Strain Engineering for Performance Enhancement in Advanced Nano Scaled SOI-MOSFETs

S. Flachowsky[a)], R. Illgen[a)], T. Herrmann[a)], A. Wei[b)], J. Höntschel[b)], M. Horstmann[b)], W. Klix[a)], and R. Stenzel[a)]

[a)] Department of Electrical Engineering,
University of Applied Sciences Dresden,
Friedrich-List-Platz 1, D-01069 Dresden, Germany

[b)] AMD Saxony LLC & Co. KG,
Wilschdorfer Landstrasse 101, D-01109 Dresden, Germany

Abstract

With the current slow-down in metal oxide semiconductor field effect transistor (MOSFET) scaling due to physical limits of gate oxides, alternative techniques are needed for next generation technologies. The integration of stress techniques into a CMOS flow results in remarkable performance improvements. A comparison between the various stress-inducing methods is given on the basis of experiments and simulations. The stress transfer mechanisms from the stress source inside the device structure are studied and various integration challenges of the multiple stress techniques are addressed. Finally, the potential for performance enhancements in future nano scaled technologies is investigated and the parasitic source/drain resistance is identified as one of the major obstacles on this way.

1. Introduction

Traditional complementary metal oxide semiconductor (CMOS) scaling has offered advantages in both device performance and density, driving higher system performance at lower cost and higher yield. However, as physical limits on device scaling are approached (e.g. the intolerable high gate leakage due to thin gate oxides), technological innovations are required to deliver continuous device performance improvements over traditional scaling. Recently, the main focus for improved transistor performance has been mobility and drive current enhancement by the application of appropriate local stress [1]. Current leading-edge CMOS technologies feature multiple process-induced stressors such as compressive and tensile overlayers, embedded-SiGe, and stress memorization techniques [2]-[6]. Other strain techniques like biaxial tensile strained substrates [7], embedded silicon-carbon layers [8],

and metal stress (strained gate and contact fill materials, [9]) are considered as potential performance boosters for next technology generations.

Strain fields are inherently localized and non-uniform. The continuous progress in CMOS scaling has required the control of these effects to achieve transistors with desired performance. Process and device simulations are indispensable to understand the interactions between the structure's topography and stress transfer into the transistor channel and help to optimize the n- and p-MOSFET simultaneously. Furthermore, there are concerns about possible stress relaxation during device processing and increasing impact of parasitic device components (resistances and capacitances) which affect the strain-induced mobility as well as device performance enhancements.

2. Strained Silicon Physics

The various strain effects of stress and strain on silicon have been studied since the 1950s. It was recognized early that when the band structure of a material is changed, many material properties are altered including band gap, effective mass, carrier scattering, mobility, diffusion of dopants, and oxidation rates [10]. The strain effects on mobility were found to be anisotropic and carrier effects are different for electrons and holes.

Fundamentally, straining the silicon lattice due to mechanical strain changes the interatomic spacing. Associated modifications in the electronic band structure and density of states contribute to changes in carrier mobility ($\mu = e\tau/m^*$) through modulated effective transport masses (m^*) and modified scattering rates ($1/\tau$). For electrons, the main reason for electron mobility enhancement under strain is the sub-band energy splitting of the conduction band whereas the lifting of the valence band degeneracy between heavy and light hole bands and a band deformation under shear stress are the dominant effects for the higher hole mobility [10]. In general, strain changes electron and hole mobility in opposite directions. Selective processes are often required to address the contradictory requirements for building strained n- and p-MOSFETs.

3. Experimental and Simulation Details

The devices in this work were fabricated using a conventional CMOS process [1] with 40 nm physical gate length featuring 1.35 nm SiON gate oxide and polysilicon as gate material. After oxide and nitride spacer formation, halo and source/drain ion implantations were done. Final activation was performed using a spike rapid thermal annealing. Silicide was formed to reduce contact resistance and the device fabrication finishes with standard back-end process.

Using SYNOPSYS Sentaurus TCAD software, 2-D finite element simulations are performed to analyze the stress distribution in various structures. Silicon was treated as an anisotropic elastic material. The simulator uses the plane-strain condition, in which the out-of-plane elongation is set to zero, which is a valid assumption for transistors with a large width. For that reason, mostly only the horizontal (σ_{xx}) and vertical (σ_{yy}) stress components are extracted, whereas the influence of the small transverse (σ_{zz}) stress component plays a minor role in mobility enhancement.

Using measured secondary ion mass spectrometry (SIMS) profiles, transmission electron microscopy (TEM) cross sections and measured transfer and output characteristics, simulation devices were constructed and calibrated in a full CMOS process flow.

4. Strained Silicon in MOSFETs

4.1. Stressor Overview

Fig. 1 gives an overview of a variety of process-induced stress techniques applicable for n- and p-MOSFETs, fabricated on silicon-on-insulator (SOI) substrate integrated simultaneously in a CMOS flow.

