Uniaxial-Process-Induced Strained-Si: Extending the CMOS Roadmap

Scott E. Thompson, *Senior Member, IEEE*, Guangyu Sun, Youn Sung Choi, and Toshikazu Nishida

***Abstract*—This paper reviews the history of strained-silicon and the adoption of uniaxial-process-induced strain in nearly all high-performance 90-, 65-, and 45-nm logic technologies to date. A more complete data set of n- and p-channel MOSFET piezoresistance and strain-altered gate tunneling is presented along with new insight into the physical mechanisms responsible for hole mobility enhancement. Strained-Si hole mobility data are analyzed using six band $k \bullet p$ calculations for stresses of technological importance: uniaxial longitudinal compressive and biaxial stress on (001) and (110) wafers. The calculations and experimental data show that low in-plane and large out-of-plane conductivity effective masses and a high density of states in the top band are all important for large hole mobility enhancement. This work suggests longitudinal compressive stress on (001) or (110) wafers and $\langle 110 \rangle$ channel direction offers the most favorable band structure for holes. The maximum Si inversion-layer hole mobility enhancement is estimated to be ~ 4 times higher for uniaxial stress on (100) wafer and ~ 2 times higher for biaxial stress on (100) wafer and for uniaxial stress on a (110) wafer.**

***Index Terms*—CMOS, enhanced mobility, strained-silicon.**

I. Introduction

THE END OF simple scaling for a solid-state device technology is not new. Scaling of the bipolar junction transistors (BJT) ended in the 1990s for various reasons: voltage, base width, and power-density limits. Material changes could have been used to further improve BJTs; however, the industry moved to CMOS devices. Now, more than a decade later, conventional CMOS is reaching its scaling limits. However, this time, there is no new device to compete or potentially replace the industry work horse. Carbon nanotubes and silicon nanowires are lead contenders but have yet to achieve commercial success in even a niche logic market (a conceivable requirement one to two decades before becoming mainstream). With the need to maintain historical performance improvements, feature-enhanced Si CMOS is now recognized as the driver for the microelectronics industry. The key feature to enhance 90-, 65-, and 45-nm technology nodes is uniaxial-process-induced stress [1]–[8]. In this paper, we look at the history of strained-Si, the physics behind some strained-Si experimental data, and the state-of-the-art of strained-Si devices in commercial production.

Manuscript received July 6, 2005; revised December 5, 2005. This work was supported by the Applied Materials Foundation, Advanced Micro Devices (AMD), Cypress Semiconductor, IBM, Intel Foundation, Semiconductor Research Corporation (SRC), Texas Instruments, Taiwan Semiconductor Manufacturing Company (TSMC), United Microelectronics Corporation (UMC), and the National Science Foundation (NSF) under Grant No. ECS-0524316. The review of this paper was arranged by Editor H. Shang.

The authors are with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: thompson@ece.ufl.edu).

Digital Object Identifier 10.1109/TED.2006.872088

II. History of Strained-Silicon

The origin of strained-Si to improve CMOS devices can be traced to thin Si layers grown on relaxed silicon–germanium (SiGe) substrates in the 1980s [9], [10]. The thin Si layer takes the larger lattice constant of the SiGe and creates biaxial tensile stress. Wafer-based substrate strain was experimentally and theoretically studied by a large number of researchers for two decades [11]. In the 1990s, two other strained-Si activities started based on process-induced strain. First, high-stress capping layers deposited on MOSFETs were investigated as a technique to introduce stress into the channel [12], [13]. Second, Gannavaram *et al.* [14] proposed SiGe in the source and drain area for higher boron activation and reduced external resistance. It was this embedded SiGe literature that prompted Intel [3] to evaluate the technology, which resulted in larger than expected device performance enhancement, which, after considerable internal debate, was later attributed to uniaxial compressive channel stress [15]. Still, neither biaxial nor uniaxial stress was immediately adopted in CMOS logic technologies for several reasons. Biaxial stress suffers from defects and performance loss at high vertical electric fields [16]. Process-induced stress requires different stress types (compressive and tensile for n- and p-channel, respectively) to simultaneously improve both n- and p-channel devices. However, inside Intel and in the industry, strain was becoming recognized as offering the best potential to enhance performance in sub-100-nm process technologies (significantly larger performance gain than high-κ gates, fully depleted silicon-on-insulator (SOI), or multi-gate devices). The only debate was on the best path to take [17] (biaxial substrate versus uniaxial-process-induced stress).

Careful analysis of the 1990's biaxial and uniaxial strained-Si experimental data suggested that the industry adopt process-induced uniaxial strain. The key observations are as follows. First, uniaxial (versus biaxial) stress provides significantly larger hole mobility enhancement at both low strain and high vertical electric field due to differences in the warping of the valence band under strain [18]. Large mobility enhancement at low strain is important since yield loss via dislocations occurs at high strain. Second, uniaxial (as compared to biaxial) stress enhanced mobility provides larger drive current improvement for nanoscale short-channel devices. This results since the uniaxial stress-enhanced electron and hole mobility arises mostly from reduced conductivity effective mass (versus reduced scattering for biaxial stress), since uniaxial shear stress provides

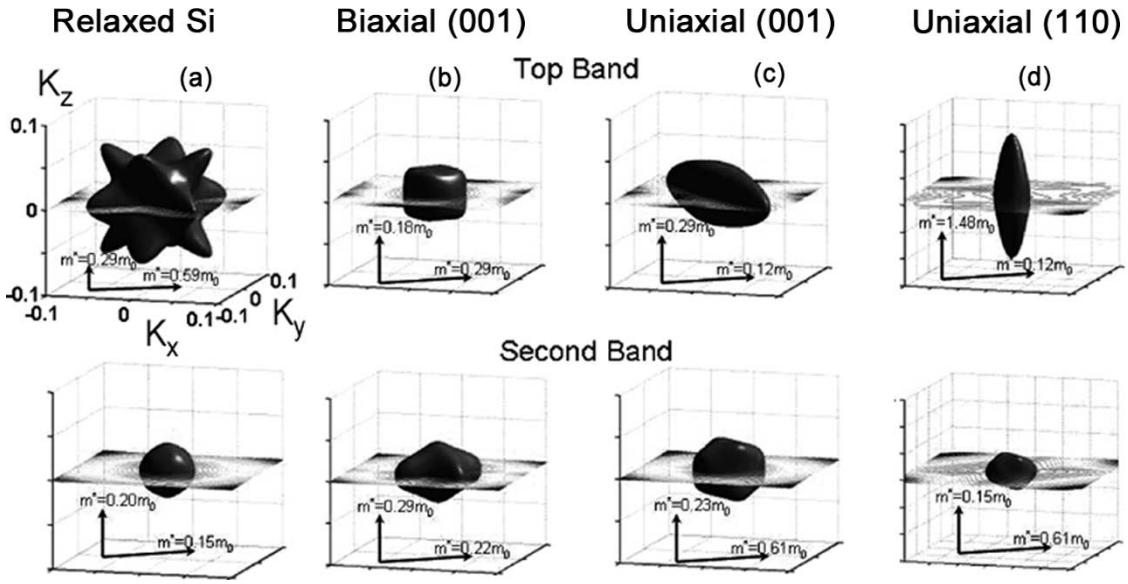


Fig. 1. Hole constant-energy band surfaces for the top band obtained from six-band $k \bullet p$ calculations for common types of 1-GPa stresses: (a) unstressed, (b) biaxial tension, (c) longitudinal compression on (001) wafer, and (d) longitudinal compression on (110) wafer (note significant differences in stress induced band warping altering the effective mass).

significant valence and some conduction band warping. Lastly, process-induced uniaxial stress causes approximately five times smaller n-channel threshold voltage shift. Since any threshold voltage shift needs to be retargeted by adjusting channel doping (for industry standard poly-Si gate devices on bulk or partially depleted SOI), the larger threshold voltage shift for wafer substrate-induced biaxial tensile stress causes approximately half of the stress-enhanced electron mobility to be lost [19]. Rarely is the stress-induced threshold voltage shift taken into account in the biaxial tensile-stress mobility data.

With advantages for process-induced uniaxial strain understood, commercial adoption into the 90-nm technology node soon followed. Two process flows were developed that independently target the stress magnitude and direction on n- and p-channel transistors. The first involved embedded and raised SiGe in the p-channel source and drain and a tensile capping layer on the n-channel device. The second uses dual stress liners: compressive and tensile Silicon Nitride (SiN) for p- and n-channel devices, respectively. Since both techniques provide large product level benefits at low cost, process-induced stress is present in nearly all high-performance logic technologies at the 90-, 65-, and 45-nm technology nodes for both microprocessor and consumer products [2], [4], [5], [8], [20]–[24]. The industry is now looking at combining various process stressors, such as compressive SiN layers, embedded SiGe, and tensile-stress shallow-trench isolation [25]. Performance gains from the various uniaxial stressors are expected to be mostly additive [1], [26].

