Integration and Optimization of Embedded-SiGe, Compressive and Tensile Stressed Liner Films, and Stress Memorization in Advanced SOI CMOS Technologies

M. Horstmann, A. Wei, T. Kammler, J. Höntschel, H. Bierstedt, T. Feudel, K. Frohberg, M. Gerhardt, A. Hellmich, K. Hempel, J. Hohage, P. Javorka, J. Klais, G. Koerner, M. Lenski, A. Neu, R. Otterbach, P. Press, C. Reichel, M.Trentsch, B. Trui, H. Salz, M. Schaller, H.-J. Engelmann, O. Herzog, H. Ruelke, P. Hübler, R. Stephan, D. Greenlaw, M. Raab, and N. Kepler

AMD Saxony LLC & Co. KG, Wilschdorfer Landstrasse 101, 01109 Dresden, Germany
Phone: +49 (0) 351 277 4548 Email: manfred.horstmann@amd.com

H. Chen, D. Chidambarrao, D. Fried, J. Holt, W. Lee, H. Nii^, S. Panda, T. Sato^, A. Waite*, S. Luning*, K. Rim, D. Schepis, M. Khare, S.F. Huang, J. Pellerin* and L.T. Su

IBM Systems & Technology Group System and Technology Group, ^Toshiba America Electronic Components, Inc., and
*AMD Corporation, at IBM Semiconductor Research and Development Center (SRDC), Hopewell Junction, NY 12533

Abstract

An optimized 4-way stress integration on partially-depleted SOI (PD-SOI) CMOS is presented. An embedded-SiGe process and a compressive-stressed liner film are used to induce compressive strain in the PMOS (PMOS "stressors"). A stress memorization process and a tensile-stressed liner film are used to induce tensile strain in the NMOS (NMOS "stressors"). With optimization, the different stress techniques are highly compatible and additive to each other, improving PMOS and NMOS saturation drive current by 53% and 32%, respectively. This improvement results in 40% higher product speed.

To demonstrate the extendibility for future transistor nodes the stress improvements were increased further resulting in record PMOS performance of IDSAT=860uA/um at 200nA IOFF (self-heating corrected) and 1V. The stress techniques are proven in AMD's 90nm manufacturing processes, and have been scaled for use in 65nm manufacturing.

Introduction

Gate oxide scaling limitations have moved the focus of transistor performance improvement to strained-Si using stress techniques. Compressive liner films and embedded-SiGe (e-SiGe) are used in manufacturing to improve PMOS drive current [1,2]. Both techniques improve hole mobility in the PMOS by inducing compressive strain in the channel -- mechanically from the compressive liner, and by a lattice-level stress transfer from the e-SiGe. For electron mobility and NMOS drive current improvement, tensile liner films and stress memorization techniques are used to induce tensile strain in the channel [1-3].

Combining the four stress techniques so that the benefits from each individual approach are fully additive, while keeping the process manufacturable and low-cost is a great challenge. In this work we have carefully studied the individual effects of each stress technique on both transistors. We demonstrate that with novel process integration, the process can be kept simple and manufacturable. And, with reduction of resistance and enhancement of surface mobility, the stressor benefits are fully additive and result in greatly enhanced transistor and product speed.

Stress Integration on SOI

Table 1 summarizes the key features of AMD's transistors at the 65nm technology node. Fig. 1 and Fig. 2 show cross section TEMs of NMOS and PMOS, respectively. The 4-way stress integration begins with introduction of SiGe into the PMOS with a highly selective cavity etch and SiGe-epitaxy. As the bulk integration [2] is not possible on thin-film SOI, we developed a novel integration scheme which embeds SiGe very close to the gate. The SiGe is grown undoped prior to the transistor implants to give optimal control of the dopant profile while keeping SiGe close to the gate. A hardmask process protects the NMOS from cavity etch and SiGe-epitaxy. Later in the process, tensile strain is "memorized" into the NMOS with a stress transfer film and anneal that transfers the stress into the Si. The stress memorization sequence is fully integrated into the process without the need of additional masking and etch steps. To improve the short channel behavior aggressive junction engineering including an aggressive spike anneal process is used. Nickel silicide (NiSi) is used for contacting the device. After silicidation, the compressive liner film is deposited and removed only from the NMOS regions, followed by the deposition of a tensile film, that is then removed only from the PMOS regions. The dual stress liner films (DSL) simultaneously apply compressive stress to PMOS and tensile stress to NMOS so that ultimate CMOS performance improvement can be achieved as reported in [1].

