Stress Memorization Technique for n-MOSFETs: Where is the Stress Memorized?

S. Flachowsky[1], R. Illgen[1], T. Herrmann[1], T. Baldauf[2], A. Wei[1], J. Höntschel[1], W. Klix[2], R. Stenzel[2], M. Horstmann[1]

[1] GLOBALFOUNDRIES Dresden Module One LLC & Co. KG, Wilschdorfer Landstraße 101, D-01109 Dresden, Germany
[2] Department of Electrical Engineering, University of Applied Sciences Dresden, Friedrich-List-Platz 1, D-01069 Dresden, Germany

Email: stefan.flachowsky@globalfoundries.com

Stress memorization techniques are widely adopted in high performance technologies to improve n-MOSFET performance. However, the exact mechanism of where the stress is memorized in the transistor has not been described yet. This study employs experimental work combined with process and device simulations to identify the area, which is responsible for the stress memorization effect. Whereas for the low temperature SMT scheme only the source/drain regions of the transistor contribute to the memory effect, in the high-temperature SMT scheme the polysilicon-gate induces an additional channel stress resulting in significant higher drain current enhancements.

1. Introduction

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].

Stress memorization techniques (SMT) by means of temporary capping layers are based on “freezing” of stress in the transistor during annealing and subsequent cooling resulting in enhanced electron mobility and n-MOSFET performance. On the other hand, p-MOSFET will be degraded by SMT and has to be excluded from SMT processing.

The exact mechanisms for the memory effect are still under discussion. Beside the question of how the stress is memorized [7], the other important issue is where the stress is memorized. All earlier work claim that the memorized stress is localized in the polysilicon-gate [8], [9]. Only recently, few publications suggest the source/drain regions as another possibility for stress memorization [6], [10].

This study employs experimental work combined with finite element simulations to identify the transistor region, which is responsible for the stress memorization effect.

2. Device Fabrication

In literature there are various approaches at which points during the process flow the stress is introduced and memorized. In general, there exist two distinct schemes [6]:

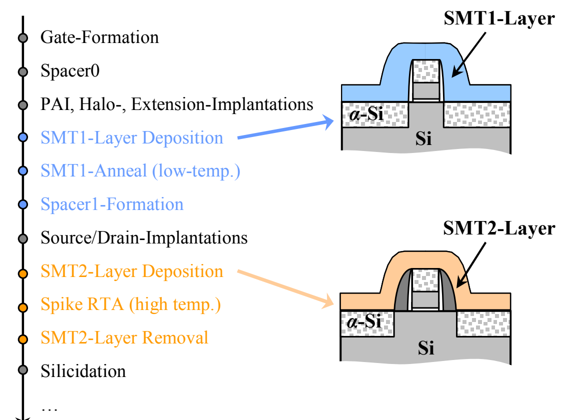

Fig. 1 Implementation of the two SMT schemes into the process flow.

- Scheme 1 (SMT1): The pre-amorphization implantation, the deposition of the SMT1-layer using a dielectric material like oxide or nitride as well as the thermal anneal at low temperatures (600°C–800°C for a few minutes) for the stress memorization process are performed early during the process flow, i.e. after the extension and halo implantations (Fig. 1). The SMT1-layer is stripped afterwards or further processed into spacers adjusting the subsequent source/drain-implantations.
- Scheme 2 (SMT2): The temporary SMT2-layer is deposited late in the process flow (after the source/drain-implantations, Fig. 1) and employs the standard spike RTA (1000°C–1100°C for a few seconds) for stress memorization followed by the SMT2-layer removal.

N-MOSFETs with gate lengths down to 40 nm were fabricated using SiON as gate dielectric and polysilicon as gate material [1] in a leading-edge 45nm SOI-technology. No other intentional stress sources like strained overlayer films or embedded silicon-carbon source/drain regions were implemented into the devices to avoid interactions of the different stressors. The SMT-layers consist of a 30nm thick nitride film with various intrinsic stress levels. The transistor channel is aligned in a <110> direction on a (001) wafer.

3. Process & Device Modeling

Using SYNOPSYS Sentaurus TCAD software [11], 2-D finite element simulations are performed to analyze the stress distribution in the devices. Silicon was treated as an anisotropic elastic material.

The energy-balance and quantum-mechanical models were used in the device simulations. The effects of ionized impurity scattering were considered via a concentration dependent mobility model. A model which accounts for effects of the lateral field dependence on mobility was employed and a perpendicular field dependence model was used to model the effects of the Si/SiO_2 interface. The strain-dependent electron mobility was calculated based on the valley-occupancy model [11] including the effective mass variation under shear strain. The strain-induced band gap narrowing was taken into account based on the deformation potential theory [11].

The process and device simulation models were calibrated using measured SIMS profiles, TEM cross sections and transfer and output characteristics of a state-of-the-art SOI transistor [1].