III. Physics and Experimental Data

When deciding on a strained-Si process flow, it is first necessary to comprehend the potential magnitude for electron versus hole mobility enhancement and whether the mobility enhancement results from reduced conductivity effective mass or scattering. Since the valence-band dispersion relationship for semiconductors depends on nearest neighbor atomic spacing, certain stress (in particular shear stress) warps the valence bands (although less so for conduction band but some warping for shear stress) [27]. The warping of the valence band provides dramatic changes to the constant-energy surfaces in k space and can lead to large hole mobility enhancement via reduced conductivity mass in the channel direction. Mobility enhancement via reduced mass (as opposed to reduced scattering) is key in nanoscale MOSFETs and often not appreciated. Only mobility enhancement from reduced mass (unlike reduced scattering) is maintained at the very short 15–20-nm channel lengths (35-nm gate length) devices currently in production [1]–[8]. A strained-Si flow, which is scalable for multiple technology nodes, thus, needs to focus on reducing the hole conductivity mass with the goal of improving the n/p ratio from ~ 2 to ~ 1. Therefore, in this section, we will focus on strain-enhanced hole mobility from reduced conductivity mass.

As a starting point, it is helpful to visualize the effect of strain on the valence-band constant-energy surfaces in k space for bulk Si. Fig. 1 shows the surfaces obtained using six band $k \bullet p$ and band parameters in [28]. The strain-altered surfaces for the top two bands are shown at 1 GPa for the common stresses of interest: longitudinal compression on (001) [4], [5], [20], [22]–[24], [29] and (110) hybrid wafer orientation [23] and biaxial tensile stress [30]. Note from the constant-energy surfaces in Fig. 1, the heavy and light hole bands lose their meaning and we label the bands (first, second, etc.) in this paper. Some important differences in the band structure under the various stresses at 500 MPa are summarized in Fig. 2 for the in-plane and out-of-plane conductivity effective masses and density of states at the band edge. We will refer to Fig. 2 in the next section during analysis of experimental data.

Before covering strain-altered hole mobility calculations, we will briefly cover a qualitative model for strained-enhanced electron mobility since the concepts are similar for electrons and holes. The important concepts to understand are

Stress	wafer	m^*_{110}/m_0 top / 2 nd	m^*_z/m_0 Top / 2 nd	$m^{2D}{}^*_{DOS}/m_0$ Top / 2 nd
Uniaxial Compression	(001)	0.13 / 0.56	0.28 / 0.22 <001>	0.34 / 0.30
Uniaxial Compression	(110)	0.13 / 0.56	0.89 / 0.16 <1-10>	0.19 / 0.35
Biaxial Tension	(001)	0.28 / 0.22	0.18 / 0.29 <001>	0.28 / 0.22

Channel direction ⊥SiO_2

$$m^* = \hbar^2 / \frac{d^2\varepsilon}{dk^2} \qquad m^{2D}{}_{DOS} = (m_x m_y)^{1/2}$$

Fig. 2. Summary of key valence-band parameters for top and second band for bulk Si under 500-MPa stress. The conductivity and density of states effective mass is listed at gamma point. Uniaxial compression is longitudinal along $\langle 110 \rangle$ channel direction (note significant differences for in-plane, out-of-plane, and density-of-states masses).

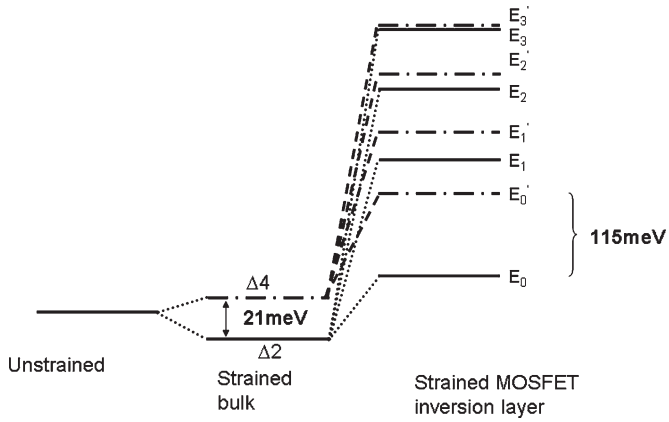


Fig. 3. Conduction valley energy-level splitting under 500 MPa of longitudinal uniaxial tensile stress: Bulk and MOSFET inversion layer (1 MV/cm). Note that energy-level splitting from inversion-layer confinement is larger than strained.

strain-induced energy-level splitting, inversion-layer quantum-confinement energy-level shifts, average mass change due to repopulation and band warping, two-dimensional (2-D) density of states, and interband scattering changes due to band splitting. All of these will be discussed in the following sections. A simple qualitative model is now presented to gain insight and to understand the more complex mathematics used elsewhere [16] and later in this paper. The electron mobility in bulk-strained-Si along $\langle 110 \rangle$ direction is determined by occupation and scattering in the Δ_2 and Δ_4 valleys and can be expressed as

$$\mu_{\text{eff}} = q \frac{\left(\tau_{\Delta 2} \frac{n_{\Delta 2}}{m_t^*} + \tau_{\Delta 4} \frac{n_{\Delta 4}}{m_l^*} \right)}{(n_{\Delta 2} + n_{\Delta 4})} \tag{1}$$

where q, n, τ, and m are the electron charge, concentration, relaxation time, and conductivity mass in the MOSFET channel direction, respectively. Strain improves the mobility by increasing the electron concentration in the Δ_2 valley. The repopulation improves the average in-plane conductivity mass (unstressed: $m_t = 0.19m_0$ versus $m_l = 0.98m_0$) and some further improvement is possible for stresses that warp the conduction valleys and lower m_t [27]. Reduced intervalley scattering by the strain-induced splitting between Δ_2 and Δ_4 plays some role (enhances long channel mobility) when the splitting becomes comparable or larger than the optical phonon energy.

In addition to a low in-plane mass, a high out-of-plane mass for the Δ_2 valley electron is equally important since carrier motion perpendicular to the SiO_2 interface (taken as the z-direction in this paper) is quantized. This quantization in addition to strain alters the position of the energy levels. The quantization leads to bands becoming subbands since only discrete wave vectors k_z are allowed. Including quantization, the total inversion-layer electron energy is given by discrete values of energy (E_n) added to the electron energy in the x- and y-directions (in the plane of the MOSFET) [31]

$$E = E_n + \frac{\hbar k_x^2}{2m_x} + \frac{\hbar k_y^2}{2m_y}. \tag{2}$$

Each step in energy is called a subband with E_n the energy of the bottom of the subband. As an example, self-consistent solution of Schrödinger and Poisson equation for 500 MPa of uniaxial tensile stress and an inversion-layer vertical field of 1 MV/cm gives the energy levels, as shown in Fig. 3. Since the subband separation is greater than kT, nearly all the electrons in most cases occupy the bottom two subbands [ground state $n = 0$ typically called E_o (from Δ_2) and $E_{o'}$ (from Δ_4)]. The ground state energy is significantly lower for the Δ_2 valleys because of the higher quantization mass ($\Delta_2 : m_z = 0.98m_0$ versus $\Delta_4 : m_z = 0.19m_0$) which leads to increased splitting between the bottom two subbands and confinement and strain splitting being additive (for the common biaxial and uniaxial tensile stress). Note, the strong confinement in an MOSFET shifts the energy levels more than the moderate $\sim$ 500-MPa stress typically used in present-day production logic technologies. Thus, a high out-of-plane mass in the bottom subband (top subband for holes) is an important requirement for the strain-altered band structure.

Lastly, in addition to a low in-plane and high out-of-plane effective mass, a high in-plane mass perpendicular to the channel direction is also important. The density of states per unit area for the quantized system is $(2/(2\pi)^2)(\sqrt{m_x m_y}/m_o) dk_x dk_y$, which results in the density-of-states mass approximated by $m_{\text{DOS}}^{2D} = \sqrt{m_x m_y}$. Though strain does not significantly alter the electron subband density of states, as discussed next, a high m_{DOS}^{2D} will be shown to be important for maintaining a hole concentration in the top subband.

Similar to strained-enhanced electron mobility, hole mobility in an inversion layer can qualitatively be described as resulting from occupation and scattering in the top two bands

$$\mu_{\text{eff}} = q \frac{\left(\tau_{\text{top}} \frac{p_{\text{top}}}{m^*_{\text{top},110}} + \tau_{2\text{nd}} \frac{p_{2\text{nd}}}{m^*_{2\text{nd},110}} \right)}{(p_{\text{top}} + p_{2\text{nd}})}. \tag{3}$$

However, hole transport is more complicated since strain significantly warps the valence band (as seen in Fig. 1) altering both the in- and out-of-plane mass and m_{DOS}^{2D}. Further, the mass changes with stress and is not constant in k space. After

the previous discussion on strain-enhanced electron transport, an advantageous strain for holes needs to warp the valence band to create both a low in-plane and high out-of-plane mass and, if possible, a large mass in the plane of the MOSFET perpendicular to the channel direction (creates a large m_{DOS}^{2D}). Band calculations and measurements to be discussed next show that uniaxial stress warps the valence band creating most of these features. In this paper, the strain-altered band structure is calculated using six band $k \bullet p$, including quantum confinement via a self-consistent solution of Schrödinger and Poisson equation [32], [33]. The mobility is calculated by a linearization of the Boltzman transport equation. The numerics confirm that the simple qualitative model captures much of the essential physics for understanding the physical mechanisms for mobility enhancement.