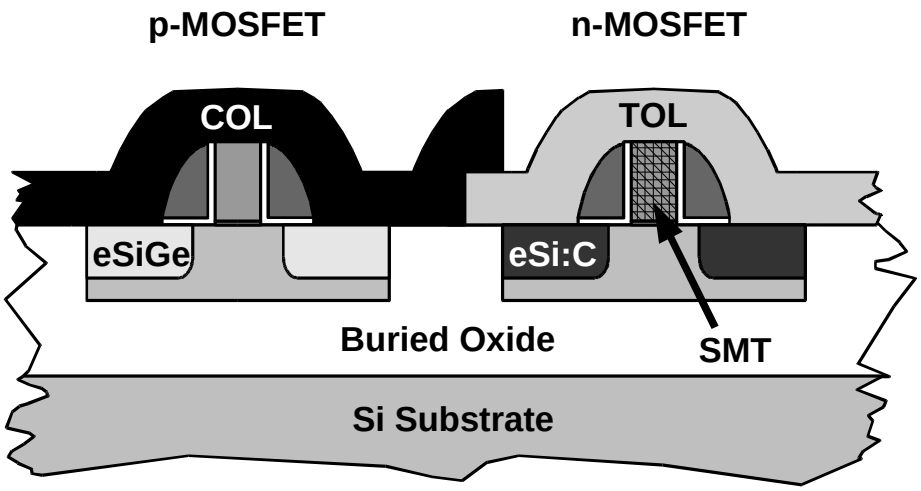


Fig. 1: Overview of the integration of various local stress techniques for n- and p-MOSFET in a CMOS process on silicon-on-insulator (SOI) substrate.

One of the most common stressor integration scheme is the stressed overlayer film. A tensile overlayer (TOL) film improves n-MOSFET drive current by enhanced electron mobility whereas a compressive overlayer (COL) film is necessary to enhance hole mobility in

p-MOSFET. Embedded source/drain stressors with different lattice constants like silicon-carbon (Si:C) or silicon-germanium (SiGe) are applied in n- and p-MOSFET, respectively. Finally, the stress memorization technique provides another strain technique, which is however only applicable for n-MOSFETs due to strain-induced enhancements in electron mobility with a simultaneous degradation in hole mobility.

4.2 Overlayer Stressor

One of the simplest methods to introduce stress in the transistor channel is to put a strained silicon nitride (Si_3N_4) layer on top of the device. Depending on the process conditions during film deposition, the intrinsic film stress can be tailored from compressive over neutral to tensile. The challenge lies in the different stress types required for the two device types. In that case masks and etch steps have to be used for the removal of tensile overlayer films (TOL) from p-MOSFET and compressive overlayer films (COL) from n-MOSFET.

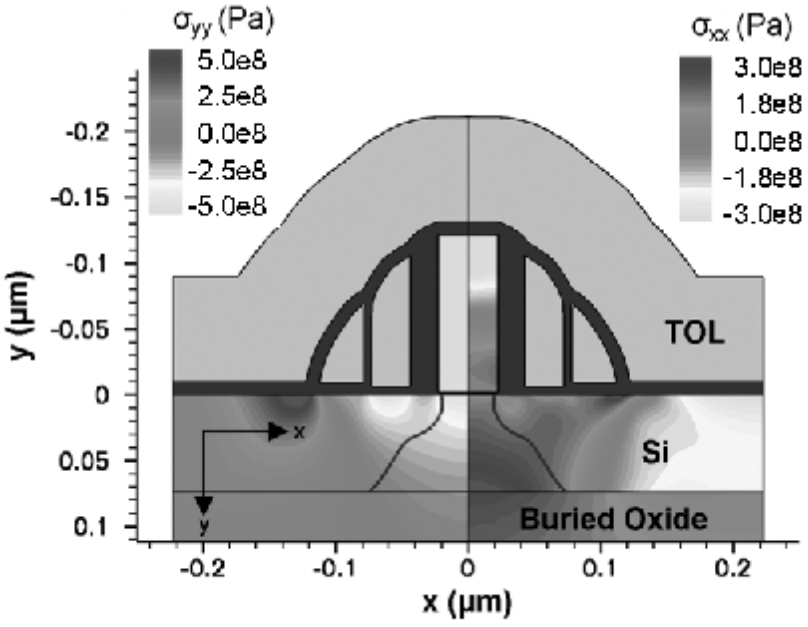


Fig. 2: Simulated vertical (σ_{yy}, left) and horizontal (σ_{xx}, right) stress distribution inside an n-MOSFET with tensile overlayer film (TOL, 1.2 GPa intrinsic stress, 80 nm thickness). Positive values indicate tensile stress, negative values are compressive.