4. Results and Discussion

4.1. Experimental Results

Fig. 2 shows the I_{on}-I_{off}-characteristics of unstrained and by SMT1 / SMT2 strained n-MOSFETs. Performance improvements of 4% and 11% can be achieved compared to the unstrained reference for SMT1 and SMT2, respectively. Other electrical parameters, like overlap capacitance, DIBL or gate oxide thickness in inversion are matched for these devices. In devices with SMT2 the strain-dependent enhanced Arsenic diffusion must be compensated through an increased offset spacer (Spacer0). A shift in the threshold voltage is observed for both SMT schemes (−22mV and −47mV) due to strain-induced electronic band modulations.

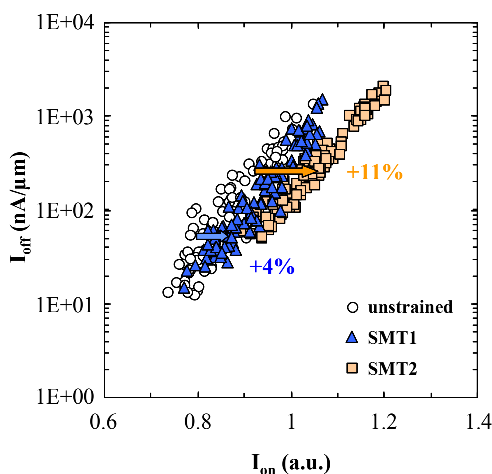

Fig. 2 I_{on}-I_{off}-characteristic of n-MOSFETs with SMT1 / SMT2.

The combination of SMT1 and SMT2 is not additive in terms of performance enhancements (Fig. 3). This already indicates, that both SMT schemes rely, at least partly, on the same mechanism. Further, the results in Fig. 3 demonstrate, that an amorphous region in the transistor is essential for SMT to work, since a recrystallization-anneal prior the SMT-anneal disables the SMT effect.

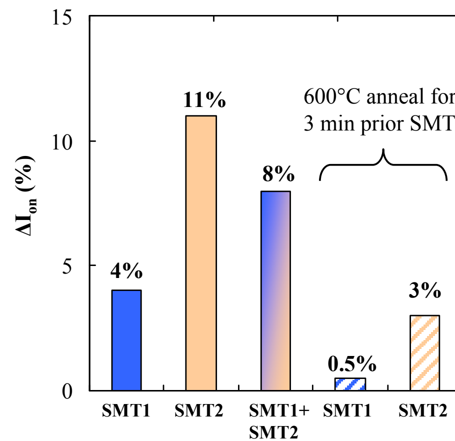

Fig. 3 Performance enhancement for a combination of SMT1 and SMT2 and with prior recrystallization-anneal.

Finally, the intrinsic stress of the SMT-nitride-layer is not correlated with the observed performance enhancement (Fig. 4) and therefore irrelevant for the discussion of the SMT effect.

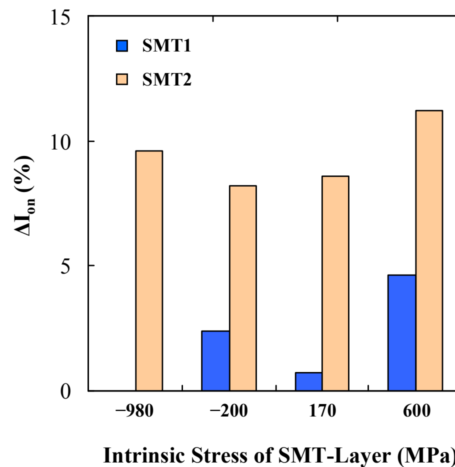

Fig. 4 Performance enhancement due to SMT1 / SMT2 as a function of the intrinsic stress level of the SMT-nitride-layer.

However, analyzing the experimental data, it is not possible to separate where the stress memorization is actually localized. Therefore finite element simulations are employed next, to clarify this question.

4.2. Simulation Results

There are three regions in the transistor which can potentially memorize the stress: the spacer, the source/drain regions and the polysilicon-gate. Each region is separately stressed with an arbitrarily stress and the impact on the electrical device performance is studied.

A) Spacer

The n-MOSFET drive current can only be enhanced if the intrinsic stress of the spacer material is assumed to be tensile (Fig. 5). However, for realistic stress values about 1–2GPa only marginal drain current enhancements around 2–3% can be achieved. The reason for the low enhancements can be found in the induced channel stress, which is not optimal for electron transport. Although the compressive vertical stress σ_{yy} is advantageous, the lateral stress σ_{xx} should be tensile for enhanced electron mobility, but is compressive. Fig. 6 shows the stress contour map for a transistor with a 2GPa tensile stressed spacer.