However, a note of caution regarding strained-Si modeling is in order. Historically, it has been difficult to predictively model strain-enhanced mobility due to the uncertainty in the inversion-layer scattering parameters [16], [34], [35] and numerical complexity that forces drastic early analytic approximation [16]. Using the conventionally accepted scattering parameters, electron and hole mobility enhancements are significantly under- and overpredicted, respectively [16], [35]. Also, the technologically important difference in the field dependence of the hole mobility enhancement for uniaxial compression and biaxial tensile stress was not predicted but first observed experimentally [18]. As a result, in Sections III-A–D, we use both modeling and experimental data to draw insight.

A. Mobility Enhancement at Low Strain

Because of the previously discussed modeling difficulties, the most effective approach at predicting and understanding strain-enhanced electron and hole mobility for the industry [16], [18] has been the empirically measured piezoresistance coefficients. To date, bulk piezoresistance coefficients have primarily been used even though differences are expected for a MOSFET resulting from inversion-layer quantization. Piezoresistance coefficients (generally extracted at low strain and low field) have the added benefit of capturing mobility enhancement primarily resulting from changes in conductivity mass. Fig. 4(a) shows a more complete set of piezoresistance coefficients extracted on industrial long n- and p-channel MOSFETs for cases of technological importance to the semiconductor industry: $\langle 110 \rangle$ and $\langle 100 \rangle$ channel orientations on (001) and (110) wafers. Since process strain is typically introduced longitudinal or perpendicular to the channel, the mechanical stress effect on the mobility is expressed as follows: $\Delta\mu/\mu \approx |\pi_{\|}\sigma_{\|} + \pi_{\perp}\sigma_{\perp}|$. $\Delta\mu/\mu$ is the fractional change in mobility, $\sigma_{\|}$ and $\sigma_{\perp}$ are the longitudinal and transverse stresses, and $\pi_{\|}$ and $\pi_{\perp}$ are the longitudinal and transverse piezoresistance coefficients expressed in Pa^{-1}, respectively. $\pi_{\|}$ and $\pi_{\perp}$ can be expressed in terms of the three fundamental cubic piezoresistance coefficients π_{11}, π_{12}, and π_{44} for a (001) wafer and five coefficients for a (110) wafer. For comparison, using the piezoresistance coefficients for bulk [36], and surface inversion layer extracted in this paper for (001) wafer and [37] for surface inversion layer on (110) wafer, Fig. 4(b) plots

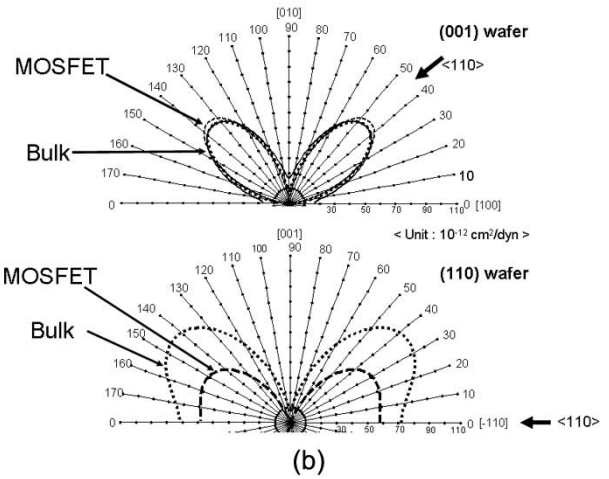
< Unit : 10^{-12} cm²/dyn >

(001) wafer	<100>		<110>	
Polarity	$\pi_{\|\|}$	$\pi_{\perp}$	$\pi_{\|\|}$	$\pi_{\perp}$
	π_{11}	π_{12}	$(\pi_{11}+\pi_{12}+\pi_{44})/2$	$(\pi_{11}+\pi_{12}-\pi_{44})/2$
n - MOSFET	-42.6/ -102	-20.7/ 53.4	-35.5 / -31.6	-14.5 / -17.6
p- MOSFET	9.1/ 6.6	-6.2/-1.1	71.7/ 71.8	-33.8 / -66.3

MOSFET / bulk

(110) wafer	<110>	
Polarity	$\pi_{\|\|}$	$\pi_{\perp}$
	$\pi_{22} = (\pi_{11}+\pi_{12}+\pi_{44})/2$	π_{12}
p- MOSFET	58.0/ 71.8	38.2 / -1.1

(a)



(b)

Fig. 4. (a) Measured long p- and n-channel MOSFET piezoresistance coefficients for (001) and (110) wafers compared to bulk Si piezoresistance. Note the larger difference for bulk and MOSFET with even the sign being different in some cases. (b) Measured longitudinal pMOSFET piezoresistance coefficient versus channel direction for (001) (this paper) and (110) [37] wafers. Note similar magnitude for piezoresistance for both wafers along $\langle 110 \rangle$. Bulk piezoresistance coefficients are shown for comparison.

the longitudinal and transverse piezoresistance coefficients for various channel directions.

A few interesting observations can be made from the data in Fig. 4(a) and (b). For the $\langle 110 \rangle$ channel, the bulk versus MOSFET coefficients are close but do vary up to $\sim$ 50% for the values extracted in this and other works [18], [38]. In some cases, large differences both in magnitude and sign occur for bulk versus MOSFET piezoresistance (for example nMOSFET $\pi_{\perp}$ for the $\langle 100 \rangle$ channel direction). The large differences result, particularly for an nMOSFET, since the quantization splitting is large (see Fig. 3) and alters the conductivity mass change that occurs for strain induced repopulation in a nMOSFET versus bulk n-Si. The hole piezoresistance coefficient is largest for longitudinal compression along $\langle 110 \rangle$ for both (001) and (110) wafers consistent with the industry primarily using $\langle 110 \rangle$ channel direction in strained-Si technologies [4], [22], [29]. The hole piezoresistance coefficients can be qualitatively understood from the band warping in Fig. 1 and (3). The large piezoresistance for longitudinal compression stress on (001) and (110) wafers with $\langle 110 \rangle$ channel results from hole repopulation into a top band with a very small mass

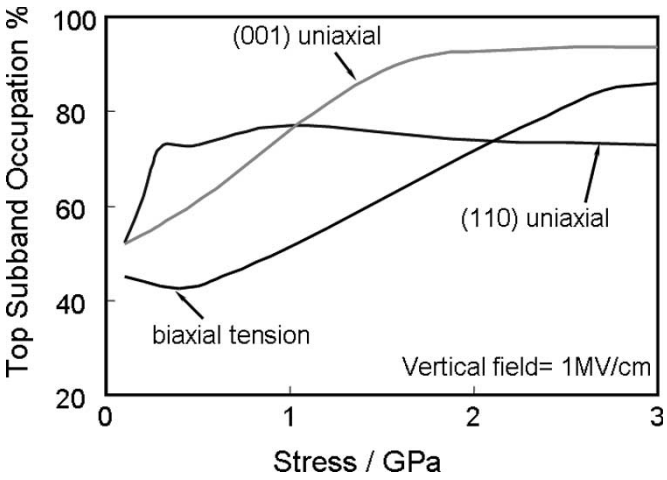


Fig. 5. Hole occupancy in top band versus stress (for three types of stress: Longitudinal compression on (001) and (110) wafers and biaxial tension). Low top-band hole occupation at larger stress for compression on (110) wafer is due to the low density of states in top subband.

($\sim 0.12m_o$ at 1 GPa). Biaxial tensile stress is less effective at enhancing hole mobility since the mass in the top band is 40% larger. An important question for the industry is the additivity of strain enhancement and the higher unstressed mobility on hybrid (110) orientated wafers. The data in Fig. 4 suggest (at least at low strain ($<$ 100 MPa) where the data in Fig. 4 is measured) that the mobility enhancement from strain and (110) orientation are mostly (but not fully) additive since the piezoresistance coefficient is $\sim$ 20% smaller on a (110) versus (001) wafer. The slightly lower piezoresistance for a (110) wafer partly results from a low density of states in the top subband due to the small in-plane mass perpendicular to the channel direction, as seen in Fig. 1 and discussed next.

To maintain a large hole population in the top band, both band splitting (by strain and confinement) and a large density of states in the top band are important. The importance of a large density of states is commonly not appreciated in some of the novel low density of states narrowband gap III-V materials [39] and carbon nanotubes [40] since not just a large mobility but a large inversion charge density is also required for large current drive. Fig. 5 plots the percentage of holes in the top band for the various stresses. Though band splitting is larger for uniaxial compression on (110) versus (000) wafers (due to confinement), the top band for the (001) wafers is more heavily populated at high stress due to its larger density of states. For completeness, at low stress, the top band on the (110) wafer is more heavily populated because of the large subband spliting due to confinement on this surface. The low density of states in k_x, k_y on (110) wafer can be inferred from the three-dimensional (3-D) constant-energy surfaces in Fig. 1. Since confinement also affects the density of states, constant-energy contours for the various 1-GPa stresses including z-direction confinement at a typical 1-MV/cm silicon vertical field are plotted in Fig. 6. The 2-D density of state mass is extracted from the constant-energy contours and shown in Fig. 7 as a function of stress. As seen in Fig. 7, approximately three times higher 2-D density of states occurs for compressive stress on (001) versus (110) wafers. The magnitude of the density of states

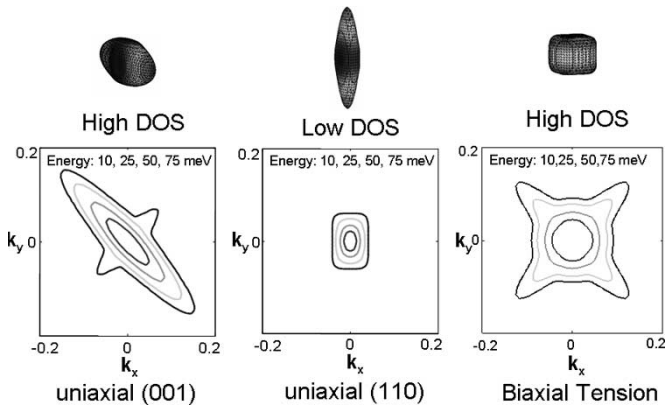


Fig. 6. 2-D constant-energy contours for the top band using six-band $k \bullet p$ calculations including quantum confinement of 1 MV/cm. Applied stress is 1 GPa in $x-y$ plane for biaxial and compressive along $\langle 110 \rangle$ channel direction for uniaxial. Similar to the constant-energy surfaces in Fig. 1, plots show the lowest 2-D density of states in k_x, k_y (inversion layer) for (110) wafer.