A TOL with intrinsic tensile stress tends to shrink and since the TOL is bound to the source/drain, gate, and spacers, the TOL shrinkage is counteracted by these regions. As a consequence, the source/drain, spacer, and gate regions are stressed, which is finally

translated into a transistor channel stress. The topography of the transistor is essential to explain the resulting stresses and can be analyzed with the help of mechanical stress simulations. A TOL (intrinsic film stress of 1.2 GPa with 80 nm film thickness) generates a tensile horizontal and compressive vertical stress component in the channel region (Fig. 2), both of them are responsible for the increase in electron mobility compared to an unstrained reference device.

The improved device performance of an n-MOSFET with TOL compared to an unstrained reference is shown in Fig. 3. For a given off-state-current (I_{off}, at V_{GS} = 0 V, V_{DS} = 1.0 V) a 12 % higher on-state-current (I_{on}, at V_{GS} = V_{DS} = 1.0 V) can be achieved for the strained transistor. Contrary, an n-MOSFET with COL (-3.5 GPa, 80 nm thick) degrades the device performance by 25 %. For p-MOSFET the inverse behavior is observed, e.g. an improvement of 35 % with COL and a degradation with TOL of about 10 %. This demonstrates the strong effort which has to be made to selectively apply various stress patterns on the different device types.

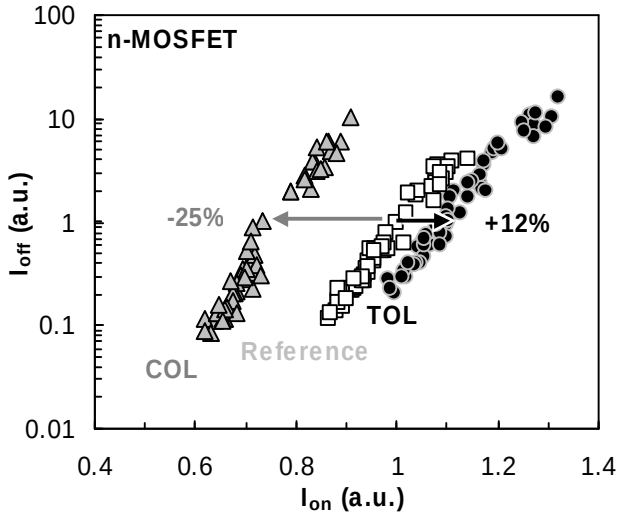
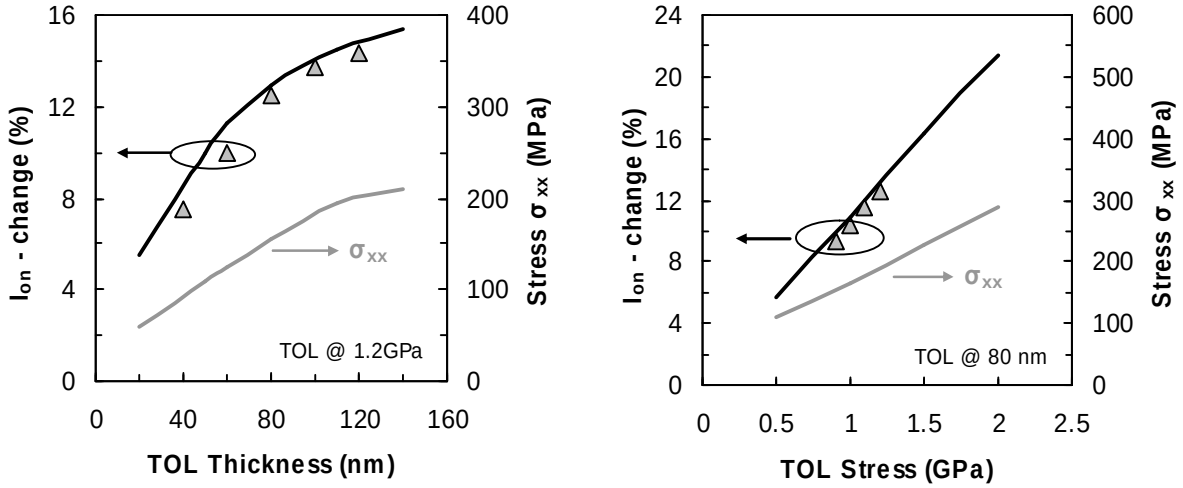


Fig. 3: Performance enhancement for TOL on n-MOSFET compared to an unstrained reference. Degraded performance results if a COL is applied to the n-MOSFET.

