Mater. Res. Soc. Symp. Proc. Vol. 913 © 2006 Materials Research Society 0913-D05-05

TCAD Modeling of Strain-Engineered MOSFETs

Lee Smith
Synopsys, Inc., 700 E. Middlefield Road, M/S A14, Mountain View, CA, 94043

ABSTRACT

The rapid rise of standby power in nanoscale MOSFETs is slowing classical scaling and threatening to derail continued improvements in MOSFET performance. Strain-enhancement of carrier transport in the MOSFET channel has emerged as a particularly effective approach to enable significant performance improvements at similar off-state leakage. In this paper we describe how strain effects are modeled within the context of TCAD process and device simulation. We also use TCAD simulations to review some of the common approaches to engineer strain in MOSFETs and to explain how strain impacts device and circuit characteristics.

INTRODUCTION

Strain engineering is rapidly becoming a ubiquitous element in modern MOSFET design [1-3]. In the current stage of technology development, much effort is being spent to model strain effects and to optimize the strain induced by various strain sources. Technology Computer Aided Design (TCAD) tools provide a convenient means of simulating the stress and strain produced during the strain-engineered process flow as well as the impact of that strain on device performance. As will be shown by example, the resulting stress and strain fields are often non-intuitive. The impact of strain on device characteristics is determined primarily through changes in the band structure. In this paper, we review how subsequent changes in carrier repopulation, effective mass, and scattering enhance, or degrade, the mobility and shift the threshold voltage for various stress configurations. In this context, optimizing the enhancement of the low-field mobility can be viewed as an exercise in band structure engineering. For high-field transport, we use Monte Carlo device simulation to investigate the impact of strain on velocity overshoot and drive current. Beyond the analysis and optimization of strain for a single device lies the next stage in strain engineering: the impact of layout. Due to the large interaction range of stress in CMOS materials, approximately 2 μm, the modeling of isolated devices is not sufficient to predict final circuit behavior. In this paper, we also review some simulation studies we have performed to investigate the impact of circuit layout on channel stress and circuit performance.

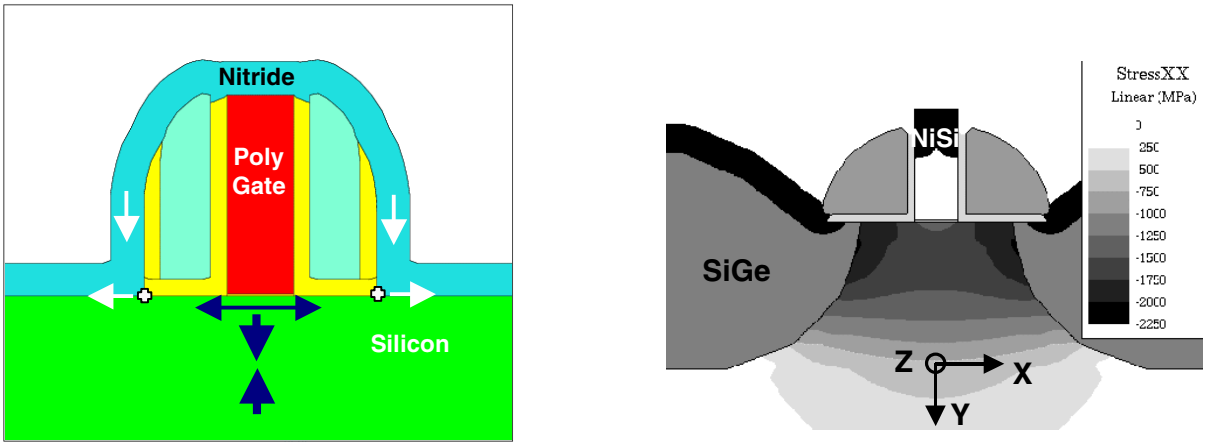
EXAMPLES OF STRAIN ENGINEERING

Strain can be engineered into a conventional MOSFET structure in many different ways. These different approaches are typically categorized as either global or a local in nature. Global approaches, such as strained-Si on relaxed SiGe, attempt to induce uniform strain throughout the

Si layer across the entire wafer [4]. While this type of approach has received a great deal of academic interest, the first approaches used in production have been of the local type [1]. Local, or process-induced, strain engineering focuses on producing stress in the channel of a single device. Figure 1 shows two of the most commonly used local approaches: a strained nitride capping layer and embedded SiGe (e-SiGe) in the source/drain regions. The stress from these approaches can be calculated throughout the process flow by solving the force equilibrium equations while considering various sources of stress such as intrinsic film stress, thermal mismatch, and lattice mismatch. Often, the resulting stress fields are non-intuitive.

For example, simulations of tensile strained nitride cap layers for NMOS consistently predict that the vertical stress component is the dominant stress in the channel, rather than the longitudinal stress component, as is commonly assumed. This vertical stress is produced by the tensile cap layer pushing the gate stack down onto the silicon substrate. The longitudinal stress is produced by the cap layer pulling the edges of the gate stack away from the channel.

