

Modeling Techniques for Strained CMOS Technology

V. Sverdlov, O. Baumgartner, T. Windbacher, and S. Selberherr

Institute for Microelectronics
TU Wien, Gußhausstraße 27-29, 1040 Wien, Austria

We review modeling techniques used to compute strain induced performance enhancement of modern MOSFETs. While p-channel MOSFETs were intensively studied, electron transport in strained structures received surprisingly little attention. A rigorous analysis of the subband structure in thin silicon films under stress is performed. Calculated subband effective masses are shown to strongly depend on shear strain and film thickness. A decrease of the transport effective mass under tensile stress in [110] direction and an additional splitting between the unprimed subbands with the same quantum number guarantees a mobility enhancement even in ultra-thin (001) silicon films. This increase of mobility and drive current combined with the improved channel control makes multi-gate MOSFETs based on thin films or silicon fins preeminent candidates for the 22nm technology node and beyond.

Introduction

The rapid increase in computational power and speed of integrated circuits is supported by the aggressive size reduction of semiconductor devices. Downscaling of MOSFETs as institutionalized by Moore's law is successfully continuing because of innovative changes in the technological processes and the introduction of new materials. Until recently, the main method of increasing complementary metal-oxide semiconductors (CMOS) transistor performance was based on geometrical scaling, which has led to an enormous success in increasing the speed and functionality of electronic devices. Although a possibility to build a MOSFET with a gate length as short as 6nm has been demonstrated (1), the semiconductor industry is facing critical challenges with ongoing scaling. Among them are the short channel effects and the high power dissipation for small transistors, and the high gate leakage current for very thin gate dielectrics. Already now, for an oxide thickness of approximately 1.5nm, gate leakage makes it difficult to maintain a sufficiently high I_{on}/I_{off} ratio, which does not degrade MOSFET performance. Rising costs of chip manufacturing are making further scaling increasingly more difficult: the feasibility of fabrication cannot be easily guaranteed, and maintaining performance and reliability becomes a severe issue. With scaling apparently approaching its fundamental limits, the semiconductor industry is facing critical challenges. New engineering solutions and innovative techniques are required to improve CMOS device performance. Strain-induced mobility enhancement is the most attractive way to increase the device speed and will certainly take a key position among other technological changes for the next technology generations. The 32nm MOSFET process technology recently developed by Intel (2) employs advanced third generation strain engineering techniques. Apart from strain, the 32nm MOSFET process technology involves new hafnium-based high-k dielectric/metal gates first introduced for the 45nm technology node (3), which represented a major change in the technological process since the invention of MOSFETs. Although alternative channel materials with a mobility higher than in silicon were already

investigated (4, 5), it is commonly believed that strained silicon will be the main channel material also for MOSFETs beyond the 32nm technology node.

In addition, new device architectures based on multi-gate structures with better electrostatic channel control and reduced short channel effects will be developed. A multi-gate MOSFET architecture is expected to be introduced for the 22nm technology node. Combined with a high-k dielectric/metal gate technology and strain engineering, a multi-gate MOSFET appears to be the ultimate device for high-speed operation with excellent channel control, reduced leakage currents, and low power budget.

In the following we shortly review stress techniques commonly used to enhance transport in CMOS FETs. We summarize the results for the relatively well understood mobility enhancement in p-channel MOSFETs. Then we briefly review the main ideas behind the two-band $\mathbf{k}\cdot\mathbf{p}$ model to describe the conduction band of silicon and analyze the subband structure in (001) ultra-thin films, calculating the subband quantization energies and effective masses. The subband dispersions are embedded into a subband Monte Carlo code in order to enable the computation of the low-field mobility. Results of the maximum current enhancement in a MOSFET at its ultimate scaling limit are analyzed.

Strain Engineering Techniques

Strain engineering technologies are based on physically stretching or/and compressing the silicon crystal lattice. The influence of strain on the intrinsic mobility of silicon was first investigated in the early 1950's (6, 7) and was revived in the beginning of the 1990's (8). A number of strain technologies have been developed. The techniques to introduce stress can be classified in two categories. Stress can be introduced in the whole substrate wafer globally. Global stress in silicon can be typically created by epitaxially growing silicon on (001) oriented SiGe substrate. Due to the crystal lattice mismatch between silicon and SiGe, silicon is symmetrically stretched along the directions [100] and [010]. In 1992 it was demonstrated that n-channel MOSFETs on a strained silicon substrate exhibit a 70% higher effective mobility (9, 10). Global strain techniques are not restricted to bulk-CMOS technology and may be integrated into silicon-on-insulator wafers by layer transfer and wafer bonding techniques. Benefits of performance for n-channel MOSFETs with ultra-thin strained silicon on SiGe-on-insulator substrate (11, 12) or strained silicon directly on insulator were successfully demonstrated with current drive enhancements of 20-25% reported (13, 14).

However, the global stress techniques did not receive so far a broad commercial appreciation and recognition. The reason is the decrease of the hole mobility in p-channel MOSFETs in biaxially strained silicon grown on (001) SiGe substrate for up to 20% germanium concentration (15, 16), regardless of the large hole mobility enhancement anticipated in the bulk (17). Since in unstrained silicon the mobility of holes is usually lower than the electron mobility, it is beneficiary to enhance the mobility of holes for CMOS applications, where similar on-currents are required in n-channel and p-channel MOSFETs. It turns out that for (001) oriented wafers the hole mobility is enhanced along the [110] direction by compressive stress, while the electron mobility in [110] direction is enhanced by tensile stress. Thus, the semiconductor industry has chosen a path on which stress is delivered to *each* n-channel and p-channel MOSFET *independently*. In these techniques stress is created locally by additional process steps (18-22). A direct way to introduce stress in a MOSFET's channel is to fill the source and drain regions with SiGe (23,24) for p-channel MOSFETs and with SiC for n-channel MOSFETs (25,26). Additional uniaxial stress is often introduced by growing compressive or tensile capping

layers (27-29). The advantage of local stress techniques is that they can be combined (30) and superimposed (31) on the same wafer.

Stress is not the only option to improve mobility. An approach called the hybrid orientation technique is based on the optimization of carrier mobility in silicon inversion layers depending on the crystal orientation and the current flow direction for both n-channel and p-channel MOSFETs. The hole mobility is 2.5 times higher on (110) oriented substrate as compared to a standard (001) wafer surface (16). The hybrid orientation technique is compatible with existing state-of-the-art strain engineering processes and is very promising for p-channel MOSFET performance enhancement. A comprehensive analysis of transport in multi-gate MOSFETs under general stress conditions is required for understanding the enhancement of device performance.

