VISTA Status Report December 2005

H. Ceric, M. Karner, A. Nentchev,
P. Schwaha, E. Ungersböck, S. Selberherr

Institute for Microelectronics
Technical University Vienna
Gußhausstraße 27-29
A-1040 Wien, Austria

Contents

1 Microstructure and Stress Aspects of Electromigration Modeling

The modifications and extensions of standard continuum models used for a description of material transport due to electromigration with models for the copper microstucture are studied. Copper grain boundaries and interfaces are modeled as a network of high diffusivity paths. Additionally, grain boundaries act as sites of vacancy recombination. The connection between mechanical stress and material transport is established for the case of strain build up induced by local vacancy dynamics and the anisotropy of the diffusivity tensor caused by these strains. High diffusivity paths are set on the surfaces of polyhedral domains representing distintcive grains. These polyhedral domains are connected by diffusive, electrical, and mechanical interface models. For a numerical solution a three-dimensional finite element method is used.

1.1 Introduction

The electromigration behavior of copper interconnects realized in damascene architecture indicates macroscopic and microscopic electromigration divergence sites. Macroscopic divergence sites exist at the cathode end of via bottoms where the barrier layer can be a blocking boundary for the electromigration flux. The sites where two or more grain boundaries intersect can be considered as microscopic electromigration divergence sites. In the cases where failures are induced far away from a via, it has been shown that their activation energies are often below the expected value for the grain boundary diffusion [1].

This is a strong indication that copper interfaces to the barrier and/or capping layer are dominant diffusion paths [1]. Considering interfacial diffusion as main contribution to electromigration was a significant simplification for modeling and simulation of both void nucleation and void evolution [2, 3]. Surface treatment aiming at strengthening the copper/capping layer interface has been successfully applied to suppress interfacial diffusion [1, 4] and to increase electromigration life time. Reducing the diffusivity at the interfaces to the level of bulk and grain boundaries diffusivities necessities modeling of the grain boundary network and the crystal orientation in the grains. Moreover, intrinsic stress, introduced by the dual damascene process, has a strong impact on the bulk and grain boundary diffusion which has also to be considered [5].

The main challenge in electromigration modeling and simulation is the diversity of the relevant physical phenomena. Electromigration induced material transport is accompanied with the material transport driven by the gradients of material concentration, mechanical stress, and temperature distribution. A comprehensive, physically based analysis of electromigration for modern copper interconnect lines serves as basis for deriving sophisticated design rules which will ensure higher steadfastness of interconnects against electromigration. In the present work we study a possible extensions of the vacancy transport model described in [2] in order to include effects of the copper microstructure and mechanical stress. characteristic features of an extended model are verified by a three-dimensional simulation example.

1.2 Theoretical Background

The most comprehensive models of electromigration and accompanying phenomena are described by Mullins [6], Korhonen *et al.* [7], Sarychev *et al.*, and Kirchheim [8]. The major ideas and concepts of these models are set here into a general framework which enables their application to simulation of realistic three-dimensional interconnect layouts.

1.3 Vacancy Continuity

The bulk chemical potential of vacancies in a stressed solid can be expressed as [9, 10],

$$\mu(\sigma, C_v) = \mu_0 + \mu(0, C_v) + \frac{1}{3} f \Omega_a \operatorname{tr}(\sigma), \qquad (1)$$

where, according to [9], the chemical potential in the absence of stress is:

$$\mu(0,C_v) = k_B T \ln\Big(\frac{C_v}{C_v^0}\Big). \quad (2)$$

C_v^0 is the equilibrium vacancy concentration in a stress free solid, μ_0 is the corresponding chemical potential, and σ is the tensor of the applied mechanical stress. A vacancy flux $\vec{J}_v$ driven by gradients of chemical potential and electromigration is given by,

$$\vec{J}_v = -\frac{C_v}{k_B T}\mathbf{D}(\mathrm{grad}\mu + |Z^*|e\mathrm{grad}\varphi). \quad (3)$$

φ is the electric potential which obeys Laplace's equation ($\Delta\varphi = 0$). Since a vacancy is a point defect with cubic symmetries and copper is an fcc crystal, the tensor of diffusivity $\mathbf{D}$ is diagonal ($\mathbf{D} = D_0\mathbf{I}$).

Vacancy transport fulfills the continuity equation,

$$\frac{\partial C_v}{\partial t} = -\mathrm{div}\vec{J}_v + G, \quad (4)$$

with G as a source function which describes the vacancy generation and annihilation process. The equations (1)-(4) model electromigration of vacancies in the perfect fcc monocrystal stressed by σ.

1.4 Mechanical Stress

Since atoms and vacancies have a different volume of about 20-40% [8], the migration and recombination of vacancies induce local stress build up.

1.4.1 Vacancy Migration

We consider a small test volume V inside the interconnect metal. If n atoms leave this volume and n vacancies enter it, due to the different volume of the single vacancy and atom ($\Omega_v/\Omega_a = f < 1$) the new volume will be,

$$V_{new} = V - n\Omega_a + nf\Omega_a. \quad (5)$$

The relative volume change in this case is

$$\frac{\delta V}{V} = -(1-f)\Omega_a\frac{n}{V} = -(1-f)\Omega_a\delta C_v, \quad (6)$$

where δC_v is the increment of the vacancy concentration. With a time derivative of (6) and the well known mechanical relationship between volume increase and strain [11]

$$\frac{\delta V}{V} = \frac{V_{new}-V}{V} = \varepsilon^m_{xx} + \varepsilon^m_{yy} + \varepsilon^m_{zz} = 3\varepsilon^m, \quad (7)$$

we obtain

$$3\frac{\partial\varepsilon^m}{\partial t} = -(1-f)\Omega_a\frac{\partial C_v}{\partial t}. \quad (8)$$

For the test volume V the vacancy continuity holds

$$-\mathrm{div}\vec{J}_v = \frac{\partial C_v}{\partial t}. \quad (9)$$

From (8) and (9) we obtain for the components of the migration strain tensor

$$\frac{\partial\varepsilon^m_{ij}}{\partial t} = \frac{1}{3}(1-f)\,\Omega_a\,\mathrm{div}\vec{J}_v\delta_{ij}. \quad (10)$$

1.4.2 Vacancy Recombination

Using the same concept as given above we calculate the new volume V_{new} as a result of production (annihilation) of n vacancies inside the initial volume,

$$V_{new} = V \pm nf\Omega_a. \quad (11)$$

Now we can express a relative volume change as,

$$\frac{\delta V}{V} = \frac{V_{new}-V}{V} = \pm f\Omega_a\delta C_v. \quad (12)$$

Using relation (7) and the time derivative we obtain

$$3\frac{\partial\varepsilon^g}{\partial t} = \pm f\Omega_a\frac{\partial C_v}{\partial t}. \quad (13)$$

The time derivative $\partial C_v/\partial t$ in this case is equal to the vacancy production/annihilation source function G. Thus the time change of the strain caused by vacancy recombination is given by,

$$\frac{\partial\varepsilon^g_{ij}}{\partial t} = \frac{1}{3}f\,\Omega_a\,G\,\delta_{ij}. \quad (14)$$