In the case of e-SiGe source/drain, the induced stress is primarily uniaxial along the longitudinal, or channel, direction. The embedded $Si_{1-x}Ge_x$ regions are fabricated by first etching a recess in the silicon substrate and then growing $Si_{1-x}Ge_x$ via selective epitaxy. Due to the lattice mismatch between $Si_{1-x}Ge_x$ and Si, the compressively strained $Si_{1-x}Ge_x$ pushes out against either end of the channel inducing a compressive, longitudinal channel stress. The stress obtained in the channel depends on many properties of the SiGe regions such as the Ge mole fraction, the SiGe recess depth and elevation height, and the shape of the SiGe regions near the channel [3].



(a) (b)

Figure 1. Examples of local stress engineering. a) Tensile strained nitride cap layer induces longitudinal tension and vertical compression into the channel. b) e-SiGe source/drain regions induce longitudinal compressive stress in the channel.

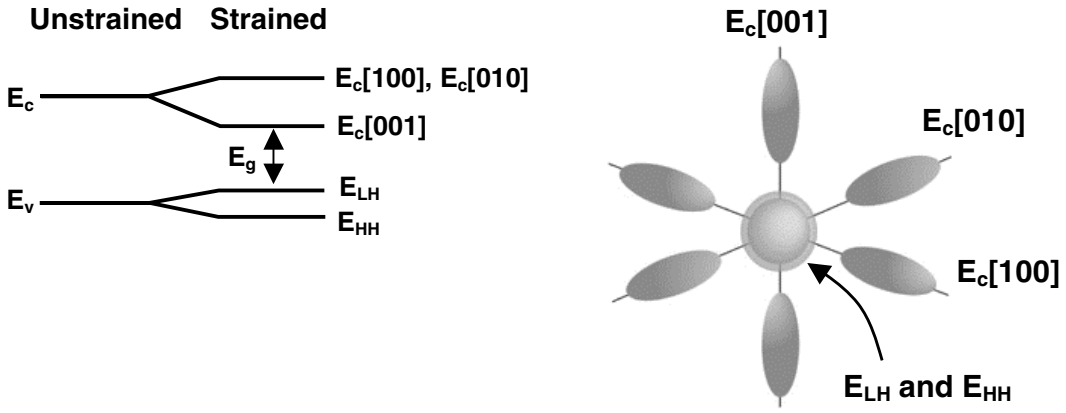


Figure 2. Band splitting in Si due to uniaxial compressive stress along [001].

MODELING THE IMPACT OF STRESS ON DEVICE BEHAVIOR

Stress modifies many aspects of device behavior such as the leakage current, threshold voltage, and mobility. The primary link between stress and these changes to device behavior is through the band structure. As depicted in Figure 2, stress, in general, splits and shifts the silicon band extrema. In the example shown here, uniaxial compressive stress along [001] splits the six Δ valleys of the conduction band into two different groups. Likewise, the heavy and light hole bands are split with the light hole band moving up in energy. These shifts cause a change in the band gap and electron affinity and a subsequent change in threshold voltage. In this example, the band gap is reduced which can lead to an increase in leakage current through increased recombination current and band-to-band tunneling.

Band splitting due to stress also alters valley repopulation and inter-valley scattering and, therefore, the mobility. Under most types of stress, the conduction valleys in silicon undergo rigid shifts with negligible change in effective mass. Recent work, however, suggests the electron effective masses can also be modified under appropriate stress [5]. Because the relaxed silicon valence band consists of degenerate valleys with a high degree of band warping, hole transport in silicon is particularly sensitive to stress. In addition to level splitting, stress significantly modulates the band curvature and the hole effective masses.

Physical model for pmos mobility

A physical approach to modeling the stress dependence of mobility starts with a detailed calculation of the band structure under stress. For example, Figure 3 shows the valence band structure of relaxed and strained Si under uniaxial, compressive stress along [110] computed using a 6-band **k·p** approach [6]. The top valence band in relaxed Si is very warped, and the band dispersion along [110] and [001] shows the familiar degenerate light hole and heavy hole bands at the Γ point. Under uniaxial [110] compressive stress, the light and heavy hole bands