Modeling of p-channel MOSFETs

Mobility enhancement in p-channel MOSFETs was intensively studied theoretically and good understanding was achieved. Due to the complex valence band structure in silicon, the consideration must be based on the 6-band $\mathbf{k}\cdot\mathbf{p}$ model (32) which takes into account the bands of heavy and light holes and the split-off band due to spin-orbit interaction. When stress is applied, splitting between the heavy and light hole bands and/or effective mass change appears. Strain can be easily incorporated into the formalism of the $\mathbf{k}\cdot\mathbf{p}$ method (33). Due to the band splitting re-population between the heavy and light holes bands is expected. Since the heavy and light holes have different conductivity masses, which may depend on the stress value, the change in their relative population introduces the mobility change. Because of the change of the density of states and band splitting, scattering also depends strongly on strain. In general, due to the complexity of the band structure, an accurate solution of the Boltzmann equation is needed in order to evaluate the mobility enhancement. Bulk in-plane and out-of-plane hole mobility in biaxially strained or, equivalently, uniaxially along [001] direction strained silicon was investigated in detail (17). Recently, hole mobility in the bulk under uniaxial [110] stress was modeled (15, 34).

The hole mobility in inversion layers for general strain conditions was carefully studied for several important substrate orientations (15, 16, 35). Mobility on a (110) unstrained surface is found roughly 2.5 times higher than on a standard (001) wafer, in agreement with earlier measurements (36). It was found that biaxial tensile stress decreases mobility on a (001) surface (15, 16). At high stress, the largest surface mobility enhancement factor was predicted for (001) oriented wafers, when transport is along the [110] direction under compressive uniaxial stress (16). However, the maximum value of mobility at saturation was found similar along the [110] direction under uniaxial compressive stress at (001) and (1-10) surfaces (16). Recently, a nearly three-fold on-current enhancement due to [110] compressive stress was predicted in strained double-gate structures with the help of a deterministic multi-subband device simulator (37).

Challenges of n-channel MOSFET Modeling

The conduction band in silicon consists of six equivalent valleys with their energy minima located close to the corresponding X -points of the first Brillouin zone. Within the usually used parabolic approximation each valley is characterized by two transversal and one longitudinal effective mass (38). At higher energy a non-parabolic isotropic correction must be included to reproduce the density of states correctly (38).

The change in the conduction band of silicon under biaxial stress is well understood (17, 39). Biaxial stress causes splitting between the six equivalent valleys.

The splitting prompts re-populations between the valleys and also reduces inter-valley scattering. These effects lead to a substantial bulk mobility enhancement due to a biaxial tensile stress. Mobility in biaxially stressed silicon both in the bulk (17) and in the electron inversion layers (40, 41) was carefully investigated.

The application of local stress techniques results in creating a uniaxially stressed silicon in the channel. Since the usual channel orientation on a (001) wafer is along [110] direction, the local stress is typically aligned with the [110] axis. Although already being used in mass production, the stress along [110] direction has received surprisingly little attention within the research community. Only recently a systematic experimental study of the mobility modification due to [110] stress was performed (42).

It was shown that the electron mobility data on (001) substrate under [110] uniaxial stress is consistent with the conductivity mass depending on the stress value, in contrast to biaxially stressed silicon, where the conductivity mass was shown to be virtually independent on the stress value (17). This effective mass dependence cannot be recovered within the effective mass approximation, and a generalization of the conduction band description in uniaxially stressed silicon is needed.

Another shortcoming of the effective mass approximation for the conduction band becomes apparent in structures with thin silicon bodies. Confining carriers within thin films reduces the channel dimension in transversal direction, which further improves gate channel control. The quantization energy in ultra-thin silicon films may reach hundreds of meV. The parabolic band approximation usually employed for subband structure calculations of confined electrons in silicon inversion layers becomes insufficient in ultra-thin films. A recent study of subband energies and transport in (001) and (110) oriented thin films reveals that even a non-parabolic isotropic dispersion relation is not sufficient to describe experimental data, and a direction-dependent anisotropic non-parabolicity must be introduced (43).

In order to overcome these difficulties, the effective mass approximation usually applied for the conduction band must be generalized. An approach based on the full band structure computed with the empirical pseudo-potential method (EPM) (44) is promising and has been recently generalized to include strain and spin-orbit interaction (45). Although it uses the realistic band structure, the method is computationally demanding and needs to be improved to include the self-consistent solution of Poisson equation.

Another approach is based on the $\mathbf{k}\cdot\mathbf{p}$ theory. Recently, a 30-band $\mathbf{k}\cdot\mathbf{p}$ method (46) was employed to investigate subbands in thin films. The method gives an accuracy comparable with the 6-band $\mathbf{k}\cdot\mathbf{p}$ method for the valence band. Since the method is developed around the Γ -symmetry point, it requires all 30 bands to obtain results for the conduction band minima located close to the edge of the first Brillouin zone..

The two-band $\mathbf{k}\cdot\mathbf{p}$ model (33,47-49) is developed in the vicinity of the X -point and thus provides a natural framework to compute the subband structure, in particular the dependences of the electron effective masses on shear strain and thickness, in thin films. In the case of a square potential well with infinite walls, which is a good approximation for the confining potential in ultra-thin films, the subband structure can be obtained analytically (50). This allows for an analysis of subband energies, effective masses, non-parabolicity, and the low-field mobility on film thickness for arbitrary stress conditions.

The peculiarity of [110] uniaxial stress is that it produces an off-diagonal element ε_{xy} of the strain tensor and results in a shear distortion of the crystal. Under shear deformation a significant change in the band structure appears. Namely, due to nonzero values of the shear deformation potential D , the degeneracy between the two lowest conduction bands at the X -points of the Brillouin zone along the [001] axis is lifted (33).

Thus, for non-zero values of ε_{xy} , an additional energy splitting $\Delta E_X = 2 |D\varepsilon_{xy}|$ between the two conduction bands appears at the X point. Due to this splitting, the effective masses in the valleys along [001] direction are substantially modified and become functions of the shear stress value. It is worth noticing that uniaxial stress along [001] direction which produces a biaxially stressed (001) substrate does not contain the shear strain component. Therefore, the influence of [110] uniaxial stress on the band structure and transport must be carefully investigated.

In the following we briefly review the main ideas behind the two-band $\mathbf{k}\cdot\mathbf{p}$ model for a valley in the conduction band of silicon. Then we shortly analyze the unprimed subband structure in (001) ultra-thin films, obtaining analytical expressions for the effective masses and non-parabolicity parameter. With these parameters the non-parabolic subband approximations for the subband dispersions are constructed. The non-parabolic subband dispersions are embedded into a subband Monte Carlo code in order to enable the computation of the low-field mobility. Results of the mobility enhancement calculations are finally analyzed.

Conduction Band in Silicon

Two-band Hamiltonian

The subband structure in a confined system must be based on accurate bulk bands including strain, where various options are available. The conduction band dispersions computed with several methods in [100] and [110] directions are compared in Fig.1a. The method based on non-local empirical pseudo-potentials from (39, 48) is the most accurate one as compared to DFT band structure results obtained with VASP (51). The $sp^3d^5s^*$ tight-binding model with parameters from (52) does not reproduce the anisotropy of the conduction band correctly. In addition, an accurate calibration of the parameters of the $sp^3d^5s^*$ model to describe the modification of the conduction band in strained silicon is performed only recently (53).