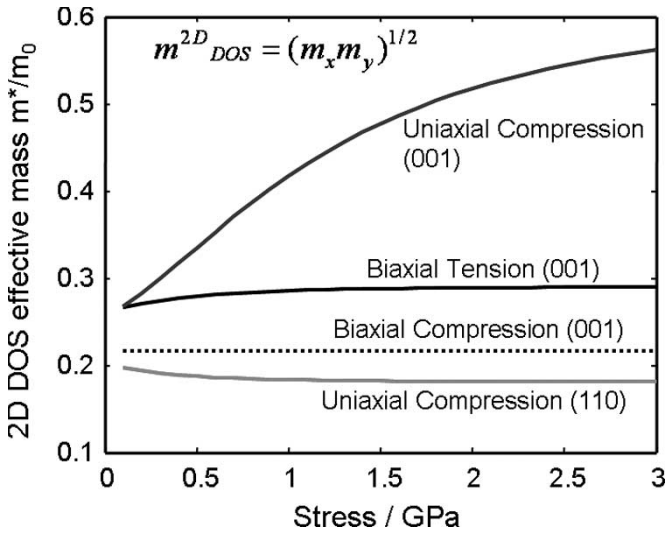


Fig. 7. Comparison of 2-D density-of-states mass for the top subbands for different stresses versus stress. Note that the uniaxial compression on (110) wafer has a very low density of states for the top subband.

will be shown in Section III-D to play an important role in the maximum mobility enhancement.

B. Stress-Altered Gate-Leakage Current

In addition to strain altering the in-plane mass and enhancing electron and hole mobility, strain also alters the out-of-plane mass and SiO_2/Si barrier height, which changes the gate tunneling current. The tunneling-current modulation by uniaxial stress has been reported in [41]. The out-of-plane mass is altered by band warping and/or repopulation and plays an important role [42]–[44] since it affects the tunneling probability. To understand the strain-altered out-of-plane mass, the strain-altered gate leakage is measured on n- and p-type MOSFETS at 1.0 V for the commonly used stresses (shown in Fig. 8). In Fig. 8, the gate current decreases for tensile and compressive stress for electrons and holes, respectively. Decreased gate leakage is observed for stresses that increase the population in a subband with an increased out-of-plane mass.

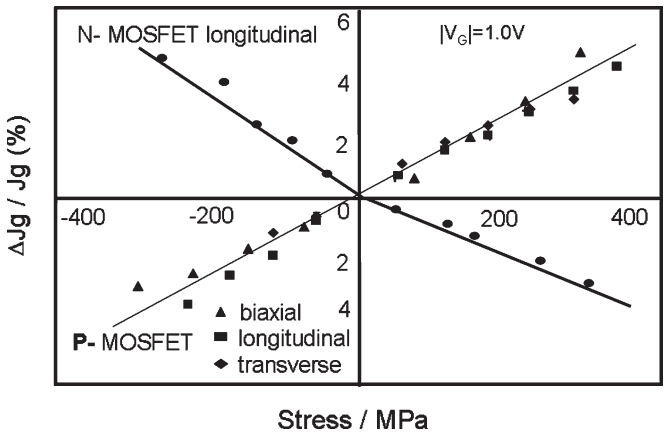


Fig. 8. Change in n- and p-MOSFET direct tunnel gate-leakage current versus stress. Note all types of compressive stress (biaxial, longitudinal, and transverse) decrease the hole tunneling current. Tensile stress decreases electron tunneling current. Industry has implemented stresses that decrease tunnel current since this provides mobility enhancement at high vertical fields.

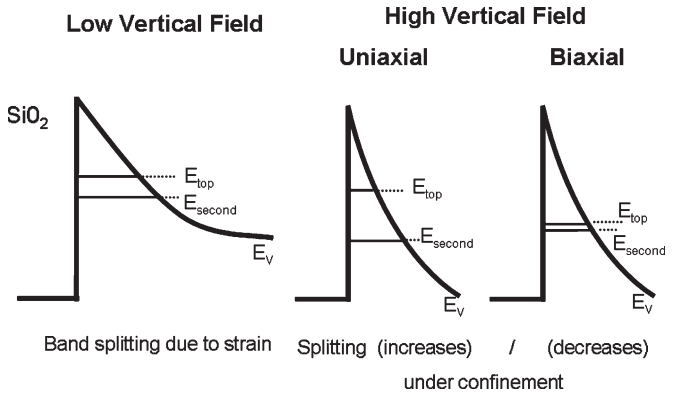


Fig. 9. Simplified schematic of valence-band splitting of strained-Si as a function of gate overdrive. Note that the net band splitting from strain and confinement is additive for uniaxial compressive stress but subtractive for biaxial tensile stress.

For examples, 1) biaxial and uniaxial tensile stresses increase the electron population in the Δ_2 valleys that have a high out-of-plane mass resulting in reduced electron tunneling current; 2) uniaxial compressive stress decreases the hole tunneling by increasing the population of holes with a higher out-of-plane mass for both (100) and (110) wafers (see Fig. 2); and 3) conversely, biaxial tensile stress creates a low out-of-plane mass for the top valence subband, which results in increased hole tunneling current (see Fig. 2). Interestingly, the gate-leakage current is observed to decrease for the types of stresses adopted by the industry (tensile and compressive stress for n- and p-MOSFETs, respectively). One reason to be discussed next is the additivity of strain and confinement splitting since the additivity and reduced leakage both depend on a large out-of-plane mass in the top subband for holes and bottom subband for electrons.

C. Mobility Enhancement at High Vertical Fields

In a MOSFET, the 2-D surface confinement in the inversion layer also shifts the valence bands and the conduction valleys [16], [45], [46]. Whether the confinement-induced shift adds to or reduces (cancels) the strain-induced splitting simply depends on the magnitude of the out-of-plane masses (valence-band splitting is more complicated but this simple model captures the essential physics) [47]. Bands or valleys with a "light" out-of-plane mass will shift more in energy relative to bands with a "heavy" mass (similar to the increasing ground state energy of a quantum well as the particle mass decreases). Hence, when the top most occupied band (or valley) has a lower out-of-plane mass compared to the next occupied band, the splitting is reduced or lost with surface confinement. Fig. 9 pictorially shows the valence-band energy-level shift with confinement for both uniaxial and biaxial stress. $E_{\rm top}$ represents the top band with large out-of-plane mass for uniaxial stress and small for biaxial stress (relative to the second band with masses given in Fig. 2). Hence, the top band will have a small shift in energy due to confinement for uniaxial stress but large shift for biaxial stress. $E_{\rm second}$ represents the second band. As seen in Fig. 9, the stress-induced band splitting $(E_{\rm top} - E_{\rm bottom})$ increases for uniaxial stress but decreases for biaxial tensile stress. Thus, although strain favors occupation of the top band for both types of stresses, confinement favors occupation of the top band for uniaxial compressive stress and the second band for biaxial tensile stress. The net band splitting from strain and confinement is additive for uniaxial compressive stress but subtractive for biaxial tensile stress. The competing effects of strain and surface confinement on the band splitting is the reason for the loss in mobility enhancement in biaxially strained-silicon p-MOSFETs at high electric fields. The undesirable light out-of-plane mass created by biaxial tensile stress occurs in other material systems, such as Ge and III-V materials, and presents a fundamental problem in using this type of strain in inversion-layer MOSFETs (dominant device type due to superior scaling properties).

The above qualitative discussion uses bands and out-of-plane masses and is presented only to help understand the physics. Correct physical treatment requires self-consistent solution to Schrödinger's and Poisson's equations to calculate the subband energy shifts in the confined MOSFET inversion layer. To show that the simple qualitative model captures the correct physics, we quantify the above discussion with quantum–mechanical calculations and confinement-induced subband splitting calculated in three ways: one [31], four [28], and six [16] band models with Schrödinger's equation. Fig. 10 shows how the splitting at a fixed stress of 500 MPa is altered versus the vertical electric field (expressed as hole concentration). For all models, surface confinement and strain band splitting are additive for uniaxial compressive stress but subtractive for biaxial tensile stress.

D. Maximum Strained Enhanced Hole Mobility

As previously discussed, the best quantitative predictor for strain-enhanced mobility comes from the piezoresistance coefficients. However, as higher levels of stress are integrated into production logic technologies, caution in using piezoresistance coefficients is needed since piezoresistance should not be expected to vary linearly with stress above $\sim$ 250–500 MPa. When evaluating the various stress options, it is also important to comprehend the maximum possible mobility enhancement.