Optimization and Resulting Transistor Improvement

Fig. 3 and Fig 4 show the improvement from the combination and optimization of all four stress techniques. The saturation drive current (IDSAT) improves by 53% for PMOS and 32% for NMOS compared to the non-stress reference while the linear drive current (IDLIN) shows an improvement of 107% for PMOS and 27% for NMOS. Measurement of on-state resistance at constant overdrive vs Lpoly show that the IDSAT and IDLIN improvements can be clearly traced down

0-7803-9269-8/05/$20.00 (c) 2005 IEEE

to higher mobility (decreased slope) and lower resistance [4]. Fig. 5 shows that the electron mobility in the NMOS is increased by the stress techniques while the resistance has been reduced through NiSi optimization. Fig. 6 shows that the hole mobility increase and resistance decrease in the PMOS are even more dramatic, due to larger available compressive stresses and the lower SiGe/NiSi barrier, respectively.

Since compressive stress enhances PMOS and degrades NMOS (and vice versa for tensile stress) the process must be optimized to minimize or eliminate the degradation of the device that does not benefit from each stress technique. For e-SiGe, the hardmask integration has been optimized to ensure that NMOS performance is not compromised for the sake of selectivity during cavity etch and SiGe growth. For stress memorization, a straightforward approach can result in a large degradation of PMOS, as shown in Fig. 7. However, this degradation can be reduced significantly by tailoring the implants and can be completely eliminated by using an optimized anneal step. For the compressive and tensile dual stress liners, which act biaxially, careful attention has been paid to layout interactions between the compressive and tensile liners. But above all, to get the maximum benefit from the combination of the stress techniques, resistance must be reduced and gate surface mobility must be improved. Otherwise resistance and poor surface mobility can significantly limit the drive current improvement for a given stress-induced mobility improvement.

Fig. 8 demonstrates that with such resistance and surface mobility improvement, fully additive stress benefits have been achieved. Table 2 summarizes the individual relative improvements for the different stress techniques, compiled from experiments in which each technique is introduced independently of the others. Fig. 8 shows that the NMOS improvements are fully additive for IDSAT, and are more than additive for IDLIN (stress techniques enhance each other). Fig. 8 also shows that PMOS improvements are more than additive for IDSAT, but are not fully additive for IDLIN. To achieve full additivity for PMOS IDLIN, further resistance reduction appears necessary.

Product Yield and Speed

At 1.0V, NMOS IDSAT=1080uA/um and PMOS IDSAT=710uA/um at IOFF=200nA/µm have been achieved for the full 4-way stress CMOS integration (Fig. 3). Correcting for self-heating the values correspond to an IDSAT of 1145uA/um for NMOS and 740uA/um for PMOS. Fig. 9 compares the yield of the different combinations of stress techniques and demonstrates that the full 4-way stress combination yields comparably to reference [1]. Fig. 10 shows that Athlon64™ product FMAX performance is improved 40% at fixed static uP current using the full 4 way stress integration.

Extendibility

The results described above have been achieved in AMD's 90nm and 65nm technologies. In order to demonstrate the extendibility for 65nm and future technology nodes we increased the performance further.

One way to increase the transistor drive current is to improve the embedded SiGe integration. Channel strain naturally increases with decreasing Lgate because the stressor is embedded. Fig. 11 shows the PMOS performance improvement for 3 different generations of SiGe integrations. The integrations are run on the same transistor setup, i.e. with the same compressive liner and stress memorization techniques as reported before. The improvement shown in Fig. 11 results in the best PMOS drive current performance reported. At 1V a PMOS drive current of IDSAT=820uA/um at IOFF=200nA has been measured. Self-heating corrected, this translates to IDSAT=860uA/um with a record low Cgate*V/Idsat performance metric of 1.25ps. In fact, this record-setting PMOS is so strong, that it has the same IDSAT as, and more IDLIN than the NMOS without stress in Fig. 3 and 4.