The $\mathbf{k}\cdot\mathbf{p}$ theory is a well established method to describe the band structure analytically. As illustrated in Fig.1a, the $\mathbf{k}\cdot\mathbf{p}$ method reproduces the band structure accurately at energies below 0.6eV, which is enough to describe the subband structure and transport properties of advanced MOSFETs. From symmetry considerations the two-band $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian of a [001] valley in the vicinity of the X point of the Brillouin zone in silicon must be in the form (33):

$$H = \left(\frac{\hbar^2 k_z^2}{2m_t} + \frac{\hbar^2 (k_x^2 + k_y^2)}{2m_l} \right) I + \left(D\varepsilon_{xy} - \frac{\hbar^2 k_x k_y}{M} \right) \sigma_z + \frac{\hbar^2 k_z k_0}{m_l} \sigma_y, \quad [1]$$

where $\sigma_{y,z}$ are the Pauli matrices, I is the 2×2 unity matrix, m_t and m_l are the transversal and the longitudinal effective masses, $k_0 = 0.15 \times 2\pi / a$ is the position of the valley minimum relative to the X point in unstrained silicon, ε_{xy} denotes the shear strain component, $M^{-1} \approx m_t^{-1} - m_0^{-1}$, and $D = 14\text{eV}$ is the shear strain deformation potential (33,47-49). The two-band Hamiltonian results in the following dispersion relations (33):

$$E = \frac{\hbar^2 k_z^2}{2m_t} + \frac{\hbar^2 (k_x^2 + k_y^2)}{2m_l} \pm \sqrt{\left(\frac{\hbar^2 k_z k_0}{m_l} \right)^2 + \delta^2}, \quad [2]$$

where the negative sign corresponds to the lowest conduction band,

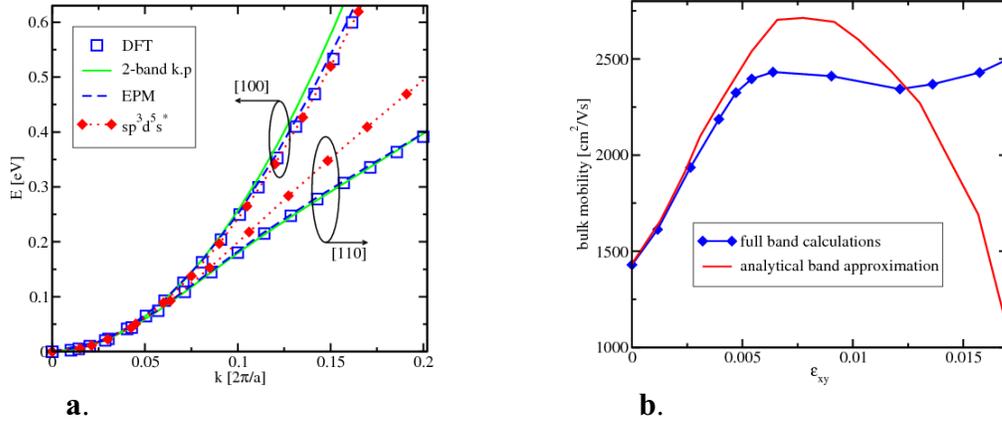


Figure 1 **a.** Comparison of bulk dispersions close at the minimum of the [001] valleys of the conduction band in [100] and [110] directions. DFT (51) and EPM (39,48) results are similar, while the $sp^3d^5s^*$ tight-binding model (52) underestimates anisotropy significantly;

b. Bulk mobility computed by accounting for the full-band structure (symbols) and approximated with non-parabolic dispersion of six valleys (solid line).

$$\delta^2 = (D\epsilon_{xy} - \hbar^2 k_x k_y / M)^2. \quad [3]$$

All moments as well as energies in [2] are counted from the X -point of the Brillouin zone. The classical parabolic approximation is obtained from [2], when coupling between the two conduction bands described by the parameter δ is neglected. The coupling between the bands is small, when the wave vectors $|k_x|, |k_y| \ll k_0 (M/m_l)^{1/2}$ and shear strain $\epsilon_{xy} = 0$. Due to band coupling the dispersion relations [2] become non-parabolic, if the shear strain component is non-zero, and/or at higher energies. In order to check the accuracy of [2] we have carried out numerical band structure calculations with the EPM with parameters from (39, 48). Excellent agreement between the two-band $\mathbf{k}\cdot\mathbf{p}$ model [1] and the EPM results was found up to an energy of 0.5eV. Equation [2] is valid in a larger range of energies compared to a parabolic dispersion relation with isotropic non-parabolic correction and can be used to determine the subband structure in thin silicon films.

Shear strain induced valley shift

It follows from [2] that the position of the conduction band minimum located at a distance k_0 from the X -point in unstrained silicon moves closer to the X -point for nonzero shear strain. Introducing dimensionless strain $\eta = m_l D \epsilon_{xy} / \hbar^2 k_0^2$, one finds for the position of the minimum (48,49):

$$k_{\min} / k_0 = \sqrt{1 - \eta^2}. \quad [4a]$$

At the same time the minimum moves down in energy:

$$\Delta E_{\min} = -\eta^2 \Delta / 4, \quad |\eta| \leq 1, \quad [4b]$$

where $\Delta = 2m_l / \hbar^2 k_0^2$ is the gap between the two conduction bands at the minimum k_0 . For $|\eta| \geq 1$ the conduction band minimum stays exactly at the X point, resulting in the following energy dependence:

$$\Delta E_{\min} = -(2|\eta| - 1)\Delta / 4, \quad |\eta| \geq 1. \quad [4c]$$

Effective masses

Shear strain ε_{xy} modifies the effective masses of the [001] valleys. The transversal mass m_t acquires two different values along (+) and across (-) tensile stress direction (48,49):

$$m_t(\eta)/m_t = [1 \pm |\eta| m_t / M]^{-1}, \quad |\eta| \leq 1; \quad [5a]$$

$$m_t(\eta)/m_t = [1 \pm m_t / M]^{-1}, \quad |\eta| \geq 1. \quad [5b]$$

The longitudinal mass m_l is expressed as

$$m_l(\eta)/m_l = [1 \pm \eta^2]^{-1}, \quad |\eta| \leq 1; \quad [6a]$$

$$m_l(\eta)/m_l = [1 \pm |\eta|^{-1}]^{-1}, \quad |\eta| \geq 1. \quad [6b]$$

Subbands in Ultra-Thin Silicon FilmsDispersion equations

For [001] silicon films the confinement potential gives an additional contribution $U(z)I$ to the Hamiltonian [1]. In the effective mass approximation described by [1] with the coefficient in front of σ_x set to zero, the confining potential $U(z)$ is known to quantize the six equivalent valleys of the conduction band of bulk silicon into the four-fold degenerate primed and the two-fold degenerate unprimed subband ladder (38). In ultra-thin films the unprimed ladder is predominantly occupied. In order to analyze the subbands, we approximate the confining potential of an ultra-thin silicon film by a square well potential with infinite potential walls. Generalization to include a self-consistent potential is straightforward though numerically involved (55).

