Strain Engineering for CMOS Devices

E. Ungersboeck, V. Sverdlov, H. Kosina, and S. Selberherr
Institute for Microelectronics, TU Wien, Gußhausstraße 27–29/E360, 1040 Wien, Austria
Phone: +43-1-58801/36022, Fax: +43-1-58801/36099,
E-mail: ungersboeck@iue.tuwien.ac.at

Abstract

This work reviews the current progress in high-mobility strained MOSFETs and covers the latest developments in strain engineering. We focus on the connections between strain, band structure, and channel mobility characteristics. We show that accurate band structure calculations are essential to understand the different mechanisms of mobility gain induced by uniaxial and biaxial strain. The transport properties of strained silicon (Si) have been investigated by solving the Boltzmann equation using the Monte Carlo method.

1. Strain Engineering Technologies

With ongoing scaling of semiconductor devices the semiconductor industry is facing several critical challenges. Among them are the high gate leakage current for very thin gate dielectrics, the difficulty to maintain a high I_{on}/I_{off} ratio, short channel effects, and the high power dissipation for small transistors. Innovative techniques such as strain engineering, shallow junction engineering, as well as low resistivity contacts, and multilayer interconnects have to be used to solve the arising problems.
It is widely believed that for the next technology nodes strain engineering takes a key position among other technological changes, since the benefits caused by mobility enhancement in the channel are comparatively big. Therefore, increasing emphasis is put on this technology to enhance chip performance.
The influence of strain on the intrinsic mobility of Si was first investigated in the early 1950's [1], [2]. While this effect was not exploited initially, the idea was revived at MIT in the early 1990's [3]. In 1992 it was first demonstrated that n-channel MOSFETs on a strained Si substrate, exhibit a 70% higher effective mobility (μ_{eff}) than those on unstrained substrates [4], [5]. Ever since semiconductor industry has adopted several different technologies to introduce strain in the Si channel of MOSFETs.
Generally, two approaches for introducing strain in MOSFETs can be identified: a global one, where stress is introduced across the entire substrate, and a local approach, where stress is engineered into the device by means of shallow-trench-isolation, epitaxial layers and/or highly stressed nitride capping layers.

1.1 Global Strain Techniques

Most of the pioneering work on strained Si focused on biaxial global strain generated by epitaxy of a thin strained Si layer on a thick relaxed SiGe virtual substrate [4], [5]. One of the drawbacks for practical processes is that subsequent process steps must comply with a low thermal budget to prevent germanium (Ge) migration. Additionally, the manufacture of the virtual substrate is expensive and prone to defects. Thus, an alternative backend approach was proposed, where the wafer is bonded to a substrate with cylindrical surface and thus uniaxial strain is introduced [6], [7], [8]. However, a drawback of the approaches based on global strain techniques in CMOS technology is that they can provide only one type of strain. Since electrons and holes are differently affected by strain, a global strain configuration such as compressive biaxial strain can be beneficial for pMOSFETs, but deteriorate the nMOSFET performance.

1.2 Local Strain Techniques

Starting from the late 1990's effects on the performance of MOSFETs with local stress arising from shallow trench isolation [9], [10], from the formation of silicide at the source/drain region [11], and from nitride contact-etch-stop layers [12], [13] were investigated. Even though process induced stress initially was not able to provide such large strain levels as global strain, the local techniques enjoyed three main advantages: (i) strain can be independently tailored to optimize performance enhancement for both, nMOSFETs, and pMOSFETs, (ii) the threshold voltage shift is smaller in uniaxially stressed MOSFETs, (iii) local stress techniques are cheaper and compatible with standard CMOS technology.
Since the local strain approach was more promising for industrial applications, the first strain engineering technologies were developed on the basis of uniaxial process induced stress.

1.3 State-of-the-Art Strain Engineering Technologies

The first generation technology transfered uniaxial compressive (tensile) stress into the Si channel by growing a local epitaxial film of SiGe (SiC) in the source and drain region of pMOSFETs (nMOSFETs) [14], [15], [16]. Using this technique impressive drive current enhancements

1-4244-0160-7/06/$20.00 ©2006 IEEE.

