Delaying Forever: Uniaxial Strained Silicon Transistors in a 90nm CMOS Technology

K. Mistry, M. Armstrong, C. Auth, S. Cea[#], T. Coan, T. Ghani, T. Hoffmann[#], A. Murthy, J. Sandford, R. Shaheed[#], K. Zawadzki, K. Zhang, S. Thompson and M. Bohr

Portland Technology Development, [#]TCAD, Intel Corp., RA1-204 Hillsboro, OR, USA

(503) 613-7988, (503) 613- 6069 FAX, Kaizad.Mistry@intel.com

Abstract

We describe the device physics of uniaxial strained silicon transistors. Uniaxial strain is more effective, less costly and easier to implement. The highest PMOS drive current to date is reported: 0.72mA/μm. Pattern sensitivity and mobility/Rext partitioning are discussed. Finally we measure inverter delays as low as 4.6pS, and show 50Mb SRAMs operational at 0.65V.

Introduction: Why Strained Silicon?

In 1965, Gordon Moore famously predicted that transistor density would periodically double [1]. As transistors enter the nanometer era, Moore, asked about the future of his eponymous law, remarked, "No exponential is forever. But you can delay forever." Strained silicon is one technique to delay forever.

As MOSFET gate lengths scale below 100nm, it is difficult to maintain high drive current due to mobility degradation. To the tyranny of the universal mobility model (mobility degrades at higher effective field) is added the malice of ionized impurity scattering: the required increase in channel doping to control short channel effects degrades mobility further [2].

The use of strain to improve mobility has been known for 50 years [3]. Biaxial tensile strained silicon has received considerable attention in the last decade as a technique for mobility enhancement [4-5], but has been difficult to implement because of misfit and threading dislocations, Ge up-diffusion, fast diffusion of S/D extensions, and cost. Also, hole mobility gain at high fields is very modest [5], because quantization lowers the separation between light and heavy hole bands [6].

Why Uniaxial Strain?

Uniaxial strain offers several advantages over biaxial strain. Electron mobility gain is similar for uniaxial and biaxial strain, since it results from the splitting of the six-fold degenerate conduction band valleys for both types of stress.

Simple calculations using peizoresistance coefficients [3] for industry standard <110> channel orientation P-MOSFETs on a (100) surface show that uniaxial stress (longitudinal compressive) offers much larger hole mobility enhancement at a given stress level than biaxial tensile strain (Fig. 1).

Hole mobility gain with uniaxial stress is maintained at high effective field as shown by us previously [7,8], as well as by wafer bending [9]. It has been suggested that mobility gain for uniaxial compression results more from band warpage than from band separation [10]. It is possible that band separation does not change with surface confinement (as seen for biaxial compression [6]). Finally, uniaxial strain may have lower surface roughness scattering due to high out-of-plane effective mass.

Two simple structures can implement uniaxial strain, avoiding the complex wafer fabrication, cost, and defects of biaxial strain. Tensile or compressive capping layers can be deposited on top of fully formed transistors, enhancing NMOS and PMOS transistors respectively [11]. One drawback of this approach is that the transistor of the opposite type can have degraded performance.

A novel structure to induce uniaxial strain is the Epitaxial Source/Drain (ESD) transistor (Fig. 2 (a)) [7,8]. After formation of the gate stack, S/D extensions and spacer, a silicon recess etch is performed. Then, selective heteroepitaxy is used to grow a strained material in the S/D regions. If the lattice spacing of this material is larger (smaller) than silicon, uniaxial compressive (tensile) strain is induced in the channel (Fig 2 (b)).

In the two uniaxial strain structures described above, the strain step is introduced late in the process flow, minimizing integration and defect challenges. Cost is lower due to an order of magnitude lower thickness of epitaxially deposited material.

Application to a 90nm CMOS Technology

Uniaxial strained silicon NMOS and PMOS transistors have been implemented in a high volume manufacturing 90nm CMOS technology [7,8]. NMOS transistors employ a tensile capping layer (Fig. 3) to induce strain and improve NMOS drive current by 10%. PMOS transistors employ selective SiGe heteroepitaxy to generate uniaxial compressive strain in the channel (Fig 4).

Record NMOS I_{DSAT} of 1.26mA/μm is achieved at a supply voltage of 1.2V and an I_{OFF}=40nA/μm (Fig 5). Astonishing PMOS I_{DSAT} of 0.72mA/μm is achieved also at 1.2V & 40nA/μm (Fig 6). The tensile capping layer improves NMOS drive current without significant PMOS loss (Fig. 7). The unique combination of NMOS & PMOS strain allows independent optimization.

Fig. 8 plots ESD PMOS I_{DLIN} gain vs. standard devices against gate length along with the simulated channel stress. Clearly the current gain is caused by strain till the device becomes Rext limited. For L_{GATE}=50nm, the mobility gain is 55% and accounts for 60% of the I_{DLIN} gain. The remainder is due to reduced Rext - in part due to a lower hole barrier at the NiSi/SiGe interface.