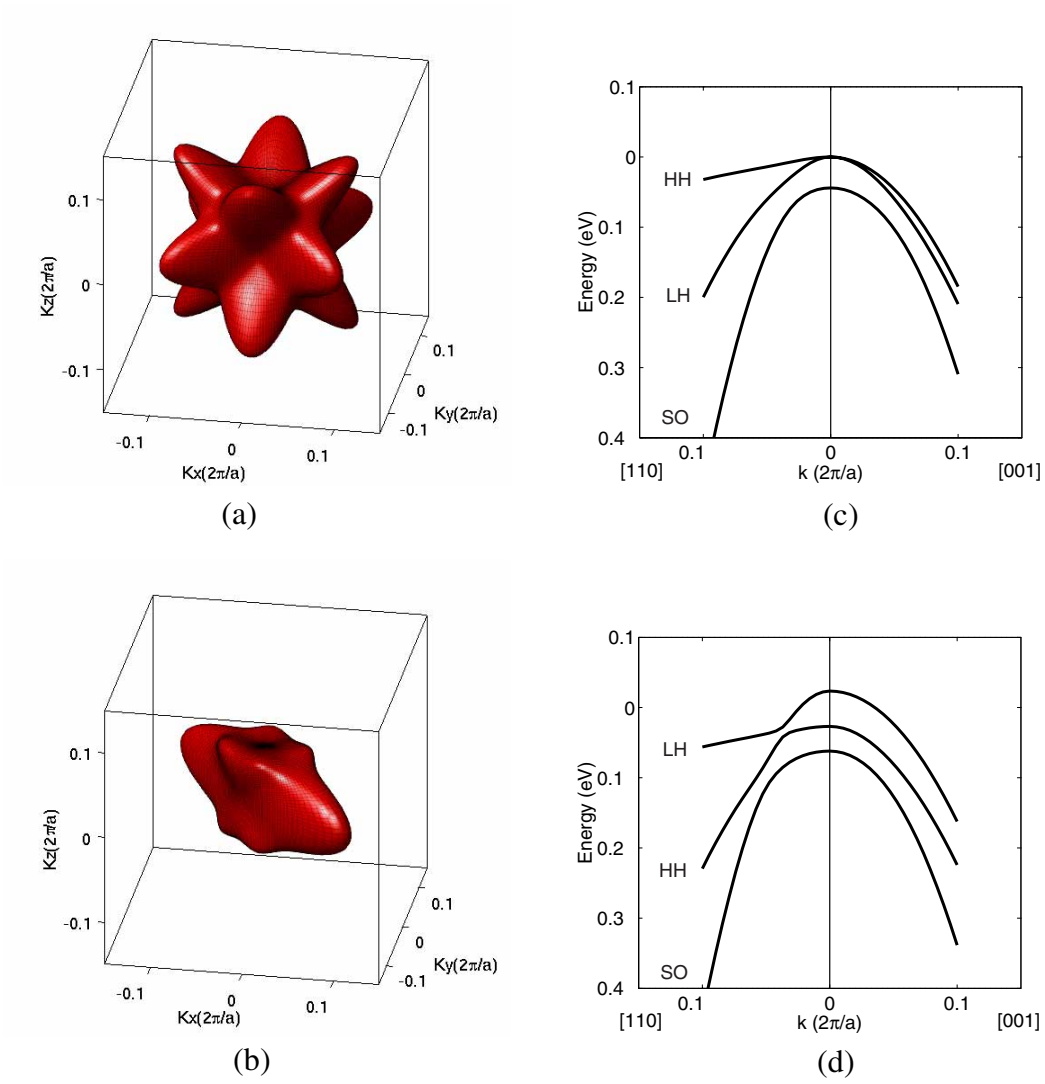


Figure 3. Bulk valence band structure for relaxed and strained Si. The stress in the strained case is 1 GPa of uniaxial, compressive stress along [110]. Figures are: Iso-energy contours throughout k-space at 50 mV below the top of the band for a) relaxed Si and b) strained Si; band dispersion along [110] and [001] for c) relaxed Si and d) strained Si.

split, and the top band simplifies to an almost single ellipsoidal band with one arm oriented along [-110]. The band dispersion along [110], the typical MOSFET channel direction, places the light hole band as the top band with a reduced mass along [110].

While mobility can be computed rigorously from the band structure and scattering rates, a simplified model often provides good accuracy and a more intuitive understanding of the important mechanisms involved [7]. As shown in Figure 4, the stress-induced changes to the top-most valence band are modeled using two ellipsoidal valleys oriented along [110] and [-110].

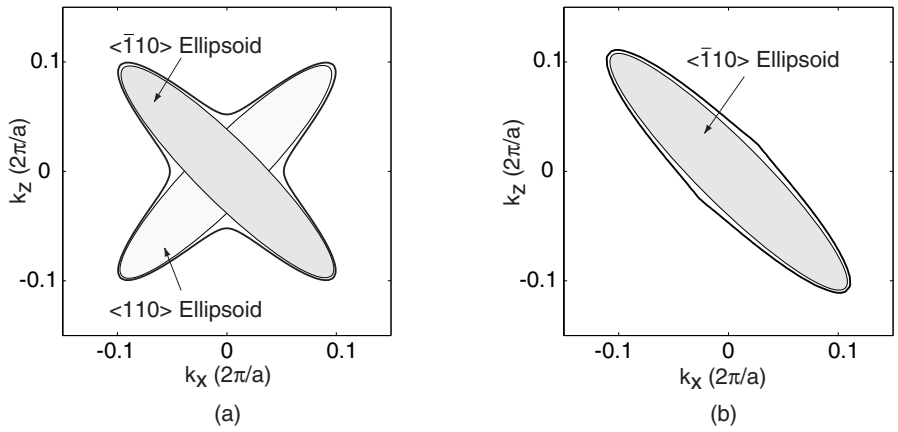


Figure 4. Thick lines show the constant energy contour of the top-most valence band at 50 mV below the top of the band for (a) relaxed and (b) uniaxially stressed Si. Also sketched are the ellipsoidal valleys used in the mobility model.