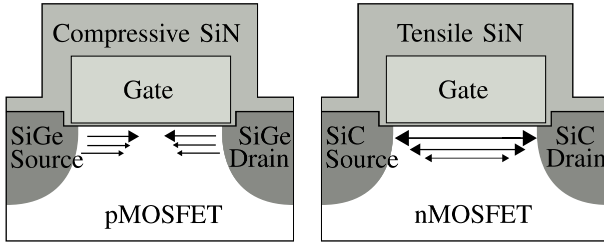

Figure 1. Combination of epitaxial source drain region and stress liner to induce compressive uniaxial stress for pMOSFET and tensile uniaxial stress for nMOSFET.

in the range of 60 to 90% have been demonstrated in short channel devices.

In the next generation a less complex technique was developed. Compressive and tensile capping layers grown on top of the transistors are used as local stressors instead of epitaxial films in the source and drain region [17], [18]. SiN layers with more than 2.0 GPa tensile and 2.5 GPa compressive stress have recently been reported allowing the introduction of more than 1.0 GPa stress in the MOSFET channel [19]. Thus, high-stress compressive films can induce channel stress comparable in magnitude to the first-generation epitaxial SiGe extensions. Interestingly, residual channel stress, present after removal of the nitride layer, was also found to improve device performance. This fact is exploited in the stress memorization technique (SMT) [20], [21].

The above mentioned techniques have been successfully combined on the same wafer, where the pMOSFET is implemented with a selective SiGe epitaxial layer to provide compressive strain, and the nMOSFET is uniaxially strained by a tensile cap film [22]. Furthermore, it has recently been demonstrated that the strain engineering techniques may not only be combined, but can also be superimposed to yield even larger mobility enhancement (Fig. 1). An embedded SiGe process and a compressively stressed liner film were used to induce compressive strain in the pMOSFET, whereas a stress memorization process and a tensile-stressed liner film are used to induce tensile strain in the nMOSFET [21].

Recently, a CMOS compatible hybrid strained SOI technology was demonstrated, utilizing an advanced method of selective biaxial-uniaxial strain hybridization. This technology enables CMOS performance scaling to the 45 nm node and beyond [23], [24].

2. Strained Si Mobility Modeling

Theoretical modeling of strain induced mobility enhancement for electrons and holes is an important issue with some critical questions still open [25], [26]. For this reason a phenomenological approach based on the empirically measured piezoresistance coefficients is commonly used to predict mobility enhancement for electrons and holes [14]. However, the validity of the piezoresistance coefficients is restricted to small strains and stresses ($< 500\,\mathrm{MPa}$), even though much higher values can be reached in current technologies [27], [19]. Furthermore, in inversion layers the enhancement of the effective mobility shows a dependence on the inversion layer concentration N_s, whereas piezoresistance coefficients yield the mobility for only one particular N_s. Additionally, stress induced enhancement of mobility was found to depend on the body thickness of UTB MOSFETs [28], thus the piezoresistance coefficients have to be calibrated for the inversion layer concentration and the Si body thickness.

To surpass the limitations of the simple piezoresistance model a physically motivated model was developed, taking into account the effect of stress on the repopulation of subbands and subband ladders, on scattering, and on the band structure. The approach is able to reproduce the experimentally observed effect of strain on the mobility enhancement $\Delta\mu$ in bulk Si and Si inversion layers. Simulations can easily be adapted for general strain conditions, substrate orientations, and channel directions.

2.1 Band Structure of Strained Si

Band structure calculations of Si under general stress conditions reveal that the change of the electron effective mass can be very pronounced [29]. This result cannot be obtained from deformation potential theory [30], [31], which is widely used for theoretical mobility calculations.

The band structure of Si was calculated using the nonlocal empirical pseudopotential method (EPM) including spin-orbit coupling [32]. Details concerning the incorporation of arbitrary strain in the band structure calculation are reported in [29]. The effective masses were extracted from the curvature of the conduction bands at the minima along various directions.

The degeneracy of the two lowest conduction bands at the six equivalent X points is preserved in unstrained Si, in biaxially strained Si layers grown on $\{001\}$-oriented $\mathrm{Si_{1-y}Ge_y}$ substrate, as well as in Si uniaxially strained/stressed along a fourfold rotation axis $\langle 100\rangle$, and the strain tensor in the crystal system contains no off-diagonal elements. The constant energy surfaces of the conduction band have a prolate ellipsoidal shape, where the semi-axes are characterized by m_l, m_t, and m_t. Here, m_l and m_t denote the longitudinal and transverse electron masses of Si, respectively.