From (10) and (14), we obtain a kinetic relation for the strain caused by vacancy migration and recombination,

$$\frac{\partial\varepsilon^v_{ij}}{\partial t} = \frac{\Omega_a}{3}\Big[(1-f)\mathrm{div}\vec{J}_v + fG\Big]\delta_{ij}. \quad (15)$$

1.4.3 Stress Equilibrium

According to [9] the general form of the mechanical equilibrium equation is

$$\sum_{j=1}^{3} \frac{\partial \sigma_{ij}}{\partial x_j} = 0, \quad \text{for} \quad i = 1,2,3. \tag{16}$$

Taking into account the strain induced by vacancy migration and recombination we obtain [12]

$$\sigma_{ij} = (\lambda \,\mathrm{tr}(\varepsilon) - B\,\mathrm{tr}(\varepsilon^v))\delta_{ij} + 2G\varepsilon_{ij}, \tag{17}$$

where λ and G are Lame's constants and $B = (3\lambda + 2G)/3$ is the bulk modulus. The strain tensor ε^v is defined by relation (15).

1.5 Anisotropic Diffusivity

In the case of a homogeneously deformed cubic crystal with strain field ε the vacancy diffusivity tensor obtains additional contributions [13]

$$D_{ij} = D_0\,\delta_{ij} + \sum_{k,l=1}^{3} d_{ijlk}\,\varepsilon_{kl}, \tag{18}$$

where d_{ijlk} is the elastodiffusion tensor. Equation (18) shows that strain causes an anisotropy of the diffusivity tensor. A comprehensive analysis of the point defect jump frequencies in a strained solid and calculation of the elastodiffusion tensor components is provided in [5].

1.6 Microstructure

The network of grain boundaries influences vacancy transport during electromigration in several different ways. The diffusion of point defects inside the grain boundary is faster compared to grain bulk diffusion due to the fact [14] that a grain boundary generally exibits 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 polycrystalline metals, grain boundaries are also recognized (together with dislocations loops) as sites of vacancy generation and annihilation [8, 15]. During the diffusion process vacancies generally seek to reach a concentration C_v^{eq} which is in equilibrium with the local stress distribution,

$$C_v^{eq} = C_v^0 \exp\Big(-f\,\frac{\mathrm{tr}(\sigma)\Omega}{3\,k_B\,T}\Big). \tag{19}$$

This tendency is supported by recombination mechanisms which are commonly modeled by a source function G in the form introduced by Rosenberg and Ohring [16],

$$G = -\frac{C_v - C_v^{eq}}{\tau}, \tag{20}$$

which means production of vacancies, if their concentration is lower than the equilibrium value C_v^{eq} and their annihilation in the opposite case. τ is the characteristic relaxation time [17]. The full understanding of the source function G is still missing but it surely has to comprise three processes: exchange of point defects between adjacent grains, exchange of point defect between grains and grain boundaries, and point defect formation/annihilation inside the grain boundaries.

1.7 Simulation Example

We consider an interconnect via realized in dual damascene architecture consisting of copper, capping, and diffusion barrier layers (Figure 1). The copper segment is split into polyhedral grains (Figure 2). For the solution of the governing equations (1)-(4) an in-house finite element method code is used. The diffusion coefficient along the grain boundaries and the copper interfaces to the capping and barrier layers is assumed to be 5000 times larger than that in the bulk regions.

The Rosenberg and Ohring recombination term G is assumed to be active only in the close vicinity of the grain boundaries. The vacancy concentration on both ends of the via is kept at the equilibrium level during simulation and all materials are assumed to be relaxed. The obtained vacancy distribution is presented in Figure 3. Consistent with experimantal results [18] the peak values of the vacancy concentration develop at the intersection lines of the grain boundaries and the capping layer.

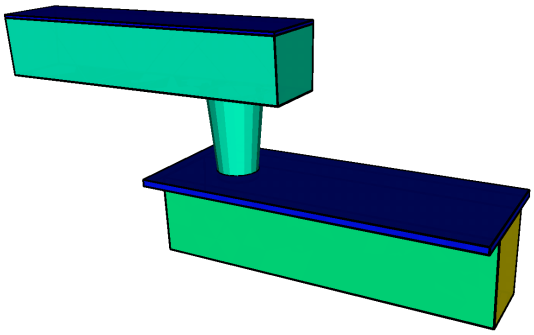

Figure 1: Typical dual-damascene layout used for simulation.

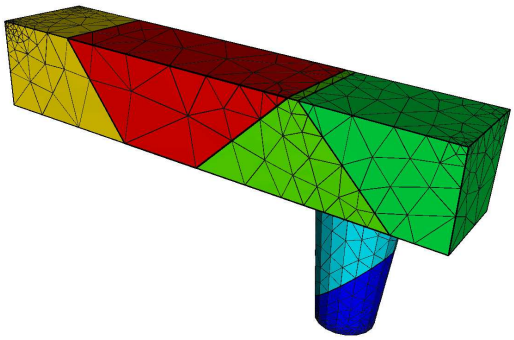

Figure 2: The copper segment is split into polyhedral grains and each polyhedron is separately meshed with initial mesh.

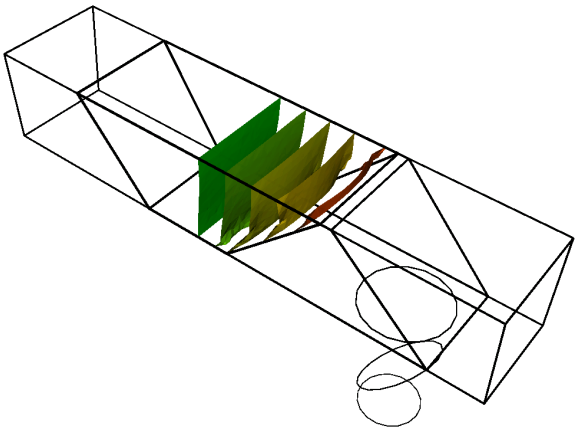

Figure 3: The peak value of vacancy concentration (displayed iso surfaces) is accumulated at the grain boundary/capping layer crossing line.

1.8 Conclusion

A careful analysis of the connection between the local vacancy dynamics and strain build-up has been carried out. The obtained relations have been coupled to an electromigration model using the concepts of stress driven diffusion and anisotropy of the diffusivity tensor.

For a correct physical handling of the grain boundary network as the network of high diffusivity paths and at the same time as sites of vacany recombination, the method of splitting of a copper segment into grain segments is introduced. The grain boundary segments are treated as simulation sub-domains connected to each other by diffusive, mechanical, and electrical interface conditions.

A dual-damascene architecture example layout is used to illustrate and verify the introduced modeling approach. The obtained simulation results qualitatively resemble the behavior observed in experimental investigations.