Because of the two-band Hamiltonian, the wave function Ψ is a spinor with the two components $|0\rangle$ and $|1\rangle$. For a wave function with space dependence in the form $\exp(ik_z z)$ the coefficients A_0 and A_1 of the spinor components are related via the equation $H\Psi = E(k_z)\Psi$. For a particular energy E there exist four solutions k_i ($i=1, \dots, 4$) for k_z of the dispersion relation [2], so the spatial dependence of a spinor component is of the form $\sum_{i=1}^4 A_i^j \exp(ik_i z)$. The four coefficients are determined by the boundary conditions that both spinor components are zero at the two film interfaces. This leads to the following dispersion equations (56):

$$\tan\left(k_1 \frac{k_0 t}{2}\right) = \frac{k_2}{\sqrt{k_2^2 + \eta^2} \pm \eta} \frac{\sqrt{k_1^2 + \eta^2} \pm \eta}{k_1} \tan\left(k_2 \frac{k_0 t}{2}\right), \quad [8]$$

where $\eta = m_t |\delta| / (\hbar k_0)^2$. If the value of

$$k_2 = \sqrt{k_1^2 + 4 - 4\sqrt{k_1^2 + \eta^2}} \quad [9]$$

becomes imaginary at high η values, the trigonometric functions in [8] are replaced by the hyperbolic ones. Special care must be taken to choose a correct branch of $\sqrt{k_2^2 + \eta^2}$ in [9]: the sign of $\sqrt{k_2^2 + \eta^2}$ must be alternated after the argument becomes zero. Introducing $y_n = (k_1 - k_2)/2$, [8] can be written in the form (56):

$$\sin(y_n k_0 t) = \pm \frac{\eta y_n \sin\left(\sqrt{\frac{1 - \eta^2 - y_n^2}{1 - y_n^2}} k_0 t\right)}{\sqrt{(1 - y_n^2)(1 - \eta^2 - y_n^2)}}. \quad [10]$$

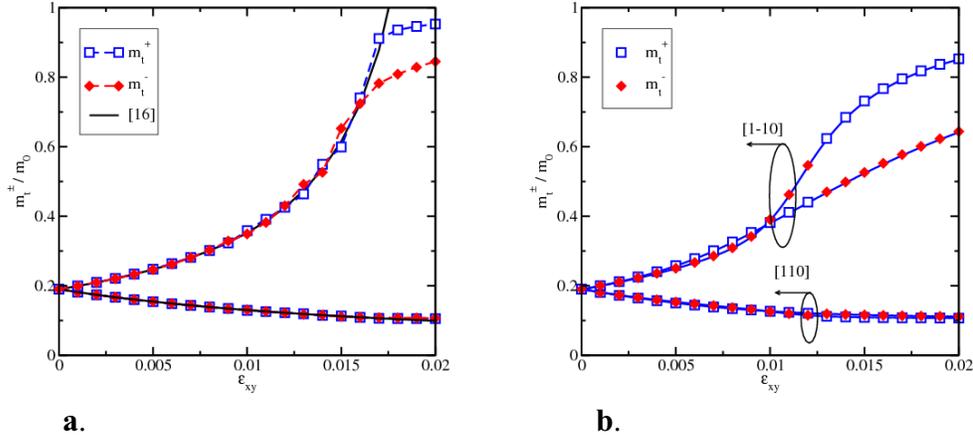


Figure 2 Effective masses of the two ground subbands. The symbols are results obtained by numerical solution of the two-band Hamiltonian [1]:

a. $t=10.86\text{nm}$; solid lines are according to [16];

b. $t=5.43\text{nm}$; solid lines are obtained by numerically differentiating subband dispersions obtained from [8].

Subband structure in thick films

We solve [10] by perturbation techniques. For small η and thick films the right-hand side in [10] can be ignored. The subband relations are found from the condition

$$y_n = \pi n / (k_0 t). \quad [11]$$

This results in the following approximate dispersion relation for unprimed subbands (16):

$$E_n(k_x, k_y) = E_n^0(k_x, k_y) - \Delta \eta^2 / [4 |1 - q_n^2|], \quad [12]$$

where $q_n = (\pi n) / (t k_0)$ and E_n^0 is the subband dispersion relation for parabolic bands:

$$E_n^0(k_x, k_y) = \frac{\hbar^2 \pi^2 n^2}{2m_t t^2} + \frac{\hbar^2 (k_x^2 + k_y^2)}{2m_l} - \frac{\hbar^2 k_0^2}{2m_l}.$$

[12] is valid when

$$(1 - q_n^2)^2 \gg \delta^2 m_l^2 / \hbar^4 k_0^4. \quad [13]$$

Equation [12] describes the subband quantization energy correction due to strain with respect to the valley minimum

$$\Delta E_n(\eta) = - \frac{\hbar^2 \pi^2 n^2}{2m_t t^2} \frac{\eta^2}{|1 - q_n^2|}, \quad [14]$$

which is obtained after taking into account the strain-induced valley minimum energy shift ΔE_{\min} . [14] can be absorbed into the quantization energy $\frac{\hbar^2 \pi^2 n^2}{2m_t t^2}$ by introducing the

longitudinal mass m_l depending on strain η and thickness t :

$$m_l(\eta, q_n) = \frac{m_l}{1 - \eta^2 / |1 - q_n^2|}. \quad [15]$$

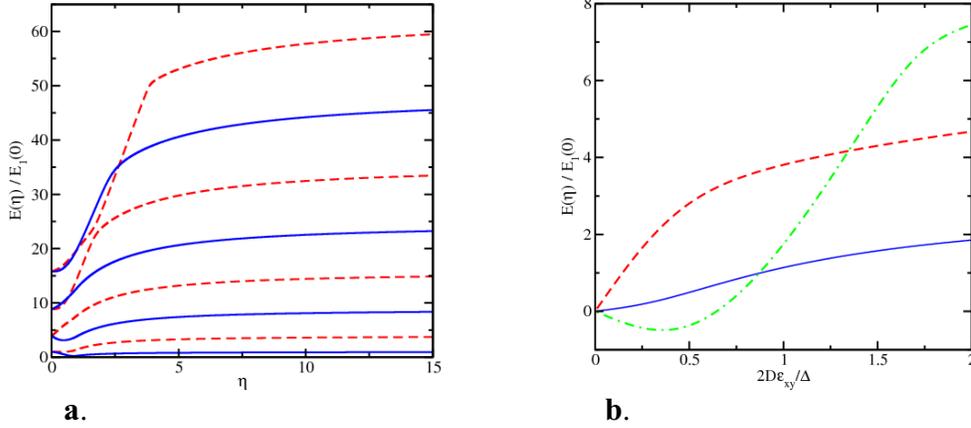


Figure 3 **a.** Normalized positions of the subband minima with respect to the strain-dependent conduction band minimum as function of dimensionless shear strain for a film of the thickness $t=3.26\text{nm}$.

b. Strain-dependent splitting between the minima of the unprimed subbands with the same n : $n=1$ -solid line, $n=2$ -dashed line, $n=3$ -dotted-dashed line.

