Two-band $k \cdot p$ model for the conduction band in silicon: Impact of strain and confinement on band structure and mobility

V. Sverdlov *, G. Karlowatz, S. Dhar, H. Kosina, S. Selberherr

Institute for Microelectronics, TU Wien, Gaußhausstr. 27–29, 1040 Vienna, Austria

1. Introduction

Continuing downscaling of transistor feature size is the key for the tremendous success of CMOS technology [1]. The scalability allows to put more transistors per unit area, while increasing transistor performance and reducing costs per operation. An anticipated performance enhancement was achieved at the expenses of the increase in subthreshold and gate leakage currents. Keeping power dissipation due to leakage currents under control forced device engineers to look for new technological solutions in order to deliver projected performance gain. At the 90 nm technology node, stress technique was introduced to enhance performance while keeping the MOSFET design intact. Since then, stress-induced mobility engineering has become a key technique to increase the low-field mobility with the dependence of the non-parabolicity parameter on the film thickness taken into account is compared with the mobility computed with the bulk value of the non-parabolicity parameter.

This reduces inter-valley scattering. In case of tensile biaxial stress applied in the (100) plane the four in-plane valleys move up in energy and become depopulated. The two out-of-plane valleys become more populated. Since electrons in the out-of-plane valleys have favorable conductivity mass and because of reduced inter-valley scattering between out-of-plane and in-plane valleys, the electron mobility is increased [3].

Biaxial stress is usually introduced globally by growing Si epitaxially on a relaxed SiGe substrate. This method, however, cannot provide comparable on-currents in n- and p-MOSFETs required by CMOS technology and is not used in mass production. Instead, the semiconductor industry employs stress techniques compatible with existing CMOS fabrication process. Stress in the channel of a MOSFET is created by using local stressors in the source and drain and additional cap layers. Although already successfully used in mass production, the technologically relevant [110] stress has received little attention within the research community. Only recently the electron mobility modification under stress was systematically investigated experimentally [4].

A shear distortion of the Si crystal lattice inherent to [110] uniaxial stress induces, apart from the nonlinear valley shift [5–7], a more pronounced modification in the conduction band. Shear strain changes substantially both the longitudinal [6,7] and transversal [4,6–8] effective masses in the out-of-plane valley minima. The decrease of the mass in the transport direction along tensile [110] stress and the valley repopulation due to the valley shifts

---

* Corresponding author.
E-mail address: sverdlov@iue.tuwien.ac.at (V. Sverdlov).

0038-1101/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
lead to the mobility enhancement [4,6,7]. The mobility increase is slightly hampered by an increase of the non-parabolicity parameter with shear strain [7].

Double-gate silicon-on-insulator (SOI) transistors with ultrathin Si body (UTB) are good candidates for the far-end ITRS roadmap scaling [9]. Superior electrostatic channel control helps reducing the leakage current and allows scaling the channel length down to 2.5 nm [10], while maintaining reasonable subthreshold slope, satisfactory DIBL, and acceptable gain. Due to size quantization mass down to 2.5 nm [10], while maintaining reasonable subthreshold slope, satisfactory DIBL, and acceptable gain. Due to size quantization mass down to 2.5 nm [10], while maintaining reasonable subthreshold slope, satisfactory DIBL, and acceptable gain.

In [001] Si films two sets of subband ladders are formed. The unprimed ladder originates from the two out-of-plane valleys and is therefore two-fold degenerate. The primed ladder is obtained from the four in-plane valleys and is four-times degenerate. Due to the large quantization mass $m_i$ the subband energies in the unprimed ladder are lower than that in the primed ladder, where the quantization mass is smaller, $m_i < m_i$. Because of the large separation between primed and unprimed subbands only the unprimed ladder is occupied in UTB FETs. The smaller transversal mass $m_i$ is the conductivity mass of unprimed subbands, which favors their superior mobility over the primed ones. If we now apply tensile uniaxial stress in the [1 1 0] transport direction, the electron mobility enhancement is due to the shear strain induced decrease of the conductivity mass [4,6,7]. The role of shear strain in determining the subband energies has not yet been analyzed.

