in their drive performance, and simulations show that the drive currents for the center and edge devices differ by 8.2%. Furthermore, if there were five devices side-by-side instead of three, the difference would increase to 14.8%. This means that the drive current of a transistor is not only layout-dependent, but it is also location-dependent. Similar experiments for NMOS devices show differences of 7.4% and 12.2% for the case of three and five side-by-side transistors, respectively.

DRAFT

4. GUIDELINES TO IMPROVE A LAYOUT IN TERMS OF PERFORMANCE

Based on the intuition developed in the previous section, we now propose some simple guidelines to optimize a given layout for stress-induced performance enhancement. Our focus is to propose rules for those aspects of the layout that a layout designer can control (such as active area length or contact placement) and not the parameters that are fixed for a technology (such as nitride layer thickness or embedded SiGe source/drain depth). Once the guidelines are presented, the end of this section discusses one other important stress

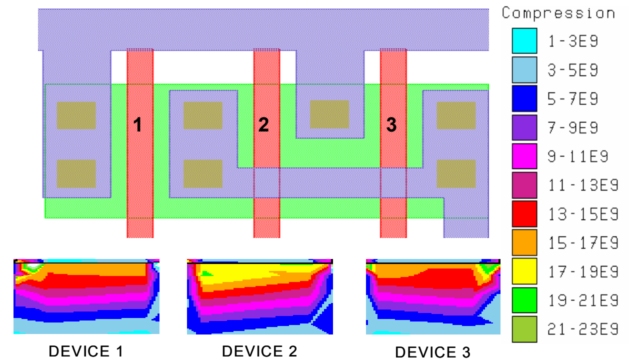


Figure 8. PMOS devices for a 3-input NAND gate and the corresponding channel stress distribution (in Pascals).

effect: the position-dependency of stress-induced performance enhancement. When mechanical stress is present in MOSFETs, matching W and L does not guarantee similar transistor performance even when neglecting process variation. Apart from W and L, the drive current is also affected by the layout parameters that influence stress: active area length, placement and number of contacts, and device context (i.e. whether the device is surrounded by other transistors or isolated by STI on one or both sides). In this paper, we have already discussed the first two parameters in great detail, while the third parameter (device context) has only been briefly mentioned (at the end of Section 3). However, since the device context or position of a transistor within a layout also affects performance, it must be accounted for by the designer, so this phenomenon is discussed in more detail after our proposed guidelines.

The following are our layout rules for improving performance for devices under stress. The key feature of these guidelines is that their application to a particular standard cell does not increase cell area; all modifications are made within the pre-existing cell boundaries.

1. ***Increase the active area in a given cell to fill up the entire cell width while obeying the DRC rules:*** This guideline is most readily applied to a compact pull-up or pull-down network (often containing an NMOS or PMOS stack) that does not use the full width of a cell. For instance, in the case of stacked transistors, the layout does not require contacts between intermediate nodes. Thus, their spacing can be significantly tighter because nodes that contain contacts need larger spacing to satisfy the technology's design rules. In the absence of stressors, it is best to minimize the active area in order to reduce the capacitance. However, in the presence of stressors, increasing active area length not only results in higher stress in the channel (and, hence, higher drive current), but it also increases the source/drain capacitances. In a given CMOS layout, increased S/D capacitance for transistors closer to the output will directly affect the output capacitance, while transistors closer to the VDD and GND rails will have a smaller affect. Hence, this guideline should be applied to cells with larger output loads, so that the change in capacitance is a small fraction of the total output capacitance. The authors would like to note that this guideline can also be extended to create high performance versions of standard cells which incur some area penalty, but are assigned optimally within a design.

2. ***Move the contacts away from gate polysilicon:*** Moving the contacts away from the channel allows more stress to be transferred by the nitride layer. For an isolated device, we recommend pulling the contacts as far away from the poly as the design rules permit. For contacts between two gates, either place them midway for identical performance enhancement of both transistors, or place them closer to the non-critical transistor (increasing stress in the critical device). Moving the contacts away will also result in a small increase in the source/drain resistance, but the increase is typically less than 5Ω, and the resulting gain in drive current outweighs the increase.