Each valley is characterized by a transverse and longitudinal effective mass. Assuming a constant mean scattering time $\langle\tau\rangle$, the mobility along the [110] direction can be modeled as:

$$\mu_{[110]} = q\langle\tau\rangle\left(\frac{f}{m^t_{[-110]}} + \frac{1-f}{m^l_{[110]}}\right) \tag{1}$$

where f is the occupancy of the [-110] valley, $m^t_{[-110]}$ is the transverse mass of the [-110] valley, $m^l_{[110]}$ is the longitudinal mass of the [110] valley and q is the electron charge. The stress-dependence of the energy level splitting between the two valleys and the effective masses are expanded as polynomials of elements of the stress tensor.

The mobility model is calibrated against measured data from wafer-bending experiments as well as from strain-engineered PMOSFETs that employ various combinations of e-SiGe source/drain and strain capping layers implemented as a compressive etch stop layer (CESL) [8, 9]. For the strain-engineered MOSFETs, the shape of the e-SiGe source/drain and the properties of the stressed cap layer were engineered to vary the compressive stress in the channel from 200 MPa to 2.0 GPa. Figure 5 shows the measured and modeled mobility gain as a function of the effective channel stress. For wafer-bending data, the channel stress is obtained directly from curvature measurements. For the strain-engineered MOSFETs, the channel stress is calculated using 2D process simulation where all intentional and unintentional stress sources as well as the stress evolution during the entire process flow are taken into account [10]. While the channel stress is primarily uniaxial along the channel, there is a non-negligible transverse component that degrades the mobility. This is treated here via an effective channel stress that represents the equivalent reduction in longitudinal stress.

The model agrees well with both the wafer-bending data and the mobility gain extracted from the strain-engineered MOSFETs. Compared to the bulk piezoresistance model [11], the

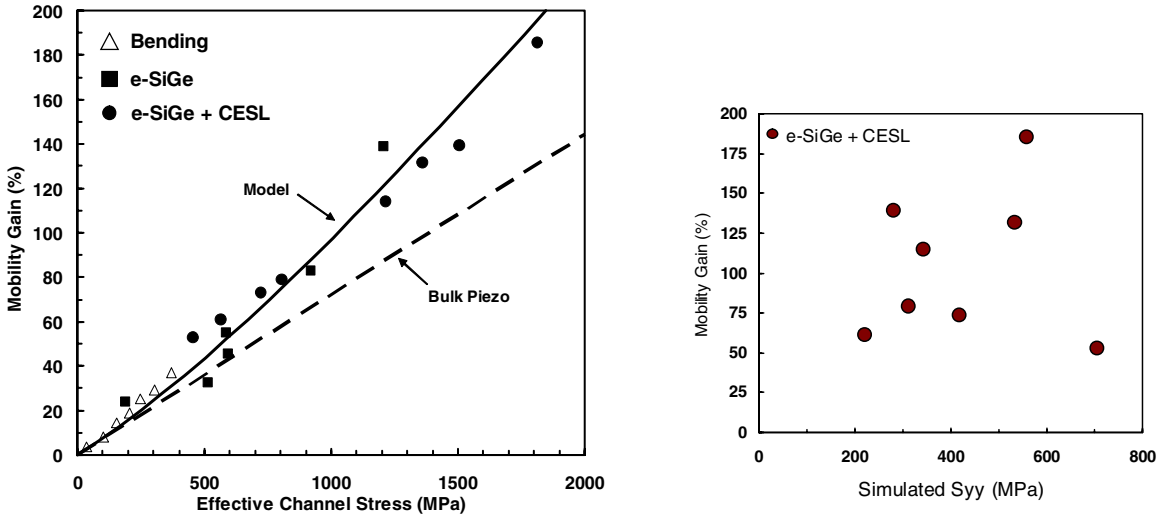


Figure 5. a) Measured and modeled hole mobility enhancement as a function of the effective channel stress. The measured data are shown as symbols, the dashed line shows bulk piezoresistance, and the solid line shows the mobility model described in the text. The bending data are from [12]. b) Measured mobility gain as a function of simulated vertical stress showing no correlation.

actual response is moderately superlinear at high stress. The stress dependence of the effective masses and band splitting that result from the calibrated fit indicate that the mobility enhancement arises from two source: repopulation of holes into the top [-110] ellipsoidal valley that has a small transverse mass along [110], and a reduction of this transverse mass as stress is increased. An almost 3x enhancement in the mobility is obtained at 2 GPa of channel stress. Importantly, no saturation in the mobility enhancement is seen even at this stress level. While it is expected that the valley occupancy and transverse mass improvement will saturate at high level splitting, further improvements in the mobility can come from the suppression of scattering.