2 Efficient Calculation of Lifetime Based Direct Tunneling Through Stacked Dielectrics

We present an efficient simulation method for lifetime based tunneling in CMOS devices through layers of high-κ dielectrics, which relies on the precise determination of quasi-bound states (QBS). The QBS are calculated with the perfectly matched layer (PML) method. Introducing a complex coordinate stretching allows artifical absorbing layers to be applied at the boundaries. The QBS appear as the eigenvalues of a linear, non-Hermitian Hamiltonian where the QBS lifetimes are directly related to the imaginary part of the eigenvalues. The PML method turns out to be a numerically stable and efficient method to calculate QBS lifetimes for the investigation of direct tunneling through stacked gate dielectrics.

2.1 Introduction

The continuous progress in the development of MOS field-effect transistors within the last decades goes hand in hand with down-scaling the device feature size. To enable further device down-scaling to the deca nanometer channel length regime, it is necessary to reduce the effective oxide thicknesses (EOT) below 2 nm, which will result in high gate leakage currents. The use of high-κ gate dielectrics provides an option to reduce the gate leakage current of future CMOS devices while retaining a good control over the inversion charge [19].

Gate dielectric stacks consisting of high-κ dielectric layers such as Si_3N_4, Al_2O_3, Ta_2O_5, HfO_2, or ZrO_2 have been suggested as alternative dielectrics. Parameter values for these materials taken from [20]-[21] are summarized in Table 1.

Apart from interface quality and reliability, the dielectric permittivity and the conduction band offset to silicon are of utmost importance as they determine the gate current density through the layer. Furthermore, at the interface to the underlying silicon substrate, an interface layer exists which is either created unintentionally during processing or intentionally deposited to improve the interface quality. Unfortunately, materials with high permittivity have a low band offset and vice versa, so that a trade-off between these parameters has to be found. However, for investigation of tunneling phenomena and especially for optimization purposes, accurate, and yet efficient simulation models are necessary.

2.2 Calculation of Direct Tunneling using a Lifetime Based Approach

Calculation of tunneling currents is frequently based on the assumption of a three-dimensional continuum of states at both sides of the gate dielectric and the conservation of parallel momentum. Then, the tunneling current can be described by the Tsu-Esaki formula [22],

$$J_{3D} = \mathrm{q} \int_{\mathcal{E}_{\min}}^{\mathcal{E}_{\max}} TC(\mathcal{E}_x, m_{\mathrm{diel}}) N(\mathcal{E}_x, m_{\mathrm{D}})\, \mathrm{d}\mathcal{E}_x, \quad (21)$$

where $TC(\mathcal{E}_x, m_{\mathrm{diel}})$ is the transmission coefficient and $N(\mathcal{E}_x, m_{\mathrm{D}})$ the supply function.

Two electron masses enter this equation: The density-of-states mass in the plane parallel to the interface, $m_{\mathrm{D}} = 2m_{\mathrm{t}}^* + 4\sqrt{m_{\mathrm{t}}^* m_{\mathrm{l}}^*}$, which, equals $2.052 m_0$ for (100) silicon with $m_{\mathrm{l}}^* = 0.92 m_0$ and $m_{\mathrm{t}}^* = 0.19 m_0$, and the electron mass in the dielectric m_{diel}, which is commonly used as a fit parameter [23].

However, in the inversion layer of a MOS-structure, the strong electric field leads to quantum confinement. Whenever electrons are confined or partially confined in movement, this gives rise to bound or quasi bound states (QBS), and the assumption of continuum tunneling is no longer valid. In the inversion layers of MOS-FETs, a major, if not the dominant, source of tunneling electrons is represented by quasi bound states [24].

The QBS tunneling current is proportional to $\sum n_i/\tau_i$ where n_i and τ_i denote the carrier concentration and the lifetime of the QBS with index i, respectively.

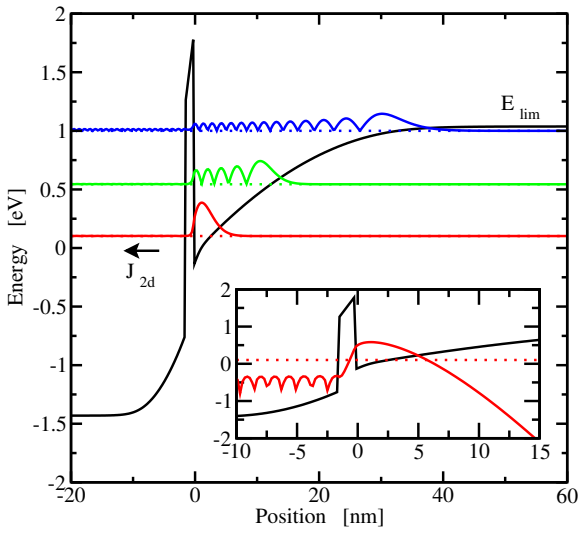

Figure 4: The potential well of an nMOS inversion layer and its eigenstates assuming closed boundary conditions. The inset displays the wave function of the first QBS on a logarithmic scale.

Table 1: Dielectric permittivity, band gap, and conduction band offset of dielectric materials.

	Permittivity κ/κ_0 [1]	Band gap $\mathcal{E}_g$ [eV]	Offset $\Delta\mathcal{E}_C$ [eV]
SiO_2	3.9	8.9 – 9.0	3.0 – 3.5
Si_3N_4	7.0 – 7.9	5.0 – 5.3	2.0 – 2.4
Ta_2O_5	23.0 – 26.0	4.4 – 4.5	0.3 – 1.5
TiO_2	39.0 – 170.0	3.0 – 3.5	0.0 – 1.2
Al_2O_3	7.9 – 12.0	5.6 – 9.0	2.78 – 3.5
ZrO_2	12.0 – 25.0	5.0 – 7.8	1.4 – 2.5
HfO_2	16.0 – 40.0	4.5 – 6.0	1.5

To take into account the tunneling current from both, continuum and quasi-bound states, (21) has to be replaced by

$$J = J_{2D} + J_{3D} = \\ = \frac{k_B T q}{\pi\hbar^2} \sum_{i,\nu} \frac{g_\nu m_\parallel}{\tau_\nu(\mathcal{E}_{\nu,i}(m_q))} \ln\left(1+\exp\left(\frac{\mathcal{E}_F-\mathcal{E}_{\nu,i}}{k_B T}\right)\right) + \\ + q\int^{\mathcal{E}_{max}} TC(\mathcal{E}_x, m_{diel}) N(\mathcal{E}_x, m_D)\, d\mathcal{E}_x. \quad (22)$$

Here, the symbols g_ν, $m_\parallel$, and m_q denote the valley degeneracy, parallel, and quantization masses respectively ($g=2$: $m_\parallel = m_t$, $m_q = m_l$ and $g=4$: $m_\parallel = \sqrt{m_l m_t}$, $m_q = m_t$), $\tau_\nu(\mathcal{E}_{\nu,i})$ is the lifetime of the quasi-bound state $\mathcal{E}_{\nu,i}$, and the integration in the Tsu-Esaki formula starts from $= E_{lim}$ as indicated in Figure 4. The following considerations are focused on the tunneling current J_{2D} originating from the QBS.