For further performance enhancements, the thickness of the stressed overlayer film can be increased which induces a higher stress into the underlying device. Another method is to increase intrinsic stress of the film itself. The effects of these two approaches are shown in Fig. 4. However, overlayer film thickness is limited due to process issues. The highly non-planar topography aggravates the control of subsequent manufacturing process steps like

lithography, material fill behavior, planarization, and contact etching though the overlayer film. Overlayer intrinsic stress is limited by deposition physics.



a) b)

Fig. 4: Experimental n-MOSFET drive current enhancement (symbols) vs. a) TOL thickness and b) vs. intrinsic TOL stress. The simulated horizontal channel stress and corresponding drive current enhancements (lines) are also shown.

The typical seen enhancements for p-MOSFET with COL (around +35 %) are significantly higher compared to improvements achieved for n-MOSFET with TOL (around +12 %). This is related to the almost three times higher intrinsic overlayer film stress possible with COL (-3.5 GPa) compared to TOL (1.2 GPa).

Further technology scaling reduces the available space between adjacent transistors reducing in turn the effectivity of the overlayer stressor. However, a reduced gate length allows the stressor to be more effective since a smaller volume (channel region) needs to be stressed by the same overlayer film. This is shown in Fig. 5, where the simulated stress in the channel continuously increases with reducing gate length L_G. Accordingly, the drive current is enhanced at least for gate lengths < 100 nm. For even smaller L_G the parasitic source/drain resistance $R_{S/D}$ limits the achievable drive current enhancement due to the voltage drop across this external resistance. The increased channel stress with reduced gate lengths on the one hand and the larger degradation from external resistance as the gate-length decreases on the other hand causes a “roll-off” of the I_{on}-gain vs. L_G curve [11].

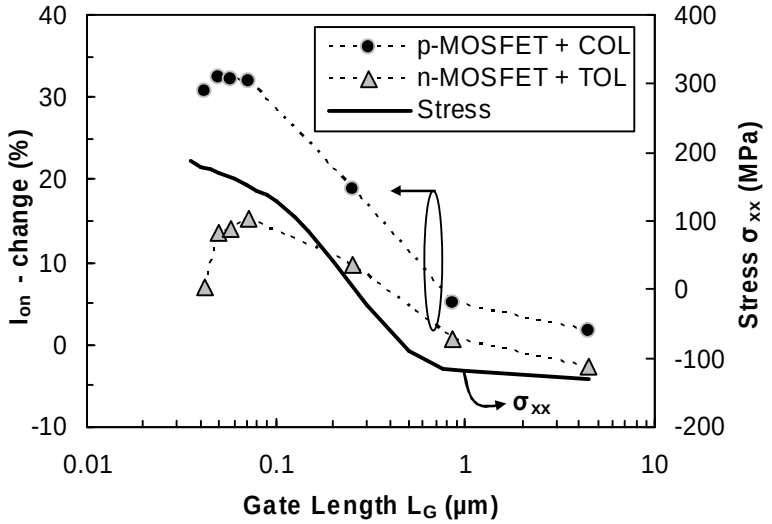


Fig. 5: Experimental drive current enhancement vs. gate lengths for n-MOSFET with TOL and p-MOSFET with COL (symbols and lines). Simulated channel stress (line only) increases continuously for reduced L_G.

4.3 Embedded Source/Drain Stressors

Epitaxial pseudomorphic growth of a material with a slightly different lattice constant than the underlying silicon substrate results in strain inside the epitaxial layer due to the lattice mismatch (Fig. 6). This is utilized for strain generation in MOSFETs by etching the source/drain regions of the transistor and refilling them with new material with a larger (smaller) lattice constant than silicon to introduce compressive (tensile) strain to the intermediate channel.

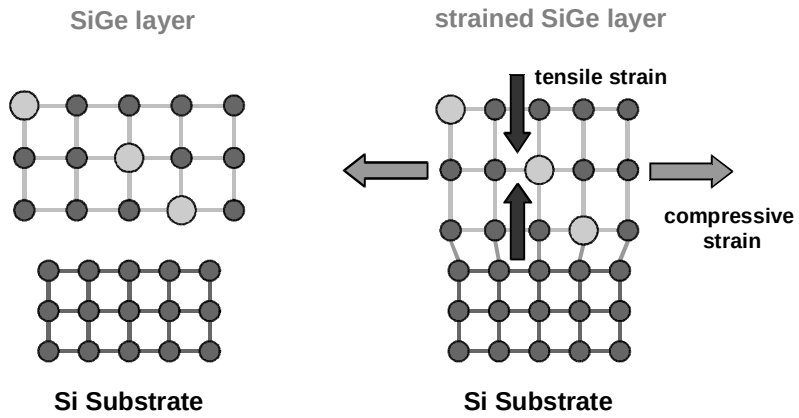


Fig. 6: Strain generation within a pseudomorphic SiGe layer grown onto a Si substrate.

