VISTA Status Report
June 2009

H. Ceric, P. Hehenberger, G. Milovanovic, V. Sverdlov, M. Vasicek, S. Selberherr

Institute for Microelectronics
TU Wien
Gußhausstraße 27-29/E360
A-1040 Wien, Austria

Contents

1 Copper Microstructure Impact on Evolution of Electromigration Induced Voids

We study the impact of microstructure on nucleation and evolution of electromigration induced voids. The grain boundaries are described with a comprehensive model which includes the dynamics of mobile and immobile vacancies in dependence of mechanical stress. The surface of an evolving void is modeled by a three-dimensional Level-Set algorithm. Simulations have shown that the constellation of grain boundaries determines the electromigration failure behavior seriously.

1.1 Introduction

Contemporary integrated circuits are often designed using simple and conservative design rules to ensure that the resulting circuits meet reliability goals. This precaution leads to reduced performance for a given circuit and metallization technology. An ultimative hope of integrated circuits designers today is to have a computer program at hand, which predicts the behavior of thin film metalizations under any imaginable condition. Due to intensive development in the last decade, physical models of electromigration have reached a level of maturity, which enables predictions of failure behavior. The cause of failure is always an electromigration induced void in the interconnect structure. We present our recent development in modeling of void nucleation and void evolution.

A new model for grain boundary physics is applied as extension and refinement of the standard electromigration continuum model. The model differentiates between mobile and immobile vacancies. Immobile vacancies are captured at grain boundaries and triple points, causing a build-up of tensile stress. High tensile stress leads to nucleation of intrinsic voids. These voids evolve through the interconnect causing a resistance change and, occasionally, a complete failure.

Different approaches have been applied to model evolving void surfaces in the last ten years. However, all of these models lack an appropriate description of the void development process, neglecting relevant physical phenomena that lead to interconnect failure. Moreover, these models are only suitable for simulations of simple two-dimensional interconnects and cannot realistically describe the void evolution mechanisms in modern complex interconnect structures. In this paper a three-dimensional Level-Set module is applied to simulate the evolving void surface. The site of void nucleation and the morphology of the evolving void accurately reproduce experimental observations.

1.2 Modeling of Grain Boundaries

The diffusion of point defects inside the grain boundary is faster compared to grain bulk diffusion due to the fact [1] that a grain boundary generally exhibits a larger diversity of point defect migration mechanisms. Moreover, formation energies and migration barriers of point defects are in average lower than those for lattice.

In 1951, Fisher published his nowadays classic paper [2] presenting the first theoretical model of grain boundary diffusion. That pioneering paper, together with concurrent measurements by Turnbull and Hoffman [3, 4], initiated quantitative studies of grain boundary diffusion in solids. The fact that grain boundaries in metals provide high-diffusivity paths was known already in the 1930s, mostly from indirect and qualitative experiments. For example, the enhanced rates of sintering, creep, discontinuous precipitation, and other processes and reactions in polycrystalline samples were attributed to an accelerated atomic transport along grain boundaries.

The grain boundary model used in this work is designed as an extension of a continuum electromigration model [5, 6]. The grain boundary is treated as a separate medium with the capability of absorbing and releasing vacancies (Fig. 1). Vacancies are trapped from both neighboring grains with the trapping rate ω_T and released to these grains with a release rate ω_R .

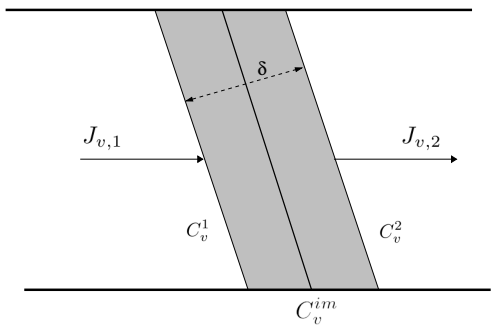

Figure 1: The fluxes $J_{v,1}$ and $J_{v,2}$ change the concentration of mobile vacancies (C_v^1,C_v^2) and immobile vacancies (C_v^{im}).

The vacancy concentration from both sides of the grain boundary is denoted as C_v^1 and C_v^2 . Correspondingly, fluxes are calculated as

$$J_{v,1} = \omega_T (C_v^{eq} - C_v^{im}) C_v^1 - \omega_R \, C_v^{im}, \qquad (1)$$

$$-J_{v,2} = \omega_T (C_v^{eq} - C_v^{im}) C_v^2 - \omega_R \, C_v^{im}. \qquad (2)$$

The vacancies are captured at the grain boundary and converted into immobile vacancies (C_v^{im}). This is mathematically expressed as

$$\frac{\partial C_v^{im}}{\partial t} = \frac{J_{v,1} - J_{v,2}}{\delta} = -\Big(\frac{\partial C_v^1}{\partial t} + \frac{\partial C_v^2}{\partial t}\Big), \quad (3)$$

where δ is the grain boundary thickness. By combining equations eq.(1), eq.(2), and eq.(3) with the vacancy balance equation, the following equation is obtained.

$$\frac{\partial C_v^\alpha}{\partial t} = -\mathrm{div}\vec{J}_{v,\alpha} + \frac{1}{\tau}\Big(C_v^{eq} - C_v^{im}\Big(1 + \frac{2\,\omega_R}{\omega_T(C_v^1 + C_v^2)}\Big)\Big), \quad (4)$$

$\alpha = 1$ and $\alpha = 2$ indicate the left and the right side of the grain boundary, respectively. $\vec{J}_{v,\alpha}$ is the vacancy transport driven by electromigration and a stress gradient. τ is expressed by

$$\frac{1}{\tau} = \frac{\omega_T(C_v^1 + C_v^2)}{\delta}. \quad (5)$$

The capacity of the grain boundary to accept trapped vacancies is expressed by the stress-dependent equilibrium concentration [5, 7]

$$C_v^{eq} = C_v^0 \exp\Big(\frac{\sigma_{nn}\Omega}{k_B\,T}\Big), \quad \sigma_{nn} = \vec{n} \cdot \bar{\bar{\sigma}} \cdot \vec{n}, \quad (6)$$

where we assume a unique equilibrium vacancy concentration C_v^0 in stress free copper, in the grain's bulk, and in the grain boundaries. $\vec{n}$ is the normal to the grain surface and $\bar{\bar{\sigma}}$ is the stress tensor.

1.3 Void Evolution

When a sufficiently high stress level is reached at some interfacial spot, where a flaw can be assumed, the failure development enters the next phase and a different modeling ansatz must be applied. Here, we have an evolving void surface shaped by two dynamic forces: the chemical potential gradient and the electron wind.

The development of fatal voids, i.e. voids which lead to a significant resistance increase, or even completely sever the line, is the ultimate cause for the electromigration induced interconnect failure [8]. Therefore, the understanding and prediction of electromigration failure behaviour can only be achieved through a detailed study of the void evolution mechanisms.

Including both contributions, electromigration and chemical potential-driven surface diffusion, gives the total surface vacancy flux $\vec{J}_s = J_s\vec{t}$, where $\vec{t}$ is the unit vector tangential to the void surface [9]

$$\vec{J}_s = -D_s\Big(eZ^*\vec{E}_s + \Omega\nabla_s\Big(\frac{\bar{\bar{\sigma}} : \bar{\bar{\varepsilon}}}{2} - \gamma_s\kappa\Big)\Big). \quad (7)$$

$\vec{E}_s = E_s\vec{t}$ is the local component of the electric field tangential to the void surface, ∇_s is the surface gradient operator, $1/2(\bar{\bar{\sigma}} : \bar{\bar{\varepsilon}})$ is the strain energy density of the material adjacent to the void surface, and κ is the curvature of the void surface. The surface diffusivity D_s is given by an Arrhenius law:

$$D_s = \frac{D_0\,\delta_s}{k_B\,T}\exp\Big(-\frac{Q_s}{k_B\,T}\Big). \quad (8)$$

δ_s is the thickness of the diffusion layer, Q_s is the activation energy for the surface diffusion, and D_0 is the pre-exponential coefficient for mass diffusion.

Numerical tracking of the surface is necessary, since it fully determines the failure dynamics. The shape of the void together with local geometry conditions cause changes in the interconnect resistance and the speed of the void growth. Thus, the migration direction is strongly influenced by the shape of a void. The three-dimensional algorithms for simulation of the void surface evolution are computationally very demanding, and currently no satisfactory solution exists.

The Level-Set method is powerful for simulating moving boundary problems, where the moving boundary is implicitly represented by the level set of a given function. The time evolution of the level set is governed by a Hamilton-Jacobi equation. A velocity field, which accounts for the physical effects acting on the moving boundary, dictates the evolution of the level set. Several algorithms and strategies to implement the Level-Set method have been proposed [10].

For the case of the void surface evolving under the influence of electromigration and the gradient of the surface energy the Hamilton-Jacobi equation for a level set ϕ is

$$\frac{\partial\phi}{\partial t} - D_s\nabla_s \cdot \Big(eZ^*\vec{E}_s - \gamma_s\Omega\nabla \cdot \Big(\frac{\nabla\phi}{|\nabla\phi|}\Big)\Big) = 0. \quad (9)$$

The diffusion coefficient D_s has to be set according to the region where the void surface lays. This region can be bulk, grain boundary, or the interface to capping or barrier layers. The role of the surface diffusivity variation is a potential cause for a huge diversity of void shapes which have been observed in experiments.

1.4 Simulation Results

The site of void nucleation and the morphology of the evolving void accurately reproduce experimental observations. The simulation is started by assuming a grain boundary network (c.f. Fig. 2) in the studied dual-damascene interconnect.

The vacancy release rate ω_R and vacancy the trapping rate ω_T are chosen in such a way that during simulation following the conditions are fulfilled:

$$\frac{2\,\omega_R}{\omega_T(C_v^1 + C_v^2)} \ll 1, \tag{10}$$

and

$$1\,\mathrm{s} < \tau = \frac{\delta}{\omega_T(C_v^1 + C_v^2)} < 2\,\mathrm{s}. \tag{11}$$

With these conditions the model for immobile vacancies eq.(1)-eq.(6) behaves analogously to a classical Rosenberg-Ohring term [11], which was already successfully applied in [12] where $\tau = 1$ s and $\tau = 2$ s is used.

All parameters for simulation are set according to experiments published in [13]. The line width is $0.18\ \mu$m, the applied current density is 1.5 MA/cm^2, and the temperature is $300\ ^\circ$C. Barrier and capping layer are Ta/TaN and SiCN, respectively. SiOC is used as interlayer dielectricum. The continuum equations eq.(4) are solved, until a stress threshold (σ_{th}) for void nucleation is reached at some triple point (c.f. Fig. 3). At this triple point an initial, small spherical void is set and the Hamilton-Jacobi equation eq.(9) is solved.

For stresses $\sigma < \sigma_{th}$ an energy barrier exists between the void embryo and a stable-growing void. If the stress is above the threshold value ($\sigma > \sigma_{th}$), the free energy monotonically decreases with void volume and the energy barrier vanishes. If we now assume an adhesion free patch with a radius of 10 nm (about 20 atoms), we obtain $\sigma_{th} \approx 344$ MPa [5].

After the void is nucleated at the triple point (c.f. Fig. 4), which is a natural free adhesion patch, it moves in electric current direction. The void encounters the second grain boundary and transforms its shape (c.f. Fig. 5). During evolution the void remains attached to the copper/capping layer interface which is also a fast diffusivity path. The strength of adhesion between copper and SiCN determines the speed of the void evolution. By choosing other capping materials and corresponding technology processes which increase adhesion, the speed of the void is reduced and the interconnect life time is increased.

It can be clearly seen in Fig. 5 that the presence of a grain boundary induces a void surface movement toward the inner part of interconnects. Here the grain boundary acts as fast diffusivity path. This result is consistent with numerous experimental observations and explains the fact that bamboo microstructures are prone to failures caused by slit-like voids [14–16].

In the further development this shape change is increased leading to a high increase in interconnect resistance (c.f. Fig. 6). The results of simulation and experiments demonstrate a decisive impact of the microstructure on the failure development [17, 18].

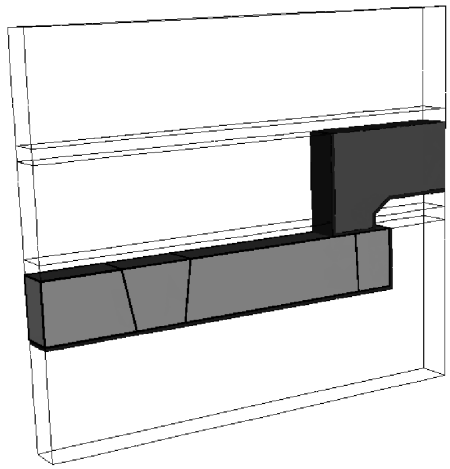

Figure 2: Dual-damascene structure used for simulation.

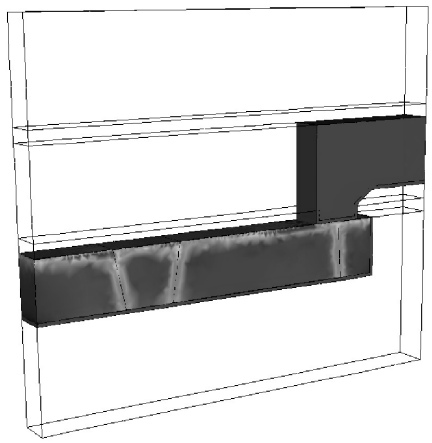

Figure 3: Peak hydrostatic tensile stress distribution (light areas) caused by electromigration.

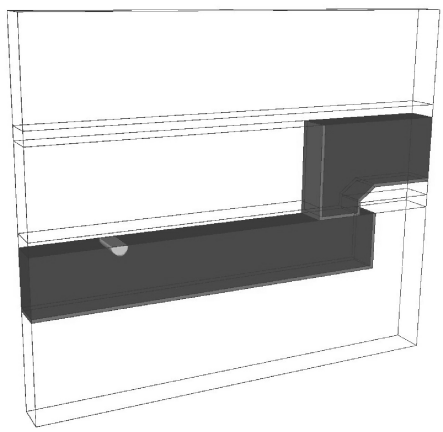

Figure 4: Initial void placed at the nucleation site.

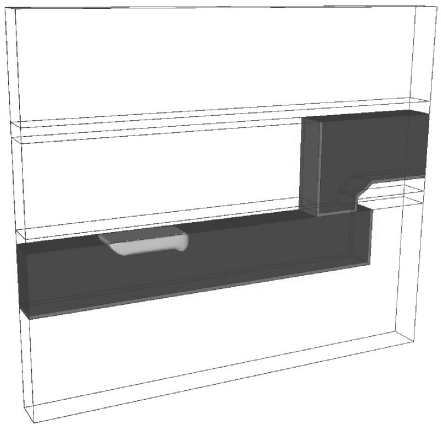

Figure 5: The void moves and encounters the second grain boundary.

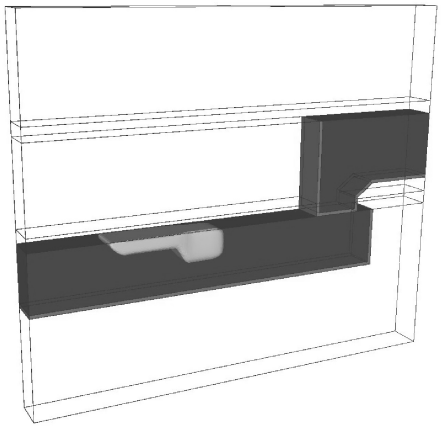

Figure 6: The shape change is enhanced as the void drifts toward the via.

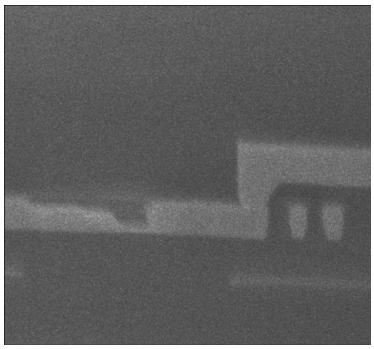

Figure 7: FIB cross-section of interconnect after an electromigration test [13] (courtesy of Dr. Lucile Arnauld).

The experimental result presented in Fig. 7 clearly shows that the applied models together with the assumed microstructure reproduce the experimentally observed void dynamics.