[12] also describes dependencies of the transversal masses on strain η , the film thickness t , and subband number n :

$$m_t^{\mp}(\eta, q_n) = m_t \left(1 \pm \eta \frac{m_t}{M} \frac{1}{|1 - q_n^2|} \right)^{-1}. \quad [16]$$

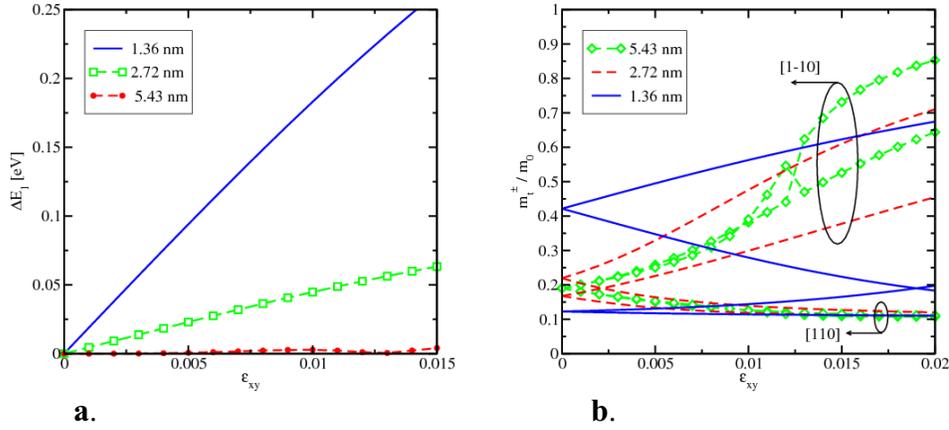
Here m_t^- is the effective mass along the direction [110] of tensile stress. The effective mass in films depends not only on strain but also on film thickness. We compare [16] (lines) with the results of the numerical solution of the two-band $\mathbf{k} \cdot \mathbf{p}$ model Hamiltonian (symbols) in a film with a thickness $t=10\text{nm}$ for the ground subband ($n=1$) (shown in Fig.2a) and the second subband ($n=2$) and find a good agreement. With the film thickness decreased and strain increased a substantial discrepancy between [16] and a numerical solution appears as shown in Fig.2b. This discrepancy is caused by the growing value of the right-hand side in [10], which cannot be ignored.

Splitting of unprimed subbands

Substituting [11] into the right-hand side of [10] and solving [10] for small strain η one obtains the following dispersion relation for the unprimed subbands n :

$$E_n^{\pm} = \frac{\hbar^2}{2m_t} \left(\frac{\pi n}{t} \right)^2 + \hbar^2 \frac{k_x^2 + k_y^2}{2m_t} \pm \left(\frac{\pi n}{k_0 t} \right)^2 \frac{|D\epsilon_{xy} - \frac{\hbar^2 k_x k_y}{M}|}{k_0 t |1 - (\pi n / k_0 t)^2|} \sin(k_0 t). \quad [17]$$

It follows that the subband degeneracy is preserved only, when the shear strain is zero and either $k_x=0$ or $k_y=0$. [17] demonstrates that the unprimed subbands are *not equivalent*. We first analyze the splitting in energy between the two unprimed subbands with the same n , which is usually called the valley splitting (16). According to [17], shear strain induces a valley splitting linear in strain, for small shear strain values (18):



a. Shear strain induced splitting of the ground subbands for several film thicknesses. In ultra-thin films the splitting is larger than kT already for moderate stress.

b. Effective masses of the two ground subbands. In ultra-thin films the effective masses of the two ground subbands are different even without stress.

$$\Delta E_n = 2 \left(\frac{\pi n}{k_0 t} \right)^2 \frac{D \varepsilon_{xy}}{k_0 t |1 - (\pi n / k_0 t)^2|} \sin(k_0 t). \quad [18]$$

The valley splitting is inversely proportional to $(k_0 t)^3$ and oscillates with the film thickness, in agreement with earlier work (16,19).

To find the valley splitting at higher strain values, [10] must be solved numerically. Results shown in Fig.3 demonstrate that the valley splitting can be effectively controlled by adjusting the shear strain and modifying the effective thickness t of the electron system. It is interesting to note that for extremely high strain values the dispersion of the lowest conduction band becomes parabolic again, and the quantization levels in a square well potential are therefore recovered in this limit. Although the value of strain in this limit is unrealistic, this result will be used to analyze dispersion relations for the primed subbands.

Uniaxial stress along [110] channel direction, which induces shear strain, is already used by industry to enhance the performance of modern MOSFETs. Therefore, its application to control valley splitting does not require expensive technological modifications. A possibility to introduce valley splitting larger than the Zeeman spin splitting makes silicon promising for future spintronic applications (20).

As seen from Fig.4a, the valley splitting in ultra-thin films can be quite large already for reasonable stress values. In this case the higher subband becomes de-populated, prompting for a mobility enhancement in (001) ultra-thin films strained along [110] direction.

Effective masses of unprimed subbands

Dispersion [17] predicts different effective masses in [110] direction for the unprimed subbands with the same quantum number n even without strain:

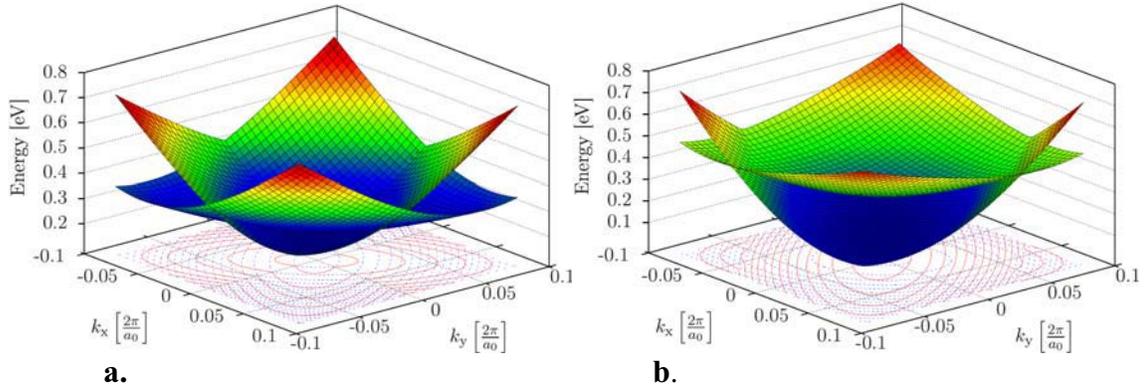


Figure 5 Dispersions of the two ground subbands for a film thickness of 1.36nm:
a. Without strain the subbands are degenerate at the minimum. The lower subband dispersion is dark blue and described by the unification of the two ellipses with different masses [19], while the second subband is described by their intersection.
b. Shear strain of 1% removes the degeneracy between the minima of the ground subbands shown in Fig.4a. The subband dispersions are now characterized by the corresponding effective masses in [110] and [1-10] direction (Fig4b).

$$m_{(1,2)} = \left(\frac{1}{m_t} \pm \frac{1}{M} \left(\frac{\pi n}{k_0 t} \right)^2 \frac{\sin(k_0 t)}{k_0 t |1 - (\pi n / k_0 t)^2|} \right)^{-1} \quad [19]$$

Numerically found values of the masses for the two ground subbands are presented in Fig.4b. The subband dispersions for the two ground subbands are shown in Fig.5. It is to note, that the subband dispersions are not equivalent. This has a profound effect on the valley splitting. Without shear strain the Landau levels in the external [001] magnetic field B are determined using the Bohr-Sommerfeld quantization conditions:

$$E_m^{(1,2)} = \hbar \omega_c \left(m + \frac{1}{2} \right) \frac{\pi}{4 \arctan(\sqrt{m_{(1,2)} / m_{(2,1)}})}, \quad [20]$$

where $\omega_c = \frac{eB}{\sqrt{m_1 m_2} c}$ is the cyclotron frequency. Therefore, the magnetic field induces a

valley splitting linear in the field strength B , in agreement with recent experimental results (20).