When the strain tensor in the crystal system contains a shear component ε_{xy}, the degeneracy of the two lowest conduction bands at the $X = \frac{2\pi}{a_0}(0,0,\pm1)$ symmetry point is lifted [33]. A nonzero component ε_{xz} and ε_{yz} lifts the degeneracy at $X = \frac{2\pi}{a_0}(0,\pm1,0)$, and $X = \frac{2\pi}{a_0}(\pm1,0,0)$, respectively. Under $\langle 110\rangle$ stress, the resulting shear strain has the effect that the constant energy surfaces of the two-fold degenerate Δ_2-valleys

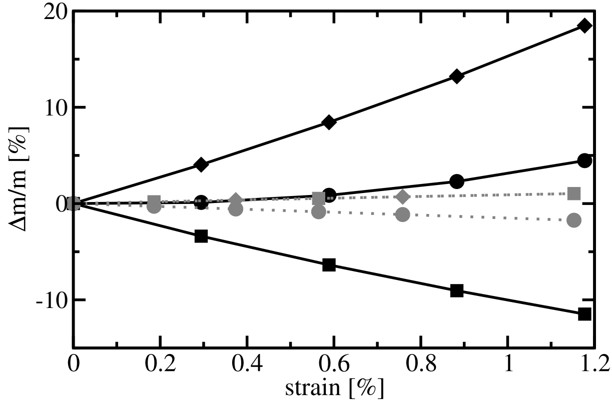

Figure 2. Relative effective mass change of the [001] X-valley under uniaxial tensile stress along [110] (solid lines) and biaxial tensile stress (dotted lines). The mass change is plotted along $[\bar{1}10]$ (diamonds), [001] (circles), and [110] (squares).

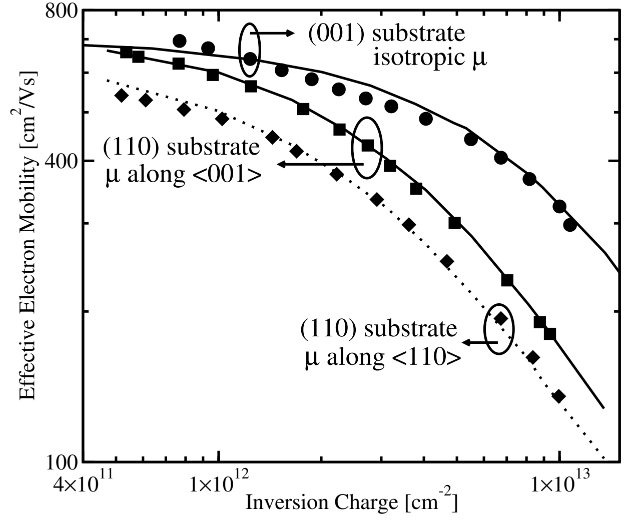

Figure 3. Simulated μ_{eff} (lines) for substrate orientation (001) and (110) compared to measurements [35], [36] (symbols). The anisotropic μ_{eff} on the (110) substrate is given along $\langle 001 \rangle$ and $\langle 110 \rangle$.

take the form of scalene ellipsoids $(m_l, m_{t,||}, m_{t,\perp})$, where $m_{t,||}$ and $m_{t,\perp}$ denote the transverse mass parallel and perpendicular to the stress direction. Band structure calculations indicate that uniaxial tensile stress along $\langle 110 \rangle$ decreases $m_{t,||}$ with respect to $m_{t,\perp}$. Simulation results are depicted in Figure 2. The observed effective mass change is consistent with a theoretical result in [33] and is in good agreement with a recently reported study [28].