The distance between the grain boundaries and the angle between the grain boundaries and the copper/capping layer interfaces are crucial for void evolution. Triple points determine the void nucleation site and during the evolution the void is shaped by the grain boundaries it encounters. In order to investigate the exact correlation between the microstructural properties and the void dynamics the statistical simulation using a large number of automatically generated microstructures is necessary.

1.5 Conclusion

The ultimate cause of electromigration failure is an intrinsic void. In this work we study void nucleation and evolution for the case, when the interconnect interface to the capping layer is the dominant material transport path and the void nucleation site is at the intersection of copper grain boundary and capping layer. A detailed, physically based grain boundary model is applied to model the void nucleation process. Void evolution is simulated using a three-dimensional Level-Set algorithm. The simulation results are successfully verified through comparison with experimental observations.

2 Do NBTI-Induced Interface States Show Fast Recovery? A Study Using a Corrected On-The-Fly Charge-Pumping Measurement Technique

Data obtained by the recently developed on-the-fly charge-pumping technique has suggested a fast initial degradation and recovery of interface states during negative and/or bias temperature stress, contrary to previously published results. By revising the analysis of the measurement setup, fast interface state creation and recovery are revealed as artifact due to a different amount of oxide traps seen during the stress and relaxation phases. From this analysis we conclude that data gathered during stress and recovery phases must not be directly compared. By properly taking the contribution of (slow) oxide charges into account, which leads to a spurious increase of the charge-pumping current during the stress phase, we demonstrate that no fast initial degradation and no fast recovery of interface states occurs. Nevertheless, the charge-pumping signal is sensitive to the continuous switching of the gate voltage into accumulation, which also accelerates interface state recovery, albeit at a slower rate. We finally conclude that both the fast initial degradation and the fast initial recovery seem to be due to oxide charges. Therefore these oxide charges need to be considered. By performing simulations with our device simulator Minimos-NT using a modified Shockley-Read-Hall model it was possible to reproduce the effect of these oxide charges. For this purpose a temperature and field activated tunneling process is assumed and results in proper agreement of measurement and simulation. A correction scheme for the on-the-fly charge-pumping measurement technique is then presented.

2.1 Introduction

The last years have seen a lively debate on whether interface states and/or oxide charges are responsible for the degradation observed during negative and/or positive bias temperature stress (NBTI, PBTI). Experimental differentiation between oxide and interface states is extremely challenging due to the rapid recovery of the degradation setting in as soon as the stress is removed. In particular, it has been observed that when after NBTI stress the device is positively biased, a considerable part of the recoverable component is lost [19–22]. Until recently, this has been explained by the detrapping of holes [20,21], while interface states have been assumed to only change their occupancy but do not recover.

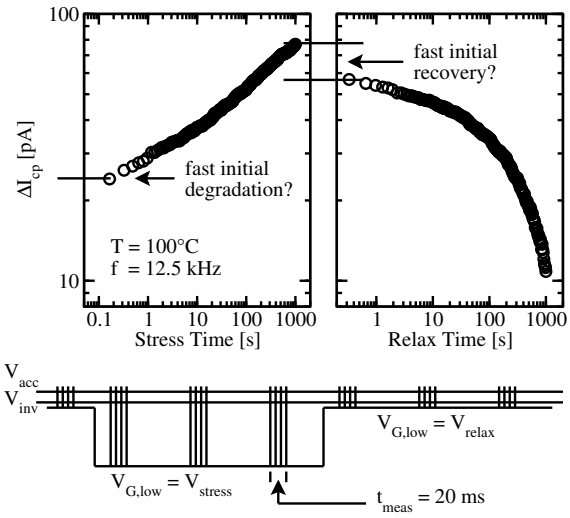

Figure 8: Typical NBTI stress and relaxation measurement of the charge-pumping current I_{cp} using the OFIT technique with a duty cycle of $50\,\%$. Each symbol in the upper graph is made up of some hundred averaged pulses as schematically displayed in the lower graph for some points. After the reference measurement $V_{\mathrm{G,low}}$ is set to V_{stress}, in order to continuously stress the device. To perform CP, V_{G} is pulsed to V_{acc} periodically. In this way the CP measurement and the application of stress are carried out consecutively. During relaxation $V_{\mathrm{G,low}} = V_{\mathrm{relax}}$. Constant slopes of the stress and relaxation pulses are ensured [23].

A quite striking result obtained with on-the-fly charge-pumping (OFIT) measurements is that in contradiction to conventional charge-pumping (CCP) measurements, OFIT data suggest a considerable amount of fast initial recovery of interface states. We note that this fast initial recovery is not explicitly measured, but is only inferred from the differences between the last stress and the first recovery measurement. Nevertheless, this misconception is a fundamental dilemma and a major issue for our understanding of BTI. Clarification of this matter is the prime requisite for the development of a reliable model. In the following we try to resolve the issue whether interface states do recover quickly ($< 1\,\mathrm{s}$) or not.

2.2 Experimental Methodology

Unfortunately, experimental validation of the fast recovery behavior is challenging because the technique that is conventionally used to directly measure interface states is the charge-pumping (CCP) technique, which inherently relies on a bias switch into accumulation. Consequently, it is unclear whether the often observed weak recovery in CP data is a consequence of the fact that interface states do not recover or whether this is an artifact of the measurement technique brought about by the strong bias switch.

The lately developed OFIT technique [24–26] has refueled the debate. As illustrated in Fig. 8, the basic difference between OFIT and CCP [27–30] is that the low-level $V_{\mathrm{G,low}}$ of the CP pulse is simultaneously used as a stress condition (for NBTI), while the actual CP measurement is performed by quickly switching back and forth between accumulation V_{acc} and stress V_{stress}. Consequently, an issue we will get back to later, the low-levels are different during stress and recovery/reference measurements.

2.3 OFIT versus CCP

As in conventional CP measurements, care has to be taken that parasitic tunneling currents and geometry effects do not pollute the measured charge-pumping current I_{cp}. The first problem is even more severe in the OFIT technique since there the low level gate voltage equals the stress voltage, resulting in excessive tunneling in thin oxides. In order to avoid these problems we also use large-area devices with thick oxides (30 nm). As shown in Fig. 9, the measured I_{cp} during stress and recovery are qualitatively identical for three completely different technologies (30 nm thick SiO_2, 3.5 nm thin SiO_2, and SiON).

Quite remarkably, continuous application of OFIT pulses (as well as CP measurements) has a dramatic impact on both the stress and the recovery characteristics. In particular, with 3 measurements per decade, I_{cp} is quasi constant during recovery, while up to 100 measurements per decade result in approximately 80% recovery of I_{cp}. Another fact worth mentioning is that the first OFIT measurement point during stress is already responsible for at least 30% of the total degradation. Likewise, the first measurement taken during recovery at 300 ms already shows 30% recovery while the rest of the recovery depends basically on the number of measurements per decade.

2.4 Analysis of the OFIT Technique

In order to deepen our understanding of the method we performed constant base-level CP measurements ($V_{\mathrm{Base}} = 2\,\mathrm{V}$) using a gradually increasing pulse amplitude ΔV_{G}. Until reaching the desired stress-level starting from $-1\,\mathrm{V}$ down to $-18\,\mathrm{V}$ the pulse slopes have to be kept constant to obtain comparable results. Constant pulse slopes ensure that the upper and lower energy boundaries of the active energy interval remain unchanged, although ΔV_{G} increases [23]. Given the requirements of a constant slope and a constant duty cycle, increasing the pulse amplitude ΔV_{G} leads to larger pulses. This means that the rise and fall times have to

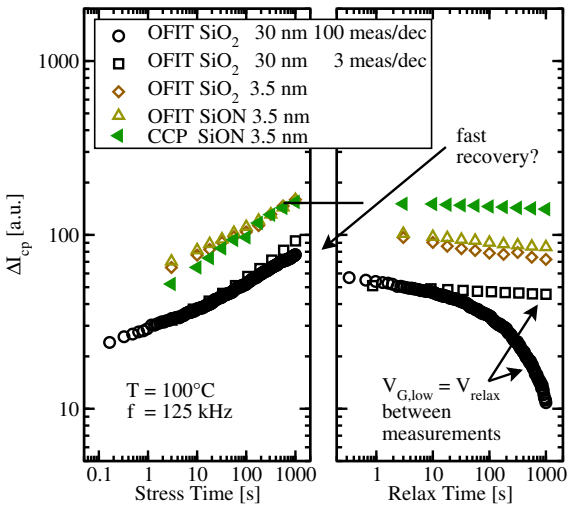

Figure 9: Comparison of OFIT and CCP results. For OFIT the offset between end of stress and beginning of relaxation is comparable for different technologies and geometries (not shown). Previously this was explained by fast recovery. This fast recovery is absent in CCP (indicated for SiON). Furthermore, continuous gate pulsing affects both stress and relaxation, causing a faster recovery of interface states with increasing number of measurements (black circles vs. black squares - both of 30 nm OFIT).

be adapted at every voltage step. Since it is inevitable to change both the pulse width and also the rise and fall times one has to ask for the potential pitfalls: Are OFIT-data obtained during stress and relaxation comparable? If that is not the case, is there some possibility to correct this nonconformity? These questions will be examined in the following.

Starting with Fig. 10 the two large arrows pointing up and down again reveal some important aspects of the temporal evolution of the pulses during a CP measurement. The charge-pumping current I_{cp} at stress conditions ($V_{\mathrm{G,low}} < -3\,\mathrm{V}$) differs a lot when compared to relaxation ($V_{\mathrm{G,low}} = -1\,\mathrm{V}$). The further the NBTI stress increases the higher the I_{cp}-signal becomes. This can be partly attributed to the desired effect of using the measurement setup to also stress the device. However, this cannot fully account for the observed behavior.

2.4.1 Dependence on Gate Voltage Low-Level

Under the assumption that only interface states N_{it} contribute to I_{cp}, I_{cp} should actually become independent of $V_{\mathrm{G,low}}$ as soon as the strong inversion regime is reached. This $I^{\mathrm{it}}_{\mathrm{cp},0}$ is marked by the dashed line in Fig. 10. However, as demonstrated previously [31, 32], I_{cp} continues to increase, albeit at a much slower rate. This increase with ΔV_{G} is routinely attributed to slower oxide traps

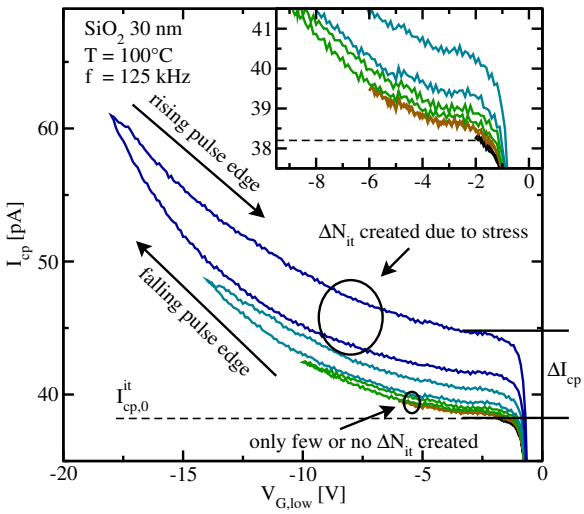

Figure 10: Charge-pumping current I_{cp} for different pulse amplitudes as observable in constant base-level CP measurements with $V_{\mathrm{Base}} = 2\,\mathrm{V}$ and a gradually increasing pulse amplitude ΔV_{G} from $V_{\mathrm{G,low}} = -1\,\mathrm{V}$ down to $V_{\mathrm{G,low}} = -17\,\mathrm{V}$. The I_{cp} shows a significant hysteresis. If the I_{cp} is evaluated at the falling pulse edge, the lower branch of the curve is traversed. Evolution of the rising pulse edges gives the upper branch. However, the contribution of slow oxide states and an additional hysteresis (marked with ΔI_{cp}) are clearly visible for increasing pulse amplitudes. This implies that depending on the pulse amplitude, I_{cp} will contain contributions in addition to the interface states. Provided only interface states are available, I_{cp} should be independent of the pulse amplitude (dashed line of $I^{\mathrm{it}}_{\mathrm{cp},0}$).

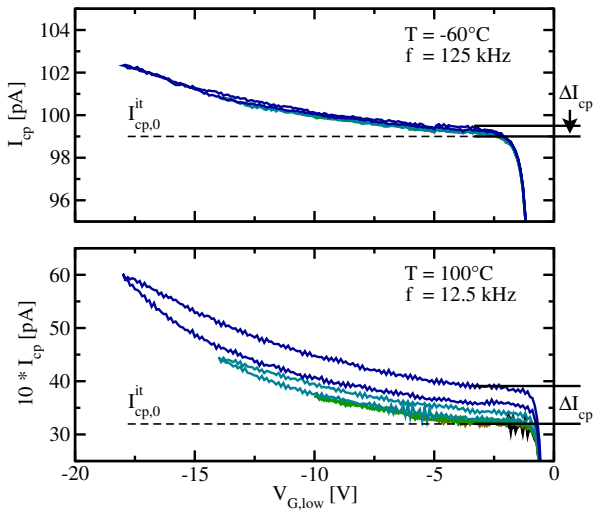

Figure 11: **Top**: At a low temperature the hysteresis is negligible (less than 1%) and the contribution of slow oxide traps is reduced. **Bottom**: At a low frequency and a high temperature the contribution of oxide traps increases due to the increased rise and fall times. A comparable if not equal part of interface states constituting ΔI_{cp} can be identified for different frequencies but equal temperature. Following these results at least part of the defects must vary with temperature or frequency. For better comparability, the data at 12.5 kHz are scaled to the reference frequency ($f_{\mathrm{ref}} = 125\,\mathrm{kHz}$).

ΔN_{ot} and $I_{\mathrm{cp}} = I^{\mathrm{it}}_{\mathrm{cp}} + I^{\mathrm{ot}}_{\mathrm{cp}}$ [29,33]. So, regardless of the amount of degradation, I_{cp} varies as function of $V_{\mathrm{G,low}}$. This fact has to be taken into account for a meaningful comparison of stress and relaxation CP data.

2.4.2 Hysteresis due to Stress

When $V_{\mathrm{G,low}}$ is lowered towards the stress voltage, as required in the OFIT technique, I_{cp} extracted from the rising and falling pulse edges start to deviate, introducing a hysteresis. The characteristic hysteresis curve is only visible for larger pulse amplitudes, indicating degradation (marked with ΔN_{it} and ΔI_{cp}) due to stress. While the impact of the oxide traps visible during medium $V_{\mathrm{G,low}}$ appears to be fully recoverable, the component causing the hysteresis is not. This can be seen in Fig. 10 where I_{cp} increases during subsequent measurements performed on the same device. We attribute this hysteresis to the creation of additional interface states due to NBTI stress at $V_{\mathrm{G,low}} = V_{\mathrm{stress}}$. Starting at $-2\,\mathrm{V}$ there is nearly no stress. The deeper the device is stressed into inversion the larger the hysteresis becomes, resulting in an increased offset for the next pulse. The total hysteresis at a certain stress level hence not only consists of the hysteresis of the current measurement but is a combination of the previous ones.

2.4.3 Hysteresis-Free Area

As displayed in the inset in Fig. 10 the very first pulses are almost free of stress (no hysteresis, $\Delta I_{\mathrm{cp}} = 0$) and hence the deviation of I_{cp} from $I^{\mathrm{it}}_{\mathrm{cp},0}$ is entirely due to oxide traps. Only a negligible amount of interface states ΔN_{it} are created by the measurement process. The hysteresis-free area will be discussed in more detail in the next section.

2.4.4 Frequency Scalability

When the experiment is repeated at a lower frequency (see bottom of Fig. 11), one finds that the interface state contribution can be scaled to the reference frequency ($f_{\mathrm{ref}} = 125\,\mathrm{kHz}$) [31]. This is compatible with the fact that the stress duration is practically independent of frequency. On the other hand, the recoverable oxide trap contribution to I_{cp} depends on frequency, consistent with the idea that the lower the frequency (corresponding to more time per pulse) the more oxide traps can contribute to I_{cp}.

2.4.5 Lower Temperature

Finally, at a low temperature (displayed at the top of Fig. 11) practically no hysteresis is introduced (no NBTI stress) and also the oxide trap contribution is reduced, consistent with the idea that these traps are due to a thermally activated tunneling mechanism [34] rather than elastic (and thus temperature-independent) hole tunneling [19].