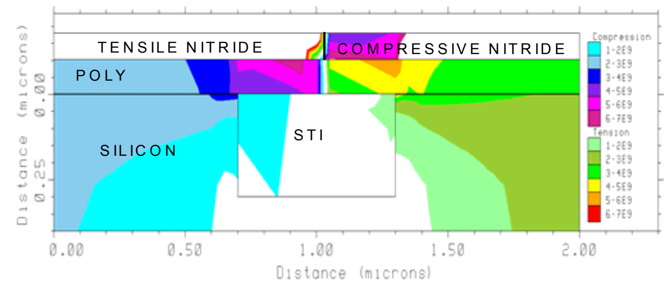


Figure 9. Sources of stress (in Pascals) for NMOS and PMOS.

3. ***In the lateral direction, move the PMOS closer to the tensile/compressive nitride interface and the NMOS away from it:*** From Figure 1, we know that the desired stress in the lateral direction is tensile for both NMOS and PMOS. Figure 9 shows the lateral stress behavior near the interface of the two nitrides (cross-section across the poly going from PMOS to NMOS over STI). The behavior is curious in the sense that there is a region of compressive stress under the tensile nitride (NMOS side) and there is a region of tensile stress under the compressive nitride (PMOS side). Therefore, if possible, it would be beneficial to move the PMOS active area into this region of tensile stress and the NMOS away from the region of compressive stress. The space for this movement is most readily available when the transistor widths are small but the cell pitch (lateral size) is large (due to pitch uniformity across standard cells); this combination of properties, for example, is common in minimum sized, simple gates (e.g. minimum size inverters, buffers, or 2-input NAND/NOR's).

DRAFT

Apart from these general guidelines for optimizing a given layout, a designer must be aware of how the channel stress is affected by the position of a device within the layout. Stress in the channel of a device depends not only upon its S/D lengths and contact placement, but also upon its surroundings. As we have shown in the previous section, devices that share their source/drain regions with other transistors have significantly higher stress (and hence drive current enhancement) than those at the edges of an active region (which are therefore bordered by STI), even for identical $L_{s/d}$ and contact placement. This difference in stress can be attributed to the effects of STI, as well as the fact that stressors for a device also affect its neighbors.

Ignoring the position-dependence of stress could lead to a number of design issues. First of all, the location of a transistor could result in an unexpected increase in drive current, resulting in smaller delay. Therefore, timing characterization should account for the maximum possible drive current increase for a device, due to its position. Overlooking this increase in current could lead to possible hold-time violations, as some gates might be faster than expected. Secondly, the position-dependent current offset could modify the noise margins of a circuit. Hence, for circuits that are sensitive to noise margins (e.g., SRAM cells, Sense Amplifiers, etc.), these deviations must be accounted for either during the design phase (for example, by guarbanding against position dependent offsets), or during the layout phase (e.g., by modifying the $L_{s/d}$'s to cancel the offsets). Finally, in certain circuits, if the strength of a transistor (in terms of drive current)

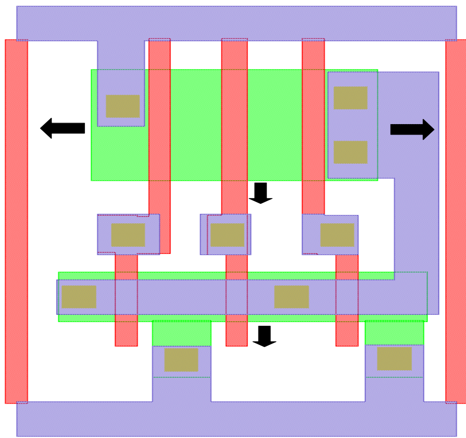


Figure 10. Layout of a 3-input NOR gate showing the scope for improvement.

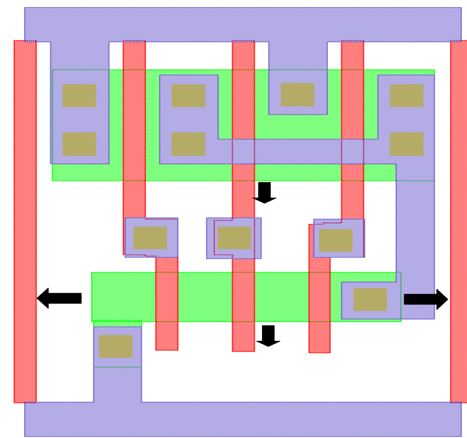


Figure 11. Layout of a 3-input NAND gate showing the scope for improvement.

is increased beyond the expected value, it could cause a substantial drop in performance. For example, any logic style where one device has to overcome another in order to change the output would be sensitive to drive strength (e.g., dynamic logic with a keeper device). In such a case, if a device that was designed to be weak was actually strengthened due to position, it could lead to significant performance degradation. All in all, designers need to be aware of the effect that position has on performance, especially if pin-to-pin delay, noise margins, or transistor strength are essential to a particular design.