Group IV alloys like SiGe for p-MOSFET and Si:C for n-MOSFET are suitable refilling materials. Embedded SiGe (eSiGe) provides strong horizontal compressive stress in the channel region (Fig. 7), which is beneficial for hole mobility in p-MOSFET. Typically, the germanium concentration is around 20-30 %. Furthermore, the smaller contact resistance of SiGe compared to Si reduces parasitic series resistance, which results in a total device performance improvement of ~30 % compared to an unstrained reference.

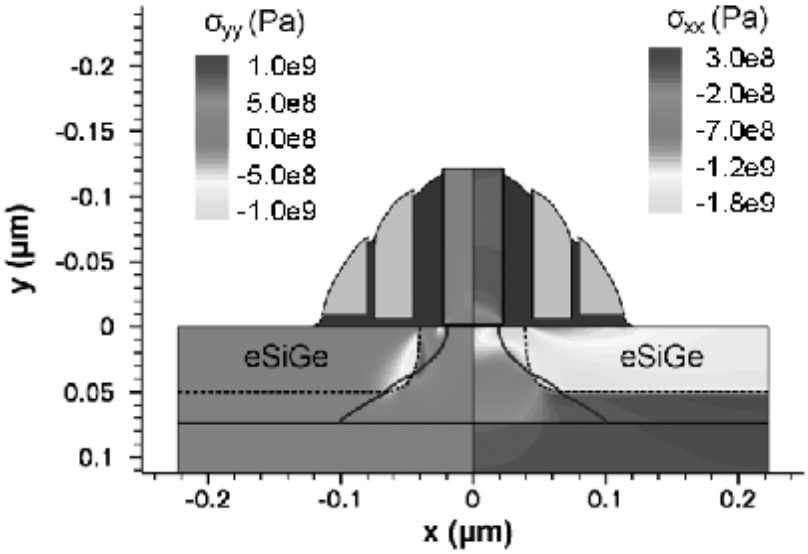


Fig. 7: Simulated vertical (σ_{yy}, left) and horizontal (σ_{xx}, right) stress distribution in a p-MOSFET with eSiGe (20 % Ge).

The induced stress distribution depends on various material and geometrical parameters, e.g. gate length, SiGe proximity to the channel, SiGe depth, and alloy composition. It has been shown that the proximity of the eSiGe to the channel is the most promising parameter to vary. Simulations confirm that a higher improvement for closer proximity is due to significantly higher stresses (in absolute value) in the channel region (Fig. 8). Although higher stresses can be incorporated by increasing the germanium content in the SiGe alloy, this approach has limitations due to process compatibility. The higher strained film is more prone to stress relaxation by forming dislocations which are detrimental for device performance and must be avoided. Careful process control is necessary to achieve maximal germanium content within the SiGe layer and thus maximal stress which can still withstand the subsequent process steps like thermal treatments, amorphizing implantations, and cleaning procedures.

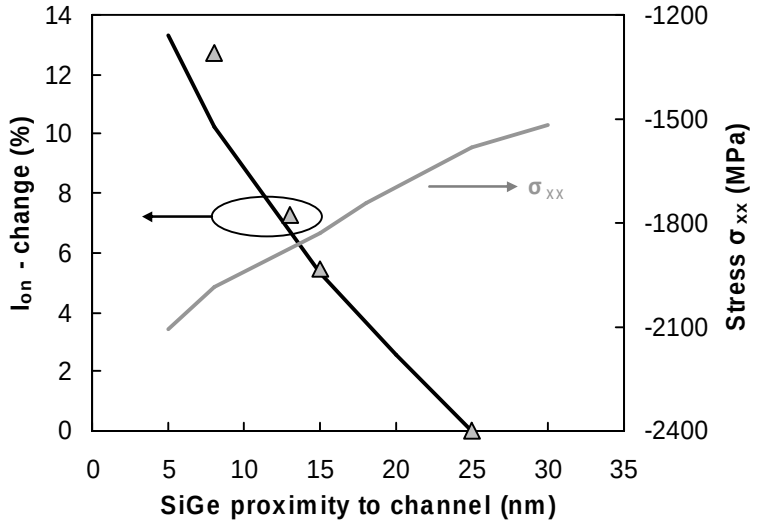


Fig. 8: Experimental (symbols) and simulated (line) drive current enhancements vs. SiGe proximity to the transistor channel for p-MOSFET with eSiGe. Simulated channel stress increases continuously in magnitude for closer SiGe to channel proximities.

Although eSiGe is meanwhile matured and widely used in volume manufacturing, its counterpart, eSi:C for n-MOSFET, is still a challenging task. Since carbon is a much smaller atom than silicon, it requires only a relatively small amount of carbon (1-2 %) to create a useful level of strain. However, the challenge lies in the very low solid solubility in silicon (<0.7 ppm) which requires metastable processes to create highly strained Si:C films.