2.2 Stress Effect on Inversion Layer Mobility

The inversion layer mobility was calculated with a Monte Carlo method taking into account phonon scattering, surface roughness scattering, and screening effects [34]. Depending on the substrate orientation the six-fold degenerate X-valley splits into up to three different subband ladders. While on (001) substrate the subbands of the lowest subband ladder (unprimed ladder) are spherical and the ladder is two-fold degenerate, on (110) substrate the lowest subbands are elliptical and the unprimed ladder is four-fold degenerate. The higher density of states and larger transport masses on (110) substrate yield a lower inversion layer mobility as compared to (001) substrate. This can be seen in Fig. 3, where we compare experimental data [35], [36] to simulation results.

We continue the analysis by investigating the origin of the electron mobility enhancement of uniaxially stressed inversion layers on (001) oriented substrate. On (001) substrate the mobility can be enhanced if stress direction and channel direction are both [110]. The reasons are the stress induced valley splitting and the stress induced effective mass change. At relatively large body thicknesses, e.g. T_{SOI}=20 nm, $\Delta\mu_{\text{eff}}$ can be understood from a combination of these two effects yielding an anisotropic μ_{eff} as compared to the unstressed system (see Fig. 4). The anisotropic μ_{eff} is consistent with recent experimental results [37].

In ultra-thin Si bodies the strong quantum confinement induces a large intrinsic valley splitting, thus stress induced valley shifts have a negligible effect on the mobility. At T_{SOI}=2.4 nm, the larger (smaller) component of the mobility parallel (perpendicular) to the stress direction [110] results from the effective mass change only, in good agreement with experimental data [28]. Additionally, in Fig. 4 the effect of uniaxial stress on the mobility with channel direction and stress direction parallel to [100] is shown. Stress along [100] lifts the degeneracy of the fourfold (primed) ladder. Since the effective mass does not change, a mobility enhancement is only observed at large T_{SOI} as a result of subband ladder repopulation only.

3. Conclusion

The latest developments in strain engineering have been reviewed. By means of Monte Carlo simulations with accurate band structure models, experimental mobility data were reproduced [28], [37], [36]. While repopulation effects can increase the bulk and inversion layer mobilities at relatively large body thicknesses ($T_{\text{SOI}} > 20$ nm), the population of higher subband ladders is intrinsically reduced by strong quantum confinement for very small body thickness ($T_{\text{SOI}} < 5$ nm). Thus, in the latter case, no mobility enhancement can be expected from stress induced repopulation effects and only the stress induced effective mass change is responsible for the experimentally observed mobility enhancement. Thus, strain engineering technologies which provide conditions to reduce the effective mass in transport direction (like

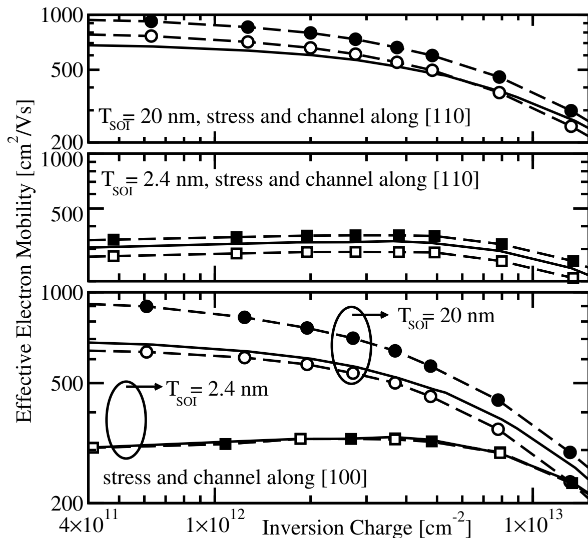

Figure 4. Simulated effective mobility for substrate orientation (001), two channel orientations, two body thicknesses of unstressed (solid lines), and 1 GPa stressed Si (dashed lines). The mobilities are plotted parallel (closed symbols) and perpendicular (open symbols) to the stress direction.

uniaxial tensile stress along ⟨110⟩) are well suited to increase mobility in UTB-MOSFETs.

Acknowledgement

This work has been partly supported by the Austrian Science Fund (FWF), projects 17285-N02 and I79-N16.