A large value of the valley splitting observed by measuring conductance through a point contact can also be attributed to the difference in the subband dispersions, in particular, to the effective mass difference [19]. Indeed, additionally confining the electron system laterally in [1-10] direction, the following dispersion relation of propagating modes within the point contact is obtained:

$$E_p^{(1,2)} = \frac{k^2 k_x^2}{2m_{(2,1)}} + \hbar \omega_{(1,2)} \left(p + \frac{1}{2} \right) + V_b. \quad [21]$$

Here $\omega_{(1,2)}^2 = \kappa / m_{(1,2)}$ and V_b is a gate voltage dependent shift of the conduction band in the point contact (59). The energy minima of the two propagating modes with the same p are

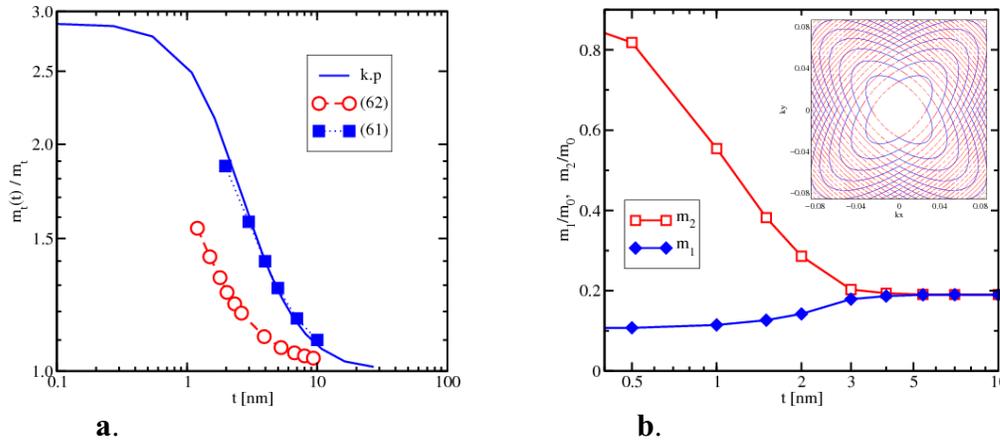


Figure 6 **a.** The thickness dependence of the effective mass of the lowest primed subbands computed with the two-band $k \cdot p$ model (solid line) is in excellent agreement with the full-band calculations (61) (filled symbols). Open symbols show calculations from (62).

b. The thickness dependence in unstrained films of the two lowest unprimed subbands. Inset: contour plots of the two lowest unprimed subbands from Fig.5a.

separated by $\Delta E_p = \hbar |\omega_1 - \omega_2|$ and they are resolved in the conductance of a point contact as two distinct steps. The difference in the effective masses [19] and, correspondingly, the valley splitting can be greatly enhanced by reducing the effective thickness t of the film.

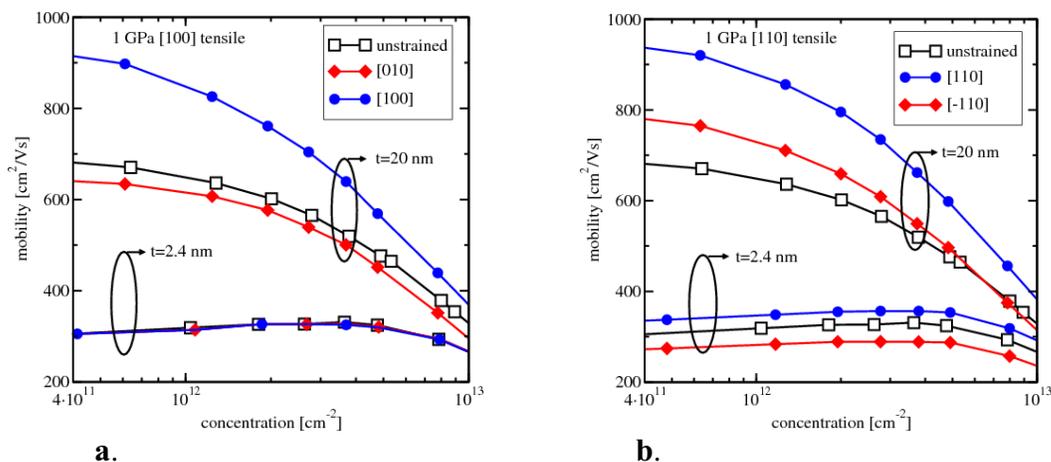
Effective mass of primed subband

Shear strain in [110] direction does not affect the primed valleys along [100] and [010] direction, except for a small shift of the minimum (60). However, recent calculations of the primed subbands based on the density functional theory (DFT) (61) and the “linear combination of bulk bands” method obtained with the empirical pseudo-potential calculations (62) reveal the dependence of the transport effective masses on silicon film thickness t . Here we briefly analyze the effective mass of the primed subbands based on the two-band Hamiltonian [1]. We assume the quantization direction along the [100] axis. By formally replacing k_0/m_1 with k_y/M and $k_x k_y/M$ with $k_z k_0/m_1$ in [1] one finds the dispersion relation and the effective masses in the primed subbands, where results of calculations are shown in Fig.6a. The two-band $k \cdot p$ results are in excellent agreement with the “linear combination of the bulk bands” method with a potential barrier of 3eV at the film interface (62), and they are also consistent with the DFT calculations (61).