We demonstrate that the two-band $k \cdot p$ model [7,8,11] not only describes accurately the dependences of the valley shifts and the effective masses on the shear strain component. By comparing the model to results from the empirical pseudopotential method (EPM) we show that the two-band $k \cdot p$ model predicts the correct energy dispersion in a wider range of momenta in the Brillouin zone capturing non-parabolicity effects due to the interaction between the two lowest conduction bands. Because the model provides an analytical expression for the energy dispersion also in the presence of shear strain, it allows to explore analytically the quantized subband dispersion in UTB FETs. An example of the dependence of subband energy on shear strain is presented. The model also predicts the dependence of the subband non-parabolicity parameter on UTB film thickness $t$. This dependence leads to a pronounced suppression of the low field mobility in UTB FETs, which helps to bring the simulated UTB FET mobility, with scattering parameters calibrated to reproduce the inversion layer mobility, closer to the measured quantities and, therefore, must be taken into account.

In the next section, we briefly outline the analytical two-band $k \cdot p$ model and verify it against the EPM numerical calculations. Then we discuss the influence of shear strain on subband structure. Finally, we investigate the dependence of the non-parabolicity parameter on film thickness and how it may affect the low-field mobility.

2. Comparison of the two-band $k \cdot p$ model with EPM results

The dispersion relations of the first two conduction bands in the vicinity of the [001] X-point can be obtained for arbitrary stress within the two-band $k \cdot p$ model [7,8,11]:

$$E_{1,2}(k) = \frac{h^2 k_x^2}{2m_i} + \frac{h^2 (k_x^2 + k_y^2)}{2m_t} + \delta E_C \pm \sqrt{ \left( \frac{\Delta k_z}{2k_0} \right)^2 + \left( \frac{D_{xy} - \frac{h^2 k_x k_y}{M}}{M} \right)^2 }.$$  

The minus sign denotes the lowest ($A_1$) and plus the second ($A_2$) conduction band. In this expression the values of energy $E(k)$ and $k_z$ are measured from the X-point, $m_i$ stands for the transversal mass of the longitudinal effective mass, $k_0 = 0.15 \times 2\pi/a$, where $a$ is the Si lattice constant, is the position of the energy minimum of (1) relative the X-point in unstrained Si. $D = 2\hbar^2 k_0^2/m_i = 0.53$ eV is the energy gap between the conduction bands $A_1$ and $A_2$ at $k_0$ and the parameter $M$ is determined by the interband matrix elements:

$$1 = \frac{1}{m_i^2} \sum_{s,s',s''=1} \frac{(p_{x})_{s_1}(p_{y})_{s_2}}{E(k_0) - E_{s''}(k_0)} + \frac{(p_{x})_{s_1}(p_{y})_{s_2}}{E(k_0) - E_{X-s''}(k_0)}.$$  

The value of $M$ is computed using the symmetric pseudopotential method at the valley minimum $k_0 = k_0$. Obtained value $M$ = 0.235m_0 close (but not equal) to $M = m_0/(1 - m_0/m_0)$ reported in [8].

The dispersion relation (1) describes the dependence of $E(k)$ on the strain tensor $\varepsilon_{x x}$, where $\alpha, \beta = x, y, z$. The shift of $E_{\alpha}$ of the [001] valley depends linearly on diagonal strain components:

$$\delta E_{\alpha} = \varepsilon_{\alpha\alpha} + \varepsilon_{\beta\beta} + \varepsilon_{\beta\alpha} + \Xi_{\alpha\beta},$$

where $\Xi_{\alpha\beta}$ is the dilatation and $\Xi_{\alpha\beta}$ the uniaxial deformation potential for the conduction band minimum.