2.5 Extrapolation of Oxide Trap Contribution

As demonstrated above, during an OFIT measurement we have to expect a distortion of I_{cp} due to oxide charges and due to the creation of defects during the low-level. In order to analyze this distortion we proceed as follows:

We determine $\hat{V}_{\mathrm{G,low}}$ to be the lowest value of $V_{\mathrm{G,low}}$ at which no hysteresis is observed. We then use the data set $V_{\mathrm{G,low}} > \hat{V}_{\mathrm{G,low}}$ to extrapolate the impact of oxide charges ΔN_{ot} down to the stress-level. It is not possible to obtain this information from the stress pulse because of the contribution of both parts: ΔN_{it} and ΔN_{ot}. Quite remarkably, the data can be fit by a quadratic polynomial, consistent with our NBTI experiments where we also observe a quadratic (E_{ox}^2) dependence of the hole-trapping component [34–36]. The hole-trapping theory developed in [34] was applied to our data and excellent agreement was obtained. The difference between the actual signal ($I_{\mathrm{cp}}^{\mathrm{it}} + I_{\mathrm{cp}}^{\mathrm{ot}}$) and the extrapolated curve in Fig. 12 finally gives ΔN_{it}.

In Fig. 12 and Fig. 13 the extraction algorithm for ΔN_{ot} and ΔN_{it} is demonstrated. Stress and relaxation pulse responses both consist of two branches, one going down (fall) and one going up (rise) as marked by arrows. In the falling branch, $V_{\mathrm{G,low}}$ varies from $0\,\mathrm{V}$ to $-17\,\mathrm{V}$. In the rising branch, $V_{\mathrm{G,low}}$ varies from $-17\,\mathrm{V}$ to $0\,\mathrm{V}$. Only pulses with constant $I_{\mathrm{cp}}^{\mathrm{rise}} - I_{\mathrm{cp}}^{\mathrm{fall}}$ (or even without a hysteresis, i.e. $I_{\mathrm{cp}}^{\mathrm{rise}} - I_{\mathrm{cp}}^{\mathrm{fall}} = 0$) are suitable to create an extrapolation guess for higher $V_{\mathrm{G,low}}$. This 'safe window' ranges from $0\,\mathrm{V}$ to $-8\,\mathrm{V}$ where both branches are indistinguishable.

The extracted components for different temperatures and frequencies are given in Fig. 14. The additionally created oxide traps ΔN_{ot} depend on frequency as well as on temperature and clearly show $V_{\mathrm{G,low}}^2 \sim E_{\mathrm{ox}}^2$ behavior. The hysteresis due to additionally created traps ΔN_{it} is independent of frequency but strongly dependent on temperature.

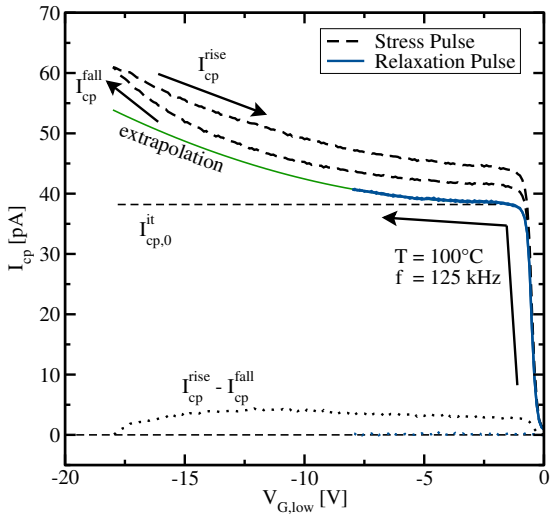

Figure 12: Charge-pumping current I_{cp} for the stress pulse ($V_{\mathrm{stress}} = -17\,\mathrm{V}$) and the relaxation pulse ($V_{\mathrm{relax}} = -8\,\mathrm{V}$), shown in Fig. 10. To unravel the contribution of oxide charges and additional interface states we look at the difference $I_{\mathrm{cp}}^{\mathrm{rise}} - I_{\mathrm{cp}}^{\mathrm{fall}}$. In the range $-8\,\mathrm{V} < V_{\mathrm{G,low}} < 0\,\mathrm{V}$, this difference is constant, implying no additional creation of interface states. From this 'safe window' we extrapolate to the minimum low-level to estimate the contribution due to oxide charges. Note that the first branches $I_{\mathrm{cp}}^{\mathrm{fall}}$ of the stress and relaxation pulse differ from each other due to pre-stress pulses between $V_{\mathrm{G,low}} = -8\,\mathrm{V}$ and $V_{\mathrm{G,low}} = -17\,\mathrm{V}$. In fact, when using fresh devices for each measurement all $I_{\mathrm{cp}}^{\mathrm{fall}}$ would coincide.

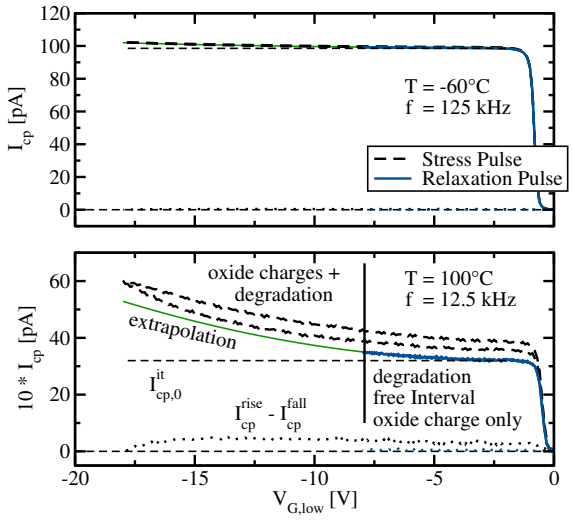

Figure 13: **Top**: Lower temperatures simplify the extrapolation due to the absence of degradation. Here the full range of pulse amplitudes can be used to verify the extrapolation down to deep inversion. Very good conformity is obtained, again indicating the absence of additional interface states in deep inversion at low temperatures. **Bottom**: Noise complicates this procedure at low frequencies. Data are scaled to $f_{\mathrm{ref}} = 125\,\mathrm{kHz}$.

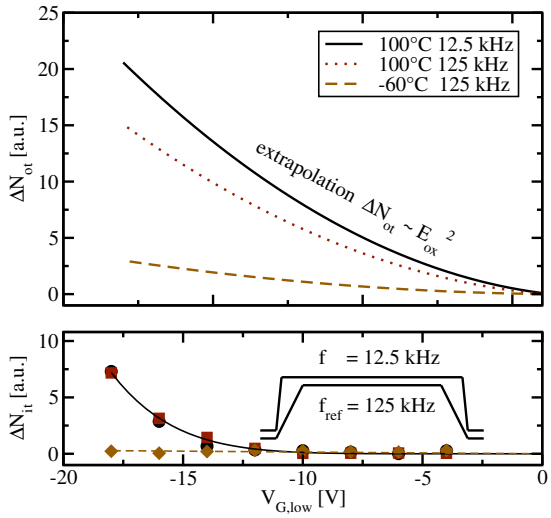

Figure 14: The extracted oxide states (**Top**) and additional interface states (**Bottom**). The change of oxide trap density ΔN_{ot} follows E_{ox}^2 and furthermore depends on the frequency as well as on the temperature. The hysteresis displayed in the previous figures are due to additionally created traps, ΔN_{it}, which are independent of the frequency but strongly dependent on the temperature.

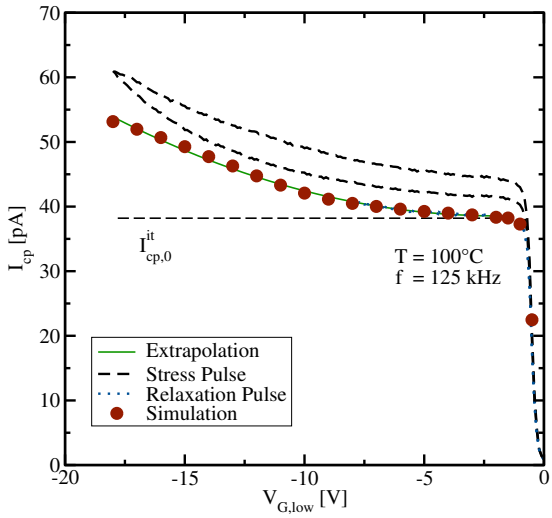

Figure 15: The contribution due to oxide traps can be well described using the model suggested in [34]. This model assumes that hole-trapping is possible via a multiphonon process implying a thermally activated barrier $\exp(-\Delta E_{\mathrm{B}}/\mathrm{k_B}T)$ and a $\exp(E_{\mathrm{ox}}^2/E_{\mathrm{ox,ref}}^2)$ field dependence. We ascertain that the simulation only describes oxide traps and interface traps without applied stress conditions. Additional interface traps due to NBTI stress are missing in the simulation because of a constant number of interface traps for each taken simulation point (solid circles).

2.6 Simulation of The Charge-Pumping Current

To approximately account for the temperature and field activated tunneling process, we use a modified Shockley-Read-Hall (SRH) model with our device simulator Minimos-NT [37]. The SRH-capture-rates are multiplied by

$$\exp\left(\frac{E_{\mathrm{ox}}^2}{E_{\mathrm{ox,ref}}^2}\right)\exp\left(-\frac{\Delta E_{\mathrm{B}}}{k_{\mathrm{B}}T}\right)$$

where E_{ox} is the electric field in the oxide, $E_{\mathrm{ox,ref}}$ is a reference value, ΔE_{B} the multiphonon emission barrier and T the temperature. ΔE_{B} can be characterized by a Gaussian distribution with the mean energy ΔE_{B}. When setting the parameters some points need to be considered in order to end up with a physically appropriate model:

1. The first exponential factor models the bias dependence. It is very sensitive to changes of $E_{\mathrm{ox,ref}}$ due to the squared exponent, leading to a very small range of valid $E_{\mathrm{ox,ref}}$ values.

2. When setting the barrier ΔE_{B} too low, the oxide traps contribute to the interface trap signal as the second factor approaches unity. Setting ΔE_{B} too high leads to very low rates, effectively eliminating the contribution of oxide traps.

3. Lastly, the distribution of ΔE_{B} determines the dependence of I_{cp} on $V_{\mathrm{G,low}}$. Increasing the mean of the distribution at ΔE_{B} increases the mean capture/emission-time constants. Since with constant-slope pulses higher pulse amplitudes ΔV_{G} require longer pulse durations, increasing the mean ΔE_{B} shifts the point from which a significant contribution of oxide traps ΔN_{ot} can be observed to higher pulse amplitudes. On the other hand, broadening the distribution of ΔE_{B} (increasing the variance) also broadens the distribution of time constants, observable as broadening the range of $V_{\mathrm{G,low}}$ where I_{cp} increases. A quadratic behavior as observed in the experiments, Fig. 15, can be well reproduced with a broad Gaussian distribution ($\Delta E_{\mathrm{B,mean}} = 1\,\mathrm{eV}, \Delta E_{\mathrm{B},\sigma} = 0.5\,\mathrm{eV}$).

Simulation results are depicted in Fig. 15. As the simulation treats the CP measurement process as stress-free, no additional interface traps due to stress are created and only the oxide-charge part is visible. With the thermally activated barrier the increasing ΔI_{cp} can be described.

2.7 Results

Based on the previous results we are now able to better understand the charge-pumping current I_{cp} measured during the OFIT sequence. The presence of additional charges contributes to the signal when the pulse amplitude ΔV_{G} is increased. A large spread of time constants slower than that of the interface states is necessary to explain the results. By assuming oxide traps with a thermally distributed activated barrier one is able to explain the measurement results with good accuracy. Whereas interface states are independent of the electric field and account for I_{cp} at low (10 kHz) and high frequencies (1 MHz) likewise due to their small time constants, the oxide traps with their assumed barrier are by far slower, only affecting I_{cp} at lower frequencies i.e. 10 kHz.

The particularly troublesome part is the application of the OFIT technique during the stress phase, where both oxide traps ΔN_{ot} and additionally created interface states ΔN_{it} add to I_{cp}. These contributions are absent during the initial reference measurements and during the OFIT recovery measurements both taken at $V_{\mathrm{G,low}} = V_{\mathrm{relax}}$. This has fundamental consequences on OFIT measurements: Initially, a reference I_{cp} is recorded. Following this reference measurement, the gate voltage low-level $V_{\mathrm{G,low}}$ is switched to stress V_{stress}. Due to the much larger ΔV_{G} now a significant contribution of $I_{\mathrm{cp}}^{\mathrm{ox}}$ is obtained.

Furthermore, with the large pulse amplitude, additional interface states are created, which is the intended effect of this OFIT measurement. However, without this the additional increase in I_{cp} due to oxide traps must not be attributed to interface states created by degradation. Consequently, we need to correct for $I_{\mathrm{cp}}^{\mathrm{ot}}$ in the measurement data. Using the mentioned extrapolation method of $\Delta N_{\mathrm{ot}} = AE_{\mathrm{ox}}^2$ we see that the 30% initial increase in I_{cp} is entirely due to oxide traps. The corrected last stress value in Fig. 16 is identical to the first value at the recovery, leading to the conclusion that no fast interface state recovery occurs.

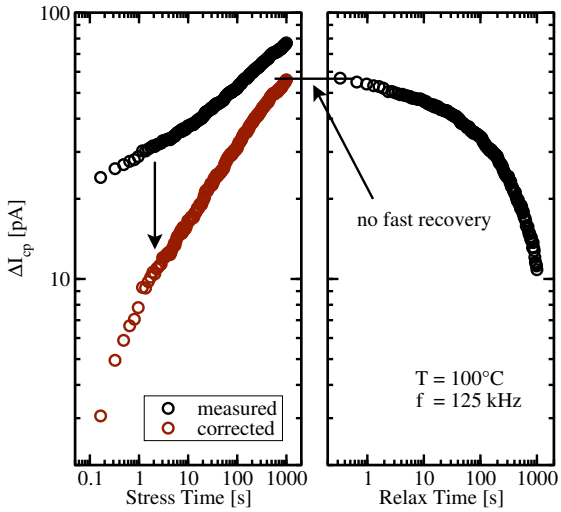

Figure 16: Oxide traps lead to a spurious increase in the charge-pumping signal during stress. Using the scheme developed for Fig. 8, a corrected I_{cp} is obtained. The smooth transition between the corrected I_{cp} during stress and the I_{cp} during recovery suggests that no fast recovery takes place.

2.8 Conclusions

We have discussed the dynamics of interface state creation in constant base-level charge-pumping measurements using a gradually increasing pulse amplitude ΔV_{G} (from inversion into deep inversion). The charge-pumping current is not constant once reaching the inversion regime, but increases due to slow oxide traps. From this analysis we conclude that data gathered during stress and recovery phases must not be directly compared. A correction scheme for the OFIT measurement technique was necessary. By properly taking the contribution of these traps into account, we demonstrate that no fast initial degradation and no fast recovery of interface states occurs. Nevertheless, the CP signal is sensitive to continuous switches of the gate voltage into accumulation which also accelerates interface state recovery, albeit at a slower rate. We finally conclude that both the fast initial degradation and the fast initial recovery are very likely to be due to oxide charges which are created by a thermally activated process.

3 Nonparabolicity Effects in Quantum Cascade Lasers

We calculate electron-LO phonon and interface roughness scattering rates in a GaAs/$Al_xGa_{1-x}As$ quantum cascade laser taking into account conduction subband nonparabolicity. In this work we investigate the Al concentration and the $\mathbf{k}_\parallel$ dependence of nonparabolicity effects. It is shown that the subband nonparabolicity may increase the electron-LO phonon scattering rates significantly, but the scattering rates are not qualitatively different from those in the parabolic approximation. Especially, we show that nonparabolicity leads to noticeable changes even for transitions involving electrons at the bottom of the subband and that this behaviour follows from the phonon wave vector and the electron phonon overlap.

3.1 Introduction

Over the past several years, solid-state lasers based on intersubband transitions in semiconductor heterostructures have proved to be very promising candidates for practical sources of radiation, particularly in the midinfrared region [38]. While the possibility of optical amplification by intersubband transitions in biased superlattices had already been predicted in 1971 [39], a successfully working quantum cascade laser (QCL) has first been reported in 1994 [40].

Understanding the physics in these heterostructures is necessary to design and fabricate new QCL's with specific properties. In this work we adress the question how the subband nonparabolicity affects the electron-LO phonon scattering rates as well as the interface roughness scattering rates in a QCL subjected to barrier characteristics.