While the mobility enhancement for vertical stress cannot be characterized via wafer-bending on planar devices, the combination of measured mobility gain and simulated stress allows for an indirect characterization of this stress component. The measured mobility gain as a function of the simulated vertical stress is shown in Figure 5b. No correlation between the mobility gain and the vertical stress is seen, indicating that the effective piezoresistance coefficient for vertical stress is small in (100) PMOS, in agreement with the bulk piezoresistance model.

Stress-enhanced mobility and effective field

When dealing with stress-enhanced mobility in MOSFETs an important consideration is the dependence of the mobility enhancement on vertical effective field. Device scaling continues to increase the effective field with values reaching over 1.5 MV/cm at the 65 nm technology

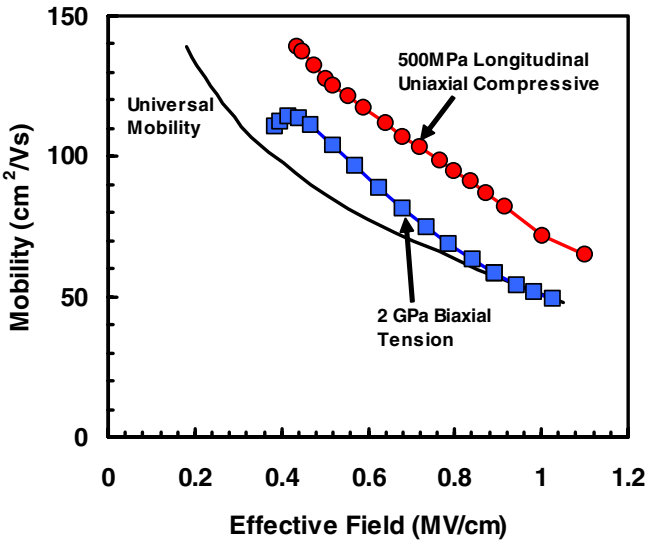


Figure 6. Effective low-field channel mobility as a function of vertical effective field for: the Si universal mobility curve [13], biaxial in-plane tension [2], and uniaxial longitudinal compression [1].

node. One view of stress enhancement techniques is that they allow the "tyranny of the Si universal mobility curve" to be broken. The goal is to find stress tensors that are able to shift the entire universal mobility curve to a higher level over the entire effective field range. Figure 6 compares the effective field dependence of the mobility enhancement for two different types of stress tensors in PMOS: biaxial, in-plane tension and uniaxial, longitudinal compression. While biaxial tension produces a healthy enhancement at low effective field, the enhancement is lost as the effective field is increased. In contrast, uniaxial, longitudinal compression is able to maintain a large mobility enhancement over the entire effective field range.

The difference in behavior between these two stress cases lies in how the subband structure changes under stress. Figure 7 shows the top two subbands for these two stress cases at low and high effective field computed using a triangular well approximation within the 6-band **k·p** approach [14]. At low effective field, both stress cases produce a healthy separation between the top two subbands which reduces inter-subband scattering and maintains a small effective mass along the transport direction. At high effective field, however, the separation between the top subbands for the biaxial case is reduced as the effects of band splitting due to stress and quantization work against each other. This leads to increased inter-subband scattering and reduced mobility. In addition to subband splitting, the effective masses at the top of the subbands along the transport direction are affected differently by the two stress cases. At high effective field, uniaxial stress produces a small effective mass while biaxial stress produces a large mass which is detrimental to mobility.