The next section discusses the results of applying our guidelines to an industrial 65nm CMOS technology standard cells.

5. APPLYING THE GUIDELINES TO STANDARD CELL LAYOUTS

This section discusses the effectiveness of applying our layout guidelines from Section 4 to standard cells from an industrial 65nm CMOS technology library. For a given layout, as shown in Section 3, a basic trade-off always exists between the source/drain length, $L_{s/d}$, and the improvement in drive current. By exploiting this trade-off, we can make faster, but leakier, versions of the standard cells with varying area increments and assign them intelligently to the critical paths in order to optimize performance. All of the active area length increments should utilize Guideline 2 from the previous section and move the contacts away from the gates. Aside from using Guidelines 1 and 2 to create our standard cell variants, we also employ Guideline 3 to achieve additional gains in drive current.

Figure 10 shows the layout for a 3-input NOR gate. It consists of three PMOS transistors in series (a 3-PMOS stack) and three NMOS transistors in parallel. This means that the source and drain of each NMOS is connected to the ground and the output, respectively, necessitating contacts at each node. The PMOS stack on the other hand, only needs one contact to VDD (at the source of the top PMOS) and one contact to the output (at the drain of the bottom PMOS). Using the classical layout methodology (where stress is ignored and capacitance is minimized), we can shrink the non-contacted S/D regions to lower the parasitic PMOS capacitance. As shown in Figure 10, the PMOS region has the capability of increasing the source/drain lengths by ~22% without affecting the overall cell area in accordance with Guideline 1. While increasing the source/drain lengths, we simultaneously adhere to Guideline 2 and shift the contacts away from the gates, maximizing performance enhancement. If we increase the active area uniformly for all transistors, drive current improves by about 12% for each PMOS. Also, there is lateral room to move the NMOS and PMOS active area in accordance with Guideline 3. This leads to further improvements of about 3% and 1.5% for NMOS and PMOS devices, respectively. Therefore, for the 3-input NOR gate, we observe overall improvements in drive current of about 13.5% for PMOS devices and about 3% for NMOS devices. Similarly, by applying Guidelines 1–3 to the layout of a 2-input NOR gate, we can achieve drive current improvements of 7.5% and 3% for the PMOS and NMOS devices, respectively.

Figure 11 shows the layout for a 3-input NAND gate. Instead of a PMOS stack, we have an NMOS stack in this case, so there is a potential to increase the NMOS active area length without affecting the cell area. In accordance with Guidelines 1 and 2, we can get an improvement of about 4% for each of the NMOS drive currents. Also, there is room for moving the active areas in accordance with Guideline 3. This leads to further improvements in NMOS and PMOS devices of about 3% and 1.5%, respectively. Overall, we can achieve a ~7% NMOS performance enhancement and a ~1.5% PMOS performance enhancement. Similarly, by applying Guidelines 1–3 to the layout of a 2-input NAND, we can obtain drive current improvements of 4.5% and 1.5% for the NMOS and the PMOS devices, respectively. Scope for such layout based improvements is found in most of the library standard cells.

Table 1 summarizes the results of applying Guidelines 1–3 to a few standard cells. It reports the percentage drive current improvement, leakage current increase, and the percentage increase in the output capacitance (assuming an F04 output loading). It also reports the leakage current increase for identical drive current improvements through V_{th} reduction. Comparing the leakage current increase for stress-aware layout optimization to V_{th} reduction re-establishes the superiority of the stress-aware layout optimization. For a 3-input NOR gate, the PMOS leakage current increased by 4X when the layout was optimized in accordance with our guidelines, while the corresponding increase for the V_{th} reduction case

DRAFT

Table 1. Summary of stress-aware layout optimization drive current improvement and trade-offs in 65nm standard cells.