GaAs/$Al_xGa_{1-x}As$ active regions grown on GaAs substrates are ideal candidates for this approach as $Al_xGa_{1-x}As$ is nearly perfectly lattice matched to the GaAs substrate.

Frequently used theoretical models describing the electron properties in a semiconductor approximate the band structure by parabolic conduction bands. Deviations from the parabolic model are not considered in quantum well (QW) calculations based on effective mass approaches [41].

Here we investigate the influence of conduction band nonparabolicity on the intersubband scattering rates due to electron-LO phonon interactions and interface roughness, considering transitions with different initial energies. In order to determine for which situations the subband nonparabolicity can not be neglected, we show results of scattering rates for different $k_\parallel$ and make comparisons with the parabolic approximation. Our results demonstrate the significance of nonparabolicity effects in heterostructure lasers even for transitions where the initial state is at the bottom of the subband, particularly for scattering rates due to electron-LO phonon interaction, in situations of large electronic confinement. We identify the physical origins of the changes induced by conduction subband nonparabolicity in dependence on the Al content and the in-plane wave vector $\mathbf{k}_\parallel$ to be found in terms of the phonon wave vector and the electron phonon overlap.

In Sec. II we discuss the formal theory of the electron-LO phonon scattering rates, the interface roughness scattering model and the inclusion of the conduction band nonparabolicity in the effective mass equation. Our results are presented and discussed in Sec. III. Finally, a summary and conclusion is given in Sec. IV.

3.2 Theoretical Model

We consider electrons in the conduction band of a QCL in an electric field applied in the direction perpendicular to the heterointerfaces. In order to describe band nonparabolicity we follow the model proposed by Nag and Mukhopadhyay [42]. The nonparabolic $E-\mathbf{k}$ relation is given by

$$E - V(x) = \frac{\hbar^2 k^2}{2m^\star(x)}(1-\gamma(x)k^2) \tag{12}$$

where $\gamma(x)$ is the nonparabolicity parameter, $m^\star(x)$ is the effective mass, and $V(x)$ is the conduction band offset.

The envelope function in the QW structure is given by

$$\Psi_{i,k}(\mathbf{x}) = \frac{1}{\sqrt{A}} e^{i\mathbf{k}_\parallel\cdot\mathbf{x}_\parallel}\Phi_i(z) \tag{13}$$

where $\mathbf{k}_\parallel$ is the in-plane wave vector, A is the cross-sectional area of the QW structure, and $\Phi_i(z)$ is the electron envelope function of the ith state. The effective mass equation for the QW structure is given by

$$\left[-\frac{\hbar^2}{2}\frac{\partial}{\partial z}\frac{1}{m^\star_\perp(z,E)}\frac{\partial}{\partial z} + V(z) + \frac{\hbar^2 k_\parallel^2}{2m^\star_\parallel(z,E)} - eFz\right]\Phi_i(z) = E_i(k_\parallel)\Phi_i(z) \tag{14}$$

where the nonparabolic effective masses along the perpendicular and parallel directions are defined as [43]

$$m^\star_\perp(x,E) \equiv \frac{m^\star(x)}{\alpha(x)}[1-(1-2\alpha(x))^{1/2}] \tag{15}$$

$$m^\star_\parallel(x,E) \equiv \frac{m^\star(x)}{(1-2\alpha(x))^{1/2}} \tag{16}$$

where $\alpha = 4\gamma(x)m^\star(x)(E - V(x))/\hbar^2$.
The Hamiltonian describing the electron-phonon interaction can be written in the form [44]

$$\hat{H}_{ep} = \sum_{i,j}\sum_{k,k'}\sum_{q}(F^{ij}_{k,k',q}c^\dagger_{i,k}b_q c_{j,k'} + F^{ij\star}_{k,k',q}c^\dagger_{j,k'}b^\dagger_q c_{i,k}) \tag{17}$$

with

$$\begin{aligned} F^{ij}_{k,k',q} &= F^{LO}_q \int \Psi^\star_{i,k}(\mathbf{x})e^{i\mathbf{q}\cdot\mathbf{x}}\Psi_{j,k'}(\mathbf{x})d\mathbf{x} \\ &= F^{LO}_q \delta_{\mathbf{k}'_\parallel,\mathbf{k}_\parallel-\mathbf{q}_\parallel}\int \Phi^\star_i(z)e^{iq_z z}\Phi_j(z)dz \end{aligned} \tag{18}$$

where $\mathbf{q}$ is the phonon wave vector, and $c^\dagger_{i,k}$ ($b^\dagger_q$) and $c_{i,k}$ (b_q) denote creation and annihilation operators of the electron (phonon), respectively. F^{LO}_q represents the coupling factor for the electron-LO phonon interaction

$$F^{LO}_q = -i\frac{e}{q}\left(\frac{\hbar\omega_{LO}}{2V\epsilon_0}(\epsilon_\infty^{-1} - \epsilon_S^{-1})\right)^{1/2} \tag{19}$$

where ϵ_∞ and ϵ_S are high-frequency and static dielectric constants, V is the volume and e is the electronic charge. The electron-LO phonon scattering rate for an electron initially in a state i to the final state j is obtained from the Fermi Golden Rule

$$W^\pm_{ij}(k_\parallel) = \frac{2\pi}{\hbar}\sum_{k'}|\langle j|\hat{H}_{ep}|i\rangle|^2\delta(E_j(k'_\parallel) - E_i(k_\parallel) \mp \hbar\omega_q) \tag{20}$$

by evaluating the matrix element in eq.(20) with the electron-phonon interaction Hamiltonian given by (17)

$$\begin{aligned} W^\pm_{ij}(k_\parallel) = -\frac{2\pi}{\hbar}\sum_{k'_\parallel,q_z}\Bigg[&\frac{\hbar\omega_{LO}}{2V}(\epsilon_\infty^{-1} - \epsilon_S^{-1}) \\ &\times\int\int\frac{e^2}{k'^2_\parallel + k^2_\parallel - 2k'_\parallel k_\parallel\cos\theta + q^2_z} \\ &\times\Phi^\star_i(z)e^{iq_z z}\Phi_j(z)d\theta dz\,(\frac{1}{2}\pm\frac{1}{2} + N_{LO}) \\ &\times\delta(E_j(k'_\parallel) - E_i(k_\parallel) \mp \hbar\omega_q)\Bigg] \end{aligned} \tag{21}$$

$\pm$ denoting emission and absorption processes, and N_{LO} is the phonon occupation number given by

$$N_{LO} = \frac{1}{e^{\hbar\omega_{LO}/k_B T} - 1} \tag{22}$$

The roughness of interfaces in a QW leads to spatial fluctuations in the width of the well, and consequently to fluctuations of the confinement energy. These fluctuations of the quantization energy act as a fluctuating potential for the motion of confined carriers [45].
The randomness of the interface is described by a correlation function at the in-plane position $\mathbf{r} = (x, y)$ which is usually taken to be Gaussian with a characteristic height of the roughness Δ, and a correlation length Λ representing a length scale for fluctuations of the roughness along the interface [46], such that

$$\langle\Delta(\mathbf{r})\Delta(\mathbf{r}')\rangle = \Delta^2 e^{-|\mathbf{r}-\mathbf{r}'|^2/\Lambda^2} \tag{23}$$

The perturbation in the potential $V(z)$ due to a position shift $\Delta(\mathbf{r})$ is given by

$$\delta V = V[z - \Delta(\mathbf{r})] - V(z) \approx -\Delta(\mathbf{r})\frac{dV(z)}{dz} \tag{24}$$

For the Ith interface, which is centered about the plane z_I and extends over the range $[z_{L,I}, z_{R,I}]$, the scattering matrix element can be defined as

$$\begin{aligned} M_{ji,I} &= \left\langle j\left|\delta V\,\mathrm{rect}(\frac{z - z_I}{z_{R,I} - z_{L,I}})\right|i\right\rangle \\ &= \frac{\varphi^2_{ji,I}}{A}\int\Delta(\mathbf{r})e^{i(\mathbf{k}_\parallel - \mathbf{k}'_\parallel)\cdot\mathbf{r}} \end{aligned} \tag{25}$$

where $|j\rangle$ and $|i\rangle$ denote the final and initial wave functions, respectively. $\varphi_{ji,I}$ is defined as

$$\varphi_{ji,I} = \int\Phi^\star_j(z)\frac{dV}{dz}\mathrm{rect}\left(\frac{z - z_I}{z_{R,I} - z_{L,I}}\right)\Phi_i(z)dz \tag{26}$$

and the rectangular function reads

$$\mathrm{rect}(z) = \begin{cases} 1, & |z| \le 0.5 \\ 0, & |z| > 0.5 \end{cases}$$

The expectation value of the square of the matrix element is given by

$$\langle|M_{ji,I}|^2\rangle = \frac{\varphi^2_{ji,I}}{A^2}\iint\langle\Delta(\mathbf{r}')\Delta(\mathbf{r})\rangle e^{i(\mathbf{k}_\parallel - \mathbf{k}'_\parallel)\cdot(\mathbf{r}-\mathbf{r}')}d\mathbf{r}'d\mathbf{r} \tag{27}$$

Making use of eq.(23) and Fermi's Golden Rule, the interface roughness induced scattering rates are given by [47]

$$\begin{aligned} W^{IR}_{ij}(k_\parallel) = &\sum_I\frac{m^\star\Delta^2\Lambda^2\mid\varphi_{ji,I}\mid^2}{\hbar^3}\delta(E_{j,k'_\parallel} - E_{i,k_\parallel}) \\ &\times\int_0^\pi e^{-(\mathbf{k}'_\parallel - \mathbf{k}_\parallel)^2\Lambda^2/4}d\theta \end{aligned} \tag{28}$$

where $\mathbf{k}'_\parallel$ and $\mathbf{k}_\parallel$ are the final and initial wave vectors, respectively, and θ is the scattering angle.

3.3 Results and Discussion

We consider a GaAs/Al$_x$Ga$_{1-x}$As QCL design which comprises a three-level sheme [48]. The conduction band profile for an electric field of 48 kV/cm is given in Fig. 17.

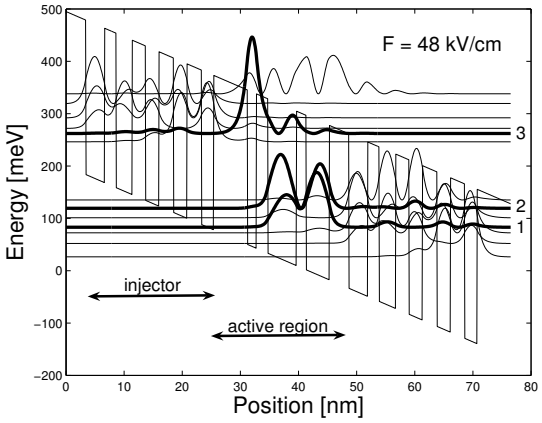

Figure 17: A schematic diagram of the conduction band profile for one and a half periods of the GaAs/$Al_{0.3}Ga_{0.7}As$ QCL for an electric field of 48 kV/cm. The layer sequence of one period, in nanometers, is: **3.4**, 3.2, **2**, 2.8, **2.3**, 2.3, **2.5**, 2.3, **2.5**, 2.1, **5.8**, 1.5, **2**, 4.9, **1.7** and 4, where normal scripts denote the wells and bold the barriers.

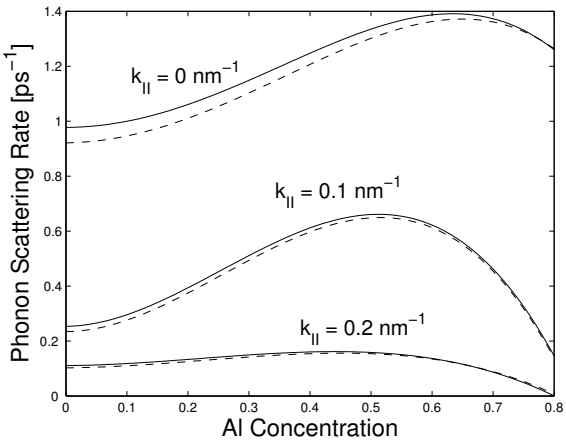

Figure 18: Intersubband electron-LO phonon scattering rate for the transition $2 \rightarrow 1$ in a GaAs/$Al_xGa_{1-x}As$ structure as a function of the Al concentration x. Solid lines contain nonparabolicity and dashed lines are for parabolic bands.

The conduction band offset is determined as a function of the aluminum concentration x by the standard approximation [49]

$$\Delta E_c = \begin{cases} 0.75x\ eV, & x \in [0, 0.45] \\ 0.75x + 0.69(x - 0.45)^2\ eV, & x \in [0.45, 1] \end{cases}$$

The expression for the nonparabilicity parameters in terms of the energy gap were obtained from Ref. [50]

$$\gamma(x) = \frac{\hbar^2}{2m^\star(x)E_g(x)} \tag{29}$$

where the energy bandgap is $E_g(x) = (1.424 + 1.247x)m_0$ in the range $0 \leq x \leq 0.45$, and $E_g(x) =$

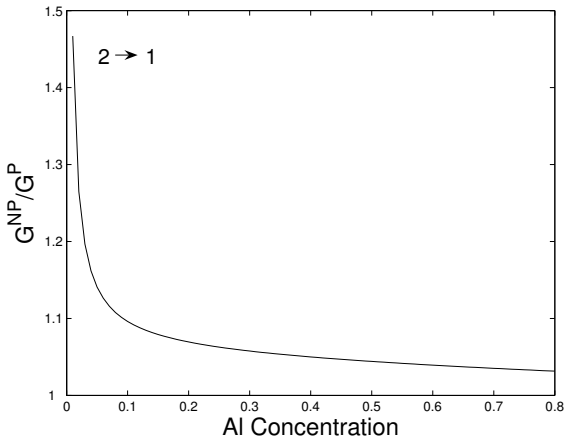

Figure 19: Ratio of the nonparabolic and parabolic electron-phonon overlap integral as a function of aluminum content.

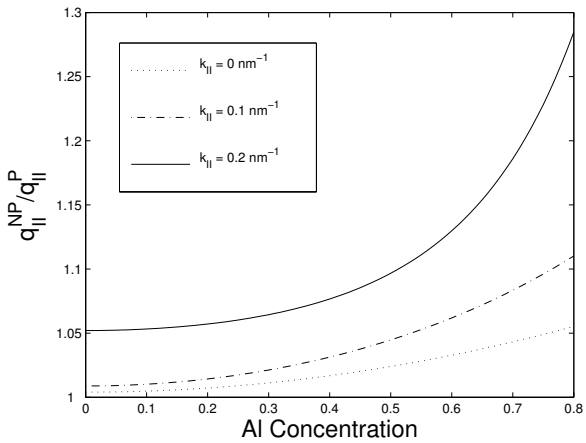

Figure 20: Ratio of the averaged nonparabolic and parabolic phonon wave vector for the intersubband transition $2 \rightarrow 1$.

$1.900 + 0.125x + 0.143x^2$ for $x > 0.45$. The material parameters used in our calculations are for $Al_xGa_{1-x}As$ [51], the effective mass $m^\star(x) = 0.067 + 0.083x$, the dielectric constants $\epsilon_S = 13.18 - 3.12x$ and $\epsilon_\infty = 10.89 - 2.73x$, and the LO-phonon energies $\hbar\omega_{LO} = 36.25 - 6.55x + 1.79x^2$ meV. The characteristic height and the correlation length of the interface roughness are taken as $\Delta = 2.83$ Å and $\Lambda = 70$ Å.

In Fig. 18 we show the calculated electron-LO phonon scattering rates for the intersubband transition $2 \rightarrow 1$ as a function of the Al content for $k_\parallel = 0$ nm^{-1}, $k_\parallel = 0.1$ nm^{-1} and $k_\parallel = 0.2$ nm^{-1}. The scattering rates with the inclusion of subband nonparabolicity are represented by solid lines and the dashed lines are for the parabolic band approximation. In general, the scattering rates are increased due to effects of nonparabolicity, except for intersubband transitions for large aluminum concentration, but the scattering rates are otherwise not qualitatively different from those in the parabolic band approx-

imation. The variations of the electron-LO phonon scattering rates due to nonparabolicity are for transitions at the bottom of the subband of almost the same order as for transitions with higher initial energies. In the region of low Alconcentration the nonparabolicity may cause a variation of about 8% for transitions initially at $k_{\|} = 0$ nm^{-1}, $\sim 10\%$ at $k_{\|} = 0.1$ nm^{-1} and $\sim 11\%$ at $k_{\|} = 0.2$ nm^{-1}.