Electron Transport in Uniaxially Stressed Films

Mobility enhancement

A multi-subband Monte Carlo method designed for small signal analysis (63) was used to evaluate the mobility in MOSFETs with a thin silicon film. The method is based on the solution of the linearized multi-subband Boltzmann equation, which is exact in the limit of vanishing driving fields. A particular advantage of the method is that it includes degeneracy effects due to the Pauli exclusion principle. Degeneracy effects are important for mobility calculations in ultra-thin films, especially at high carrier concentrations. The multi-subband method uses the subband wave functions and subband energies. They can be found by solving the Schrödinger equation and the Poisson equation self-consistently for each value of the gate voltage. The wave functions are then employed to evaluate



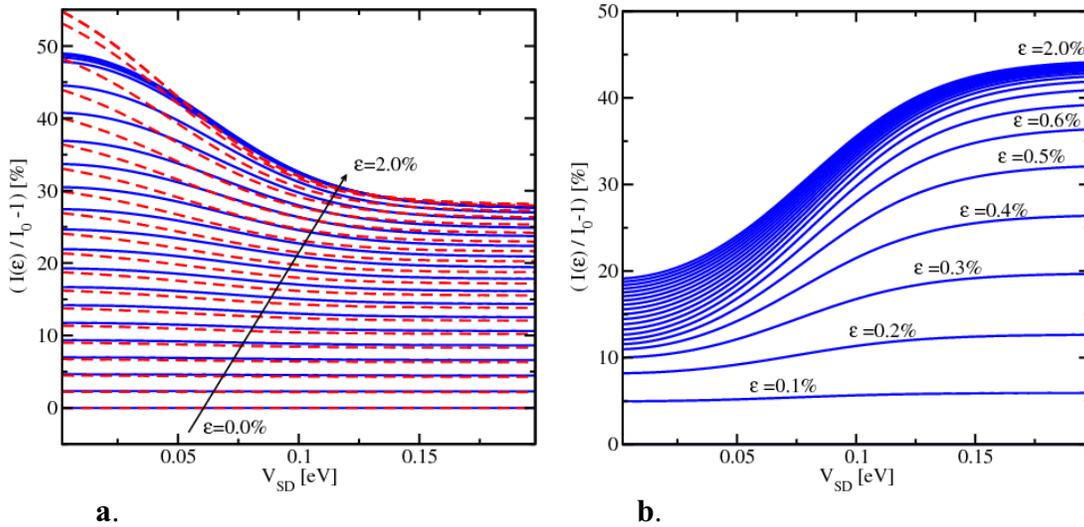
a. The mobility enhancement in [100] direction due to [100] tensile stress of 1GPa is due to de-population of the primed subbands with an unfavourable effective mass in transport direction. In ultra-thin films, where the primed subbands are already de-populated, the mobility remains unchanged.

b. The mobility enhancement due to tensile stress of 1GPa in [110] transport direction due to the transport mass decrease is preserved even in ultra-thin films.

the scattering rates. Scattering with phonons and surface roughness is included. The surface roughness at the two thin film interfaces is assumed to be equal and uncorrelated. The parameters of the Gaussian surface roughness correlation function were calibrated to reproduce the universal mobility curve of Takagi (64) in the inversion layer. The same parameters are then used for mobility calculations in thin film MOSFETs.

Fig.7a shows the mobility in (001) silicon films for two film thicknesses under tensile uniaxial stress of 1GPa in [100] direction. The stress shifts up and de-populates the two [100] primed valleys with unfavorable mass m_1 in the transport direction providing the mobility enhancement in the stress direction for a 20nm film. In [010] transport direction the [100] primed valleys have a smaller transversal mass. Therefore, the de-population of the [100] primed subbands due to [100] tensile stress has a detrimental effect on the mobility in [010] direction. In ultra-thin films the primed subbands are nearly de-populated already without stress due to the large separation in energy between primed and unprimed subbands in a silicon film with a thickness of 2.4nm. Therefore, additional shifting up in energy of the [100] valleys does not have any effect on the mobility in ultra-thin films (Fig. 7a).

Apart from shifting the primed subbands with unfavorable transport masses in the (001) plane up in energy and de-populating them, tensile stress in [110] direction generates a shear component which modifies the transport effective masses of the unprimed subbands (Fig.4b). The decrease of the effective masses in [110] direction induced by shear strain becomes more pronounced with the film thickness reduced guaranteeing the mobility enhancement even in ultra-thin films, as demonstrated in Fig.7b. However, the density of states effective mass $m^* = \sqrt{m_t^+ m_t^-}$ in unprimed subbands increases with shear strain. This results in higher scattering rates which deteriorate the benefits of the thickness-enhanced transport mass decrease at higher stress values, as displayed in Fig.7b. However, the mobility enhancement remains substantial.



a. Figure 8 **a.** Current enhancement as a function of drain bias for several shear strain values for a ballistic MOSFET with a silicon body thickness $t=10.86\text{nm}$. The enhancement is mostly due to the effective mass modification [16] as indicated by dashed lines.

b. The same as in Fig8a but for a body thickness $t=1.36\text{nm}$. The enhancement of the on-current at saturation is larger due to an additional strain-induced splitting between the unprimed subbands shown in Fig.4a.

Current enhancement in ultimate MOSFETs

Calculated subband parameters are used to evaluate transport in a ballistic MOSFET [65], which is sometimes considered as the ultimate limit of scaling [66]. Although not ballistic and backscattering [67] cannot be completely ignored, current devices operate already at approximately 70% ballisticity. It is thus expected that with scaling continuing the next generation MOSFETs would operate even closer to the ballistic limit. Therefore, the results below demonstrate the upper estimate on how effective strain can be for ultra-scaled MOSFETs provided the ballistic limit has been achieved.

In thick films, the enhancement in the linear regime is larger than in saturation. Fig.8a shows the current increase with the exact values for the effective masses (Fig4b) and with the masses [16], which demonstrates that the enhancement in a $t=10.9\text{nm}$ film is due to the effective mass modifications with strain. For small strain values the current enhancement is due to the [110] mass decrease. However, for larger strain the density of states effective mass starts rapidly increasing. For a fixed gate voltage it lowers the chemical potential, thus reducing the current.

In thin silicon films the behavior of the enhancement is reversed as seen in Fig.8b. The reason is the strain-induced energy splitting between the unprimed subbands with the same quantum number [56]. The splitting increases with strain and is particularly large in thin films. Because of this splitting the density of states at low energies decreases with strain. Since the carrier concentration is fixed by the gate voltage, the decrease of the density of states prompts rise of the chemical potential and thus velocity and current.

Conclusion

A rigorous analysis of the subband structure in thin silicon films is performed. The thickness dependence of the effective mass of primed subbands calculated within the two-band $\mathbf{k}\cdot\mathbf{p}$ model is in agreement with earlier full-band calculations. It is demonstrated that within the two-band $\mathbf{k}\cdot\mathbf{p}$ model the unprimed subbands with the same quantum number n are not equivalent. A large splitting between the unprimed valleys of ultra-thin films can be introduced by a shear strain component. Calculated subband effective masses are shown to depend on shear strain and thickness simultaneously. Interestingly, the effective masses of the two unprimed valleys are different in ultra-thin silicon films even without strain. This results in a linear dependence of the subband splitting on the magnetic field strength and leads to large subband splitting in a laterally confined electron system in a point contact.

The mobility enhancement in strained MOSFETs with ultra-thin silicon films is investigated by a subband Monte Carlo method. The method based on the solution of the linearized Boltzmann equation includes the carrier degeneracy exactly. Transport in thin films is determined by the subband structure, in particular by the effective masses. The decrease of the transport effective mass and additional splitting between the unprimed subbands induced by the shear strain component is the reason for current enhancement even in ultra-thin (001) silicon films. This mobility and drive current increase combined with the improved channel control makes multi-gate MOSFETs based on thin films or silicon fins preeminent candidates for the 22nm technology node and beyond.