References

[1] H. H. Hall, J. Bardeen, and G. L. Pearson, Physical Review **84**, 129 (1951).
[2] C. S. Smith, Physical Review **94**, 42 (1954).
[3] E. Fitzgerald *et al.*, J.Appl.Phys. **59**, 811 (1991).
[4] J. Welser, J. Hoyt, and J. Gibbons, in Proc. Intl. Electron Devices Meeting (1992), pp. 1000–1002.
[5] J. Welser, J. Hoyt, and J. Gibbons, IEEE Electron Device Lett. **15**, 100 (1994).
[6] R. Belford, J.Electron.Mater. **30**, 807 (1950).
[7] B. M. Haugerud, L. A. Bosworth, and R. E. Belford, J.Appl.Phys. **94**, 4102 (2003).
[8] L. Bera *et al.*, IEEE Electron Device Lett. **27**, 58 (2006).
[9] G. Scott *et al.*, in Proc. Intl. Electron Devices Meeting (1999), pp. 827–830.
[10] T. Matsumoto *et al.*, in Proc. Intl. Electron Devices Meeting (2002), pp. 663–666.
[11] A. Steegen, M. Stucchi, A. Lauwers, and K. Maex, in Proc. Intl. Electron Devices Meeting (1999), pp. 497–500.
[12] S. Ito *et al.*, in Proc. Intl. Electron Devices Meeting (2000), pp. 247–251.
[13] A. Shimizu *et al.*, in Proc. Intl. Electron Devices Meeting (2001), pp. 433–436.
[14] S.-E. Thompson *et al.*, IEEE Trans.Electron Devices **51**, 1790 (2004).
[15] Q. Ouyang *et al.*, in VLSI Symp. Tech. Dig. (2005), pp. 28–29.
[16] D. Zhang *et al.*, in VLSI Symp. Tech. Dig. (2005), pp. 26–27.
[17] S. Pidin *et al.*, in Proc. Intl. Electron Devices Meeting (2004), pp. 213–216.
[18] C. Sheraw *et al.*, in VLSI Symp. Tech.Dig. (2005), pp. 12–13.
[19] R. Arghavani *et al.*, IEEE Electron Device Lett. **27**, 114 (2006).
[20] C. C. Hao *et al.*, in VLSI Symp. Tech.Dig. (2004), pp. 56–57.
[21] M. Horstmann *et al.*, in Proc. Intl. Electron Devices Meeting (2005), pp. 233–236.
[22] C.-H. Jan *et al.*, in Proc. Intl. Electron Devices Meeting (2005), pp. 60–63.
[23] A.-Y. Thean, L. Prabhu, V. Vartanian, and M. Ramon, in Proc. Intl. Electron Devices Meeting (2005), pp. 515–518.
[24] H. Yin *et al.*, Appl.Phys.Lett. **87**, 061922 (2005).
[25] M. V. Fischetti, F. Gamiz, and W. Hänsch, J.Appl.Phys. **92**, 7320 (2002).
[26] M. V. Fischetti and Z. Ren, J.Appl.Phys. **94**, 1079 (2003).
[27] L. Washington *et al.*, IEEE Electron Device Lett. **27**, 511 (2006).
[28] K. Uchida, T. Krishnamohan, K. Saraswat, and Y.Nishis, in Proc. Intl. Electron Devices Meeting (2005), pp. 135–138.
[29] E. Ungersboeck *et al.*, in 11th International Workshop on Computational Electronics Book of Abstracts (2006), pp. 141–142.
[30] J. Bardeen and W. Shockley, Physical Review **80**, 72 (1950).
[31] C. Herring and E. Vogt, Physical Review **101**, 944 (1956).
[32] M. M. Rieger and P. Vogl, Physical Review B **48**, 14276 (1993).
[33] G. L. Bir and G. E. Pikus, Symmetry and Strain Induced Effects in Semiconductors (Wiley, New York, 1974).
[34] E. Ungersboeck and H. Kosina, in Proc. Simulation of Semiconductor Processes and Devices (Tokio, Japan, 2005), pp. 311–314.
[35] K. Uchida, J. Koga, and S. Takagi, in Proc. Intl. Electron Devices Meeting (2003), pp. 805–808.
[36] G. Tsutsui *et al.*, in Proc. Intl. Electron Devices Meeting (2005), pp. 747–750.
[37] H. Irie, K. Kita, K. Kyuno, and A. Toriumi, in Proc. Intl. Electron Devices Meeting (2004), pp. 225–228.