For larger aluminum content the effects due to subband nonparabolicity for transitions initially at the bottom of the subband are even higher than for initial conditions away from it.
The nonparabolicity induced increase in the density of final states and of the quantum confinement, resulting in a larger electron-phonon overlap, and the influence of the subband on the phonon wave vector are mainly responsible for the variations in the transition rates. Since these variations are negligible for large x, the confinement decreases with the increase of the Al concentration and the effects of the nonparabolicity on the overlap integral become weaker. This is illustrated in Fig. 19, where the ratio of the nonparabolic and parabolic electron-phonon overlap for the intersubband transition $2 \rightarrow 1$ is shown.
Since the electron-phonon interaction has a strong dependence on the phonon wave vector ($\propto 1/q$) it is crucial to investigate its behavior due to nonparabolicity in order to understand the influence of nonparabolic subbands on the electron-LO phonon scattering rates. Fig. 20 shows that for nonparabolic subbands the phonon wave vector is larger than for parabolic subbands and that this effect becomes more pronounced for regions with higher in-plane wave vector of the electron $\mathbf{k}_{\|}$. Thus the electron couples more weakly to the phonon for nonparabolic subbands yielding a decrease of the electron-LO phonon

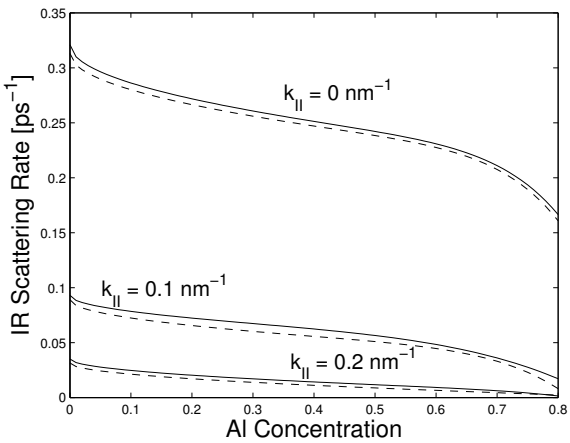

Figure 21: Interface roughness induced scattering rate for the transition $2 \rightarrow 1$ as a function of the Al content in the barrier. Solid lines are transition rates with subband nonparabolicity and dashed lines are for parabolic bands.

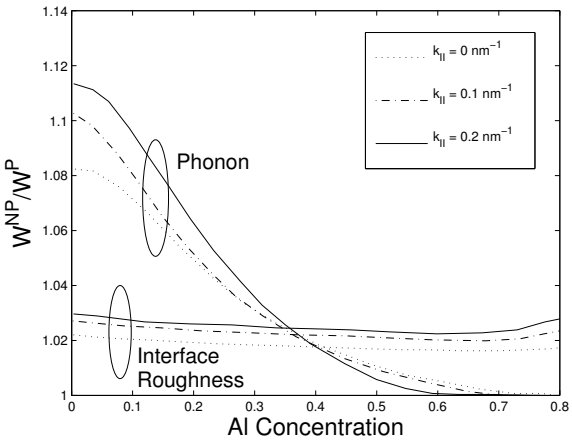

Figure 22: Ratio of the nonparabolic and parabolic electron-LO phonon and interface roughness induced scattering rates as a function of the Al content.

scattering rates, which is more affected in case of larger phonon wave vectors. Although in situations of weak confinement the role of the phonon wave vector becomes more important, the decrease of scattering rates resulting from larger phonon wave vectors is compensated by the electron-phonon overlap integral. Hence these rates are generally larger for nonparabolic subbands. This explains why the variations in

the electron-LO phonon scattering rates due to inclusion of nonparabolicity with higher initial electron in-plane vectors are of almost the same order as for $k_{\|} = 0$ nm^{-1} and in some situations of weak confinement even smaller.
The calculated scattering rates due to interface roughness as a function of Al concentration are illustrated in Fig. 21. We show results for transitions where the initial state is at the bottom of the subband and away from it. In general, we observe that the variations on the interface roughness induced scattering rates due to nonparabolicity are approximately constant for all values of x. Numerically estimated we obtain variations of the order of 2-3 %. The effects of subband nonparabolicity on scattering rates due to electron-LO phonon interaction and interface roughness are better illustrated in Fig. 22 where we display the ratio of the parabolic and nonparabolic scattering rates.

3.4 Conclusion

In conclusion, we have calculated the scattering rates of the intersubband transition $2 \rightarrow 1$ due to electron-LO phonon interaction and interface roughness in a GaAs/Al$_x$Ga$_x$As QCL at three different initial state conditions, namely at the bottom of the subband $k_{\|} = 0$

nm^{-1}, $k_{\parallel} = 0.1\ \mathrm{nm}^{-1}$ and $k_{\parallel} = 0.2\ \mathrm{nm}^{-1}$. We find that the electron-LO phonon scattering rates are in general increased, except in situations of low confinement, which is equivalent to a large aluminum concentration. In particular, in situations of low aluminum concentration the results indicate the importance of including nonparabolicity in the conduction band due to a noticeable increase of the electron-LO phonon scattering rates even for transitions at the bottom of the subband, as the variations are of almost the same order as for initial states with larger electron in-plane vectors. We have shown that these subband nonparabolicity effects can be understood in terms of the electron-phonon overlap and the phonon wave vector.

4 Valley Splitting in Thin Silicon Films from a Two-Band k·p Model

4.1 Introduction

The rapid increase in computational power and speed of integrated circuits is supported by the continuing size reduction of semiconductor devices' feature size. Thanks to constantly introduced innovative changes in the technological processes the miniaturization of MOSFETs institutionalized by Moore's law successfully continues. The $45\mathrm{nm}$ MOSFET process technology by Intel [52] and recently introduced $32\mathrm{nm}$ technology [53] involve new high-k dielectric/metal gates and represents a major change in the technological process since the invention of MOSFETs. Although alternative channel materials with a mobility higher than in Si were already investigated [54], [55], it is believed that Si will still be the main channel material for MOSFETs beyond the $32\mathrm{nm}$ technology node. 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 one of the most attractive solutions to increase the device speed. It will certainly maintain its key position among possible technological innovations for the future technology generations. In addition, new device architectures based on multigate 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 $22\mathrm{nm}$ 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 highspeed operation with excellent channel control, reduced leakage currents, and low power budget. Confining carriers within a thin film reduces the channel dimension in transversal direction, which further improves gate channel control. At the same time the search for post-CMOS device concepts has accelerated. Spin as a degree of freedom is promising for future nanoelectronic devices and applications. A concept of a racetrack memory recently proposed in [56] is based on the controlled domain wall movement by spin-polarized current in magnetic nanowires. Silicon, the main element of microelectronics, possesses several properties attractive for spintronic applications. Silicon is composed of nuclei with predominantly zero spin and is characterized by small spin-orbit coupling. In a recent ground-breaking experiment coherent spin transport through an undoped silicon wafer of $350\mu\mathrm{m}$ length was demonstrated [57]. The experiment was possible due to a unique injection and detection technique of polarized spins delivered through thin ferromagnetic films. Spin coherent propagation at such long distances makes the fabrication of spinbased switching devices likely already in the near future. Spin-controlled qubits may be thought of as a basis for upcoming logic gates. However, the conduction band of silicon contains six equivalent valleys, which is a source of potentially increased decoherence. For successful applications the degeneracy between the valleys must be removed and become larger than the spin Zeeman splitting. Shubnikov-de-Haas measurements in an electron system composed of thin silicon films in Si-SiGe heterostructures reveal that the valley splitting is small [58]. At the same time, recent experiments on the conductivity measurements of point contacts created by confining a quasi-twodimensional electron system in lateral direction with the help of additional gates deposited on the top of the silicon film demonstrate a splitting between equivalent valleys larger than the spin splitting [58]. In this work we demonstrate that a large valley splitting in the confined electron system can be induced by a shear strain component. Our analysis is based on the two-band **k·p** model for the conduction band in silicon. The parabolic band approximation usually employed for subband structure calculations of confined electrons in Si inversion layers is insufficient in ultra-thin Si films. The two-band **k·p** model includes strain and is shown to be accurate up to energies of $0.5\mathrm{eV}$. This model can therefore be used to describe the subband structure in thin silicon films, where the subband quantization energy may reach a hundred meV.

We first describe the subband structure in a thin unstrained silicon film. We demonstrate that the peculiarities of the subband dispersion obtained within the two-band **k·p** model result in a linear dependence of the valley splitting on the magnetic field. We show that a large valley splitting is observed in experiments on conduction quantization through a quantum point contact in [110] direction, but the splitting is suppressed in [100] point contacts. Finally, we demonstrate that the valley splitting is greatly enhanced in films strained in [110] direction.

4.2 Two-Band k·p Model

The closest band to the lowest conduction band Δ_1 near its minimum is the second conduction band Δ_2. These two bands are degenerate exactly at the X point. Since the minimum of the lowest conduction band in unstrained silicon is only $k_0 = 0.15 \cdot 2\pi/a$ away from the X point, where a is the lattice constant of unstrained silicon, the two bands must be included on equal footing in order to describe the dispersion around the minimum. More distant bands separated by larger gaps are included in the second order **k·p** perturbation theory [62], which results in the following two-band **k·p** Hamiltonian:

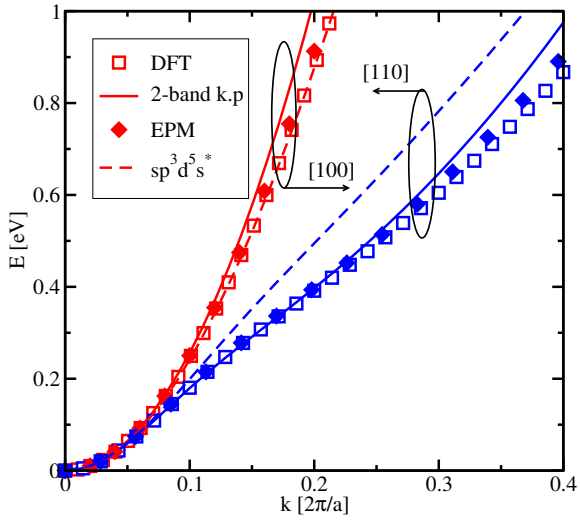

Figure 23: Comparison of the conduction band of silicon computed with the density functional theory (DFT) [59], the empirical pseudo-potential method (EPM) [60], the $sp^3d^5s^*$ tight-binding method [61], and the two-band **k·p** model [62]. The two-band **k·p** model is accurate up to an energy of $0.5\,\mathrm{eV}$.

$$H = \left(\frac{\hbar^2 k_z^2}{2m_l} + \frac{\hbar^2(k_x^2+k_y^2)}{2m_t}\right) I + \left(D\varepsilon_{xy} - \frac{\hbar^2 k_x k_y}{M}\right)\sigma_x + \frac{\hbar^2 k_z k_0}{m_l}\sigma_z \tag{30}$$

where $\sigma_{x,z}$ are the Pauli matrixes, I is the 2×2 unity matrix, m_t and m_l are the transversal and the longitudinal effective masses, respectively, ε_{xy} denotes the shear strain component, $M^{-1} \approx m_t^{-1} - m_0^{-1}$, and $D = 14\,\mathrm{eV}$ is the shear strain deformation potential [62]. This is the only form of the Hamiltonian in the vicinity of the X point allowed by symmetry considerations [62]. The two-band Hamiltonian eq.(30) results in the following dispersion relations:

$$E(\vec{k}) = \frac{\hbar^2 k_z^2}{2m_l} + \frac{\hbar^2(k_x^2+k_y^2)}{2m_t} \pm \sqrt{\left(\frac{\hbar^2 k_z k_0}{m_l}\right)^2 + \delta^2}\,, \tag{31}$$

where the negative sign corresponds to the lowest conduction band,

$$\delta = \left(D\varepsilon_{xy} - \frac{\hbar^2 k_x k_y}{\mathrm{M}}\right) \quad . \tag{32}$$

The energy E and k_z are counted from the X-point. A comparison of eq.(31) with the band structure of [001] valleys obtained numerically with the density functional theory (DFT) [59], empirical pseudopotential method (EPM) [60], and with the tightbinding $sp^3d^5s^*$ model [61] in [100] and [110] direction is shown in Fig. 23. The **k·p** model accurately describes the dispersion relation up to energies of about $0.5\mathrm{eV}$. Therefore, the **k·p** Hamiltonian eq.(30) can be used to describe the subband structure in thin silicon films and inversion layers.

4.3 Subband Dispersion in [001] Thin Silicon Films

For [001] silicon films the confinement potential gives an additional contribution $U(z)I$ to the Hamiltonian eq.(30). In the effective mass approximation described by eq.(30) 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. In ultra-thin films the unprimed ladder is predominantly occupied and must be considered. The term with σ_x in eq.(30) couples the two lowest conduction bands and lifts the two-fold degeneracy of the unprimed subband ladder. The additional unprimed subband splitting, or the valley splitting, can be extracted from the Shubnikov-de-Haas oscillations and is typically in the order of a few tens μeV [58]. However, the valley splitting is greatly enhanced in a laterally confined two-dimensional electron gas [58]. The valley splitting is usually addressed by introducing a phenomenological intervalley coupling constant at the silicon interface [63]. Here we investigate the valley splitting based on the two-band **k·p** model eq.(30) without introducing any additional parameters. We approximate the confining potential of an ultrathin silicon film by a square well potential with infinite potential walls. This is sufficient for the purpose to analyze the valley splitting in a quasi-twodimensional gas due to interband coupling. Generalization to include a self-consistent potential is straightforward though numerically involved [64]. Because of the two-band Hamiltonian, the wave function Ψ is a spinor with the two components $|0>$ and $|1>$. For a wave function with space dependencein a form $e^{ik_z z}$ 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 $k_i (i = 1, \cdots, 4)$ for k_z of the dispersion relation eq.(31), so the spatial dependence of a spinor component α is in the form $\sum_{i=1}^{4} A_\alpha^i e^{ik_i z}$. 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 equations:

$$\tan(k_1 \frac{k_0 t}{2}) = \frac{k_2}{\sqrt{k_2^2+\eta^2} \pm \eta} \frac{\sqrt{k_1^2+\eta^2} \pm \eta}{k_1} \cdot \tan(k_2 \frac{k_0 t}{2}) \tag{33}$$

where $\eta = m_l|\delta|/(\hbar k_0)^2$. The value of

$$k_2 = \sqrt{k_1^2 + 4 - 4\sqrt{k_1^2+\eta^2}} \tag{34}$$

becomes imaginary at high η values. Then the trigonometric functions in eq.(33) are replaced by the hyperbolic ones. Special care must be taken to choose a/the correct

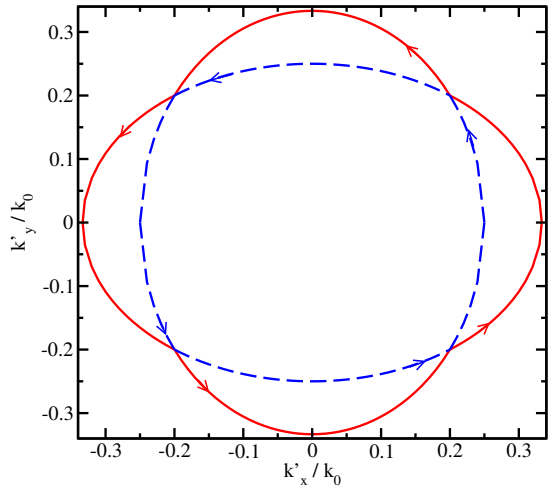

Figure 24: Quasi-classical orbits of electron motion in a magnetic field corresponding to the subband dispersion eq.(36). Kinematic moment components k'_x, k'_y are in $<110>$ and $<1-10>$ directions. The difference between the quasiclassical orbits causes the valley splitting eq.(37) in a magnetic field perpendicular to the electron system.

branch of $\sqrt{k_2^2+\eta^2}$ in eq.(34): the sign of $\sqrt{k_2^2+\eta^2}$ must be alternated after the argument becomes zero. Introducing $y_n = (k_1-k_2)/2$, eq.(33) can be written in the form:

$$\sin(y_n k_0 t) = \pm \frac{\eta\, y_n \sin\left(\frac{1-\eta^2-y_n^2}{1-y_n^2} k_0 t\right)}{\sqrt{(1-y_n^2)(1-\eta^2-y_n^2)}} \tag{35}$$

For small values of the parameter η we obtain from eq.(35) the following dispersion relation for the unprimed subbands n:

$$\begin{aligned} E_n^{\pm} =& \frac{\hbar^2}{2m_l}\left(\frac{\pi n}{t}\right)^2 + \frac{\hbar^2(k_x^2+k_y^2)}{2m_t} \\ &\pm \left(\frac{\pi n}{k_0 t}\right)^2 \frac{|D\varepsilon_{xy} - \frac{\hbar^2 k_x k_y}{M}|}{k_0 t|1-(\pi n/k_0 t)^2|}\sin(k_0 t) \end{aligned} \tag{36}$$

eq.(36) demonstrates that the unprimed subbands are not necessarily degenerate and degeneracy is preserved only, when shear strain is zero and either $k_x = 0$ or $k_y = 0$.