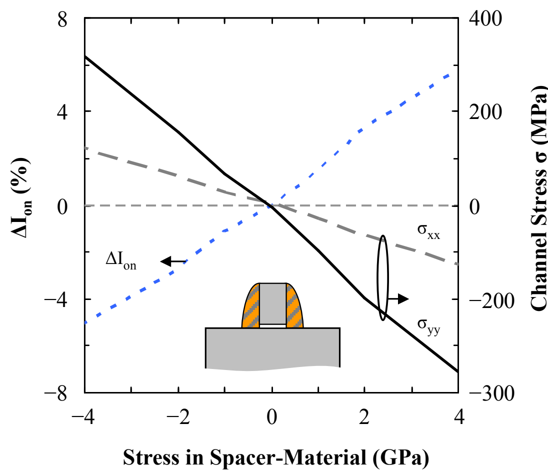

Fig. 5 Drain current change and lateral (σ_{xx}) and vertical (σ_{yy}) channel stress as a function of spacer stress.

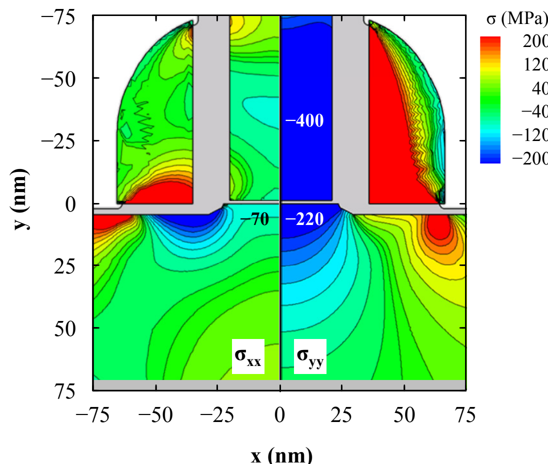

Fig. 6 Lateral σ_{xx} (left) and vertical σ_{yy} (right) stress caused by a spacer with 2GPa intrinsic stress.

B) Source/Drain Regions

Based on TEM images (inset in Fig. 7) the source/drain regions are assumed to contain the memorized stress. Tensile strained source/drain regions induce a lateral tensile and a vertical compressive channel stress (Fig. 7), which is exactly what is needed for enhanced electron mobility and drive current improvement. For 1GPa intrinsic stress in the source/drain regions the drain current increases about 4–5% (Fig. 7). The lateral stress is the dominant stress component in the channel, which is about 270MPa (Fig. 8).

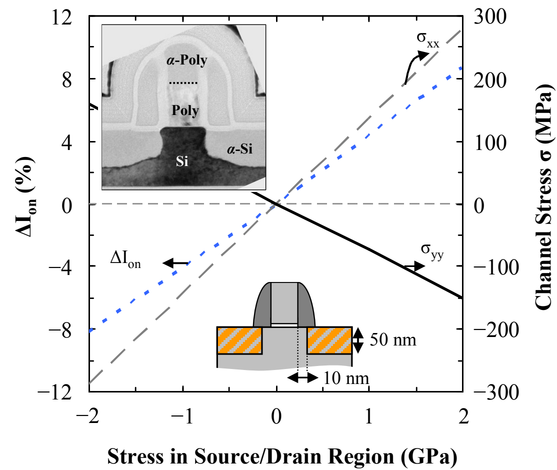

Fig. 7 Drain current change and lateral (σ_{xx}) and vertical (σ_{yy}) channel stress as a function of the source/drain region stress.

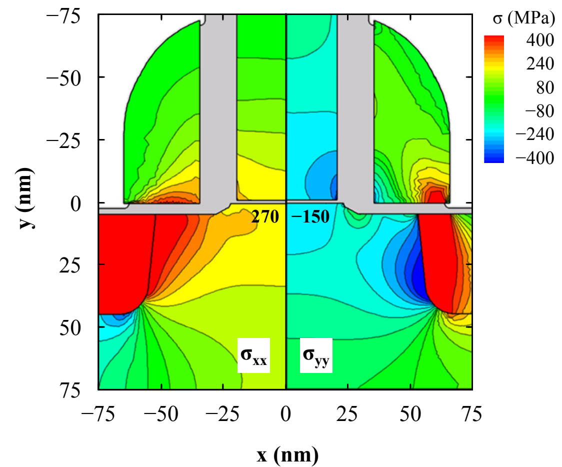

Fig. 8 Lateral σ_{xx} (left) and vertical σ_{yy} (right) stress caused by a source/drain region with 1GPa intrinsic stress.

C) Polysilicon-Gate

The intrinsic stress in the polysilicon-gate has to be compressive to induce a stress pattern in the transistor channel, which leads to enhanced drive current (Fig. 9). In order to achieve a drive current enhancement of 11% as seen in the experimental data for

SMT2, the intrinsic polysilicon-gate stress has to be chosen to −2.4GPa. This is very high, but not unrealistic. The lateral and vertical stress in the transistor for −2GPa intrinsic polysilicon-gate stress is shown in Fig. 10.