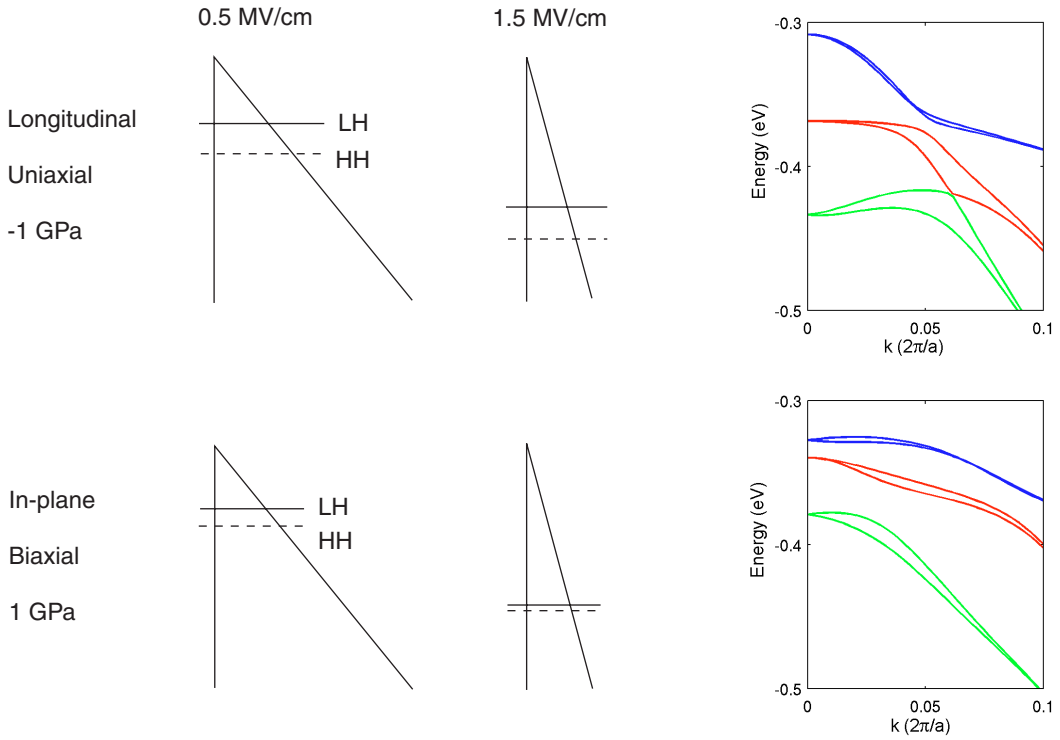
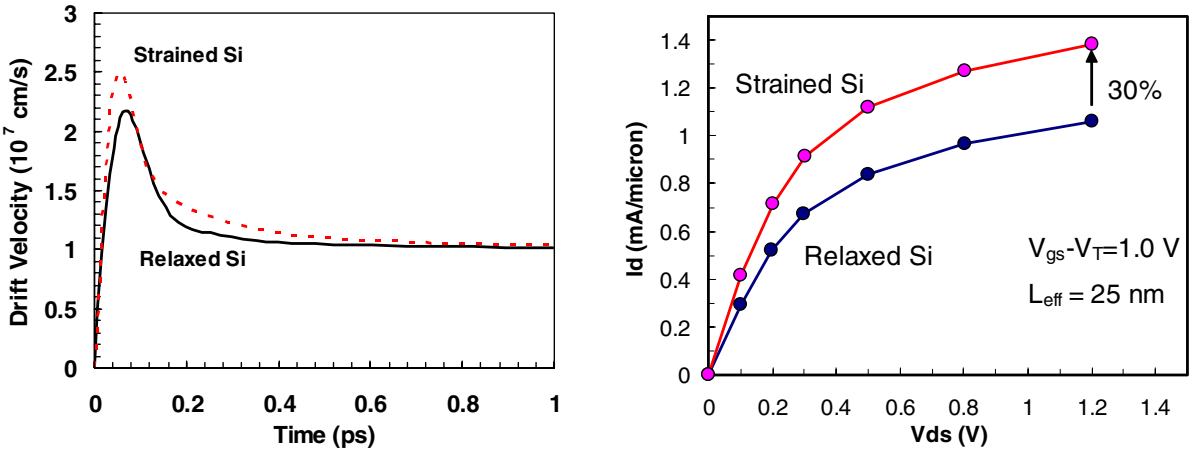


Figure 7. Comparison of subband behavior under uniaxial and biaxial stress in PMOS under low and high effective field. The band dispersion shows the top six subbands along [110] at high effective field. The biaxial stress case produces a large effective mass in the top subband which is detrimental to mobility.

High-field transport under stress

While enhancement of the low-field mobility is an important objective for stress engineering, the ultimate goal is to enhance the final device performance in terms of improved drive current. In addition to increased vertical effective field, device scaling also increases the longitudinal driving field from source to drain. This pushes devices further into the quasi-ballistic regime. Competing effects such as velocity saturation and velocity overshoot now become important in determining the final drive current and how the low-field mobility enhancement is translated into drive current enhancement.

Monte Carlo device simulation provides a direct means of investigating the impact of stress on high field transport. Both Monte Carlo simulations and experiments of high-field transport in strained Si show that the saturation velocity is not significantly enhanced by stress. In contrast, simulations suggest that velocity overshoot is strongly affected. Figure 8 compares the simulated transient velocity overshoot effect in bulk Si for relaxed and biaxally strained Si with a stress equivalent to a strained-Si layer grown on relaxed $Si_{0.8}Ge_{0.2}$. Repopulation of carriers into valleys with a small effective mass along the transport direction leads to significant enhancement of the velocity in the overshoot regime. The simulation of a 25 nm gate length NMOSFET shows a drive current enhancement of 30% as compared to a 75% enhancement of



(a) (b)

Figure 8. Monte Carlo device simulation of relaxed Si and strained Si on $Si_{0.8}Ge_{0.2.}$ a) Comparison of transient velocity overshoot behavior in n-type Si. b) Comparison of drain characteristics for a 25 nm gate-length NMOSFET.

the low-field mobility. This ratio of approximately 0.5 between drive-current gain and the low-field mobility gain has also been seen experimentally [15]. We have extracted a similar ratio between the drive-current gain and the low-field mobility gain for PMOS under uniaxial stress [9].