Initial results showed a large difference in I_{DSAT} for minimum pitch vs. wider pitch ESD PMOS transistors. With process optimization (Fig. 9, Process 3) the difference can be minimized.

90nm Circuit Results

Strained silicon results in fast ring oscillators. Fig. 10 shows measured F.O.=1 R.O. delay vs. transistor I_{OFF} at 1.2V. For high VT devices at (I_{OFF-N}+I_{OFF-P})=80nA/μm, the delay is 5.5pS. For low VT devices at (I_{OFF-N}+I_{OFF-P})=800nA/μm, the delay is 4.6pS.

Fig. 11 shows schmoo for a 50Mb SRAM using a 1.0μm^2 6-T SRAM cell. The SRAM is fully operational down to V_{DD}=0.65V. Low voltage SRAMs are key to long battery life in mobile applications. Strained silicon enables improved VCCmin for a given cell area because the strain-enhanced PMOS pullup helps balance V_T mismatch in the NMOS pulldown. Strained silicon thus enables both performance & power scaling.

Conclusions

Uniaxial strained silicon has been implemented in a high volume manufacturing 90nm logic technology for the first time, with impressive performance results and improved power scaling.

References

[1] G. E. Moore, *Electronics,* pp. 114-117, 1965.
[2] T. Ghani et al., *Symp. VLSI Tech. Dig.*, 2000, pp. 174-175.
[3] C. S. Smith, *Phys. Rev.*, Vol. 94, pp. 42-49, 1954.
[4] J. Welser, et al., *IEDM Tech. Dig.*,1000-1003, 1992.
[5] J. Hoyt, et al., *IEDM Tech. Dig.*, pp. 23-26, 2002.
[6] M. Fischetti et al., *J. Appl. Phys.*, vol. 94, pp. 1079-1095, 2003.
[7] S. Thompson et al., *IEDM Tech. Dig.*, pp. 61-64, 2002.
[8] T. Ghani et al., *IEDM Tech. Dig.*, pp. 978-980, 2003.
[9] R.E. Belford et al., *DRC Conf Dig.*, pp.41-42, 2002.
[10] M. Giles et al., submitted to *2004 VLSI Symp Tech.*.
[11] S. Ito et al., *IEDM Tech. Dig.*, pp. 247-250, 2000.

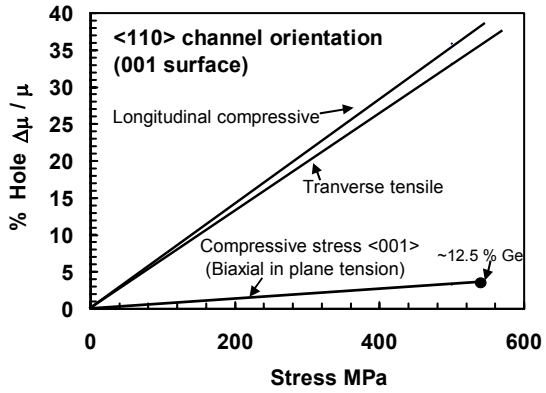
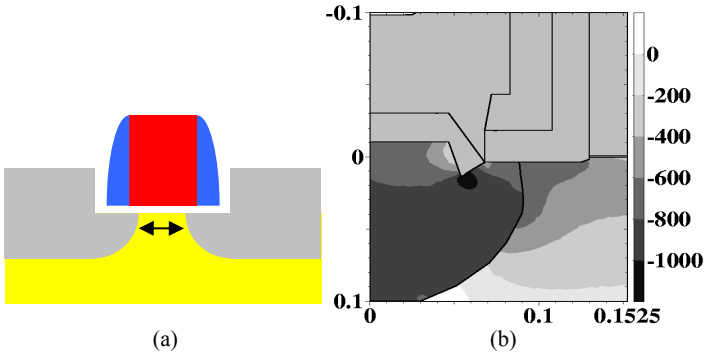


Fig. 1 Hole mobility gain vs. stress from the piezo coefficients in [3].



(a) (b)

Fig. 2 (a) Epitaxial S/D Transistor structure. (b) Stress simulations: the resulting stress is dominantly uniaxial along the [110] current flow direction X and Y-axes show dimensions in μm. Contours show stress in MPa.

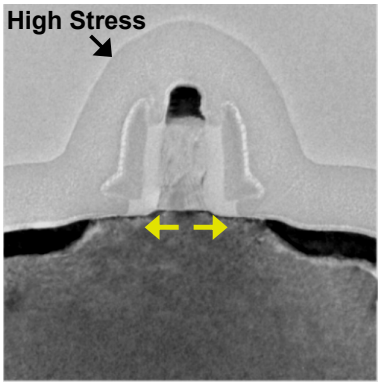


Fig. 3 TEM of NMOS transistor showing high tensile stress nitride overlayer.