4.4 Valley Splitting in a Magnetic Field

For zero shear strain the Landau levels in an othogonal magnetic field B are determined from eq.(36) using the Bohr-Sommerfeld quantization conditions:

$$E_m^{1,2} = \hbar\omega_c\left(m+\frac{1}{2}\right)\frac{\pi}{4\arctan\left(\sqrt{m_{(1,2)}/m_{(2,1)}}\right)}\,, \tag{37}$$

$$m_{(1,2)} = \left(\frac{1}{m_t} \pm \frac{1}{M}\left(\frac{\pi n}{k_0 t}\right)^2 \frac{\sin(k_0 t)}{k_0 t|1-(\pi n/k_0 t)^2|}\right)^{-1} \tag{38}$$

and

$$\omega_c = \frac{eB}{\sqrt{m_1 m_2}c}$$

is the cyclotron frequency, e is the electron charge, and c is the speed of light. According to eq.(37), the valley splitting $|E_m^1 - E_m^2|$ is linear regarding the magnetic field. In order to obtain the linear dependence, two conditions must be satisfied:

(i) no shear strain and

(ii) the cyclotron energy is smaller than the

subband quantization energy. Both conditions are satisfied in a biaxially stressed silicon film of $10\mathrm{nm}$ thickness on SiGe used in [58] for magnetic fields as strong as $1\mathrm{T}$. It follows from eq.(37) and eq.(39) that the valley splitting can be several tens μeVs, which is consistent with the experiment [58].

4.5 Valley Splitting in a Point Contact

We consider a point contact in [110] direction realized by confining an electron system of a thin silicon film laterally by depleting the area under additional gates. Without strain the low-energy effective Hamiltonian in the point contact can be written as:

$$H_{(1,2)} = \frac{\hbar^2 {k'_x}^2}{2m_{(2,1)}} + \frac{\hbar^2 {k'_y}^2}{2m_{(1,2)}} + \frac{1}{2}\kappa {x'}^2 + V_b\,, \tag{39}$$

where the primed variables are along the [110] and [1-10] axes, the effective masses are determined by eq.(39), κ is the spring constant of the point contact confinement potential $V(x') = \kappa {x'}^2/2$ in [1-10] direction, and V_b is a gate voltage dependent conduction band shift in the point contact [65]. The dispersion relation of propagating modes within the point contact is written as:

$$E_p^{(1,2)} = \frac{\hbar^2 {k'_x}^2}{2m_{(2,1)}} + \hbar\omega_{(1,2)}\left(p+\frac{1}{2}\right) + V_b\,, \tag{40}$$

where $\omega_{(1,2)}^2 = \kappa/m_{(1,2)}$. Since the energy minima of the two propagating modes with the same p are separated, they are resolved in the conductance experiment through the point contact as two distinct steps. The valley splitting is $\Delta E_p = \hbar|\omega_1 - \omega_2|$. The difference in the

effective masses eq.(39) and, correspondingly, the valley splitting can be greatly enhanced by reducing the effective thickness t of the quasi-two-dimensional electron gas which is usually the case in a gated electron system, when the inversion layer is formed. In a [100] oriented point contact without strain the effective Hamiltonian is

$$H^{\pm} = \frac{\hbar^2(k_x^2+k_y^2)}{2m_t} + \left(\frac{\pi n}{k_0 t}\right)^2 \frac{|\frac{\hbar^2 k_x k_y}{M}|\sin(k_0 t)}{k_0 t|1-(\pi n/k_0 t)^2|} + \frac{\kappa}{2}x^2 \,. \tag{41}$$

Due to symmetry with respect to k_y the subband minima in a point contact are always degenerate. For this reason the valley splitting in [100] oriented point contacts is greatly reduced.

4.6 Valley Splitting by Shear Strain

It follows from eq.(36) that shear strain induces a valley splitting linear in strain for small shear strain values and depends strongly on the film thickness [66]:

$$\Delta E_n = 2\left(\frac{\pi n}{k_0 t}\right)^2 \frac{D\varepsilon_{xy}}{k_0 t|1-(\pi n/k_0 t)^2|}\sin(k_0 t)\,. \tag{42}$$

For higher strain values eq.(33) must be solved numerically. Results shown in Fig.3 demonstrate that valley splitting can be effectively controlled by adjusting the shear strain and modifying the effective thickness t of the electron system. Uniaxial stress along [110] channel direction, which induces shear strain, is already used by industry to enhance the performance of MOSFETs. Therefore, its application to control valley splitting does not require expensive technological modifications.

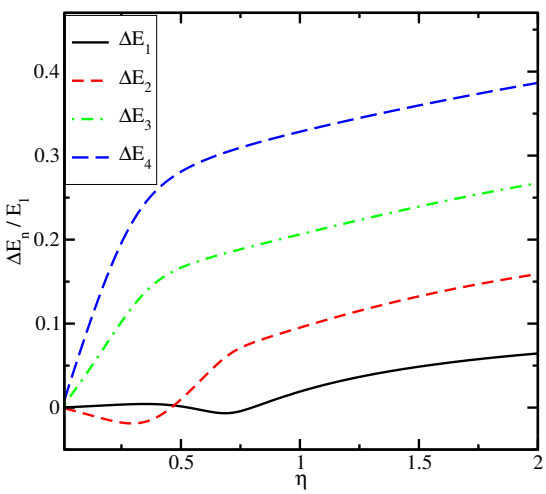

Figure 25: Splitting between the unprimed subband energies, or the valley splitting, in a $6.5\,\mathrm{nm}$ thick silicon film as a function of shear strain. The splitting values are normalized to the energy of the ground subband without strain. The value $\eta = 1$ corresponds to the shear strain value $\varepsilon_{xy} = 0.016$. The value of valley splitting may alternate its sign, in accordance to eq.(36).

4.7 Conclusion

The unprimed valley structure in (001) silicon thin films has been analyzed within the two-band **k·p** model. It is shown that the two-fold degeneracy of the unprimed subbands can be lifted leading to the socalled valley splitting which is proportional to the strength of the perpendicular magnetic field. The valley splitting can be enhanced in <110> oriented point contacts, while it is suppressed in a $<100>$ point contact. Finally, the valley splitting can be controlled and made larger than the Zeeman splitting by shear strain.This makes silicon very attractive for spintronic applications.

5 Consistent Higher-Order Transport Models for SOI MOSFETs

We have developed a two-dimensional non-parabolic macroscopic transport model up to the sixth order. To model higher-order transport parameters with as few simplifying assumptions as possible, we apply an extraction technique from Subband Monte Carlo simulations followed by an interpolation within these Monte Carlo tables through the whole inversion layer. The impact of surface-roughness scattering as well as quantization on the transport parameters is inherently considered in the Subband Monte Carlo data. These tables are used to model higher-order mobilities as well as the macroscopic relaxation times as a function of the effective field and the carrier temperature. We have studied the influence of the inversion layer concentration on higher-order transport parameters within high fields and show the behavior of these parameters in a quantized system of a UTB SOI MOSFET.

5.1 Introduction

For engineering applications macroscopic transport models based on Boltzmann's transport equation (BTE) like the drift-diffusion (DD) model or the energy transport model (ET) are very efficient compared to the time consuming Monte Carlo (MC) simulation [67]. However, with further decrease of the device dimensions into the deca-nanometer regime both the DD, and the ET model become more and more inaccurate [68]. Investigations have demonstrated [69, 70] that higher-order macroscopic models can cover the gate length range of $100\,\mathrm{nm}$ down to about $25\,\mathrm{nm}$. In the analysis of macroscopic models up to the sixth order, it is essential to describe the transport parameters, namely the carrier mobility μ_0, the energy-flux mobility μ_1, the energy relaxation time τ_1, the second-order energy-flux mobility μ_2, and the second-order energy relaxation time τ_2 with as few simplifying assumptions as possible. A rigorous study of the behavior of these parameters in the bulk case has been already carried out using bulk MC tables [69]. The use of this bulk data set for modeling transport in MOSFET devices is problematic due to the importance of surface roughness scattering and quantization in the inversion channel. In [71] surface roughness scattering on the carrier mobility has been investigated using the semiempirical Matthiesen rule, but a rigorous study has not been performed yet.
In order to take these important inversion layer effects as well as non-parabolic bands for high fields into account, we have developed a Subband Monte Carlo (SMC) [72, 73] table based 2D electron gas six moments transport model. It is now possible to study the influence of the inversion layer effects on higher-order transport parameters.

5.2 Model

In order to derive 2D higher-order macroscopic models like the six moments model one has to multiply the BTE

$$\partial_t f + \mathbf{v}\cdot\nabla_{\mathrm{r}} f - \frac{\mathbf{F}}{\hbar}\cdot\nabla_{\mathrm{k}} f = (\partial_t f)_{\mathrm{coll}}\,, \tag{43}$$

with special weight functions [74] and integrate over the $\mathbf{k}$-space. $f(\mathbf{r},\mathbf{k},t)$ is the distribution function, $\mathbf{F}$ is the driving force, and $\mathbf{v}\,(\mathbf{k})$ is the group velocity in the BTE. The moments in 2D space are defined as

$$x(\mathbf{r},t) = \frac{2}{(2\pi)^2}\int_0^\infty X(\mathbf{k}) f(\mathbf{r},\mathbf{k},t)\mathrm{d}^2k = n(\mathbf{r},t)\langle X(\mathbf{k})\rangle\,, \tag{44}$$

with $x(\mathbf{r},t)$ as the macroscopic values together with the microscopic counterpart $X(\mathbf{k})$. $n(\mathbf{r},t)$ is the carrier concentration. Substituting the microscopic weights with $\epsilon^i\,(\mathbf{k})$ and $\mathbf{v}\,(\mathbf{k})\,\epsilon^i\,(\mathbf{k})$ one will obtain the even, scalar-valued moments, and the odd, vector-valued moments, respectively. For the six moments model i is in the range of $i \in [0,2]$. The scattering operator of the BTE is modeled using the macroscopic relaxation time approximation [75]. For the even moments the macroscopic relaxation time ansatz looks like

$$\partial_t\langle X\rangle_{\mathrm{coll}}^{\mathrm{even}} \approx -n\frac{\langle\epsilon^i\rangle - \langle\epsilon^i\rangle_0}{\tau_i}\,, \tag{45}$$

and for the odd moments

$$\partial_t\langle X\rangle_{\mathrm{coll}}^{\mathrm{odd}} \approx -n\frac{\langle\mathbf{v}\epsilon^i\rangle}{\tau_i}\,. \tag{46}$$

$\langle\epsilon^i\rangle_0$ are the even moments in equilibrium. Further approximations during the derivation are the diffusion approximation, which states that the anti-symmetric part of the distribution function is much smaller than the symmetric one [76], the modeling of the tensorial components, and the closure relation [77]. The transport model will be formulated in terms of $w_i = \langle\epsilon^i\rangle$ and $\mathbf{V}_i = \langle\mathbf{v}\epsilon^i\rangle$ [69]. The general conservation equation reads

$$\partial_t\,(nw_i) + \nabla\cdot(n\mathbf{V}_i) - i\mathbf{F}\cdot n\mathbf{V}_{i-1} = -n\frac{w_i - w_{\mathrm{eq}}}{\tau_i}\,, \tag{47}$$

whereas the general fluxes are defined as

$$n\mathbf{V}_i = -\frac{\mu_i H_{i+1}}{2\mathrm{q}}\left(\nabla\,(nw_{i+1}) - n\mathbf{F}w_i\frac{2+iH_i}{H_{i+1}}\right)\,. \tag{48}$$

H_i are the non-parabolicity factors, and are defined in the 2D space as

$$U_i = \frac{1}{2}\mathrm{tr}\,(\mathbb{U}_i) = \frac{1}{2}w_i H_i\,, \tag{49}$$

with $\mathbb{U}_i = \langle \mathbf{v} \otimes \mathbf{p} \epsilon^i \rangle$ as the energy tensors [77]. The even moments w_i of the six moments model are expressed as $\mathrm{k_B} T_n$, $2\,(\mathrm{k_B} T_n)^2\,\beta$, and $6\,(\mathrm{k_B} T_n)^3\,\beta^c$. β is the kurtosis and denotes the deviation from a heated Maxwellian distribution function

$$\beta = \frac{1}{2}\frac{w_2}{w_1^2}\,. \tag{50}$$

c is an integer in the range $c \in [0, 3]$ [69]. To take the quantization as well as surface roughness scattering into account we extract the transport parameters τ_i and μ_i from a self-consistent coupling between a SMC simulator and a Schrödinger-Poisson (SP) solver. In the SMC simulator we consider non-parabolic bands, quantization effects, phonon induced scattering as well as surface roughness scattering [72, 78]. The SP solver incorporates the quantum confinement in inversion layers. Our device simulator Minimos-NT [37] calculates the effective field through the channel of a device and extracts higher-order transport parameters from the SMC tables. The mobilities as well as the relaxation times are now a function of the effective field and the carrier temperature $\mu_\mathrm{i}(E_\mathrm{eff}, T_n)$, $\tau_\mathrm{i}(E_\mathrm{eff}, T_n)$.

5.3 Results

As an example device a fully depleted SOI MOSFET with a Si film thickness of $4\,\mathrm{nm}$ in $\langle 100 \rangle$ direction, and a donor doping concentration of $10^{20}\,\mathrm{cm}^{-3}$ in the source and the drain regions, and with an acceptor doping in the channel of $10^{16}\,\mathrm{cm}^{-3}$ has been investigated. An electric field is applied in $\langle 010 \rangle$ direction.

In Fig. 26 we show higher-order mobilities as a function of the inversion layer concentration $\mathrm{N_{Inv}}$. We point out that in bulk MOSFETs for the low-field case the mobility of our simulation fits the measurement data of Takagi [79] while a significant reduction is observed in the quantized 4 nm channel region [80]. Furthermore it is demonstrated that for high lateral fields and high $\mathrm{N_{Inv}}$ the mobilities converge to the same value. This is not the case for the relaxation times. In Fig. 27 and Fig. 28 the energy relaxation times and the second-order energy relaxation time as a function of $\mathrm{N_{Inv}}$ are plotted for different lateral fields, respectively. For high $\mathrm{N_{Inv}}$ and for low fields the relaxation times increase compared to the high lateral field case where the relaxation times are constant. This can be explained with Fig. 29 where the first subband occupation as a function of $\mathrm{N_{Inv}}$ of the unprimed, primed, and double primed valley for lateral fields of $50\,\mathrm{kV/cm}$, and $200\,\mathrm{kV/cm}$ is shown. Due to the fast increase of the occupation number of the first subband in the unprimed valley at $50\,\mathrm{kV/cm}$ compared to the high-field case, where the occupation is constant, the change in the higher-order relaxation times increases as well for high N_Inv. Fig. 30 shows effective fields for different bias points through an SOI MOSFET with a gate length of $40\,\mathrm{nm}$. With the effective fields and the SMC tables one can model higher-order transport parameters through a device, as presented in Fig. 31. A profile of the carrier, the energy-flux, and the second-order energy flux mobilities has been pointed out.