MODELING THE IMPACT OF LAYOUT ON STRESS AND CIRCUIT BEHAVIOR

While local stress sources such as strained capping layers and e-SiGe source/drain are optimized to induced stress into the channel of a single MOSFET, the stress fields that they produce can extend quite a distance beyond the target transistor. Even unintentional stress sources such as STI can induce significant stress over a 2 μm range. These stress proximity effects (SPE) make the final stress in the channel of a MOSFET depend on the environment in which the MOSFET is placed, i.e. on the circuit layout.

As an example of SPE we consider the impact of layout density on the performance of inverters in a ring oscillator. As diagrammed in Figure 9, in a sparse layout a single, isolated transistor is surrounded by a large area of STI. In a dense layout, transistors are closely nested and separated by a small amount of STI. Detailed analysis of the stress fields produced by these layouts was performed using 3D process simulation [16].

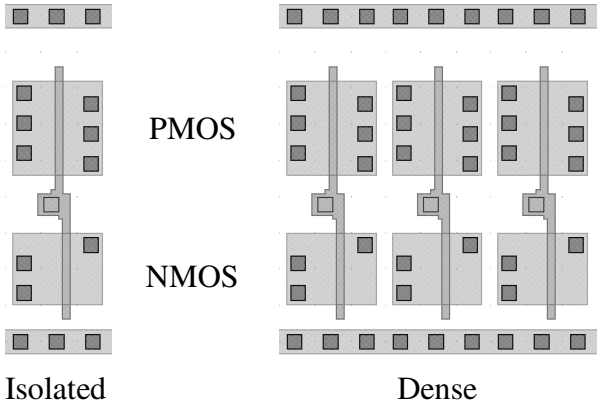
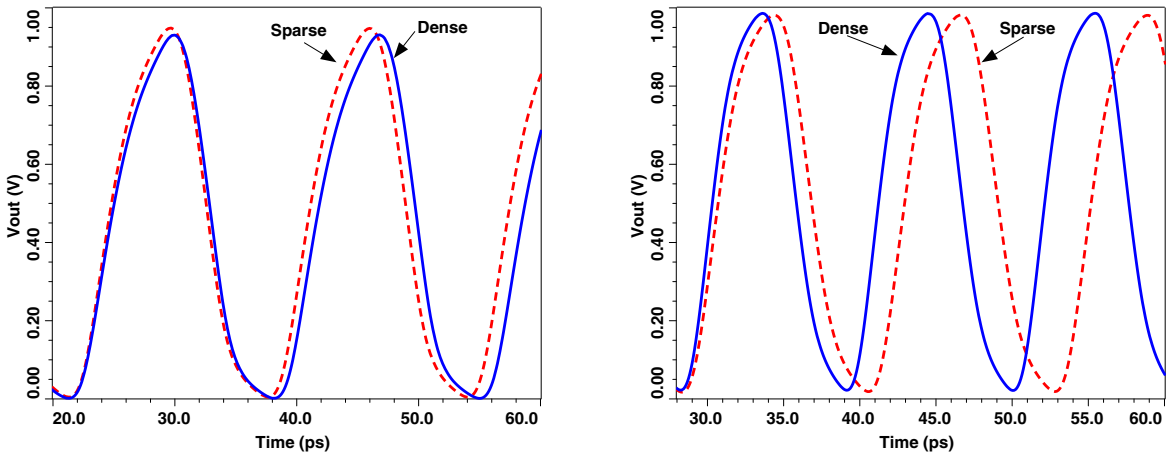


Figure 9. Layout for inverters in isolated and dense environments.

We first consider MOSFETs without any engineered stress sources. In this case the STI produces unintentional biaxial compressive stress in the isolated transistors. This has little effect on the PMOSFET mobility but degrades the performance of the NMOSFET. In the dense layout the small amount of STI between the transistors is unable to generate significant stress and, therefore, the longitudinal stress component is suppressed. This is beneficial for the NMOSFET mobility but detrimental for the PMOSFET performance. The net impact of these layout-induced changes to stress on circuit performance was investigated by simulating the response of a 3-stage ring oscillator. The device output characteristics shown in Figure 10 were computed using energy balance device simulation to consider the impact of stress on both the low-field mobility and velocity overshoot. Overall, the PMOSFET response to layout determines the overall change in circuit response, causing the oscillator frequency to slow when changing from a sparse to dense layout.

Adding an engineered stress source to the PMOSFET completely changes the layout dependence. Including e-SiGe source/drain into the PMOSFET induces beneficial stress into the channel. The amount of channel stress that can be generated, however, depends on the environment around the e-SiGe stressor. Because of the relative softness of STI, a large area of STI can act as a stress relaxor and reduce the amount of channel stress obtained. The effectiveness of the e-SiGe stressor is therefore degraded in the sparse layout. In the dense layout, little stress relaxation occurs because of the limited area of STI along the longitudinal direction. The change in oscillator performance between these two layouts is shown in Figure 10. The enhanced PMOSFET performance in the dense layout now greatly improves the overall oscillator speed as compared to the sparse layout.



(a) (b)

Figure 10. Comparison of ring oscillator behavior for sparse and dense layouts. a) STI as the only stress source. b) STI and e-SiGe as stress sources.