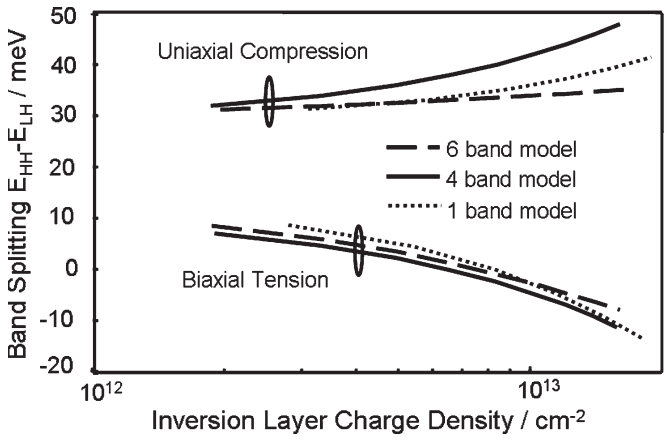


Fig. 10. Valence energy band splitting calculated using three different models versus inversion charge density for longitudinal compression and biaxial tension stress. Note that all models show the net band splitting from strain, and confinement is additive for uniaxial compressive stress but subtractive for biaxial tensile stress.

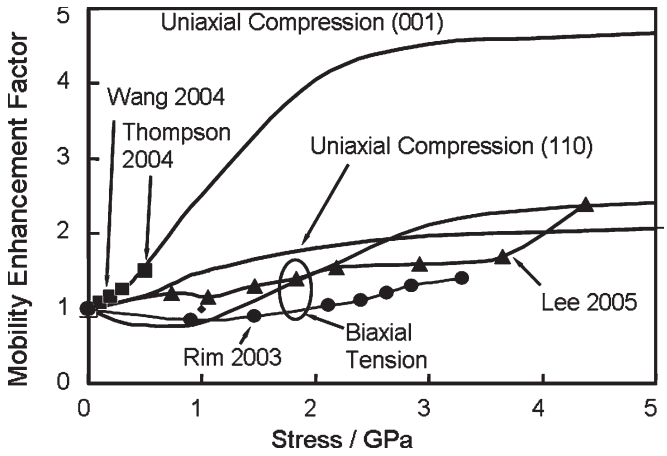


Fig. 11. Calculated and experimental data for longitudinal compressive and biaxial tensile-stress-enhanced mobility versus stress (Biaxial stress $= \sigma_X + \sigma_Y$). Note that the maximum predicted Si inversion-layer hole mobility enhancement is estimated to be ~ 4 times higher for uniaxial stress on (100) wafer and ~ 2 times higher for biaxial stress on (100) wafer and for uniaxial stress on a (110) wafer.

This section provides such a prediction for uniaxial-process-induced stress on (001) and (110) wafers relying heavily on experimental data due to the uncertainty in scattering rates. To date, unlike biaxial stress [48], limited data exist for the maximum mobility possible for uniaxial stress. In this paper, we used a set of scattering parameters that fit the experimental data for hole mobility enhancement under biaxial tensile stress [16]. The calculations include acoustic and optical phonon and surface roughness scattering. This set of scattering parameters shows that the dominant mechanism responsible for biaxial tensile-stress mobility enhancement (at large stress) is reduced optical phonon scattering. Acoustic phonon scattering is only slightly altered due to the changes in the density of states. Surface roughness scattering is slightly changed by stress but uncertainty exists in the literature [16], [49], [50], and more work is needed especially for the (110) substrate. The calculations, in this paper, for biaxial stress are consistent with the previous work [16], although in this paper, Schrödinger's and Poisson's equations are solved self-consistently. The model fit to the biaxial tensile-stress experimental data is shown in Fig. 11.

Using the same scattering parameters, mobility enhancement for uniaxial stress on (001) and (110) wafers is calculated, as shown in Fig. 11, and compared to uniaxial stress data from [11], [22], and [38]. The mobility calculations use the full six-band subband structure and Kubo–Greenwood linearization of the Boltzmann equation [16]. Where data exist (0 to ~ 600 MPa for uniaxial stress), the model shows good agreement. The maximum predicted Si inversion-layer hole mobility enhancement is estimated to be ~ 4 times higher for uniaxial stress on (100) wafer and ~ 2 times higher for biaxial stress on (100) wafer and for uniaxial stress on a (110) wafer. The larger maximum mobility enhancement on a (001) wafer results from the high density of states in the top band, as discussed previously but scattering differences also play a role. Scattering differences for various substrate orientations and stresses should be expected as captured in analytical scattering expressions.

First, for the acoustic phonon in the 2-D inversion layer, the scattering time $\tau_{\rm ac}$ is expressed as [16], [49], [51]

$$\frac{1}{\tau_{\rm ac}} = \frac{D_{\rm ac}^2 b_{\rm mn} m_{\rm DOS}^{2D} k_B T}{\hbar^3 \rho u_l^2} \propto \left(m_{\rm DOS}^{2D}\right) \tag{4}$$

where $D_{\rm ac} = 3.1$ eV [52] is the acoustic deformation potential constant of the valence band, $m_{\rm DOS}^{2D}$ is the density-of-state effective mass, ρ is the density, and u_l is the longitudinal sound velocity. The constant $b_{\rm mn} = \int_0^w dz |\psi_k^m(z)|^2 \cdot |\psi_k^n(z)|^2$ is the form factor that defines the transition from initial state m to final state n, and $2/b_{mm}$ represents the effective well width for the m-th subband [50]. Since the acoustic phonon energy is very small compared to the subband splitting, the acoustic phonon scattering mainly occurs via intraband scattering. Thus, stress-induced band splitting only weakly affects the acoustic phonon scattering time [18], [53]. As seen from (4), an increased density-of-states will decrease the acoustic phonon scattering time, which is proportional to $m_{\rm DOS}^{2D}$. For uniaxial stress on a (100) wafer with a high density of states in the top band, this slight negative effect on mobility (at least for uniaxial stress on (001) wafer) is offset by the high hole density in the top band having a light conductivity mass in the channel direction.

Second, the optical phonon scattering time $\tau_{\rm op}$ is [18], [50], [51], [53]

$$\frac{1}{\tau_{\rm op}} = \frac{m_{\rm DOS}^{2D} b_{\rm mn} D_{\rm op}^2}{2\rho\omega_0\hbar^2} \times \left[N_q \times \frac{1 - f(\varepsilon + k_B\Theta - \Delta E)}{1 - f(\varepsilon)} + (N_p + 1) \times \frac{1 - f(\varepsilon - k_B\Theta - \Delta E)}{1 - f(\varepsilon)} \right] \tag{5}$$

where ΔE is the band splitting energy, $D_{\rm op} = 10.5 \times 10^8$ eV/cm [54] is the optical deformation potential constant of the valence band, $f(\varepsilon)$ is Fermi–Dirac distribution function at energy ε, $\Theta = 735$ K is the Debye temperature, and

$k_B\Theta = \hbar\omega_0 = 63$ meV is the optical phonon energy [51], and $N_q = [\exp(\hbar\omega_0/k_BT) - 1]^{-1} = [\exp(\Theta/T) - 1]^{-1}$ is the number of phonons from Bose–Einstein statistics. From (5), strain does not significantly alter hole intervalley scattering until the subband splitting (ΔE) is greater than the optical phonon energy, $k_B\Theta = \hbar\omega_0 = 63$ meV. For subband splitting larger than the optical phonon energy, greater than 1 GPa of stress is required on the (100) wafer since strain induced valence-band splitting is less than the conduction band. Also, the correlation between the topmost two subbands $b_{\rm mn}$ under compressive uniaxial stress (as compared to biaxial tensile stress) is smaller due to the higher band splitting (strain and confinement being additive). This causes the scattering rate to be less for compressive uniaxial than biaxial stress. Lastly, in addition to strain splitting the band, a high out-of-plane mass (for top versus second band) causes large subband splitting and reduces the interband optical phonon scattering rate. This is an important factor in the reduced scattering for (110) versus (001) surface devices.

Third, the surface roughness scattering relaxation time $\tau_{\rm sr}$ [16], [50] can be expressed as

$$\frac{1}{\tau_{\rm sr}} = \frac{q^2 E_{\rm eff}^2 m_{\rm DOS}^{2D}}{2\pi\hbar^3} \int_0^{2\pi} S(q)\,(1 - \cos(\theta))\,d\theta \qquad (6)$$

where $E_{\rm eff}$ is the transverse effective electric field in the inversion layer and $S(q) = \pi L^2\Delta^2/[1 + (q^2L^2/2)]^3$ is the power spectrum of the roughness at the interface. L is the correlation length ($L = 2.6$ nm) and Δ is the average step height ($\Delta = 0.4$ nm). Differences in $\tau_{\rm sr}$ for various stresses and substrates result from changes in the density-of-states and location of the inversion layer charge from the SiO_2 interface. There is also a fair amount of uncertainty in surface roughness scattering particularly on a (110) wafer since the commonly used universal mobility versus effective oxide field $E_{\rm eff}$ applies only to the (100) substrate [55], [56]. However, one can conclude since the effective well width depends heavily on the out-of-plane effective mass for each subband, the top subband for a (110) devices, having a very large out-of-plane effective mass (see Fig. 2), will lead to carriers significantly closer to the interface and greater surface roughness scattering.