Fig. 32 is a comparison of the carrier temperature with the second-order temperature $\theta = \beta T_n$. The deviation of these two temperatures increases for high drain

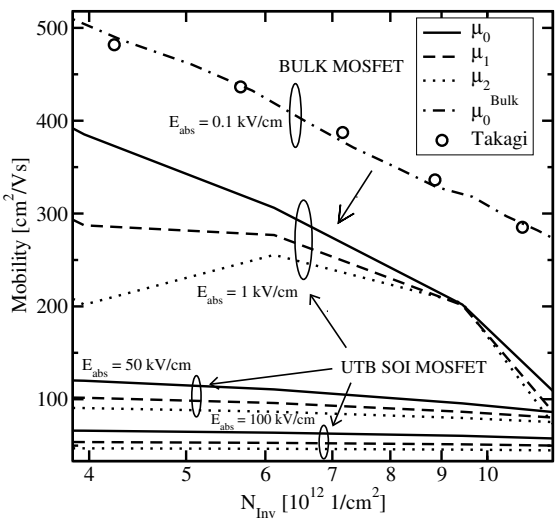

Figure 26: The higher-order mobilities as a function of the inversion layer concentration $\mathrm{N_{Inv}}$ for different lateral fields $\mathrm{E_{abs}}$ are plotted. For high fields the difference of the mobilities decreases. For low fields in a bulk MOSFET the carrier mobility is equal to the measurement data of Takagi.

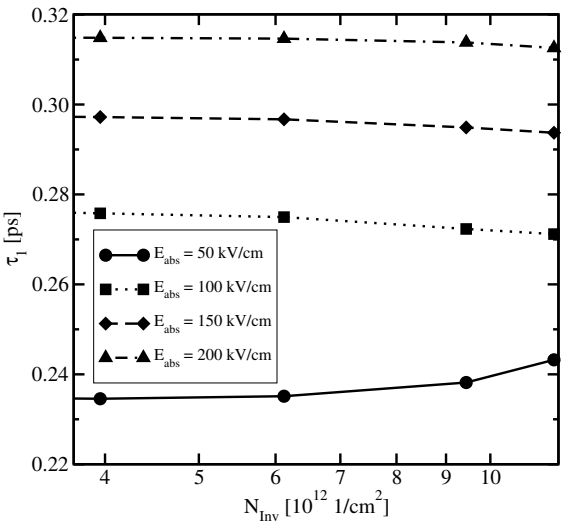

Figure 27: The energy relaxation time as a function of the inversion layer concentration for different lateral fields is shown. For a field of $50\,\mathrm{kV/cm}$ and a high $\mathrm{N_{Inv}}$, τ_1 increases compared to high-fields, where the energy relaxation time is more or less constant.

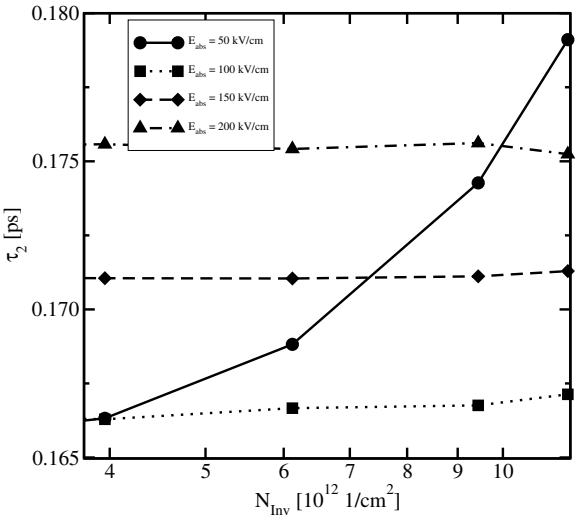

Figure 28: The second-order relaxation time is shown as a function of the inversion layer concentration for different lateral fields. The increase of τ_2 at $50\,\mathrm{kV/cm}$ is as well higher than at the other fields. The change at high N_{Inv} and low fields of the second-order relaxation time compared to the other fields is higher than in the energy relaxation time.

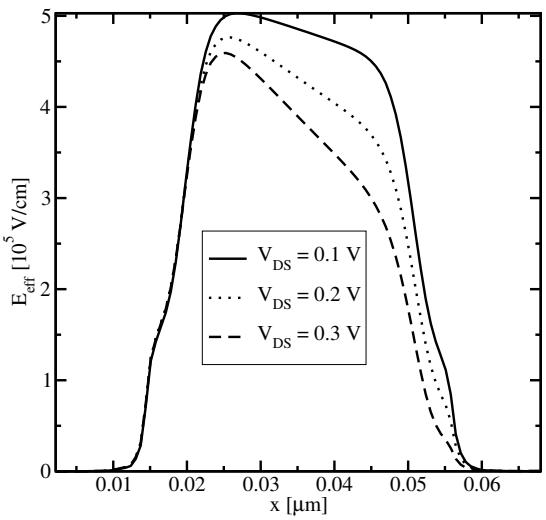

Figure 30: The effective field throughout the whole device for drain voltages of $0.1\,\mathrm{V}$, $0.2\,\mathrm{V}$, and $0.3\,\mathrm{V}$. With the effective field and an interpolation between SMC tables, higher-order transport parameters can be modeled throughout the whole device.

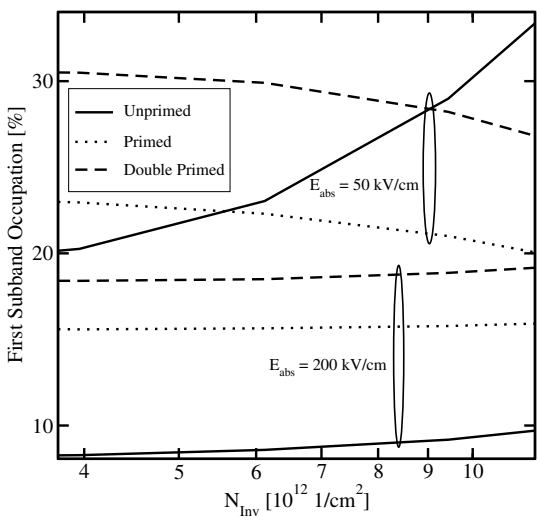

Figure 29: The first subband occupation of the unprimed, primed, and double primed valley as a function of the inversion layer concentration for fields of $50\,\mathrm{kV/cm}$ and $100\,\mathrm{kV/cm}$. Due to the light mass of the unprimed valley in transport direction the subband occupation number is higher than in the other valleys.

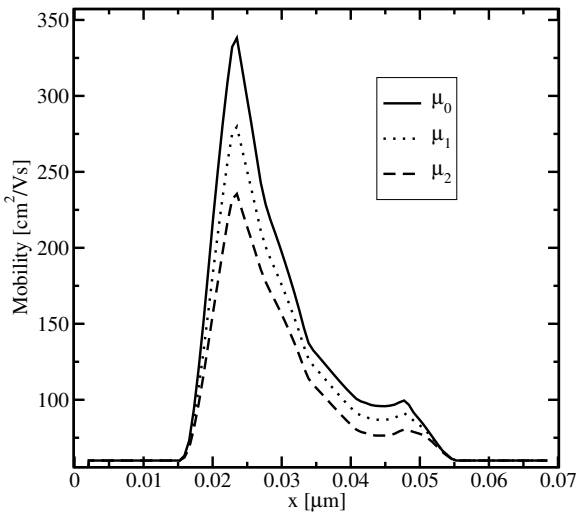

Figure 31: The carrier, the energy-flux as well as the second-order energy flux mobilities are plotted through a $40\,\mathrm{nm}$ gate length SOI MOSFET. A drain voltage of $0.3\,\mathrm{V}$ has been applied.

voltages due to the deviation of the distribution function from a Maxwellian distribution. The maximum peak is at the channel drain junction where hot electrons from the channel meet cold electrons from the drain. Fig. 33 shows velocity profiles calculated with the DD, ET, and the six moments (SM) models of an SOI MOSFETs with a gate lengths of $40\,\mathrm{nm}$ and $60\,\mathrm{nm}$. In the $40\,\mathrm{nm}$ device the ET model considerably overestimates the velocity (and thus the drain current) with a velocity three times as high as the DD model, whereas for the $60\,\mathrm{nm}$ device it is just two times as high [68]. For large devices the velocity of the ET and the SM converge to the value of the DD model. In Fig. 34 we point out that with further increase of the gate lengths the difference between the output currents decreases. For large devices all transport models yield the same result.

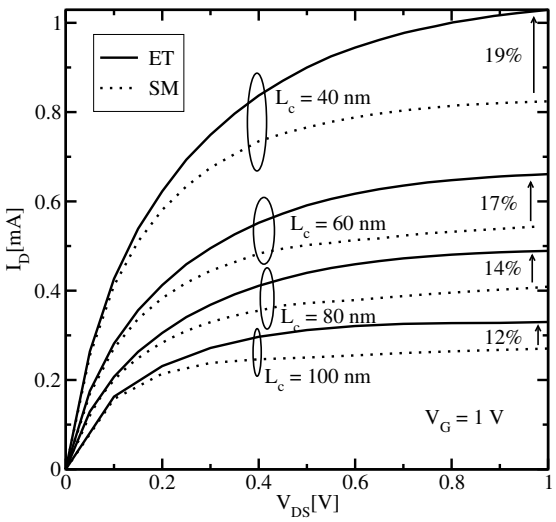

Figure 34: Output characteristics calculated with the ET and the SM model for SOI MOSFETs with 40 nm, 60 nm, 80 nm, and 100 nm gate lengths. A gate voltage of 1 V is applied. For long channel devices the ET and the SM model yield the same result.

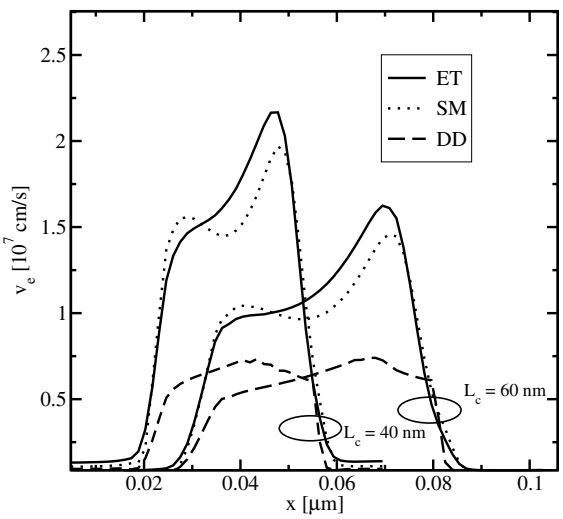

Figure 33: Velocity profile calculated with the DD, ET, and the SM model for a 40 nm and 60 nm device. The velocity of the ET and SM model is smaller in the 60 nm than in the 40 nm device. DD predicts velocities independent from the gate length. Due to the quantization and surface roughness scattering the velocity in the DD model is smaller than the saturation velocity ($\approx 10^7\,\mathrm{cm/s}$) in the bulk.

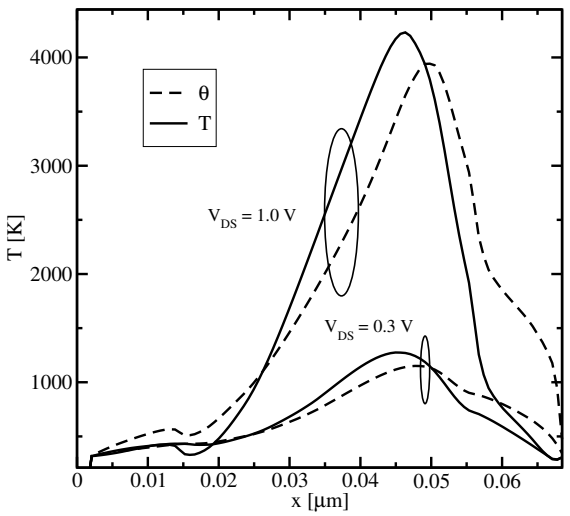

Figure 32: The second-order temperature $\theta = \beta T_n$ in comparison to the carrier temperature T_n for $V_{\mathrm{DS}} = 0.3\,\mathrm{V}$ and $V_{\mathrm{DS}} = 1.0\,\mathrm{V}$. The deviation of the two temperatures has its maximum in the drain region, where the hot electrons from the channel meet cold electrons in the drain region. The difference of θ and T_n for $V_{\mathrm{DS}} = 1.0\,\mathrm{V}$ is higher than at $V_{\mathrm{DS}} = 1.0\,\mathrm{V}$.

5.4 Conclusion

We have studied the behaviour of higher-order transport parameters in inversion layers. A method based on an interpolation between SMC tables for modeling a two-dimensional electron gas has been developed. This approach allows the investigation of UTB SOI devices including the influence of surface-roughness scattering and quantization within higher-order moment models. A generalized set of equations has been used to derive a novel 2D six moments model. First the influence of different inversion layers on higher-order transport parameters within high fields has been shown. We found a significant change in the relaxation times for high inversion layer concentrations and low fields compared to high fields. Furthermore higher-order mobilities converge for high inversion layer concentrations to the same value. Second we have pointed out the importance of using higher-order moment models for the modeling of devices with a gate length in the deca-nanometer regime.

References

[1] M. R. Sorensen, Y. Mishin, and A. F. Voter. Diffusion Mechanisms in Cu Grain Boundaries. *Physical Review B*, 62(6):3658–3673, 2000.

[2] J. C. Fisher. Calculation of Diffusion Penetration Curves for Surface and Grain Boundary Diffusion. *Journal of Applied Physics*, 22(1):74–77, 1951.

[3] R. W. Balluffi. Grain Boundary Diffusion Mechanisms in Metals. *Metallurgical Transactions A*, 13:2069–2095, 1982.

[4] R. W. Balluffi. Grain Boundary Structure and Kinetics. *ASM, Metals Park*, 1980.

[5] H. Ceric, R. L. de Orio, J. Cervenka, and S. Selberherr. A Comprehensive TCAD Approach for Assessing Electromigration Reliability of Modern Interconnects. *IEEE Trans. Mat. Dev. Rel.*, 9(1):9–19, 2009.

[6] M. E. Sarychev and Yu. V. Zhitnikov. General Model for Mechanical Stress Evolution During Electromigration. *Journal of Applied Physics*, 86(6):3068 – 3075, 1999.

[7] H. Ceric, R. L. de Orio, J. Cervenka, and S. Selberherr. Stress-Induced Anisotropy of Electromigration in Copper Interconnects. *to be published in Stress-Induced Phenomena in Metallization, AIP*, 2009.

[8] R. J. Gleixner, B. M. Clemens, and W. D. Nix. Void Nucleation in Passivated Interconnect Lines: Effects of Site Geometries, Interfaces, and Interface Flaws. *J. Mat. Res.*, 12:2081–2090, 1997.

[9] D. R. Fridline and A. F. Bower. Influence of Anisotropic Surface Diffusivity on Electromigration Induced Void Migration and Evolution. *Journal of Applied Physics*, 85(6):3168–3174, 1999.

[10] J. A. Sethian. *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision and Materials Science*. Cambridge University Press, 1999.

[11] R. Rosenberg and M. Ohring. Void Formation and Growth During Electromigration in Thin Films. *Journal of Applied Physics*, 42(13):5671–5679, 1971.

[12] V. Sukharev, E. Zschech, and W. D. Nix. A Model for Electromigration-Induced Degradation Mechanisms in Dual-Inlaid Copper Interconnects: Effect of Microstructure. *Journal of Applied Physics*, 102(5):530501–530514, 2007.

[13] L. Doyen, E. Petitprez, P. Waltz, X. Federspiel, L. Arnaud, and Y. Wouters. Extensive Analysis of Resistance Evolution due to Electromigration-Induced Degradation. *Journal of Applied Physics*, 104:123521–123525, 2008.

[14] Z. Suo, W. Wang, and M. Yang. Electromigration Instability: Transgranular Slits in Interconnects. *Applied Physics Letters*, 64(15):1944–1946, 1994.

[15] M. Mahadevan and R. Bradley. Simulations and Theory of Electromigration-Induced Slit Formation in Unpassivated Single-Crystal Metal Lines. *Physical Review B*, 59(16):11037–11046, 1999.

[16] E. Arzt and W. D. Nix. A Model for the Effect of Line Width and Mechanical Strength on Electromigration Failure of Interconnects with "Near-Bamboo" Grain Structures. *J. Appl. Phys.*, 6:731–736, 1991.

[17] C. Ryu, K.-W. Kwon, A. L. S. Loke, H. Lee, T. Nogami, V. M. Dubin, R. A. Kavari, G. W. Ray, and S. Wong. Microstructure and Reliability of Copper Interconnects. *IEEE Elec. Dev.*, 46:1113–1120, 6.

[18] E. Arzt, O. Kraft, W. D. Nix, and J. E. Sanchez. Electromigration Failure by Shape Change of Voids in Bamboo Lines. *Journal of Applied Physics*, 76(3):1563–1571, 1994.