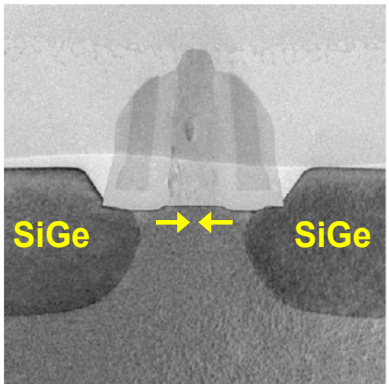


Fig. 4 TEM of PMOS showing SiGe heteroepitaxial S/D inducing uniaxial strain.

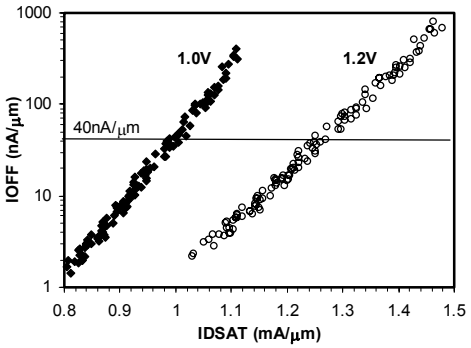


Fig. 5 NMOS Ion/Ioff curve at 1.2V & 1.0V. At 1.0V & I_{OFF}=40nA/μm, I_{DSAT}=1.01mA/μm

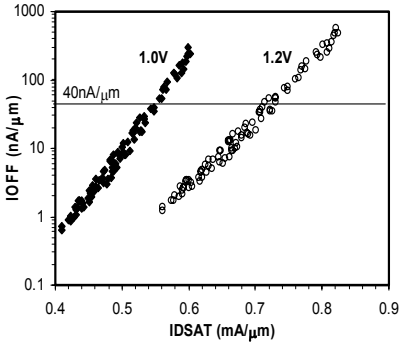


Fig. 6 PMOS Ion-Ioff at 1.2V & 1.0V. At 1.0V & I_{OFF}=40nA/μm, I_{DSAT} =0.55mA/μm.

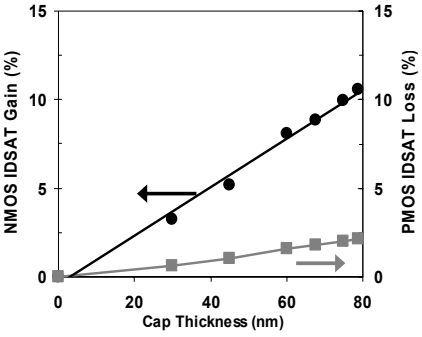


Fig. 7 N & P I_{DSAT} vs. nitride capping layer thickness.

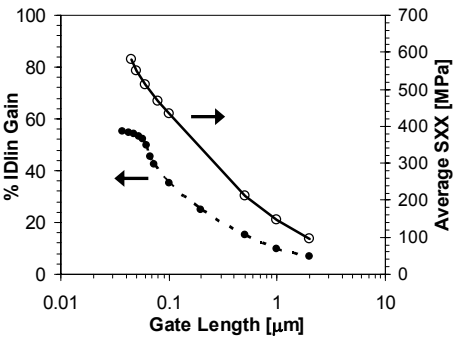


Fig. 8 Avg. PMOS channel stress (SXX) and % I_{DLIN} gain [corrected for (V_G-V_T)] vs. L_{GATE}

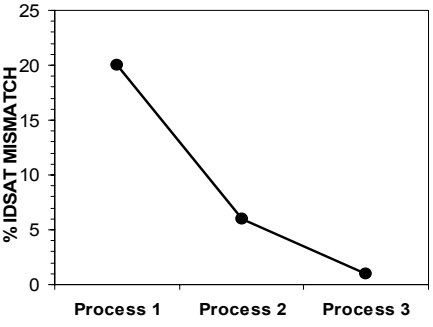


Fig. 9 PMOS I_{DSAT} mismatch for minimum pitch vs. 1.5X min pitch.

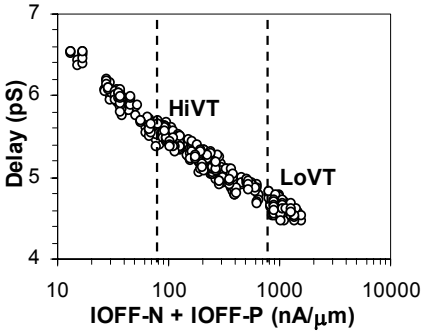


Fig 10 Ring Oscillator delay for fanout=1 vs. sum of NMOS and PMOS I_{OFF}.

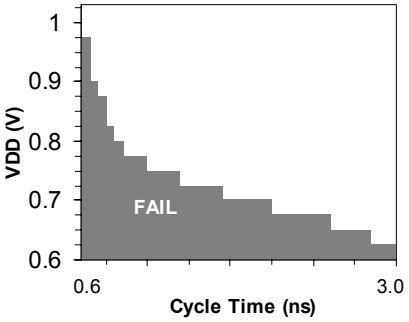


Fig. 11 Schmoo of 50Mb SRAM operating down to V_{DD}=0.65V.