However, as strain-induced mobility increases, external resistance at some point limits the improvement. For example, the 3rd generation SiGe integration in Fig. 11 could have delivered a 65% IDSAT improvement and 120% IDLIN improvement without a compressive liner, relative to control without a compressive liner. With a compressive liner it brings only 50% IDSAT improvement and 80% IDLIN improvement relative to a control with a compressive liner, since resistance limits the improvement from stress-induced mobility increase. All future stress increases will have to come with novel methods to reduce external resistance. In addition such a strong PMOS drive current improvement require an optimization of the product circuitry for the resulting NMOS to PMOS relationship.

Conclusion

An advanced PD-SOI technology with transistors greatly enhanced by four integrated stress techniques has been shown. NMOS and PMOS IDSAT improvements of 32% and 53% as well as excellent product speed and yield, have been achieved. Record PMOS drive currents have been measured by further optimizing the embedded SiGe integration. The developed stress techniques are highly manufacturable for 90nm and 65nm technology nodes, and are extendible to future technology nodes.

Acknowledgement

We would like to thank P.-O. Hansson, A. Samoilov, E. Sanchez, O. Luckner, and S. Weiher-Telford from Applied Materials, Inc., Buchenstr. 12, 01097 Dresden, Germany, for the assistance regarding SiGe epitaxial growth.

References

[1] H. Yang, et al., IEDM 2004, p. 1075
[2] P. Bai, et al., IEDM 2004, p. 657.
[3] C.H. Chen, et al., VLSI 2004, p56.
[4] K. Rim , et al., IEDM 2002, p. 43.

Feature	nm
STI Pitch	220
GOX Thickness	1.3
SOI Thickness	77
BOX Thickness	145
Gate Length	40
Spacer Width	45
Gate Pitch	250

Table 1. shows AMD 65nm technology transistor features.

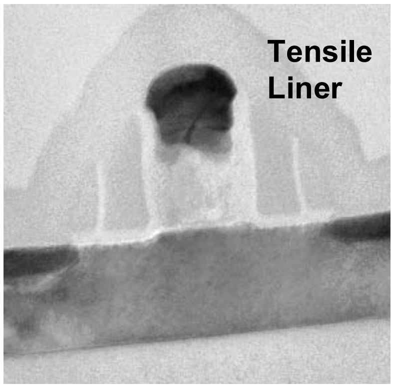


Fig. 1. NMOS with stress memory and tensile stressed liner on SOI (NMOS stressors).

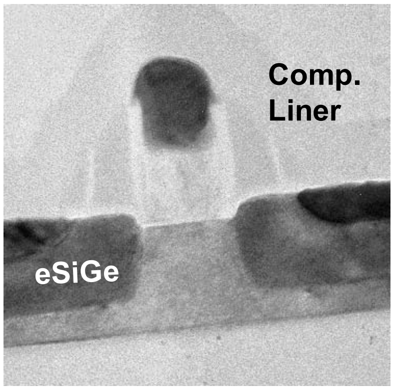


Fig. 2. PMOS with embedded-SiGe and compressive stressed liner on SOI (PMOS stressors).