Besides an epitaxial approach similar to SiGe, another method was proposed, which relies on carbon ion implantation into the pre-amorphized source/drain regions with a subsequent anneal to recrystallize the amorphous Si:C film [12]. To bring the carbon on substitutionell lattice sites and to keep it there throughout the device fabrication process is the crucial point, otherwise carbon precipitates into carbide cluster which relax strain and deteriorate the electrical properties of the film. So far only minor device performance improvements (~6 %, [8]) have been achieved and further work is necessary.

With further scaling the distance between two adjacent gates is more and more reduced leaving less space for the embedded source/drain stressor. The smaller stressor volume generates less stress, however, as the gate length is also scaled, the stress in the channel is steadily increasing for nano scaled devices as already pointed out previously for the overlayer stressors. But a stress increase does not necessarily transfer into mobility and drive current enhancements because, again, the parasitic resistance is limiting the stress-induced enhancements for gate lengths smaller <50 nm (Fig. 9). One important difference between p-MOSFET with eSiGe compared to n-MOSFET with eSi:C is the reduction in

parasitic $R_{S/D}$ due to the reduced contact resistance in SiGe simultaneously to the induced channel stress. This provides higher p-MOSFET drive current gains for similar stresses in ultra-short channels due to the delayed impact of the detrimental parasitic $R_{S/D}$.

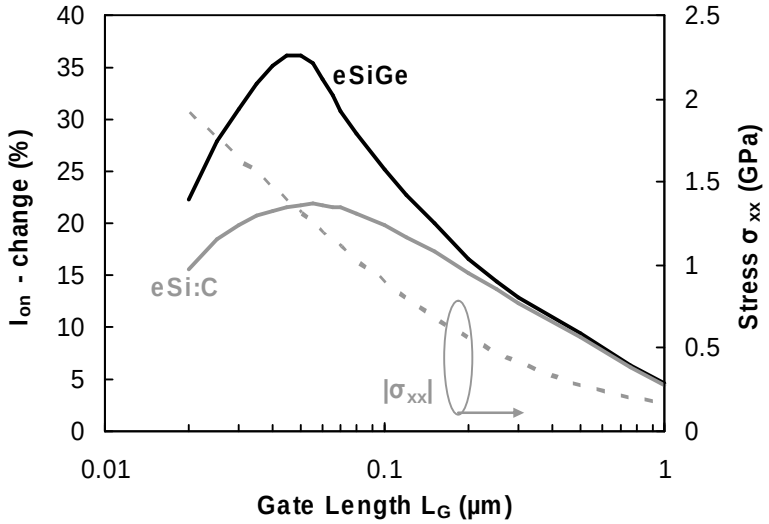


Fig. 9: Simulated drive current enhancement vs. gate length for n-MOSFET with eSi:C and p-MOSFET with eSiGe at comparable channel stress (dotted line) for both embedded stressors.

4.4 Stress Memorization

Stress memorization techniques (SMT) by means of stressed spacer material or temporary capping layers are based on a “freezing” of stress during recrystallization of amorphous source/drain areas and poly-Si-gates. The reasons for this effect are still under discussion. Some results indicate that SMT is related to a “freezing” of amorphization-related expansion of polysilicon [5]. Other publications try to explain SMT by “freezing” of temperature-related changes in the properties of the source/drain regions [6]. Maybe SMT is a combination of both, however, always multiple materials with differences in the thermal expansion coefficient and mechanical properties are involved which interact during temperature annealing to generate and memorize the strain. Interestingly, the SMT effect results in channel strain which is beneficial for electron transport in n-MOSFET (e.g. horizontal tensile and/or vertical compressive strain) which brings up to ~20 % I_{on}-I_{off} improvement depending on the applied scheme. On the other hand, p-MOSFET will be degraded by SMT and has to be excluded from SMT processing. One concern with SMT is an enhanced dopant diffusion probably caused by strain-dependent diffusion. It is experimentally observed, that the Arsenic dopants

diffuse more during standard spike RTA if SMT is applied, which requires a readjustment of the dopant profiles to compensate for the otherwise degraded short-channel behavior.