Cell Name	Percentage drive current improvement by layout optimization		Increase in leakage current by layout optimization		Increase in leakage current for identical drive current improvement by V_{th} reduction		Percentage increase in output capacitance with a F04 output loading
	NMOS	PMOS	NMOS	PMOS	NMOS	PMOS	
3-input NOR	3%	13.5%	1.22X	4.02X	1.31X	9.20X	2.74%
2-input NOR	3%	7.5%	1.22X	2.24X	1.31X	3.52X	1.92%
3-input NAND	7%	1.5%	1.98X	1.10X	2.36X	1.53X	1.85%
2-input NAND	4.5%	1.5%	1.45X	1.10X	1.68X	1.53X	1.30%

was 9.2X. The increase in NMOS leakage for a 3-input NAND gate was found to be 2X for stress-based layout optimization, and 2.4X for the case of V_{th} reduction. Application of Guideline 1 increased the S/D capacitance since we increased the $L_{s/d}$, but as shown in Table 1, this increase was very small (<3% if we assume an FO4 output loading).

As mentioned in Section 5, the position of a device within a layout also affects its stress, and, therefore, the drive current. This position-dependent drive current enhancement can significantly hurt the performance for some circuits. This fact was verified using the circuit shown in Figure 12, which contains the schematic and partial layout for a basic domino implementation of a 2-input AND gate. Keeper device P2 is a weak PMOS that is used to hold the high state at node N during the evaluation period of the clock, so that N is not discharged by the NMOS leakage currents. The keeper, P2, should be sized large enough to replace NMOS leakage current and sustain a high voltage at node N, but at the same time it should small enough so that the pull-down network can discharge node N quickly to minimize short-circuit current.

Figure 12 also shows two possible layout scenarios for the three PMOS transistors. In one case P2 is located between P1 and P3, while in the other case P1 is in the middle. As shown in Section 3, for the two scenarios the drive current for P2 differs by about 8%. This means that the first scenario has higher drive current for keeper P2 than the expected value. As the keeper fights against the pull-down stage, there is a performance loss. HSPICE based simulations show that the time taken to discharge node N increases by about 12%. This performance loss can worsen for a more aggressively sized case. For these HSPICE simulations, we approximated the drive current increase due to stress by changing the relevant mobility numbers in the transistor models.

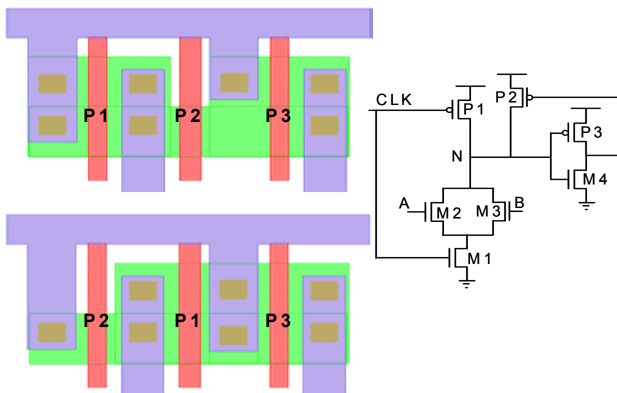


Figure 12. Basic Domino gate and two possible layouts for the PMOS.

6. Conclusions

In this paper, we proposed standard cell layout guidelines for optimizing stress-induced device performance enhancement. We studied the dependence of drive current improvement on layout parameters like source/drain length, and contact placement. We found that we could enhance the performance of any given layout by increasing the active area length. Furthermore, in most cases, we observed that the layout could be modified to enhance performance without increasing the cell area. Hence, based on our observations, we devised a set of guidelines for improving the layout. By applying these guidelines to standard cells from a 65nm industrial library we showed that there is ample scope for performance enhancement. Results show that we get an average performance enhancement of 6% and 4.4% for NMOS and PMOS drive currents, respectively, without increasing the cell area. The average increase in leakage was found to be 2.23X and 1.47X for PMOS and NMOS devices, respectively.

References

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

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

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

[4] W. H. Lee et al., “High performance 65 nm SOI technology with enhanced transistor strain and advanced-low-K BEOL,” Proc. IEDM, 2005.

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

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

[7] K. Ota et al., “Novel locally strained channel technique for high performance 55nm CMOS,” Proc. IEDM, pp. 27-30, 2002.

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

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

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

[11] S. Wolf, “Silicon Processing for the VLSI Era,” Lattice Press, 1995.

[12] A. Chandrakasan et al., “Design of High-Performance Microprocessor Circuits,” IEEE press, 2001.

[13] Manual, Davinci 3D TCAD, Version 2005.10.

[14] Manual, Synopsys TSUPREM4, Version 2007.03.

DRAFT