The shear strain component $\varepsilon_{xy}$, which is created by [1 1 0] uniaxial stress, lifts the degeneracy of the conduction bands $A_1$ and $A_2$ at the X-point ($k_0 = 0$). The gap $|E_{1}(0) - E_{2}(0)| = 2D_{xy}$ opens, where the shear deformation potential $D = 14$ eV is introduced. From (1) it follows that the minimum $k_{min}$ along the $k_z$ axis of the [001] valley depends on shear strain. First, it moves closer to the X point:

$$k_{min} = k_0 \sqrt{1 - \eta^2}, \quad |\eta| < 1,$$  

where a dimensionless shear strain $\eta$ is introduced as

$$\eta = \frac{2D_{xy}}{D}.$$  

For $|\eta| > 1$ the valley minimum rests at the X-point: $k_{min} = 0$. The minimum of the [001] valley also moves down in energy with respect to the remaining four fold degenerate valleys. For $|\eta| < 1$ the dependence of the valley shift is quadratic [7,6,5], while it becomes linear for $|\eta| > 1$:

$$\delta E_{\alpha \beta} = \left\{ \begin{array}{ll} -\frac{\eta^2}{2}, & |\eta| < 1, \\ -\frac{\eta}{2(\eta^2 - 1)}, & |\eta| > 1. \end{array} \right.$$  

The analytical conduction band model (1) allows to obtain the dependences of the effective masses on shear strain $\varepsilon_{xy}$ at the band minimum the mass $m_1$ across and $m_2$ along the [1 1 0] stress direction are:

$$\frac{m_1(\eta)}{m_2} = \left\{ \begin{array}{ll} \frac{1}{1 - \eta^2}, & |\eta| < 1, \\ \frac{1}{1 - \eta^2/\eta^2}, & |\eta| > 1. \end{array} \right.$$  

$$\frac{m_2(\eta)}{m_1} = \left\{ \begin{array}{ll} \frac{1}{1 - \eta^2}, & |\eta| < 1, \\ \frac{1}{1 - \eta^2/\eta^2}, & |\eta| > 1. \end{array} \right.$$  

The analytical expression (1) predicts dispersions not only close to the band minimum. Comparison of the analytical $k_0$ dispersion (1) to the results of the numerical empirical pseudopotential calcula-
EPM data, the value from [12] are used. In order to fit the analytical dispersion to the shear strain. For EPM calculations the pseudopotential parameters up to the energy of approximately 0.5 eV from the valley minimum. The band degeneracy lifting at the X-point is dictated by symmetry [11] and is therefore exact.

For shear strain values up to the large interval ~0.08, which corresponds to a shear strain of 11%, the analytical dispersion (1) (lines) reproduces the EPM data (symbols) excellently in a large interval |k_x - k_0| < 0.25(2π/a) around the minimum value k_0 even at large strain ε_0 = 2%, when the valley minimum is at the X-point. Therefore, we conclude that (1) describes the k_x dispersion up to the energy of approximately 0.5 eV from the valley minimum. We also notice that the dispersion relation (1) describes accurately the band degeneracy lifting at the X-point, where the form (1) is dictated by symmetry [11] and is therefore exact.

In order to validate the use of (1) in a larger portion of the Brillouin zone, we now compare the dispersions in xy plane. Results of (1) and numerical EPM calculations are shown in Fig. 2. Panel (a) displays a good agreement between the analytical and numerical dispersions up to 0.5 eV from the valley minimum in unstressed Si. Panel (b) demonstrates a good agreement of analytical band dispersions up to 0.5 eV from the valley minimum in unstressed Si. Panel (c) displays a good agreement between the analytical and numerical potential with infinite walls at the Si/SiO_2 interfaces. The wave function in the quantization direction is counted from the minimum value of the wave function in the quantization direction K_z [2π/a].