4.5 Strained Substrates

All these aforementioned local strain techniques are strongly dependent on the transistor layout which needs to be optimized to gain the maximum performance benefit. Global strained substrates like biaxial tensile strained sSOI-wafers [7] introduce high and uniform strain levels directly into the transistor channel and this strain is independent on layout variation and thus scaling. However, an individual stress tailoring to CMOS devices is difficult and mainly n-MOSFET benefits from the biaxial tensile strain whereas p-MOSFET delivers only some enhancements at very high strain levels. Electron mobility enhancements up to 120 % can be achieved with n-MOSFETs build on sSOI wafers. This results mainly from increased valley splitting and, thus, reduced carrier scattering, although recent studies suggest a reduced surface scattering in biaxial strained inversion layers to be partly responsible for the high mobility improvement [13]. However, with further gate length scaling a strong reduction in drive current enhancement from +80 % at L_G = 1 µm down to ~10 % at L_G = 40 nm is observed (Fig. 10). Both, experimental and simulation results show that the reason for this behavior is the drastically increased parasitic source/drain resistance for sSOI devices (about +50 % to ~300 Ω·µm), degrading the strain-induced drive current enhancement in short channel devices. Furthermore a strong threshold voltage reduction (~ -100 mV) occurs due to band modulations in biaxial tensile strained silicon which requires higher channel doping to compensate for the threshold voltage shift. However, this causes increased impurity scattering which degrades carrier mobility and counteracts the effect of strain. These two facts, increased parasitic source/drain resistance and threshold voltage reduction, make this stress technique beside other integration issues (process complexity, higher costs) challenging and complicate an implementation into volume manufacturing.

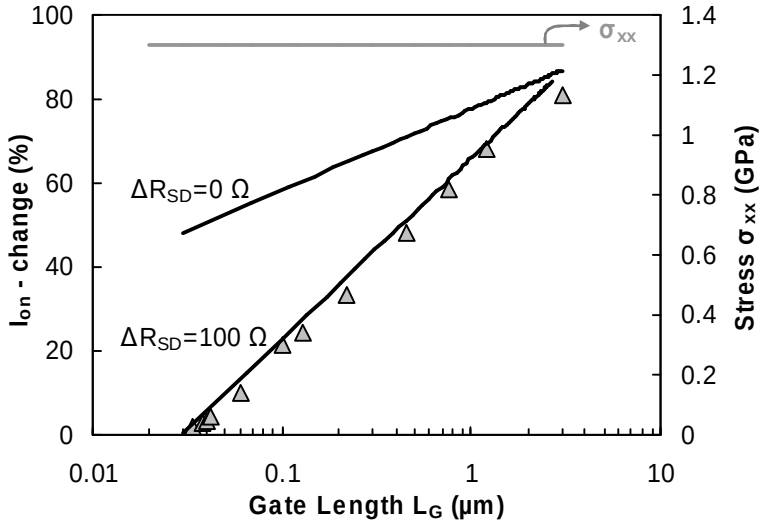


Fig. 10: Experimental n-MOSFET drive current enhancement vs. gate length L_G on sSOI substrate compared to unstrained SOI devices (symbols). Also shown are the simulated channel stress, which stays constant, and the corresponding drive current enhancements (lines) for two parasitic source/drain resistance values.

4.6 Stressor Comparison

One open question with these stress techniques is their additivity to achieve even higher strain levels and performance improvements. Since the various approaches generate different pattern of horizontal (σ_{xx}), vertical (σ_{yy}) and transversal stress (σ_{zz}), see Fig. 11, differences in the carrier mobility enhancement and electric field dependence occur.

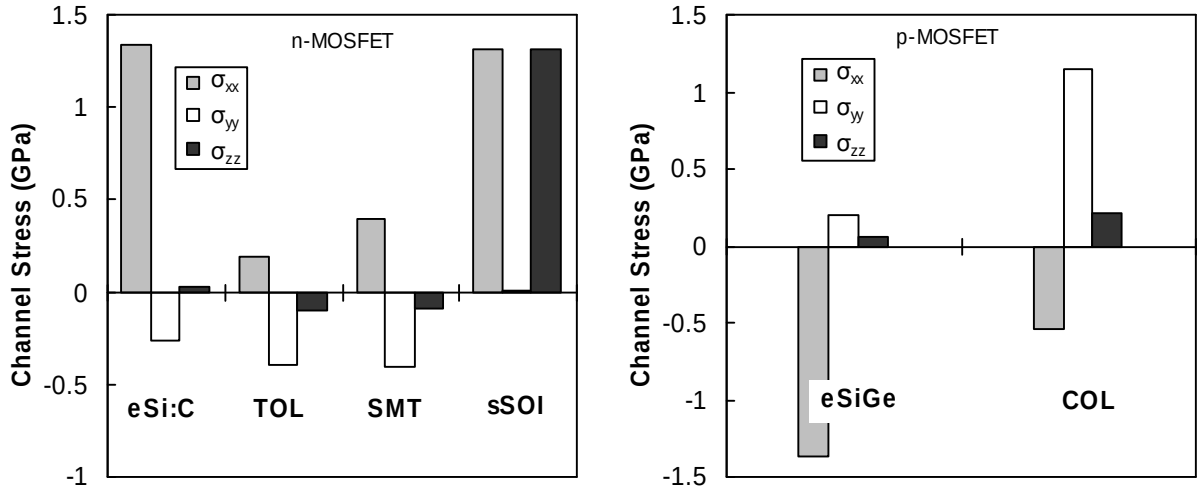


Fig. 11: Comparison of simulated horizontal σ_{xx} (grey bars), vertical σ_{yy} (white), and transversal σ_{zz} (black) stress components for various stress techniques.