Within our simulation framework the QBS are obtained from the single particle, time-independent, effective mass SCHRÖDINGER equation:

$$-\frac{\hbar^2}{2}\nabla\cdot\left(\tilde{m}^{-1}\nabla\Psi(\mathbf{x})\right) + V(\mathbf{x})\Psi(\mathbf{x}) = \mathcal{E}\Psi(\mathbf{x}). \quad (23)$$

Several methods have been proposed to calculate the quasi-bound states and their respective lifetimes [25]. In a first approximation the energy levels of the QBS can be estimated by the eigenvalues of the Hamiltonian of the closed system as displayed in Figure 4. Since closed boundaries are assumed, no information about the broadening and the associated QBS lifetimes is available. It is to note that bound states cannot carry any current, since their wavefunctions Ψ fulfill the relation: $\Psi\nabla\Psi^* - \Psi^*\nabla\Psi = 0$.

A semi-classical approximation based on corrected closed-boundary eigenvalues, which uses a classical formulation of the lifetime (escape time) is pointed out in [26]. However, using the closed-boundary eigenvalues for the calculation of open-boundary QBS lifetimes seems to be questionable.

A more rigorous way to apply open boundary conditions to (23) is the quantum transmitting boundary method (QTBM) [27] where a computationally intensive scanning of the derivative of the phase of the reflection coefficient [25] or the reflection coefficient itself [28] yields the desired QBS lifetimes. These methods are especially demanding in the presence of strong confinement (high lifetimes).

2.3 Perfectly Matched Layer Method

Recently, a method based on absorbing boundary conditions (known as the Perfectly Matched Layer (PML) method) for SCHRÖDINGER's equation has been applied for band structure calculations in III-V heterostructure devices [29]. In the present work the PML formalism which is often used in

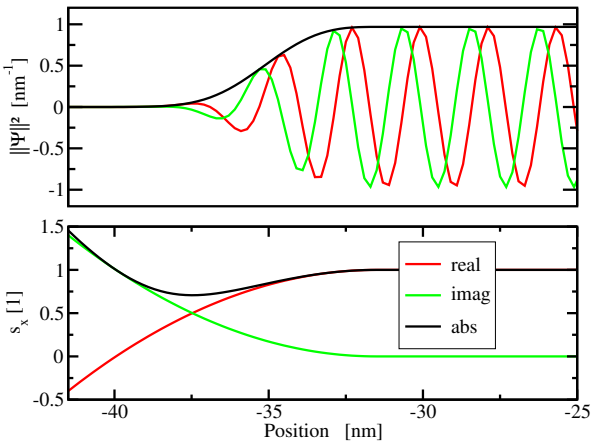

Figure 5: The wave function of the first QBS and the complex stretching function are displayed in the perfectly matched layer region as well as its transition to the physical region.

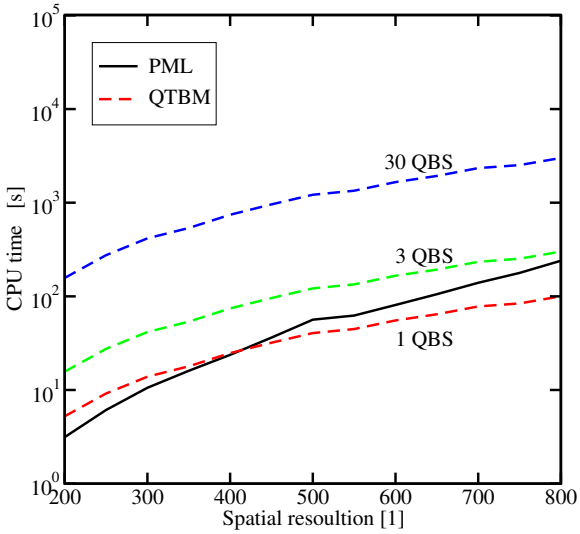

Figure 6: Comparison of the CPU time demand for the PML, and the QTB methods.

electromagnetics, has been applied to determine the energy levels and the lifetime broadening of QBS in MOS inversion layers. In contrast to the QTBM, the Hamiltonian of the system is still linear. Thus, all QBS are calculated in one step and no iteration or scanning procedures are needed.

The basic principle is to add non-physical absorbing layers at the boundary of the simulation region (physical region). This procedure prevents reflections at the boundary of the physical region. The artificial absorbing layers allow the application of Dirichlet boundary conditions, and the QBS are determined by the eigenvalues of the non-Hermitian Hamiltonian of the system. This yields the desired QBS which are the eigenstates of the open system, although Dirichlet boundary conditions are applied. The absorbing property of the PML region is achieved by introducing stretched coordinates

$$\tilde{x} = \int_0^x s_x(\tau)\,\mathrm{d}\tau \tag{24}$$

in (23). The evaluation of the gradient operator ∇ in one dimension yields:

$$\frac{\partial}{\partial \tilde{x}} = \frac{1}{s_x(x)}\frac{\partial}{\partial x}. \tag{25}$$

In the artificial layers the stretching function $s_x(x)$ is given as $s_x(x) = 1 + (\alpha + \imath\beta)x^n$, with $\alpha = 1$, $\beta = 1.4$, and $n = 2$, while it is unity in the physical region as displayed in Figure 5. Adding absorbing layers at the boundary of the physical simulation region, the Hamiltonian becomes non-Hermitian and admits complex eigenvalues $\mathcal{E} = \mathcal{E}_r + \imath\mathcal{E}_i$. The QBS lifetimes are related to the imaginary parts of the eigenvalues as $\tau_i = \hbar/2\mathcal{E}_i$.

To better clarify the PML method, let us assume a constant potential $V(z)$ in the PML region. Then, within this region, the wave function can be written as a plane wave $\Psi(x) = \Psi_0 \exp(\imath\tilde{k}_x x)$ with the wave vector $\tilde{k}_x = k_x/s_x$. Considering two points in the PML region x_1, $x_2 = x_1 + \mathrm{d}x$ the wave vector at the point x_2 can be approximated as

$$k_x(x_2) \approx \frac{s_x(x_2)}{s_x(x_1)} k_x(x_1) = (1 + (\alpha + \imath\beta)\,\mathrm{d}x)\,. \tag{26}$$

Therefore, the parameter α scales the phase velocity of the plane wave, while β acts as a damping parameter. Since this damping coefficient is greater than zero in the absorbing region, the envelope of the wave functions decay to zero, as can be seen in Figure 5.

These parameters, as well as the thickness of the absorbing layer can be varied over a wide range with virtually no influence on the results, as long as there are no reflections at the boundaries. However, to achieve this goal, the complex stretching function and its first derivative have to be continuous.

In the gate region, using QTBM or assuming closed boundary conditions results in a superposition of two plane waves in opposite directions, which can bee seen in the inset of Figure 4. In contrast, when using PML, there are no reflected waves. The wave function is a traveling wave with a constant envelope function. In the absorbing layer, the wave functions are gradually decaying to zero (Figure 5). The QBS, however, are reproduced correctly.