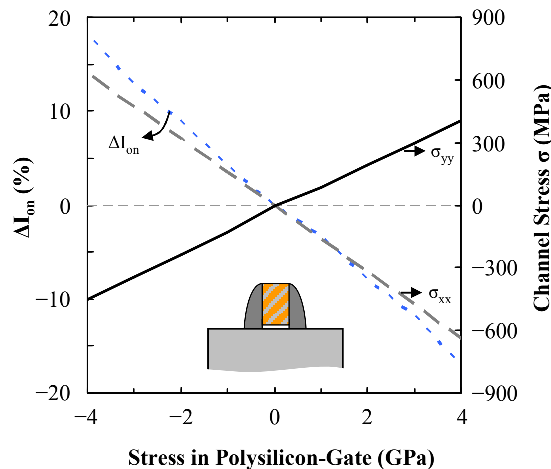

Fig. 9 Drain current change and lateral (σ_{xx}) and vertical (σ_{yy}) channel stress as a function of polysilicon-gate stress.

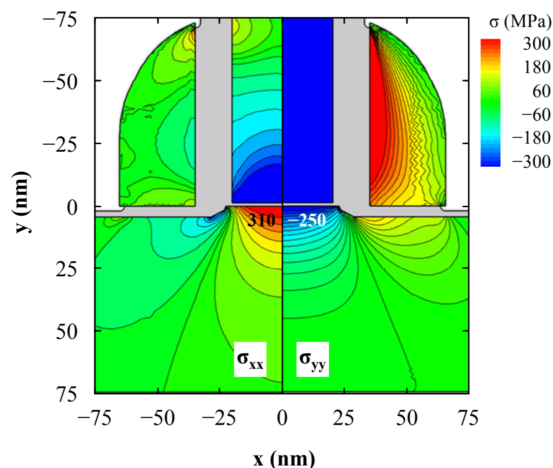

Fig. 10 Lateral σ_{xx} (left) and vertical σ_{yy} (right) stress caused by a polysilicon-gate with −2GPa intrinsic stress.

4.3. Discussion

The spacers are unlikely for stress memorization since the simulation results predict only marginal drain current enhancements due to the unfavorable stress distribution in the transistor channel. The other two options, stress memorization in the source/drain regions or in the polysilicon-gate, are more convenient to explain the SMT effect.

From the experiments it follows that for both SMT schemes an amorphous region in the transistor must exist. As can be seen from the inset in Fig. 7, the amorphous region of the polysilicon-gate is restricted to the upper part of the polysilicon-gate. Due to the large distance of this area to the channel it will not change the stress in the channel of the transistor significantly. Also, structural deformations in the non-amorphized polysilicon-gate, caused e.g. by grain boundary dynamics, do not take place at the low temperatures used in the SMT1 scheme. Therefore, only the source/drain regions remain as a reasonable location for stress memorization in SMT1.

The much higher temperatures during SMT2 processing allow plastic deformation due to polygrain-size changes [12] even in the non-amorphized region of the gate, inducing a compressive stress in the polysilicon. This additional stress from polysilicon-gate results in the much higher drain current enhancements seen with SMT2 compared to SMT1.

5. Conclusion

The annealing of a transistor capped with a nitride layer memorizes stress in different parts of the transistor even after removal of the cap-layer. This results in enhanced electron mobility and n-MOSFET performance. Whereas for SMT1 the memory effect originates from the source/drain regions, SMT2 benefits from additional stress caused by the deformed polysilicon-gate and thus higher drain current enhancements are possible.

Acknowledgments

This project was funded by the German Federal Ministry of Education and Research registered under Funding No. 13N10346.

References

[1] M. Horstmann *et al.*: IEDM 2005, pp. 233–236
[2] H.S. Yang *et al.*: IEDM 2004, pp. 1075–1077
[3] M. Wiatr *et al.*: RTP Conf. 2007, pp. 19–29
[4] A. Wei *et al.*: SSDM 2005, pp. 32–33
[5] C. Ortolland *et al.*: VLSI 2006, pp. 78–79
[6] A. Wei *et al.*: VLSI 2007, pp. 216–217
[7] L.S. Adam *et al.*: SISPAD 2005, pp. 139–142
[8] K. Ota *et al.*: IEDM 2002, pp. 27–30
[9] C. Chen *et al.*: VLSI 2004, pp. 56–57
[10] C. Ortolland *et al.*: IEEE Transactions on Electron Devices 56(8), pp. 1690–1697, 2009
[11] Sentaurus Process & Device User's Manual, Release Z-2007.03, Synopsys Inc., 2007
[12] A. Eiho *et al.*: VLSI 2007, pp. 218–219