IV. State-of-the-Art Strained-Silicon Transistors

This section describes three techniques used in commercial 90- and 65-nm logic technologies to introduce uniaxial stress into the Si channel. The techniques in production include high-stress tensile and compressive SiN capping layers and selective epitaxial SiGe deposited in recessed/raised source and drains. Future techniques for process stress or mobility enhancement include tensile shallow trench oxide, embedded SiC for n-MOSFETs and hybrid orientated (110) wafers [57]. One key scaling advantage of process stress is the increasing channel stress for decreasing channel length. Fig. 12 shows the longitudinal channel stress for a fixed SiN layer versus channel length obtained using FLOOPS [58]. As seen in Fig. 12, reducing the channel length below $\sim$ 100–150 nm causes a dramatic increase

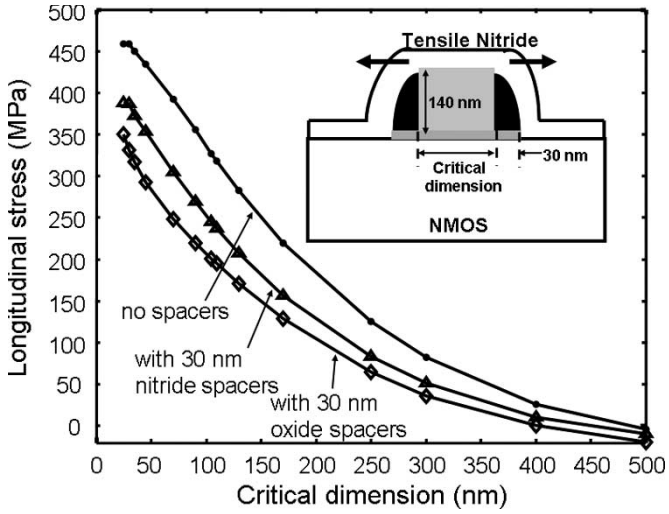


Fig. 12. Stress in the center of channel for 70-nm 1.6-GPa SiN versus gate length. Note that the process-induced-stress has the unique feature of increased stress as the channel length is scaled (reduced).

in the channel stress and helps process-induced stress scaling for several technology nodes after the first introduction at 90 nm [15], [22].

The cost to implement the first and second generation process stressors is low. Process strain only adds a few percent to wafer cost since comparatively few new process steps are added (typically existing steps are modified to introduce strain). Several process flows exist to introduce the epitaxial SiGe into a MOSFET [6], [22], [25]. The first consists of the steps shown in Fig. 13(a). The source/drains are etched creating a silicon recess. Next, SiGe (for p-channel) or SiC for n-channel is epitaxially grown in the source and drain. First generation embedded SiGe used $\sim$ 17% Ge to create $\sim$ 500 MPa of channel stress. Future generations bring the SiGe closer to the channel and will likely increase the Ge concentration [6], [21]. Locating the SiGe closer to the channel will require reduced midsection thermal cycles to prevent any boron or Ge out diffusion from the SiGe into the channel. To date, a maximum of $\sim$ 900 MPa of stress has been created with embedded SiGe, and impressive current improvements from 60%–90% have been demonstrated on short devices ($\sim$ 35 nm) [6], [21].

The second flow [6], [25] integrates the SiGe before the source and drain, which has some advantages. The SiGe is closer to the channel, which significantly increases the channel stress. The removal of the poly-Si gate hardmask without spacer loss also has a larger process window in this integration flow. For additional performance, the SiGe can be *in situ* doped with boron but this requires a low thermal cycle midsection to prevent boron out-diffusion.

One scaling concern for embedded SiGe is reduced SiGe volume in the source/drain [59] on future technology nodes starting with the 45-nm generation. Reduced SiGe volume makes it more difficult to increase the channel stress especially on short channel devices. However, even with reduced SiGe volume at the 45-nm node, depositing a high-compressive SiN on top of the SiGe will create > 1 GPa of channel stress and mobility enhancement $> 200\%$.

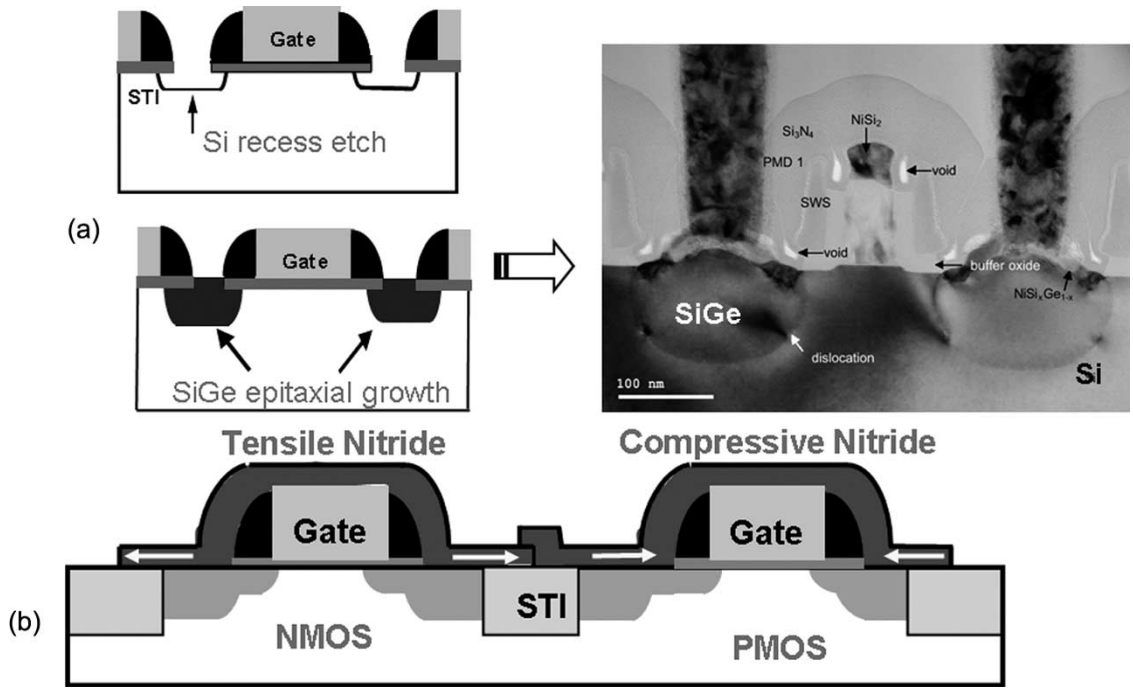


Fig. 13. (a) Strained-Si p-channel MOSFET process flow for the representative stacked gate transistor and transmission electron microscopy cross-sectional view (source: Chipworks) and (b) dual stress liner process architecture with tensile and compressive silicon nitride capping layers.

Instead of embedded SiGe, dual stress liners (tensile and compressive capping layers) [5] are also being widely adopted [2], [6], [7], [22]. The advantages of a dual stress liner flow over epitaxial SiGe are reduced process complexity and integration issues. Recent progress in increasing stress of SiN films to $\sim$ 3.0 GPa for compressive and $\sim$ 2.0 GPa tensile [1] increases the attractiveness of this option. The capping films are introduced either as a sacrificial layer before source and drain anneal [1], [5], [60] or as a permanent layer post-salicide [Fig. 13(b)]. With 2–3-GPa stress in the SiN, a comparable performance to the first generation SiGe has been demonstrated. The process flow consists of a uniform deposition of a high tensile SiN liner post-silicidation over the entire wafer followed by patterning and etching the film off p-channel transistors. Generally, a thin-etch stop layer is used under the liner to prevent any damage to the silicide. With highly selective etches, the etch stop layer can be $<$ 50 A, which only slightly degrades the stress transfer into the channel. Next, a highly compressive SiN layer is deposited, and this film is patterned and etched from n-channel regions. Design rules need to account that at n/p boundaries, the nitride stress layer is cut, which relaxes the stress at distances of approximately a few tenths of micrometers. Also, similar to the SiGe volume-scaling issue, as the space between the stacked gates decreases, it becomes harder to transfer stress into the channel with the stress capping layers, and higher-stress films are needed.

Capping layers can also introduce strain into the silicon channel via a stress memorization of the poly-Si gate [7]. In this approach, a highly tensile nitride capping layer acts as a temporary stressor. The flow consists of the following steps: 1) poly-Si gate amorphization; 2) deposition of a high-stress SiN layer on top of the poly-Si gate; 3) recrystallization of the poly-Si gate during source/drain anneal; and 4) removal of the SiN layer. After the removal of the poly-Si capping layer, some compressive stress remains in the vertical direction since the stress nitride prevents the poly-Si regrowth to expand upward. The vertical stress introduced via the poly-Si gate into the Si channel enhances N-channel transistor mobility by $\sim$ 10 % without degrading or enhancing pMOSFETs.

V. Conclusion

The physics and advantages of uniaxial-process-induced stress are becoming understood. Calculations and experimental data show that low in- and large out-of-plane conductivity effective masses, and a high density of states in the top band are all necessary for large hole and electron mobility enhancement. Paths to implement greater than 1 GPa of process stress have been identified and mobility enhancements of over 200% expected. Uniaxial process stress is making Si a high-mobility semiconductor with mobility competitive with unstrained III-V materials. With such large improvements in the channel mobility, the reducing external resistance will need to be an important focus at the 45-nm technology node and beyond.