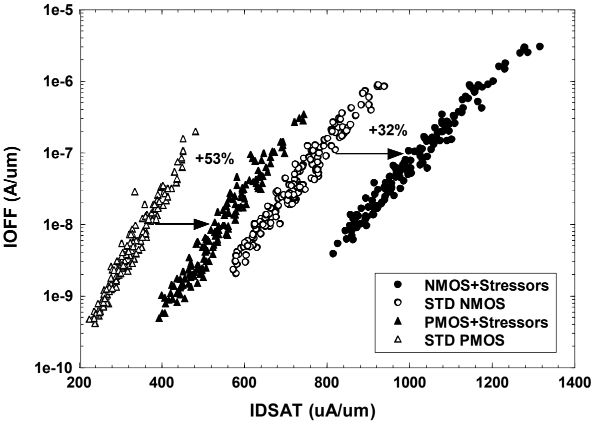


Fig. 3. IOFF vs IDSAT at VDD=1.0V for NMOS and PMOS with and w/o all stressors

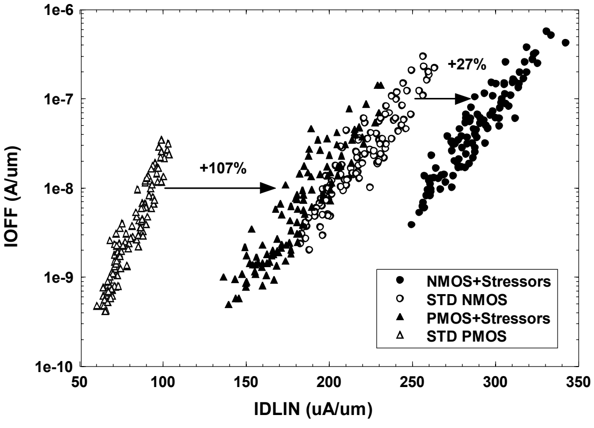


Fig. 4. IOFF vs IDLIN (VDS=0.1V) at VDD=1.0V for NMOS and PMOS with and w/o all stressors

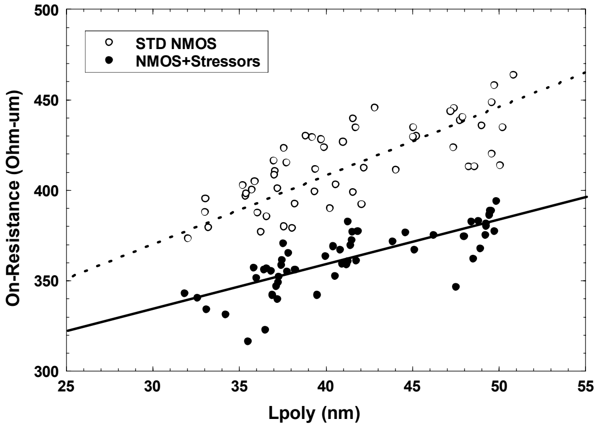


Fig. 5. NMOS on-state resistance measurement at constant overdrive vs Lpoly showing mobility and resistance improvement.

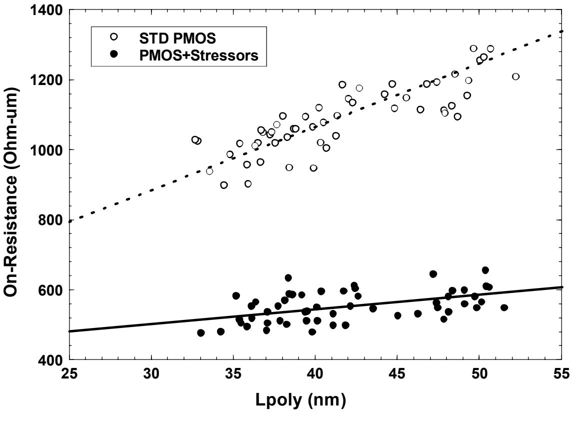


Fig. 6. PMOS on-state resistance measurement constant overdrive vs Lpoly showing mobility and resistance improvement.

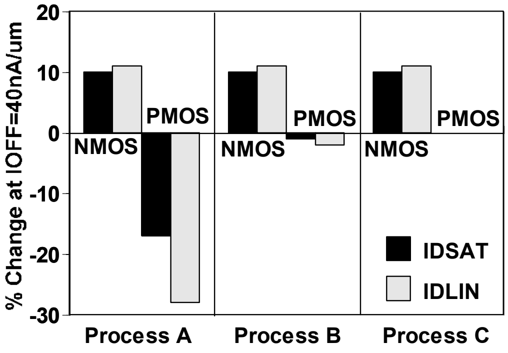
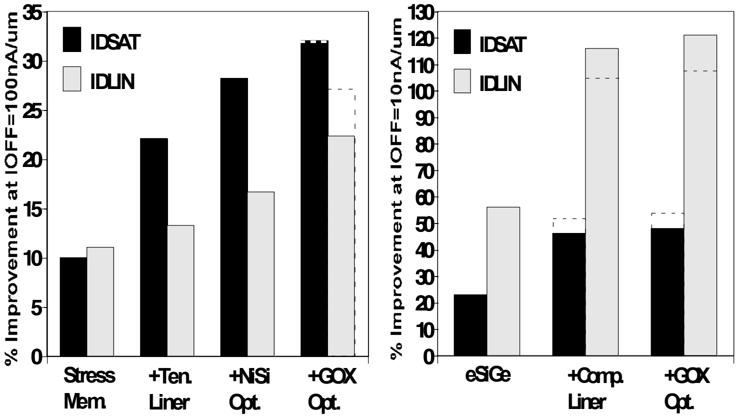


Fig. 7. Stress memory can be optimized to minimize PMOS degradation. Process A shows the direct implementation of stress memory w/o optimization Process B has optimized implants, Process C has a optimized anneal.