Subband structure within the parabolic effective mass approximation with strain dependent masses (4) and (5) is added to panel (c). Agreement between all the three methods is good for energies close to the valley minimum. At energies larger than 40 meV the parabolic approximation becomes less accurate, while the analytical two-band k·p model (1) closely follows the numerical dispersion. Because the k·p model is valid in a larger portion of the Brillouin zone, we use the model to investigate the dispersion of the quantized subbands in UTB FETs in the next section.

3. Subband structure in strained UTB FETs

Shear strain may affect the subband energies in UTB Si films. In order to estimate this effect, we assume that the potential inside the Si film of thickness t can be approximated by the square well potential with infinite walls at the Si/SiO_2 interfaces. The wave function in the conduction band is taken of the form:

$$
\psi(x, y, z) = A \sum_{j=1}^{2} (-1)^{j-1} \exp(ik_x z) \exp(ik_x x + ik_y y).
$$

It has the usual plane wave structure in the x, y plane with fixed k_x, k_y. The values k_x of the wave function in the quantization direction must be determined from the conditions that the wave function must vanish at the interfaces:
\[ \psi(x, y, \pm t/2) = 0 \]

and that the corresponding energies are equal:

\[ E(k_x, k_y, k_z) = E(k_x, k_y, -k_z) = E. \]

The values of \( k_z \) are found from (1), when it is resolved for \( k_z \).

The corresponding equation is

\[ h^2 k^2_z - 4(m_E + h^2 k^2_z) + 4m^2_l(E^2 - \delta^2) = 0, \quad (8) \]

where \( \delta \) is defined as:

\[ \delta = h^2 k_z k_x / M - D_{xy}. \quad (9) \]

The solution for \( k_z \) is easily found when \( \delta = 0 \):

\[ k_z^2 = \left( k_0 \pm \sqrt{2m_E E / h^2 + k_0^2} \right)^2. \quad (10) \]

By choosing the values \( k_{1,2} = k_0 \pm \sqrt{2m_E E / h^2 + k_0^2} \), the wave function (7) becomes zero at the interfaces if

\[ \sqrt{2m_E E / h^2 + k_0^2} = \pi n / l, \quad n = 1, 2, 3, \ldots \]

from which the standard subband dispersions are

\[ E^0_n(k_x, k_y) = E^0_n + \frac{h^2 (k_x^2 + k_y^2)}{2m_l}, \quad (11) \]

where

\[ E^0_n = \frac{h^2 \pi^2 n^2}{2m_l t^2} - \frac{\Delta}{4} \]

are the subband energies relative to the X-point.

By analogy, for \( \delta \neq 0 \) one looks for a solution of (8) in the form:

\[ k_{1,2} = \xi \pm \frac{\pi n}{l}. \quad (12) \]

The subband energies are

\[ E_n(k_x, k_y) = E^0_n(k_x, k_y) - \frac{\delta^2}{\Delta} \frac{1}{1 - \frac{\pi n}{l t}}. \quad (13) \]

while \( \xi \) depends on \( \delta \):

\[ \xi = k_0 \sqrt{1 - \frac{4\delta^2}{\Delta^2} \frac{1}{1 - \frac{\pi n}{l t}}} \quad (14) \]

The obtained expressions (13) and (14) are valid when (8) has got four real roots, which results in the conditions

\[ \frac{\pi^2 n^2}{k_0^2 t^2} \leq \frac{1 - 2|\delta|}{\Delta}, \]

\[ \frac{\pi^2 n^2}{k_0^2 t^2} \geq \frac{1 - 2|\delta|}{\Delta}. \quad (15) \]

We note that the dispersion relation (13) may become less accurate close to its validity region defined in (15) due to the used ansatz (7) for the wave function. In this approximation the two valleys with the dispersion (1) are considered to be independent. Coupling between the valleys introduces the valley splitting in a confined system and prevents from finding simple analytical expressions for subband dispersions. At the same time the valley splitting does not significantly affect the subband effective masses, at least for small and moderate strain values. In order to obtain the analytical expression for the subband effective masses and the non-parabolicity parameter we neglect the valley splitting in this study.