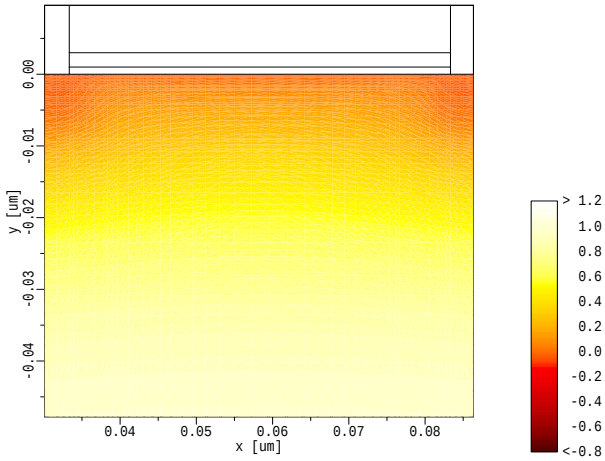

Figure 7: The band edge energy [eV] of the nMOS device with a stacked gate dielectric evaluated at a gate bias of 1.5 V and a drain voltage of 0 V.

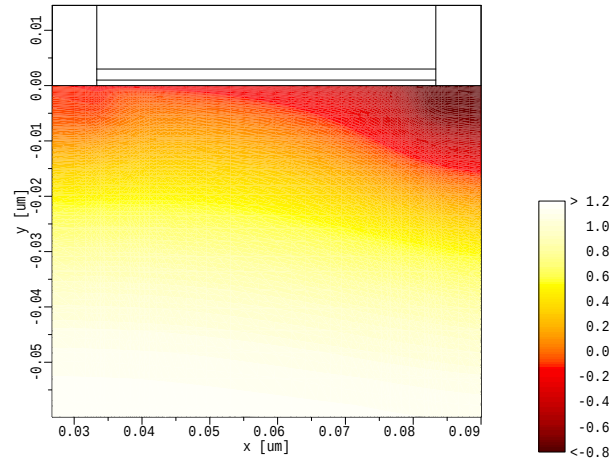

Figure 8: The band edge energy [eV] of the nMOS device with a stacked gate dielectric evaluated at a gate bias of 1.5 V and a drain voltage of 0.6 V.

For an arbitrary potential well a comparison between the PML method and the established methods has been carried out in [30]. Very good agreement between the established QTBM and the PML formalism has been obtained. Furthermore, the computational effort of the PML and QTBM approaches was compared.

Figure 6 shows the CPU time necessary to calculate 1, 3, and 30 quasi-bound states with the QTB and PML methods as a function of the spatial resolution. For the QTBM, an equidistant grid in energy space was used to determine the lifetime broadening of the QBS. Although the dimension of the system increases due to the additional points in the PML region, the computational effort of the PML method has shown to be in almost all cases lower than that of the QTBM.

2.4 Application to Device Simulation

With the described method, the gate leakage currents of nMOS transistors with a gate length of 50nm have been evaluated. The gate current density has been evaluated for a stacked SiO_2-Si_3N_4 and a single SiO_2 layer gate dielectric having nearly the same EOT. A doping of N_A=3 $\times$ 10^{17}cm^{-3} in the bulk and N_D=1 $\times$ 10^{19}cm^{-3} in the poly-silicon gate was assumed. For the investigation of gate leakage currents in the whole device, the conduction band edge has been acquired from the device

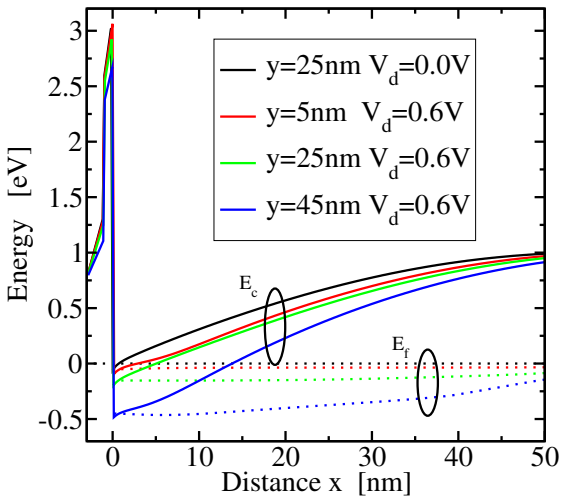

Figure 9: Cuts of conduction band edge energy of the nMOS transistor. The y-coordinate is relative to the beginning of the gate contact.

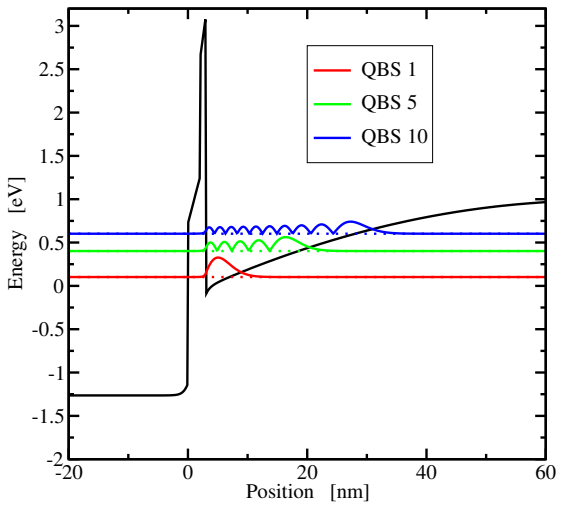

Figure 10: Potential barrier and eigenstates assuming open boundary conditions using the PML technique.

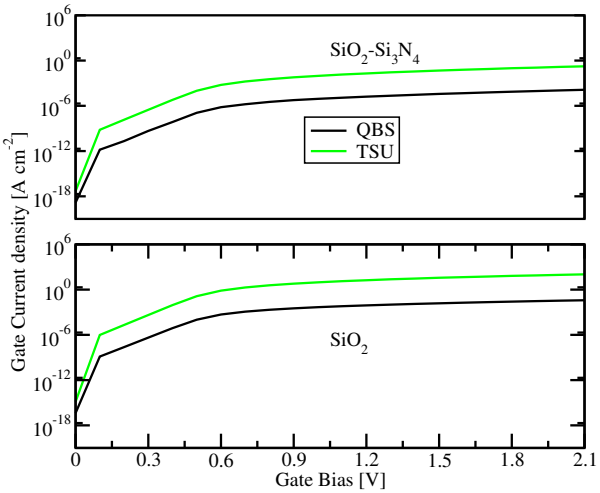

Figure 11: The gate current density for a single SiO_2 layer as well as for a stacked SiO_2-Si_3N_4 dielectric calculated from the Tsu-Esaki formula and the lifetime based approach.

simulator MINIMOS-NT [31]. It is displayed for strong inversion at a gate bias of 1.2 V and V_{DS}=0.0V in Figure 7, and at drain bias of 0.6 V in Figure 8. Several one-dimensional cuts through the simulation region are shown in Figure 9.

As a post-processing step on these cuts the QBS energy levels and the related lifetimes have been evaluated using the PML formalism. Based on an accurate computation of the QBS lifetimes, the tunneling current has been estimated according to (22). For the stacked gate dielectric some of the extracted quasi-bound states are shown in Figure 10 considering the transversal mass as the quantization mass at V_{GB}=1.2V. The energy levels, the QBS lifetimes, and their contribution to the total current density are listed in Table 2.

Table 2: The QBS of the MOS capacitor for a gate bias of 1.2V, the corresponding lifetimes, and their contribution to the total gate current density.