Nevertheless, the generated stresses add up to each other when combined and the corresponding drive current enhancements follow more or less the same trend. For n-MOSFETs the sum in drive current enhancements is slightly lower than the combination of each stress technique would suggest. In contrast, in the case of p-MOSFET the drive current shows a super linear increase with increasing stress (Table 1).
Higher stresses from the more effective p-MOSFET stressors (COL and eSiGe) compared to n-MOSFET and a higher mobility-strain correlation for holes compared to electrons are the reasons that the p-MOSFET drive currents are now approaches the current level of n-MOSFETs, removing the historical 1:2 ratio of p- and n-MOSFET performance difference.

Table 1 Summary of the stress-induced drive current enhancements for various stress techniques in 40 nm n- and p-MOSFET. Note, that the number for eSi:C is discussible, as this stress technique is not matured yet. Also, sSOI is not applied in manufacturing so far and is not compatible with SMT. Therefore, the total number for n-MOSFET comes from a combination of TOL and SMT only.

I_{on}-gain	**n-MOSFET**	**p-MOSFET**
eSi:C	(6 %)	-
TOL	12 %	-
SMT	10 %	-
sSOI	(10 %)	-
COL	-	35 %
eSiGe	-	25 %
Total Theory	**22 %**	**60 %**
Total Experiment	**20 %**	**72 %**

5. Summary

Strain techniques have been proven as indispensable features in state-of-the-art CMOS technologies and will continuously be used in future generations. Various methods to induce stress into the transistor channel are presented with a discussion of their individual integration issues. It is shown that a combination of strain techniques is beneficial to increase the performance enhancement. However, with further scaling another concern arises which is the increasing impact of parasitic resistances and capacitances for short channel devices which hinder the strain-induced enhancements to be transferred into device and circuit improvements. Apart from higher strain through more effective or even new stressors, an investigation of novel silicides, high-mobility channel materials, and alternative device architectures are required to alleviate this problem.

Acknowledgements

This project was funded by the German Federal Ministry of Education and Research, registered under funding number 01M3167B. The author named in the publication bears responsibility for all published contents.

References

[1] Horstmann, M.: Integration and Optimization of Embedded-SiGe, Compressive and Tensile Stressed Liner Films, and Stress Memorization in Advanced SOI CMOS Technologies. IEDM Tech. Dig. 2005, pp. 233-236

[2] Yang, H. S.: Dual Stress Liner for High Performance sub-45nm Gate Length SOI CMOS Manufacturing. IEDM Tech. Dig. 2004, pp. 1075-1077

[3] Wiatr, M.: Review on Process-induced Strain Techniques for Advanced Logic Technologies. 15th Conference on Rapid Thermal Processing 2007, pp. 19-29

[4] Wei, A.: Combining embedded and overlayed compressive stressors in advanced SOI CMOS technologies. Int. Conf. on Solid State Devices and Materials 2005, pp. 32-33

[5] Ortolland, C.: Multiple Stress Memorization Technique (SMT) Optimization for 45nm CMOS. Proc. VLSI Symposium 2006, pp. 78-79

[6] Wei, A.: Multiple Stress Memorization In Advanced SOI CMOS Technologies. Proc. VLSI Symposium 2007, pp. 216-217

[7] Rim, K.: Fabrication and Mobility Characteristics of Ultra-thin Strained Si Directly on Insulator (SSDOI) MOSFETs. IEDM Tech. Dig. 2003, pp. 49-52

[8] Ren, Z.: On Implementation of Embedded Phosphorus-doped SiC Stressors in SOI nMOSFETs. Proc. VLSI Symposium 2008, pp. 172-173

[9] Auth, C.: 45nm High-k + Metal Gate Strain-Enhanced Transistors. Proc. VLSI Symposium 2008, pp. 128-129

[10] Chidambaram, P.R.: Fundamentals of Silicon Material Properties for Successful Exploitation of Strain Engineering in Modern CMOS Manufacturing. IEEE Transactions on Electron Devices, Vol. 53, No. 5, 2006, pp. 944-964

[11] Flachowsky, S.: Gate length scaling trends of drive current enhancement in CMOSFETs with dual stress overlayers and embedded-SiGe. Material Sciences Engineering B, Vol. 154-155, 2008, pp. 98-101

[12] Liu, Y.: Strained Si Channel MOSFETs with Embedded Silicon Carbon Formed by Solid Phase Epitaxy. Proc. VLSI Symposium 2007, pp. 44-45

[13] Hadjisavvas, G.: The Origin of Electron Mobility Enhancement in Strained MOSFETs. IEEE Electron Device Letters, Vol. 28, No. 11, 2007, pp. 1018-1020