References

[1] A. Al-Bayati, L. W. L. Q. Xia, M. Balseanu, Z. Yuan, and M. Kawaguchi, "Production processes for inducing strain in CMOS channels," in *Semiconductor Fabtech*, 26th ed. London, U.K.: Bernard Henry, Trans-World House, 2004, pp. 84–88.

[2] V. Chan, R. Rengarajan, N. Rovedo, W. Jin, T. Hook, P. Nguyen *et al.*, "High speed 45 nm gate length CMOSFETs integrated into a 90 nm bulk technology incorporating strain engineering," in *IEDM Tech. Dig.*, San Francisco, CA, 2003, pp. 3.8.1–3.8.4.

[3] A. Murthy, R. S. Chau, T. Ghani, and K. R. Mistry, *Semiconductor Transistor Having a Stressed Channel*. Santa Clara, CA: Intel, 2005.

[4] T. Ghani, S. E. Thompson, M. Bohr *et al.*, "A 90 nm high volume manufacturing logic technology featuring novel 45 nm gate length strained-silicon CMOS transistors," in *IEDM Tech. Dig.*, San Francisco, CA, 2003, pp. 11.6.1–11.6.3.

[5] H. S. Yang *et al.*, "Dual stress liner for high performance sub-45 nm gate length SOI CMOS manufacturing," in *IEDM Tech. Dig.*, 2004, pp. 1075–1077.

[6] P. R. Chidambaram, B. A. Smith, L. H. Hall, H. Bu, S. Chakravarthi, Y. Kim *et al.*, "35% drive current improvement from recessed-SiGe drain extensions on 37 nm gate length PMOS." in *VLSI Symp. Tech. Dig.*, Honolulu, HI, 2004, pp. 48–49.

[7] C. Chien-Hao, T. L.Lee, T. H. Hou, C. L. Chen, C. C. Chen, J. W. Hsu, K. L. Cheng, Y. H. Chiu, H. J. Tao, Y. Jin, C. H. Diaz *et al.*, "Stress memorization technique (SMT) by selectively strained-nitride capping for sub-65 nm high-performance strained-Si device application," in *VLSI Symp. Tech. Dig.*, 2004, pp. 56–57.

[8] Y. C. Liu, J. W. Pan, T. Y. Chang, P. W. Liu, B. C. Lan, C. H. Tung, C. H. Tsai *et al.*, "Single stress liner for both NMOS and PMOS current enhancement by a novel ultimate spacer process," in *IEDM Tech. Dig.*, Washington, DC, 2005.

[9] H. M. Manasevit, I. S. Gergis, and A. B. Jones, "Electron mobility enhancement in epitaxial multilayer Si–$Si_{1-x}Ge_x$ alloy films on (100) Si," *Appl. Phys. Lett.*, vol. 41, no. 5, pp. 464–466, Sep. 1982.

[10] R. People, J. C. Bean, D. V. Lang, A. M. Sergent, H. L. Stormer, K. W. Wecht, R. T. Lynch, and K. Baldwin, "Modulation doping in Ge_xSi_{1-x}/Si strained layer heterostructures," *Appl. Phys. Lett.*, vol. 45, no. 11, pp. 1231–1233, Dec. 1984.

[11] K. Rim *et al.*, "Fabrication and mobility characteristics of ultra thin strained-Si directly on insulator (SSDOI) MOSFETs," in *IEDM Tech. Dig.*, San Francisco, CA, 2003, pp. 3.1.1–3.1.4.

[12] A. Shimizu *et al.*, "Local mechanical-stress control (LMC): A new technique for CMOS-performance enhancement," in *IEDM Tech. Dig.*, San Francisco, CA, 2001, pp. 19.4.1–19.4.4.

[13] S. Ito *et al.*, "Mechanical stress effect of etch-stop nitride and its impact on deep submicron transistor design," in *IEDM Tech. Dig.*, San Francisco, CA, 2000, pp. 247–250.

[14] S. Gannavaram, N. Pesovic, and C. Ozturk, "Low temperature (800 °C) recessed junction selective silicon-germanium source/drain technology for sub-70 nm CMOS," in *IEDM Tech. Dig.*, 2000, pp. 437–440.

[15] S. Thompson *et al.*, "A 90 nm logic technology futuring 50 nm strained silicon channel transistors, 7 layers of Cu interconnects, low-κ ILD, and 1 μm^2 SRAM cell," in *IEDM Tech. Dig.*, San Francisco, CA, 2002, pp. 61–64.

[16] M. V. Fischetti *et al.*, "Six-band $k \bullet p$ calculation of the hole mobility in silicon inversion layers: Dependence on surface orientation, strain, and silicon thickness," *J. Appl. Phys.*, vol. 94, no. 2, pp. 1079–1095, Jul. 2003.

[17] G. Fitzgerald, "A quick primer on strained silicon," *Electron. Eng. Times*, 2004.

[18] S. E. Thompson, G. Sun, K. Wu, J. Lim, and T. Nishida, "Key differences for process-induced uniaxial vs. substrate-induced biaxial stressed Si and Ge channel MOSFETs," in *IEDM Tech. Dig.*, 2004, pp. 221–224.

[19] J. G. Fossum and W. Zhang, "Performance projections of scaled CMOS devices and circuits with strained Si-on-SiGe channels," *IEEE Trans. Electron Devices*, vol. 50, no. 4, pp. 1042–1048, Apr. 2003.

[20] D. Zhang, B. Y. Nguyen, T. White, B. Goolsby, T. Nguyen, V. Dhandapani *et al.*, "Embedded SiGe S/D PMOS in thin body SOI substrate with drive current enhancement," in *VLSI Symp. Tech. Dig.*, 2005, pp. 26–27.

[21] P. Bai *et al.*, "A 65 nm logic technology featuring 35 nm gate lengths, enhanced channel strain, 8 cu interconnects layers, low-κ ILD and 0.57 μm^2 SRAM cell," in *IEDM Tech. Dig.*, 2004, pp. 657–660.

[22] S. E. Thompson *et al.*, "A logic nanotechnology featuring strained silicon," *IEEE Electron Device Lett.*, vol. 25, no. 4, pp. 191–193, Apr. 2004.

[23] Q. Quyang, M. Yang, J. Holt, S. Panda, H. Chen, H. Utomo *et al.*, "Investigation of CMOS devices with embedded SiGe source/drain on hybrid orientation substrates," in *VLSI Symp. Tech. Dig.*, 2005, pp. 28–29.

[24] S. Pidin, T. Mori, K. Inoue, S. Fukuta, N. Itoh, E. Mutoh *et al.*, "A novel strain enhanced CMOS architecture using selectively deposited high tensile and high compressive silicon nitride films," in *IEDM Tech. Dig.*, Tokyo, Japan, 2004, pp. 213–216.

[25] W.-H. Lee, A. Waite, H. Nii, H. M. Nayfeh, V. McGahay, H. Nakayama, D. Fried, H. Chen *et al.*, "High performance 65 nm SOI technology with enhanced transistor strain and advanced-low-κ BEOL," in *IEDM Tech. Dig.*, Washington, DC, 2005.

[26] R. Arghavani, L. Xia, H. M'Saad, M. Balseanu, G. Karunasiri, A. Mascarenhas, and S. E. Thompson, "A reliable and manufacturable method to induce a stress of $>$ 1 GPa on a p-channel MOSFET in high volume manufacturing," *IEEE Electron Device Lett.*, vol. 27, no. 2, pp. 114–116, Feb. 2006.

[27] K. Uchida, T. Krishnamohan, K. C. Saraswat, and Y. Nishi, "Physical mechanisms of electron mobility enhancement in uniaxial stressed MOSFETs and impact of uniaxial stress engineering in ballistic regime," in IEDM Tech. Dig., San Francisco, CA, 2006.

[28] C. Y.-P. Chao and S. L. Chuang, "Spin-orbit-coupling effects on the valence-band structure of strained semiconductor quantum wells," *Phys. Rev. B, Condens. Matter*, vol. 46, no. 7, pp. 4110–4122, Aug. 1992.

[29] M. D. Giles, M. Armstrong, C. Auth, S. M. Cea, T. Ghani, T. Hoffman, and R. Kotlyar *et al.*, "Understanding stress enhanced performance in Intel 90 nm technology," in *VLSI Symp. Tech. Dig.*, Honolulu, HI, 2004, pp. 118–119.

[30] K. Rim, J. Welser, J. L. Hoyt, and J. F. Gibbons, "Enhanced hole mobilities in surface-channel strained-Si p-MOSFETs," in *IEDM Tech. Dig.*, San Francisco, CA, 1995, pp. 517–520.

[31] F. Stern, "Self-consistent results for n-type Si inversion layers," *Phys. Rev. B, Condens. Matter*, vol. 5, no. 12, pp. 4891–4899, Jun. 1972.

[32] R. Oberhuber, G. Zandler, and P. Vogl, "Subband structure and mobility of two-dimensional holes in strained Si/SiGe MOSFET's," *Phys. Rev. B, Condens. Matter*, vol. 58, no. 15, pp. 9941–9948, Oct. 1998.