Examples of quantization for \( \delta = 0 \) and \( \delta = 53 \text{ m eV} \) for a film of the thickness \( t = 4.9 \text{ nm} \) are shown in Fig. 3. By means of Fig. 3 the solution of (8) in the form (12)–(14) allows a simple graphic interpretation: the energy \( E_n \) and the position \( \xi \) are determined from the condition:

\[ E(\xi + \frac{\pi n}{l}) = E(\xi - \frac{\pi n}{l}) = E_n. \]

Depending on the parameters, Eq. (8) may have either four or two real roots. In case of four roots, they are combined in pairs (12) around the corresponding minimum (14) in such a way that within the pair each member has the group velocity of opposite sign. In case of only two real roots, the minimum \( \xi = 0 \), and the subband energy is determined by (1) with \( k_x = \pi n / l t \). These solutions may exist only for finite \( \delta \) when the gap at the X-point is opened.

Eq. (13) provides the non-parabolic dispersion relation of the unprimed subbands. In particular, the subband energy in strained Si film is shifted with respect to \( E^0_n \):

\[ E_n(\eta) = E^0_n - \frac{\Delta}{4} \eta^2 \frac{1}{1 - \frac{\pi n}{l t}}. \quad (16) \]

Taking the bulk band shift (3) into account (16) can be rewritten for \( (\pi n / k_0 l)^2 \ll 1 \) as

\[ E_n(\eta) = \pi^2 \frac{n^2}{2m(\eta l)^2} - \Delta E_{\text{shear}}, \quad (17) \]

where \( m(\eta) \) is defined by (6). Shear strain also modifies the in-plane effective mass \( m_1 \), making it anisotropic. Interestingly, the modification of the effective mass also depends on UTB thickness \( t \). As in the bulk case, the principal axis of anisotropy are [110] and [1 10]. The mass across \( m_{n1} \) and along \( m_{n2} \) the [1 10] stress direction now are

\[ \frac{m_{n1}(\eta, l)}{m_1} = \left( 1 - \frac{\eta}{4} \frac{m_1}{M} \frac{1}{1 - \frac{\pi n}{l t}} \right)^{-1} \quad (18) \]

\[ \frac{m_{n2}(\eta, l)}{m_1} = \left( 1 + \frac{\eta}{4} \frac{m_1}{M} \frac{1}{1 - \frac{\pi n}{l t}} \right)^{-1} \quad (19) \]

Expressions (13), (18) and (19) are subject to the conditions \( |\eta| \leq 1 \) and (15).

4. Dependence of the subband non-parabolicity parameter on UTB FET thickness

For vanishing shear strain (\( \epsilon_{xy} = 0 \)) expression (13) provides a non-parabolic dispersion relation for unprimed subbands as

\[ E_n(k_x, k_y) = \frac{h^2 \pi^2 n^2}{2m_l k^2} + \frac{h^2 (k_x^2 + k_y^2)}{2m_l} - \frac{h^4 k_x^2 k_y^2}{M^2 \Delta} \left( 1 - \frac{\pi n}{l t} \right) - \frac{\Delta}{4}. \quad (20) \]

Besides the usual parabolic dependence on \( k_x, k_y \), the dispersion (20) contains a fourth order term, which modifies the energy dependence of the density of states. Such deviation is usually accounted for by the non-parabolicity parameter \( \gamma_n \). An isotropic non-parabolic dispersion including the non-parabolicity parameter is usually taken in the form:

\[ \frac{h^2 (k_x^2 + k_y^2)}{2m_l} = E_n(k_x, k_y) (1 + \gamma_n E_n(k_x, k_y)). \quad (21) \]