QBS	$\mathcal{E}_r$ [eV]	τ_l [s]	J_G [A cm^{-2}]
1	0.054	2.1×10^{-4}	3.2×10^{-3}
2	0.210	8.5×10^{-5}	2.0×10^{-5}
3	0.326	3.7×10^{-5}	5.1×10^{-8}
5	0.507	8.5×10^{-6}	1.9×10^{-10}

The resulting IV-characteristics as a function of the gate voltage for zero drain bias of the two structures are compared in Figure 11. It can be seen that the gate current leakage of the stacked dielectric is considerably smaller. Furthermore, we have to point out that the Tsu-Esaki approach overestimates the gate current leakage under inversion conditions. Thus, the use of the more sophisticated lifetime based approach is mandatory for accurate modeling of direct tunneling through stacked gate dielectrics under inversion conditions.

2.5 Conclusion

We presented an efficient approach for the estimation of lifetime based tunneling currents through stacked gate dielectrics. The lifetimes of quasi bound states (QBS) have been evaluated with the perfectly matched layer (PML) formalism. The traditional approach requires a computationally very demanding scanning procedure.

The QBS lifetimes appear as the complex eigenvalues of a non-Hermitian Hamiltonian. Since the equation to be solved is linear, highly efficient algorithms are available. Moreover, the PML approach was used to evaluate QBS in the conduction band on several cuts of the MOS inversion

layer and its contribution to the total gate leakage current was determined. For typical device parameters, the QBS tunneling is the dominant tunneling component. The PML formalism represents an efficient and numerically stable method to determine QBS. Therefore, it is appropriate for integration in a device simulator for the investigation of direct tunneling phenomena.

3 On-Chip Interconnect Simulation of Parasitic Capacitances in Periodic Structures

Parasitic capacitances determine significantly the performance of todays complex on-chip interconnect structures. This article handles the implementation of periodic boundary conditions for electrical field calculation and parasitic capacitance extraction in interconnect structures. Various examples show the impact of these convenient boundary conditions.

3.1 Introduction

The numerical capacitance calculation uses commonly the energy method $C = \frac{U^2}{2W}$, where U is the voltage applied between the conductors and W is the electric energy stored in the interior of the dielectric $\mathcal{V}_D$. W can be derived from the electric field $\vec{E}$:

$$W = \int_{\mathcal{V}_D} \vec{E} \cdot \vec{D} \mathrm{d}V = \varepsilon_0 \int_{\mathcal{V}_D} \vec{E} \cdot (\tilde{\varepsilon_r} \cdot \vec{E}) \mathrm{d}V. \quad (27)$$

$\vec{E}$ is given by $\vec{E} = -\vec{\nabla}\varphi$ and φ is the solution of the equation:

$$\vec{\nabla} \cdot (\tilde{\varepsilon_r} \cdot \vec{\nabla}\varphi) = -\frac{\rho}{\varepsilon_0}. \quad (28)$$

The insulators are free of electric charge ($\rho = 0$). Therefore, the solution of (28) is completely extracted from the data, defined on the boundary $\partial\mathcal{V}_D$. On one part of this boundary Dirichlet conditions ($\varphi(\vec{r}) = \varphi_0$) are applied and on the other part homogeneous Neumann conditions ($\vec{n}.\vec{\nabla}\varphi(\vec{r}) = 0$) [32].

For the numerical solution of the involved partial differential equation (28) the finite element method (FEM) [33] is applied. The simulation area is discretized in tetrahedrons. The algorithm for the linear algebraic equations arising from the finite element discretization is based on the iterative conjugate gradient method, which uses incomplete Cholesky preconditioning technique to speed up the iteration convergence [34]. The accurate numerical calculation of boundary value problems such as (28) requires an appropriately fine discretization. In large areas this leads to generation of a lot of simulation nodes, which means sizeable memory consumption and unacceptably long duration of the simulation process.

Often the interconnects represent regular structures (i.e. on-chip interconnect buses, DRAM cells) which can be described through mirroring and periodic spatial continuation of a given subspace [35]. A large simulation domain can be composed from numerous sub-domains [36].

A smart simplification is to simulate only in this small sub-domain by applying boundary conditions corresponding to the way how the whole domain is composed from the sub-domain pattern (by mirroring or periodic extension). Therefore the simulation duration and memory consumption are decreased heavily. To use this feature the simulation software has to provide mirroring and periodic functionality. The mirroring can be easily accomplished by applying homogeneous Neumann boundary conditions at the mirroring surface. However, the periodic boundaries require special treatment.

We define two faces $\mathcal{A}_{1p} \subset \partial\mathcal{V}$ and $\mathcal{A}_{2p} \subset \partial\mathcal{V}$ as periodic boundary, if:

- Each node from $\mathcal{A}_{1p}$ is uniquely mapped to another node from $\mathcal{A}_{2p}$.
- If $\vec{r}_{1i}$ is the position pointer to a simulation point of $\mathcal{A}_{1p}$ and $\vec{r}_{2i}$ is the position pointer to the corresponding simulation point of $\mathcal{A}_{2p}$ then $\varphi(\vec{r}_{1i}) = \varphi(\vec{r}_{2i})$ for each point of $\mathcal{A}_{1p}$ and $\mathcal{A}_{2p}$.
- Each node of $\mathcal{A}_{1p}$ has its own neighbor nodes and the neighbor nodes of the corresponding node from $\mathcal{A}_{2p}$.

Although each two corresponding periodic points are separated in the space, due to the periodic condition, they should behave as if they were attached to each other.

3.2 Domain Discretization

The discretization of the volume of interest is usually the first step of the finite element analysis. In this case the simulation domain is subdivided into a number of small volume elements, in our case tetrahedral elements. Therefore the surface is broken into a number of triangular elements.

For the sake of clarity this paragraph is restricted to parallelepiped structures. Of course periodic boundary conditions can be applied to an arbitrary pair of faces with unique bidirectional node mapping. If at two opposite parallel faces periodic boundary conditions are applied, the grid generated has to guarantee that the surface grids at these faces are identical. Two different three-dimensional grid generation approaches are used in our interconnect simulation software *Smart Analysis Programs*.

The first one is a fully unstructured grid generation approach which uses the program delink [37, 38]. The second approach is a layered based method which extends two-dimensional grid generator [39] into the third dimension by means of linear extrusion. Both approaches do not fulfill the above mentioned requirements for periodic boundaries a priori. To extend the grid generator for periodic boundaries an iterative approach is used.

At first the simulation domain is meshed without any special treatment for periodic boundaries. Afterwards the periodic boundary faces are checked for conformity. If they are not conform the periodic nodes are merged. Therefore at the periodic surfaces new points are generated which are fed into the grid generator as additional input.

These steps are repeated until conformity is reached. In the layer based meshing approach this iteration procedure must only be applied to the two-dimensional grid generation process. The conformity of the side walls is preserved by the following extrusion step.

In the fully unstructured meshing method the conformity of the nodes on the periodic faces is not sufficient, because the same set of boundary nodes can lead to different boundary meshes (at least for cospherical points). Therefore also edge conformity has to be guaranteed. Because of these additional difficulties the layer based grid generation method is preferred for problems with periodic boundaries.