[33] C. G. VandeWalle and R. M. Martin, "Theoretical calculations of heterojunction discontinuities in the Si/Ge system," *Phys. Rev. B, Condens. Matter*, vol. 34, no. 8, pp. 5621–5634, Oct. 1986.

[34] M. V. Fischetti *et al.*, "On the enhanced electron mobility in strained-silicon inversion layers," *J. Appl. Phys.*, vol. 92, no. 12, pp. 7320–7324, Dec. 2002.

[35] S. Takagi, J. L. Hoyt, J. J. Welser, and J. F. Gibbons, "Comparative study, of phonon-limited mobility of two-dimensional electrons in strained and unstrained Si metal-oxide-semiconductor field-effect transistors," *J. Appl. Phys.*, vol. 80, no. 3, pp. 1567–1577, Aug. 1996.

[36] C. S. Smith, "Piezoresistance effect in germanium and silicon," *Phys. Rev.*, vol. 94, no. 1, pp. 42–49, Apr. 1954.

[37] D. Colman, R. T. Bate, and J. P. Mize, "Mobility anisotropy and piezoresistance in silicon p-type inversion layers," *J. Appl. Phys.*, vol. 39, no. 4, pp. 1923–1931, Mar. 1968.

[38] E. Wang *et al.*, "Quantum mechanical calculation of hole mobility in silicon inversion layers under arbitrary stress," in *IEDM Tech. Dig.*, 2004, pp. 147–150.

[39] V.-C. Lee, "Electron-mediated indirect interaction in narrow-band-gap semiconductors," *Phys. Rev. B, Condens. Matter*, vol. 44, no. 19, pp. 10892–10894, Nov. 1991.

[40] J. W. Mintmire and C. T. White, "Universal density of states for carbon nanotubes," *Phys. Rev. Lett.*, vol. 81, no. 12, pp. 2506–2509, Sep. 1998.

[41] W. Zhao, A. Seabaugh, V. Adams, D. Jovanovic, and B. Winstead, "Opposing dependence of the electron and hole gate currents in SOI MOSFETs under uniaxial strain," *IEEE Electron Device Lett.*, vol. 26, no. 6, pp. 410–412, Jun. 2005.

[42] Y. T. Hou *et al.*, "Direct tunneling hole currents through ultrathin gate oxides in metal-oxide-semiconductor devices," *J. Appl. Phys.*, vol. 91, no. 1, pp. 258–264, Jan. 2002.

[43] S. Rodriguez, J. A. Lopez-Villanueva, I. Mechor, and J. E. Carceller, "Hole confinement and energy subbands in a silicon inversion layer using the effective mass theory," *J. Appl. Phys.*, vol. 86, no. 1, pp. 438–444, Jul. 1999.

[44] A. Haque and K. Alam, "Accurate modeling of direct tunneling hole current in p-metal-oxide-semiconductor devices," *Appl. Phys. Lett.*, vol. 81, no. 4, pp. 667–669, Jul. 2002.

[45] L. Shifren, X. Wang, P. Matagne, B. Obradovic, C. Auth, and S. Cea, "Drive current enhancement in p-type metal-oxide-semiconductor," *Appl. Phys. Lett.*, vol. 85, no. 25, pp. 6188–6190, Dec. 2004.

[46] S. Tiwari, M. V. Fischetti, P. M. Mooney, and J. J. Welser, "Hole mobility improvement in silicon-on-insulator and bulk silicon transistors using local strain," in *IEDM Tech. Dig.*, 1997, pp. 939–941.

[47] H. Nakatsuji, Y. Kamakura, and K. Taniguchi, "A study of subband structure and transport of two-dimensional holes in strained-Si p-MOSFETs using full-band modeling," in *IEDM Tech. Dig.*, 2002, pp. 727–730.

[48] K. Rim *et al.*, "Fabrication and mobility characteristics of ultra-thin strained Si directly on insulator (SSDOI) MOSFETs," in *IEDM Tech. Dig.*, 2003, pp. 49–52.

[49] A. Pirovano, A. L. Lacaita, G. Zandler, and R. Oberhuber, "Explaining the dependences of the hole and electron mobilities in Si inversion layers," *IEEE Trans. Electron Devices*, vol. 47, no. 4, pp. 718–724, Apr. 2000.

[50] F. Gamiz, J. B. Roldan, and J. A. Lopez-Villanueva, "Surface roughness at the Si/SiO_2 interface in fully depleted silicon-on-insulator inversion layers," *J. Appl. Phys.*, vol. 86, no. 12, pp. 6854–6863, Dec. 1999.

[51] K. Seeger, *Semiconductor Physics: An Introduction*. Berlin, Germany: Springer-Verlag, 2004.

[52] K. Takeda, A. Taguchi, and M. Sakata, "Valence-band parameters and hole mobility of Ge–Si alloys-theory," *J. Phys. C.*, vol. 16, no. 12, pp. 2237–2249, 1983.
[53] K. Matsuda, H. Nakatsuji, and Y. Kamakura, "Strain effect on the final state density-of-state for hole scattering in silicon," in *Tech. Proc. Nanotechnol. Conf. and Trade Show*, 2003, pp. 186–189.
[54] T. Yamada and D. K. Ferry, "Monte Carlo simulation of hole transport in strained $Si_{1-x}Ge_x$," *Solid State Electron.*, vol. 38, no. 4, pp. 881–890, Apr. 1995.
[55] S. Yamakawa, H. Ueno, K. Taniguchi, C. Hamaguchi, K. Miyatsuji, K. Masaki, and U. Ravaioli, "Study of interface roughness dependence of electron mobility in Si inversion layers using the Monte Carlo method," *J. Appl. Phys.*, vol. 79, no. 2, pp. 911–916, Jan. 1996.
[56] M. Lundstrom, *Fundamentals of Carrier Transport*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2000.
[57] M. Yang, V. Chan, and S. H. Ku, "On the integration of CMOS with hybrid crystal orientations," in *VLSI Symp. Tech. Dig.*, 2004, pp. 160–161.
[58] S. M. Cea, M. Armstrong, C. Auth, T. Ghani, M. D. Giles, T. Hoffmann, R. Kotlyar *et al.*, "Front end stress modeling for advanced logic technologies," in *IEDM Tech. Dig.*, 2004, pp. 963–966.
[59] T. Ueno, H. S. Rhee, S. H. Lee, H. Lee, D. S. Shin, Y. S. Jin, S. Maeda, and N. I. Lee, "Dramatically enhanced performance of recessed SiGe source-drain PMOS by In-Situ etch and regrowth technique (InSERT)," in *VLSI Symp. Tech. Dig.*, 2005, pp. 24–25.
[60] S. Pidin, T. Mori, K. Inoue, S. Fukuta, N. Itoh, E. Mutoh, K. Ohkoshi, R. Nakamura, K. Kobayashi, K. Kawamura, T. Saiki, S. Fukuyama, S. Satoh, M. Kase, and K. Hashimoto, "A novel strain enhanced CMOS architecture using selectively deposited high tensile and high compressive silicon nitride films," in *IEDM Tech. Dig.*, 2004, pp. 213–216.

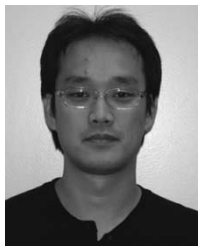
Scott E. Thompson (S'87–M'87–SM'05) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of Florida, Gainesville, in 1985, 1987, and 1992, respectively.

He joined Intel Corporation in 1992 and worked on transistor design, circuit design, and process integration on Intel's 0.35-μm through 90-nm logic technologies. He was an Intel Fellow and Director in The Logic Technology and Manufacturing Group. He was responsible for the next-generation process integration and transistor design. He joined the Electrical and Computer Engineering Department, University of Florida, Gainesville in January 2004. He has published 30 papers and holds 13 patents all relating to transistor design.

Dr. Thompson has received three Intel Achievement awards for his work on transistor design.

Guangyu Sun was born in Shandong, China. He received the B.S. degree in applied physics from the University of Science and Technology of China, Hefei, China, in 1996, and the M.S. degree in electrical engineering from the University of Florida, Gainesville, in 2001, where he is currently working toward the Ph.D. degree in electrical engineering.

He has been a Research Assistant with Dr. Thompson since 2004. His current research focuses on theoretical modeling of strain effect on silicon and germanium MOSFETs.



Youn Sung Choi received the B.S. degree in material science and engineering, and the M.S. degree in physics from Yonsei University, Seoul, Korea, in 1998 and 2002, respectively. He is currently working toward the Ph.D. degree in electrical engineering at the University of Florida, Gainesville.

He was with the Semiconductor Research and Development Center, Samsung Electronics, Korea, as a Process Development Engineer from 2002 to 2004. During that period, he was involved in the dry etch process development of NAND Flash Memory. His current research involves electrical measurements and modeling of strained-Si.

Toshikazu Nishida received the B.S. degree in engineering physics, and the M.S. and Ph.D. degrees in electrical and computer engineering at the University of Illinois at Urbana–Champaign, IL, in 1983, 1985, and 1988 respectively.

He is an Associate Professor in the Department of Electrical and Computer Engineering (ECE) and an Affiliate Associate Professor in the Department of Mechanical and Aerospace Engineering (MAE) at the University of Florida, Gainesville. His current research interests include solid-state physical sensors and actuators, transducer noise, strained semiconductor devices, and reliability physics of semiconductor devices. He is also a member of the Interdisciplinary Microsystems Group (IMG).