NMOS	Mobility	IDSAT	IDLIN
Stress Mem.	35%	10%	11%
Tens. Liner	27%	11%	2%
NiSi Opt.	-	5%	3%
GOX Opt.	8%	3%	5%
PMOS	**Mobility**	**IDSAT**	**IDLIN**
eSiGe	70%	23%	56%
Comp. Liner	90%	19%	39%
NiSi Opt.	-	-	-
GOX Opt.	4%	1%	2%

Table 2. shows the Mobility, IDSAT, and IDLIN improvements for each stress technique. Table compiled from experiments where each stress technique/improvement is done separately from the others.



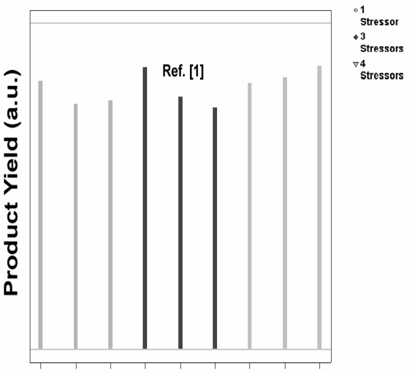
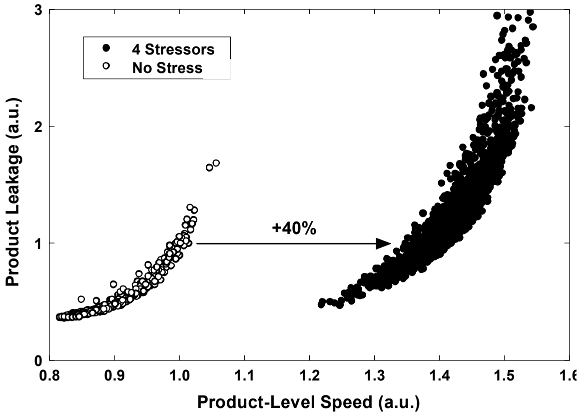


Fig. 8. Percentage NMOS (left) PMOS (right) improvements from Table 2 as individual stress techniques are added. Dotted lines indicate any deviations from the values predicted from Table 2, based on experiments where multiple stress techniques are combined. NMOS stress techniques are fully additive for IDSAT, more than additive for IDLIN with gate oxide optimization. PMOS stress techniques are more than additive for IDSAT, but not fully additive for IDLIN as soon as the compressive liner is added.

Fig. 9. Yield prior to production ramp for stress memorization (1-stressor), plus compressive and tensile liners stress, plus embedded-SiGe (4-way stressor).



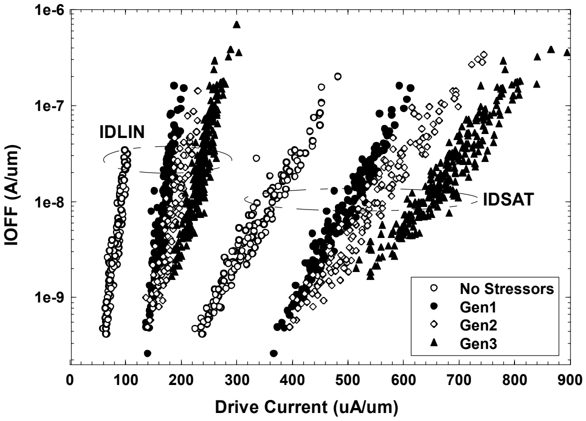


Fig. 10. Normalized Athlon64™ product-level speed (FMAX) versus leakge (static uP current) showing improvement with all 4 way stress techniques (4 stressors) integrated and optimized into the product

Fig. 11. PMOS IDSAT and IDLIN for 3 different generations of e-SiGe integrations on the same transistor with the same compressive liner.