3.3 Assembling

If due to the discretization N points are created, the electric potential φ in $\mathcal{V}_D$ is approximated by the sum:

$$\varphi \approx \tilde{\varphi} = \sum_{j=1}^{NU} c_j N_j(\vec{r}) + \sum_{j=NU+1}^{N} c_j N_j(\vec{r}). \qquad (29)$$

The shape function N_j has the value 1 only on the node j (for $\vec{r} = \vec{r}_j$). N_j is different from 0 only in the elements directly attached to the node j and is 0 otherwise. Therefore the weighting factor c_j represents the potential on the node j.

The points which do not belong to $\partial\mathcal{V}_{D1}$ are the unknown nodes and are numbered by $1 \leq j \leq NU$. The Dirichlet (known) nodes (at $\partial\mathcal{V}_{D1}$) are numbered by $NU+1 \leq j \leq N$. Using (29) FEM leads to a linear equation system for the unknown c_j.

In general the grid generation software does not order the simulation nodes as in (29). To implement the desired node ordering a supplemental auxiliary index array with the length N is allocated. This additional index array is used by the assembling procedure. The first NU entries of this index array refer to the nodes in $\mathcal{V}_D$ without $\partial\mathcal{V}_{D1}$.

The remaining entries refer to the nodes on the Dirichlet boundary $\partial\mathcal{V}_{D1}$ (from $NU+1$ till N). The additional index assignment of the simulation nodes gives advantages to the implementation of the periodic boundary conditions. Each two corresponding points of the plains $\mathcal{A}_{1p}$ and $\mathcal{A}_{2p}$ get the same index in the additional index array.

Thus, they are assembled to the same row in the linear equation system. Due to the element-by-element processing of the simulation volume each periodic point has not only its neighbor nodes but it is also connected to the neighbor nodes of the corresponding periodic point.

3.4 Conceptual Formulation

In the presented example which is considered part of a representative on-chip bus structure the simulation area consists of a SiO_2 rectangular parallelepiped with two parallel conductors inside as shown in Figure 12. The z axis is oriented such that the x y and z axes build a right aligned coordinate system. On one conductor a voltage of 1 V is applied, while the other is set to 0 V. The conductors in this simulation area are shown from another viewing angle in Figure 13.

On the interface between the dielectric and the electrodes Dirichlet conditions are applied, which comply with the potential of the electrodes. At the boundaries parallel to the xy plane homogeneous Neumann conditions are applied. For the remaining outer faces different combinations of homogeneous Neumann conditions and periodic boundaries are investigated. Homogeneous Neumann conditions influence the electric field in the simulation domain, as if the simulation domain would be mirrored with respect to the face, at which the homogeneous Neumann conditions are applied.

It is convenient to combine opposite faces to periodic boundaries. In this case the electric field in the simulation area is influenced as if the simulation area would be shifted along the direction perpendicular to these faces by the corresponding length of the simulation area. That way the simulation domain appears as if it could be a part of a structure, which is constructed by mirroring of the simulation domain along the z direction and by periodic or mirrored spatial continuation of the simulation domain along the x and y directions. Such a structure could be for instance an interconnect bus. The total capacitance is the sum of the capacitances from all parts.

3.5 The Electric Field

The simulation results are evaluated by visualization of the electric field using VTK[40]. The differences between the investigated cases are well shown by the potential iso faces and by the angles between these potential iso faces and the outer

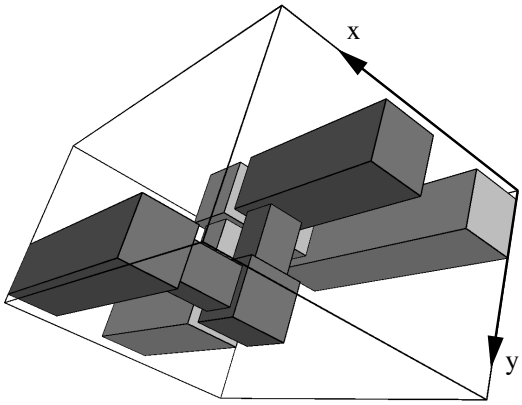

Figure 12: The Simulation Area.

boundaries. The simulated potential distribution confirms with the expected one under consideration of the corresponding boundary conditions.

3.5.1 Mirroring along the x and y axes

At the boundary faces of the simulation area parallel to yz and xz planes homogeneous Neumann conditions are specified. Therefore the simulation domain is mirrored in $\vec{x}$, $-\vec{x}$, $\vec{y}$ and $-\vec{y}$ directions. The mirror planes are the planes, at which homogeneous Neumann conditions are defined. The potential distribution and the corresponding iso faces are shown in Figure 14 and Figure 15. As expected the potential iso faces are perpendicular to the outer bounds. The iso faces should wrap round an electrode and its imaginary mirror images. Therefore iso faces can be seen only between the two electrodes in

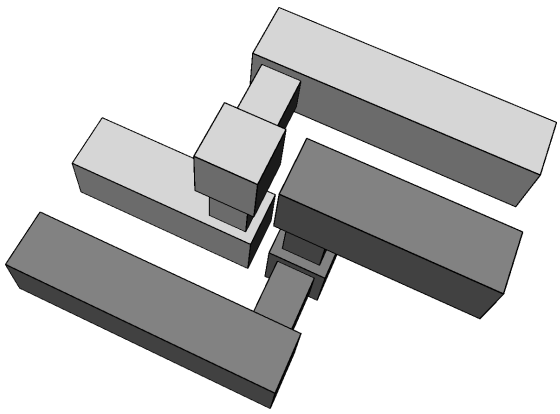

Figure 13: The Electrodes.

the simulation area. The other cases should be compared to this one to indicate the changes caused by the substitution of the Neumann boundaries for the periodic boundaries.

3.5.2 Periodicity along the x and Mirroring along the y Axis

In this case, at the boundary faces of the simulation area parallel to the yz plane a periodic condition is defined. That is as if the simulation domain would be moved in $\vec{x}$ and $-\vec{x}$ direction by the length of its x dimension. In $-\vec{y}$ and $\vec{y}$ direction the simulation domain is mirrored. This behavior can be observed in Figure 16 and Figure 17.

The electric field looks like, as if the one boundary parallel to yz would be directly connected to the opposite one. The stamp of the one of the electrodes which is lying on the one of the periodic faces can be seen on the other periodic face. The iso faces are not any more perpendicular to the boundaries parallel to the yz plane. Of course they are still perpendicular to the boundaries parallel to the xz plane.

3.5.3 Mirroring along the x and Periodicity along the y Axis

An interconnect bus structure is modeled, which is built by mirroring the simulation domain from Figure 12 along the x axis and by periodic spatial iteration along the y axis. This case is similar to the previous one. As shown in Figure 18 and Figure 19 the potential iso faces are not perpendicular to the boundaries parallel to the xz plane, but are still perpendicular to the boundaries parallel to the yz plane. This time the boundaries parallel to the xz plane seem to be connected to each other.