SUMMARY

Strain engineering is now a critical part of modern MOSFET design. TCAD process simulation provides a convenient means of analyzing the interaction of multiple stress sources during the process flow. The results produced by these complex interactions as well as the relaxation induced by STI are often not intuitive. In terms of device behavior, strain engineering can be viewed as an exercise in band engineering. The change in band structure with strain provides a physical basis for modeling strain-induced mobility enhancement and for identifying particular stress tensors that are beneficial for enhancing device performance. As device scaling continues, strain effects on quasi-ballistic transport will become more important in setting the ratio of drain current enhancement to low-field mobility enhancement. Due to the large range of stress proximity effects in CMOS materials, the modeling of isolated devices is not sufficient to predict final circuit behavior. The impact of circuit layout on channel stress will need to be considered as well.

ACKNOWLEDGMENTS

The author would like to thank Victor Moroz, Xiaopeng Xu, Dipu Pramanik, and Fabian Bufler from Synopsys for useful discussions and simulations and Faran Nouri and Lori Washington from Applied Materials, and Geert Eneman from IMEC, for useful discussions and device data.

REFERENCES

1. S. E. Thompson, M. Armstrong, C. Auth, S. Cea, R. Chau, G. Glass, T. Hoffman, J. Klaus, Z. Ma, B. Mcintyre, A. Murthy, B. Obradovic, L. Shifren, S. Sivakumar, S. Tyagi, T. Ghani, K. Mistry, M. Bohr and Y. El-Mansy, *IEEE Electron Device Lett.* **25,** 191 (2004).
2. K. Rim, J. Chu, J.H. Chen, K.A. Jenkins, T. Kanarsky, K. Lee, A. Mocuta, H. Zhu, R. Roy, J. Newbury, J. Ott, K. Petrarca, P. Mooney, D. Lacey, S. Koester, K. Chan, D. Boyd, M. Ieong, H.-S. Wong, *Symposium on VLSI Technology, Digest of Technical Papers* **2002**, 98 (2002).
3. F. Nouri, P. Verheyen, L. Washington, V. Moroz, I. De Wolf, M. Kawaguchi, S. Biesemans, R. Schreutelkamp, Y. Kim, M. Shen, X. Xu, R. Rooyackers, M. Jurczak, G. Eneman, K. De Meyer, L. Smith, D. Pramanik, H. Forstner, S. Thirupapuliyur and G. S. Higashi, *IEDM Tech. Dig*. **2004**, 1055 (2004).
4. J. Welser, J.L. Hoyt and J.F. Gibbons, *IEEE Electron Device Lett.* **15**, 100 (1994).
5. K. Uchida, T. Krishnamohan, K.C. Saraswat and Y. Nishi, *IEDM Tech. Dig*. **2005**, 129 (2005).
6. T. Manku and A. Nathan, *J. Appl. Phys.* **73**, 1205 (1993).
7. B. Obradovic, P. Matagne, L. Shifren, X. Wang, M. Stettler, J. He and M. D. Giles, *IWCE Tech. Dig*., 26 (2004).
8. L. Smith, V. Moroz, G. Eneman, P. Verheyen, F. Nouri, L. Washington, M. Jurczak, O. Penzin, D. Pramanik and K. De Meyer*, IEEE Electron Device Lett.* **26**, 652 (2005).
9. L. Washington, F. Nouri, S. Thirupapuliyur, G. Eneman, P. Verheyen, V. Moroz, L. Smith, X. Xu, M. Kawaguchi, T. Huang, K. Ahmed, M. Balseanu, L. Xia, M. Shen, Y. Kim, R. Rooyackers, K. De Meyer and R. Schreutelkamp, *IEEE Electron Device Lett.,* accepted for publication (2006).
10. V. Moroz, N. Strecker, X. Xu, L. Smith and I. Bork, *Material Science in Semiconductor Processing* **6**, 27 (2003).
11. C. S. Smith, *Phys. Rev*. **94,** 42 (1954).
12. E. Wang, P. Matagne, L. Shifren, B. Obradovic, R. Kotlyar, S. Cea, J. He, Z. Ma, R. Nagisetty, S. Tyagi, M. Stettler and M. D. Giles, *IEDM Tech. Dig*. **2004**, 147 (2004).
13. S. Takagi, A. Toriumi, M. Iwase and H. Tango, *IEEE Trans. on Electron Devices* **41**, 2357 (1994).
14. S. Rodriguez, J. A. Lopez-Villanueva, I. Melchor and J. E. Carceller, *J. Appl. Phys.* **86**, 438 (1999).
15. A. Lochtefeld and D. Antoniadis, *IEEE Electron Device Lett*. **22**, 591 (2001).
16. V. Moroz, G. Eneman, P. Verheyen, F. Nouri, L. Washington, L. Smith, M. Jurczak, D. Pramanik and X. Xu, *International Conference on the Simulation of Semiconductor Processes and Devices (SISPAD)* **2005**, 143 (2005).