Acknowledgments

The authors thank F.Schanovsky for providing the data from DFT calculations. This work was supported in part by the Austrian Science Fund FWF, projects P19997-N14 and I79-N16.

References

1. B.Doris *et al.*, *IEDM* 2002, pp.267-270.
2. S.Natarjan, *et al.*, *IEDM* 2008, pp.941-943.
3. K.Mistry, *et al.*, *IEDM* 2007, pp.247-250.
4. M.K.Hudati, *et al.*, *IEDM* 2007, pp.625-628.
5. R.Chau, *ISDRS* 2007, p.3.
6. H.H.Hall, *Phys.Rev.* **84**, 129 (1951).
7. C.S.Smith, *Phys.Rev.* **94**, 42 (1954).
8. E.Fitzgerald, *et al.*, *Appl.Phys.Lett.* **59**, 811 (1991).
9. J.Welser, *et al.*, *IEDM* 1992, pp.1000-1002.
10. J.Welser, *et al.*, *IEEE Electron Device Lett.* **15**, 100 (1994).
11. B.Ghyselen, *et al.*, *Solid-State Electron.* **48**, 1285 (2004).
12. M.Sadaka, *et al.*, *IEEE SOI Conference* 2004, pp.209-211.
13. K.Rim, *et al.*, *IEDM* 2003, pp.49-52.
14. F.Andrieu, *et al.*, *Solid-State Electron.* **50**, 566 (2006).
15. E.Wang, *et al.*, *IEEE T-ED* **53**, 1840 (2006).
16. G.Sun, *et al.*, *J.Appl.Phys.*, **102**, 084501 (2007).
17. M.V.Fischetti and S.E.Laux, *J.Appl.Phys.* **80**, 2234 (1996).
18. G.Scott, *et al.*, *IEDM* 1999, pp.827-830.
19. T.Matsumoto, *et al.*, *IEDM* 2002, pp.663-666.
20. A.Steegen, *et al.*, *IEDM* 1999, pp.497-500.
21. S.Ito, *et al.*, *IEDM* 2000, pp.247-251.

22. A.Shimizu, *et al.*, *IEDM* 2001, pp.433-436.
23. Q.Ouyang, *et al.*, *VLSI Symp.* 2005, pp.28-29.
24. P.Bai, *et al.*, *IEDM* 2004, pp.657-660.
25. K.-W.Ang, *et al.*, *IEDM* 2004, pp.1069-1071.
26. K.-W.Ang, *et al.*, *Appl.Phys.Lett.* **86**, 093102 (2005).
27. S.Pidin, *et al.*, *IEDM* 2004, pp.213-216.
28. C.Sheraw, *et al.*, *VLSI Symp.* 2005, pp.12-13.
29. R.Arghavani, *et al.*, *IEEE Electron Device Lett.* **27**, 114 (2006).
30. C.-H.Jan, *et al.*, *IEDM* 2005, pp.60-63.
31. M.Horstmann, *et al.*, *IEDM* 2005, pp.233-236.
32. J.M.Luttinger and W.Kohn, *Phys.Rev.* **97**, 869 (1955).
33. G.L.Bir and G.E.Pikus, *Symmetry and Strain-Induced Effects in Semiconductors*, J.Wiley & Sons, NY, 1974.
34. Y.Sun, *et al.*, *J.Appl.Phys.* **101**, 104503 (2007).
35. M.V.Fischetti, *et al.*, *J.Appl.Phys.* **94**, 1079 (2003).
36. T.Sato, *et al.*, *J.Appl.Phys.* **8**, 588 (1969).
37. A.T.Pham, *et al.*, *IEDM* 2008, pp.895-898.
38. T.Ando, *et al.*, *Rev.Mod.Phys.* **54**, 437 (1982).
39. M.Rieger and P.Vogl, *Phys.Rev. B* **48**, 14275 (1993).
40. M.V.Fischetti, *et al.*, *J.Appl.Phys.* **92**, 7320 (2002).
41. S.I.Takagi, *et al.*, *J.Appl.Phys.*, **80** 1597 (1996).
42. K.Uchida, *et al.*, *IEDM* 2005, pp.135-138.
43. K.Uchida, *et al.*, *IEDM* 2006, pp.1019-1021.
44. D.Esseni and P.Palestri, *Phys.Rev.B* **72**, 165342 (2005).
45. V.Sverdlov, *et al.*, *IWCE* 2009, pp. 49-53.
46. D.Rideau, *et al.*, *Phys.Rev. B* **74**, 195208 (2006).
47. J.C.Hensel, *et al.*, *Phys.Rev.* **138**, A225 (1965).
48. E.Ungersboeck, *et al.*, *IEEE T-ED* **54**, 2183 (2007).
49. V.Sverdlov, *et al.*, *ESSDERC* 2007, pp.386-389.
50. V.Sverdlov, *et al.*, *Sol.State Electron.* **52**, 1563 (2008).
51. VASP (Vienna Ab-initio Simulation Program) G.Kresse and J.Hafner, *Phys.Rev. B* **47**, 558 (1993); *ibid.* **B 49**, 14251 (1994); G.Kresse and J.Furthmüller, *Phys.Rev. B* **54**, 11169 (1996); *Computs.Mat.Sci.* **6**, 15 (1996).
52. T.B.Boykin, *et al.*, *Phys.Rev.B* **69**, 115201 (2004).
53. Y.M.Niquet *et al.*, *IWCE* 2009, pp.293-296.
54. Y.M.Niquet *et al.*, arXiv:0902.0491.
55. O.Baumgartner, *et al.*, *EUROSOI* 2009, pp.57-58.
56. V.Sverdlov and S.Selberherr, *Sol.State Electron.* **52**, 1861 (2008).
57. M.Friessen, *et al.*, *Phys.Rev.B* **75**, 115318 (2007).
58. S.Goswami, *et al.*, *Nature Physics* **3**, 41 (2007).
59. B.J.van Wees, *et al.*, *Phys.Rev.Lett.*, **60**, 848 (1988).
60. D.Rideau, *et al.*, *Sol.State Electron.*, **53**, 452 (2009).
61. A.Martinez, *et al.*, *IEEE T-Nano*, **8**, 159 (2009).
62. J.van der Steer, *et al.*, *IEEE T-ED* **54**, 1843 (2007).
63. V.Sverdlov, *et al.*, *Sol.State Electron.* **51**, 299 (2007).
64. S.-I.Takagi, *et al.*, *IEEE T-ED* **41**, 2357 (1994).
65. K.Natori, *J.Appl.Phys.* **48**, 4878 (1994).
66. K.Likharev, in *Nano and Giga Challenges in Electronics*, p.27, Elsevier, 2003.
67. M.Lundstrom and Z.Ren, *IEEE T-ED* **49**, 133 (2002).