Assuming the non-parabolic term in (20) to be small, we average it, as in [13], over the equienergy surface \( E^0_n(k_x, k_y) = E \). The obtained isotropic dependence on \( k_x^2 + k_y^2 \) is re-expressed via \( E = E^0_n(k_x, k_y) \), using again the smallness of the non-parabolic term. As a result
one obtains the following expression for the non-parabolicity parameter:

\[ \alpha_n(t) = \alpha_0 \cdot \frac{1}{\sqrt{m_t}} \cdot \frac{1}{k_0^2} \cdot \frac{t}{k_0^2} \cdot \frac{t^2}{k_0^2} \]

(22)

where

\[ \alpha_0 = \frac{1}{2} \left( \frac{m_t}{M} \right)^2 \]

The estimated value of \( \alpha_0 = 0.63 \text{ eV}^{-1} \) is close to the phenomenological value \( \alpha_0 = 0.5 \text{ eV}^{-1} \) [13] obtained experimentally. It is interesting to note that the non-parabolicity is determined by the strength of interaction of the lowest conduction band to the closest band, which is inversely proportional to the gap \( \Delta \). The dependence of the non-parabolicity parameter on the film thickness \( t \) is shown in Fig. 3.

Fig. 4. Dependence of the non-parabolicity parameter on the film thickness for the three lowest unprimed subbands.

5. Low-field mobility in UTB Si films

Through the modification of the density-of-states, the non-parabolicity parameter affects the scattering rates, and therefore the mobility of the system. As example we consider the mobility in a double-gate FET with a thin Si body. The subband energies and the corresponding wave functions are calculated from the Schrödinger equation using the parabolic approximation also for the [001] valleys with the effective mass \( m_1 \) defined by (6) and \( m_t \) by (18) and (19), respectively. The Schrödinger equation is solved self-consistently with the Poisson equation for each value of the effective field [14]. The wave functions obtained are used to evaluate the scattering rates. Our transport calculations account for electron–phonon [13,15] and surface roughness scattering. For the surface roughness scattering matrix elements we use the original formulation by Prange and Nee [16–19]. A Gaussian correlation function [20] for the surface roughness is assumed. The zero-field Monte Carlo algorithm [21,22] which accounts for the Pauli blocking factor in the scattering rates was used to evaluate the low-field mobility.

Results of the mobility simulations in an UTB FET, with and without the dependence of the non-parabolicity parameter on the Si film thickness taken into account, are shown in Fig. 5. For

\[ \frac{\mu(t) - \mu(\alpha_0)}{\mu(\alpha_0)} \]

Fig. 5. Relative mobility correction due to film thickness dependence of the non-parabolicity parameter.
...if the carrier concentration is not too large. An increase of the non-parabolicity parameter, and relative corrections are about 7% for \( t = 3 \text{ nm} \) and 15% for \( t = 2.5 \text{ nm} \). The mobility correction is about 7% for \( t = 3 \text{ nm} \) and it reaches up to 20% at high carrier concentration for \( t = 2.5 \text{ nm} \). The mobility decrease due to the thickness dependent non-parabolicity parameter helps to bring the simulated low field mobility in UTB FETs closer to its experimental value.

Otherwise the mobility simulated with the surface roughness parameters calibrated to reproduce the universal mobility curve in inversion layers [23] is too high [24] compared to the measurement [25].

6. Conclusion

By comparing with numerical pseudopotential calculations, we demonstrated that the two-band model is valid in a larger portion of the Brillouin zone as compared to the parabolic approximation. The two-band \( k \cdot p \) model is then used to analytically describe the subband dispersion in ultra-thin Si films, including strain. The model provides an analytical expression for the thickness-dependent non-parabolicity parameter in the unprimed subbands. Finally, the low-field mobility with the dependence of the non-parabolicity parameter on the film thickness taken into account is compared with the mobility computed with the bulk value of the non-parabolicity parameter, and relative corrections are about 7% for \( t = 3 \text{ nm} \) and 15% for \( t = 2.5 \text{ nm} \).

Acknowledgement

This work has been supported by the Austrian Science Fund FWF, Projects I79-N16 and P19997-N14.

References