[19] M. Denais, V. Huard, C. Parthasarathy, G. Ribes, F. Perrier, N. Revil, and A. Bravaix. Interface Trap Generation and Hole Trapping under NBTI and PBTI in Advanced CMOS Technology with a 2-nm Gate Oxide. *T-DMR*, 4:715–722, 2004.

[20] V. Huard, M. Denais, and C. Parthasarathy. NBTI Degradation: From Physical Mechanisms to Modelling. *Microelectron. Reliab.*, 46(1):1–23, 2006.

[21] V. Huard, C. Parthasarathy, N. Rallet, C. Guerin, M. Mammase, D. Barge, and C. Ouvrard. New Characterization and Modeling Approach for NBTI Degradation from Transistor to Product Level. *Proc. Intl.Electron Devices Meeting*, 2007.

[22] T. Grasser, B. Kaczer, and W. Goes. An Energy-Level Perspective of Bias Temperature Instability. *Proc. Intl.Rel.Phys.Symp.*, 2008.

[23] Th. Aichinger and M. Nelhiebel. Charge Pumping Revisited - The Benefits of an Optimized Constant Base Level Charge Pumping Technique for MOSFET Analysis. In *IIRW*, 2007.

[24] M.-F. Li, D.M. Huang, C. Shen, T. Yang, W.J. Liu, and Z.Y. Liu. Understand NBTI Mechanism by Developing Novel Measurement Techniques. *T-DMR*, 8(1):62–71, 2008.

[25] W.J. Liu, Z.Y. Liu, D.M. Huang, C.C. Liao, L.F. Zhang, Z.H. Gan, W.S. Wong, C. Shen, and M.-F. Li. On-The-Fly Interface Trap Measurement and its Impact on the Understanding of NBTI Mechanism for p-MOSFETs with SiON Gate Dielectric. In *IEDM*, 2007.

[26] Z.Y. Liu, D.M. Huang, W.J. Liu, C.C. Liao, L.F. Zhang, , Z.H. Gan, W.S. Wong, and M.-F. Li. Comprehensive Studies of BTI Degradation in SiON Gate Dielectric CMOS Transistors by New Measurement Techniques. In *IRPS*, 2008.

[27] P. Heremans, J. Witters, G. Groeseneken, and H.E. Maes. Analysis of the Charge Pumping Technique and its Application for the Evaluation of MOSFET Degradation. *T-ED*, 36:1318–1335, 1989.

[28] G.V.d. Bosch, G.V. Groeseneken, P. Heremans, and H.E. Maes. Spectroscopic Charge Pumping: A New Procedure for Measuring Interface Trap Distributions on MOS Transistors. *T-ED*, 38:1820–1831, 1991.

[29] R.E. Paulsen and M.H. White. Theory and Application of Charge Pumping for the Characterization of $\mathrm{Si}-\mathrm{Si0_2}$ Interface and Near-Interface Oxide Traps. *T-ED*, 41:1213–1216, 1994.

[30] D. Bauza. Rigorous Analysis of Two-Level Charge Pumping: Application to the Extraction of Interface Trap Concentration versus Energy Profiles in Metal-Oxide-Semiconductor Transistors. *Journal of Applied Physics*, 94:3239–3248, 2003.

[31] G. Groeseneken, H.E. Maes, N. Beltran, and R. F. de Keersmaecker. A Reliable Approach to Charge-Pumping Measurements in MOS Transistors. *T-ED*, 31:42–53, 1984.

[32] J.S. Brugler and P. Jespers. Charge Pumping in MOS Devices. *T-ED*, 16:297–302, 1969.

[33] Th. Aichinger, M. Nelhiebel, and T. Grasser. On the Energy Dependence of Oxide Trap Recovery after NBTI Stress. In *IRPS*, 2009.

[34] T. Grasser, B. Kaczer, W. Gös, Th. Aichinger, Ph. Hehenberger, and M. Nelhiebel. A Two-Stage Model for Negative Bias Temperature Instability. In *IRPS*, 2009.

[35] D.M. Fleetwood. Fast and Slow Border Traps in MOS Devices. In *RADECS*, 1995.

[36] B. Kaczer, T. Grasser, Ph.J. Roussel, J. Martin-Martinez, R. O'Connor, B.J. O'Sullivan, and G. Groeseneken. Ubiquitous Relaxation in BTI Stressing-New Evaluation and Insights. *Proc. Intl.Rel.Phys.Symp.*, 2008.

[37] IμE. *MINIMOS-NT 2.1 User's Guide*. Institut für Mikroelektronik, Technische Universität Wien, Austria, 2004. http://www.iue.tuwien.ac.at/software/minimos-nt.

[38] C. Gmachl, F. Capasso, D. L. Sivco, and A. Y. Cho. Recent Progress in Quantum Cascade Lasers and Applications.

[39] R.F. Kazarinov and R.A. Suris. Possibility of the Amplification of Electromagnetic Waves in Semiconductors. *Reports on Progress in Physics*, 5(4):707–709, 1971.

[40] Jerome Faist, Federico Capasso, Deborah L. Sivco, Carlo Sirtori, Albert L. Hutchinson, and Alfred Y. Cho. Quantum Cascade Laser. *Science*, 264(5158):553–556, 1994.

[41] N. Schildermans, M. Hayne, V.V. Moshchalkov, A. Rastelli, and O.G. Schmidt. Nonparabolic Band Effects in GaAs/$\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{As}$ Quantum Dots and Ultrathin Quantum Wells. *Physical Review B*, 72:115312–1–115312–5, 2005.

[42] B.R. Nag and S. Mukhopadhyay. Energy Levels in Quantum Wells of Nonparabolic Semiconductors. *Physica Status Solidi B*, 175:103–112, 1993.

[43] S. Panda, B.K. Panda, and S. Fung. Effect of Conduction Band Nonparabolicity on the Dark Current in a Quantum Well Infrared Detector. *Journal of Applied Physics*, 107:043705–1 – 043705–8, 2007.

[44] F. Rossi and T. Kuhn. Ultrafast Phenomena in Photoexcited Semiconductors. *Reviews of Modern Physics*, 74(3):895–950, 2002.

[45] U. Penner, H. Rücker, and I.N. Yassievich. Theory of Interface Roughness Scattering in Quantum Wells. *Semiconductor Science and Technology*, 13:709–713, 1998.

[46] S. Tsujino, A. Borak, E. Müller, M. Scheinert, C.V. Falub, H. Sigg, D. Grützmacher, M. Giovannini, and J. Faist. Interface-Roughness-Induced Broadening of Intersubband Electroluminescence in p-SiGe and n-GaInAs/AlInAs Quantum-Cascade Structures. *Applied Physics Letters*, 86:062113–1–062113–3, 2005.

[47] T. Unuma, M. Yoshita, T. Noda, H. Sakaki, and H. Akiyama. Intersubband Absorption Linewidth in GaAs Quantum Wells Due to Scattering by Interface Roughness, Phonons, Alloy Disorder, and Impurities. *Journal of Applied Physics*, 93(3):1586–1597, 2003.

[48] C. Sirtori, P. Kruck, S. Barbieri, P. Collot, J. Nagle, M. Beck, J. Faist, and U. Oesterle. GaAs/$\mathrm{Al_xGa_{1-x}As}$ Quantum Cascade Lasers. *Applied Physics Letters*, 73(24):3486–3488, 1998.

[49] P.J. Turley and S.W. Teitsworth. Electronic Wave Functions and Electron-Confined-Phonon Matrix Elements in GaAs/$\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{As}$ Double-Barrier Resonant Tunneling Structures. *Physical Review B*, 44(7):3199–3210, 1991.

[50] D.F. Nelson, R.C. Miller, and D.A. Kleinman. Band Nonparabolicity Effects in Semiconductor Quantum Wells. *Physical Review B*, 53(14):7770–7773, 1987.

[51] S. Adachi. GaAs, AlAs, and $\mathrm{Al_xGa_{1-x}As}$: Material Parameters for Use in Research and Device Applications. *Journal of Applied Physics*, 58(3):R1–R29, 1985.

[52] K. Mistry, C. Allen, C. Auth, B. Beattie, D. Bergstrom, M. Bost, M. Brazier, M. Buehler, A. Cappellani, R. Chau, C. . Choi, G. Ding, K. Fischer, T. Ghani, R. Grover, W. Han, D. Hanken, M. Hattendorf, J. He, J. Hicks, R. Huessner, D. Ingerly, P. Jain, R. James, L. Jong, S. Joshi, C. Kenyon, K. Kuhn, K. Lee, H. Liu, J. Maiz, B. McIntyre, P. Moon, J. Neirynck, S. Pae, C. Parker, D. Parsons, C. Prasad, L. Pipes, M. Prince, P. Rarade, T. Reynolds, J. Sandford, L. Shifren, J. Sebastian, J. Seiple, D. Simon, S. Sivakumar, P. Smith, C. Thomas, T. Troeger, P. Vandervoorn, S. Williams, and K. Zawadzki. A 45nm Logic Technology with High-k+ Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm dry patterning, and 100% Pb-Free Packaging. In *IEDM*, 2007.

[53] S. Natarajan, M. Armstrong, M. Bost, R. Brain, M. Brazier, C. . Chang, V. Chikarmane, M. Childs, H. Deshpande, K. Dev, G. Ding, T. Ghani, O. Golonzka, W. Han, J. He, R. Heussner, R. James, I. Jin, C. Kenyon, S. Klopcic, S. . Lee, M. Liu, S. Lodha, B. McFadden, A. Murthy, L. Neiberg, J. Neirynck, P. Packan, S. Pae, C. Parker, C. Pelto, L. Pipes, J. Sebastian, J. Seiple, B. Sell, S. Sivakumar, B. Song, K. Tone, T. Troeger, C. Weber, M. Yang, A. Yeoh, and K. Zhang. A 32nm Logic Technology Featuring 2'nd-Generation High-k + Metal-Gate Transistors, Enhanced Channel Strain and $0.171\mu m^2$ SRAM Cell Size in a 291Mb Array. In *IEDM*, 2008.

[54] M. K. Hudait, G. Dewey, S. Datta, J. M. Fastenau, J. Kavalieros, W. K. Liu, D. Lubyshev, R. Pillarisetty, W. Rachmady, M. Radosavljevic, T. Rakshit, and R. Chau. Heterogeneous Integration of Enhancement Mode in0.7Ga 0.3As Quantum Well Transistor on Silicon Substrate Using Thin ($\leq 2\mu\mathrm{m}$) Composite Buffer Architecture for High-Speed and Low-Voltage (0.5V) Logic Applications. In *IEDM*, 2007.

[55] R. Chau. Challenges and Opportunities of Emerging Nanotechnology for VLSI nanoelectronics. 2007.

[56] Stuart S. P. Parkin, Masamitsu Hayashi, and Luc Thomas. Magnetic Domain-Wall Racetrack Memory. *Science*, 320(5873):190–194, 2008.

[57] I. Appelbaum, B. Huang, and D. J. Monsma. Electronic Measurement and Control of Spin Transport in Silicon. *Nature*, 447:295–298, 2007.

[58] S. Goswami, K. A. Slinker, M. Friesen, L. M. McGuire, J. L. Truitt, C. Tahan, L. J. Klein, J. O. Chu, P. M. Mooney, D. W. van der Weide, R. Joynt, S. N. Coppersmith, and M. A. Eriksson. Controllable Valley Splitting in Silicon Quantum Devices. *Nature Physics*, 3:41–45, 2007.

[59] Vienna Ab-Initio Simulation Package (VASP). http://cms.mpi.univie.ac.at/vasp.

[60] E. Ungersboeck, S. Dhar, G. Karlowatz, V. Sverdlov, H. Kosina, and S. Selberherr. The Effect of General Strain on the Band Structure and Electron Mobility of Silicon. *IEEE T-ED*, 54(9):2183–2190, 2007.

[61] T. B. Boykin, G. Klimeck, M. Friesen, S. N. Coppersmith, P. von Allmen, F. Oyafuso, and S. Lee. Valley Splitting in Low-Density Quantum-Confined Heterostructures Studied Using Tight-Binding Models. *Physical Review B*, 70(16):165325–1–165325–12, 2004.

[62] G.L. Bir and G.E. Pikus. *Symmetry and Strain-Induced Effects in Semiconductors*. J.Wiley & Sons, New York - Toronto, 1974.

[63] M. Friesen, S. Chutia, C. Tahan, and S.N. Coppersmith. Valley Splitting Theory of SiGe/Si/SiGe Quantum Wells. *Physical Review B*, 75(11):115318, 2007.

[64] O. Baumgartner, M. Karner, V. Sverdlov, and H. Kosina. Numerical Study of the Electron Subband Structure in Strained Silicon UTB Devices. In *Workshop of the Thematic Network on Silicon on Insulator Technology, Devices, and Circuits (EUROSOI)*, 2009.

[65] B. J. van Wees, H. van Houten, C. W. J. Beenakker, J. G. Williamson, L. P. Kouwenhoven, D. van der Marel, and C. T. Foxon. Quantized Conductance of Point Contacts in a Two-Dimensional Electron Gas. *Physical Review Letters*, 60(9):848–850, 1988.

[66] V. Sverdlov and S. Selberherr. Electron Subband Structure and Controlled Valley Splitting in Silicon Thin-Body SOI FETs: Two-Band $\mathbf{k}\cdot\mathbf{p}$ Theory and Beyond. *Solid-State Electronics*, 52(12):1861–1866, 2008.

[67] F. Gamiz, J.A. Lopez-Villanueva, J.B. Roldan, J.E. Carceller, and P. Cartujo. Monte Carlo Simulation of Electron Transport Properties in Extremely Thin SOI MOSFET's. *T-ED*, 45(5):1122–, 1998.

[68] T. Grasser, C. Jungemann, H. Kosina, B. Meinerzhagen, and S. Selberherr. Advanced Transport Models for Sub-Micrometer Devices. In *Proc. SISPAD*, 2004.

[69] T. Grasser, R. Kosik, C. Jungemann, H. Kosina, and S. Selberherr. Nonparabolic Macroscopic Transport Models for Device Simulation Based on Bulk Monte Carlo Data. *Journal of Applied Physics*, 97(9):093710–1 – 093710–12, 2005.

[70] M. Vasicek, J. Cervenka, M. Karner, M. Wagner, and T. Grasser. Parameter Modeling for Higher-Order Transport Models in UTB SOI MOSFETs. In *Proc. 12th IWCE*, 2007.

[71] B. Neinhüs, C.D. Nguyen, C. Jungemann, and B. Meinerzhagen. A CPU Efficient Electron Mobility Model for MOSFET Simulation with Quantum Corrected Charge Densities. In *Proc.ESSDERC*, Cork, Ireland, 2000. Frontier Group.

[72] G. Karlowatz, E. Ungersboeck, W. Wessner, and H. Kosina. Full-Band Monte Carlo Analysis of Electron Transport in Arbitrary Strained Silicon. In *Proc. SISPAD*, 2006.

[73] M. Karner, A. Gehring, S. Holzer, M. Pourfath, M. Wagner, H. Kosina, T. Grasser, and S. Selberherr. VSP - A Multi-Purpose Schrödinger-Poisson Solver for TCAD Applications. In *Proc. 11th IWCE*, 2006.

[74] M. Lundstrom. *Fundamentals of Carrier Transport*, volume X of *Modular Series on Solid State Device*. Addison-Wesley, 1990.

[75] K. Blotekjaer. Transport Equations for Electrons in Two-Valley Semiconductors. *T-ED*, ED-17(1):38–47, 1970.

[76] T. Grasser. Non-Parabolic Macroscopic Transport Models for Semiconductor Device Simulation. *Physica A*, 349(1/2):221–258, 2005.

[77] T. Grasser, H. Kosina, M. Gritsch, and S. Selberherr. Using Six Moments of Boltzmann's Transport Equation for Device Simulation. *Journal of Applied Physics*, 90(5):2389–2396, 2001.

[78] M. Vasicek, M. Karner, E. Ungersboeck, M. Wagner, H. Kosina, and T. Grasser. Modeling of Macroscopic Transport Parameters in Inversion Layers. In *Proc. ISDRS*, 2007.

[79] S.-I. Takagi, A. Toriumi, M. Iwase, and H. Tango. On the Universality of Inversion Layer Mobility in Si MOSFET's: Part I–Effects of Substrate Impurity Concentration. 41(12):2357–2362, 1994.

[80] A. Khakifirooz and D. Antoniadis. On the Electron Mobility in Ultrathin SOI and GOI. *EDL*, 25(2):80–82, 2004.