3.5.4 Periodicity along the x and y Axes

In this case periodic boundaries are applied at the faces parallel to the xz and yz planes. That way the structure of Figure 12 is spatially continued periodically in $\vec{x}$, $-\vec{x}$, $\vec{y}$ and $-\vec{y}$ directions. The potential distribution and the corresponding iso faces are shown in Figure 20 and Figure 21. Each boundary parallel to the z axis looks so as if it would be connected to the opposite one. They also lie not any more perpendicular to the potential iso faces. Each electrode is placed among the second electrode in the simulation area and the imaginary periodic images. The iso faces in the simulation area look correspondingly.

As mentioned above the periodic boundary condition is not applied at the faces parallel to the xy plane. For the specific simulation domain of Figure 12 in order to have a continuous interconnect structure in z direction only mirroring (homogeneous Neumann boundary) can be applied.

3.6 Capacitance

Table 3 shows the extracted capacitance values depending on the different boundary conditions applied. C_0 is the capacitance between the conductors of Figure 12 if at all boundary faces homogeneous Neumann conditions are applied. The capacitance of the three other cases is relative to it.

Table 3: The Capacitance Values Extracted Using Different Boundary Conditions.

x mirroring and y mirroring	C_0
x periodic and y mirroring	$1.33C_0$
x mirroring and y periodic	$1.07C_0$
x periodic and y periodic	$1.69C_0$

The smallest capacitance occurs if no periodic boundaries are specified. The biggest capacitance is in the case of x and y periodicity. The calculated capacitance values refer only to the small simulation area as defined in Figure 12. This simulation area is used to construct the whole domain. The capacitance of the whole area is the capacitance of the small simulation domain multiplied by the number of all small simulation domains needed to construct the complete structure.

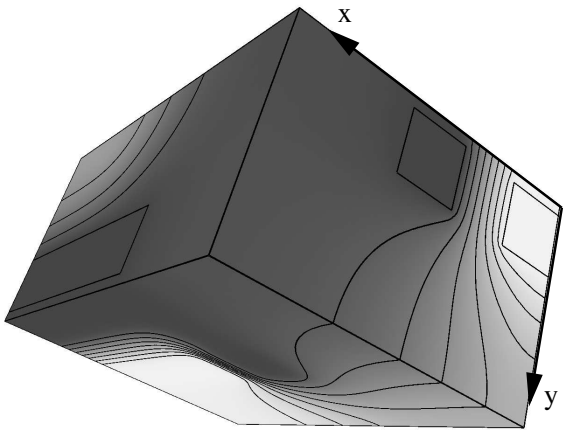

Figure 14: The Potential Distribution without Periodicity.

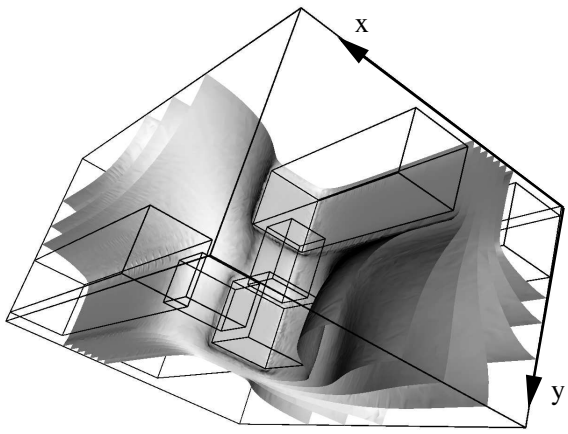

Figure 15: The Iso Faces of the Potential Distribution without Periodicity.

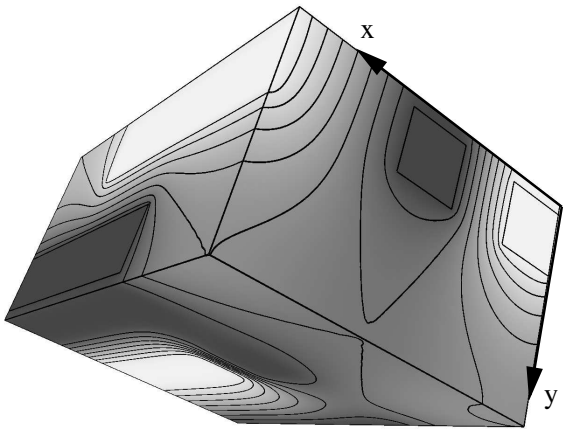

Figure 16: The Potential Distribution with x Periodicity.

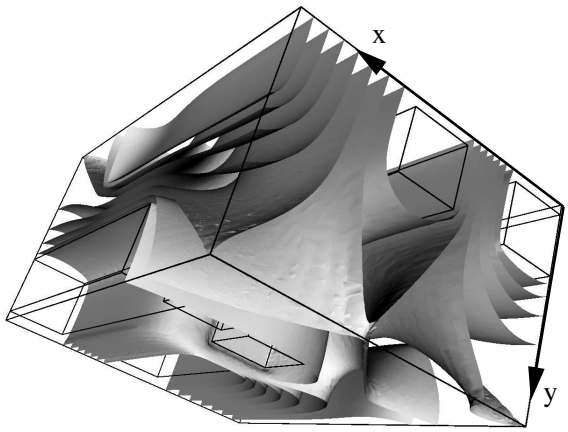

Figure 17: The Iso Faces of the Potential Distribution with x Periodicity.

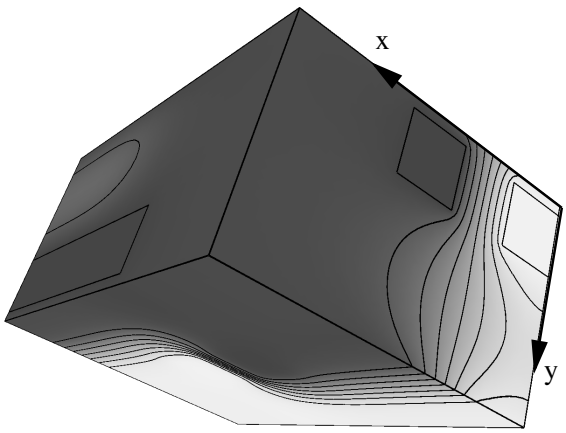

Figure 18: The Potential Distribution with y Periodicity.

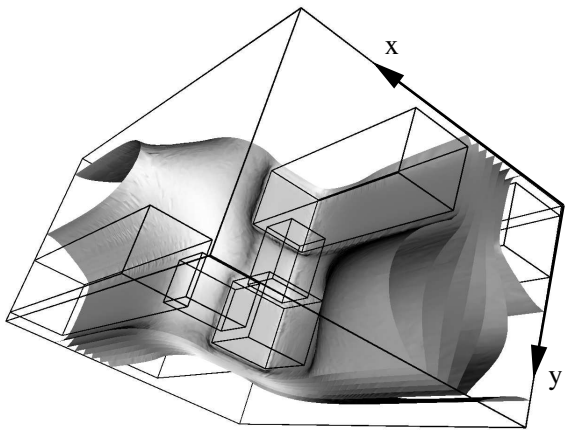

Figure 19: The Iso Faces of the Potential Distribution with y Periodicity.

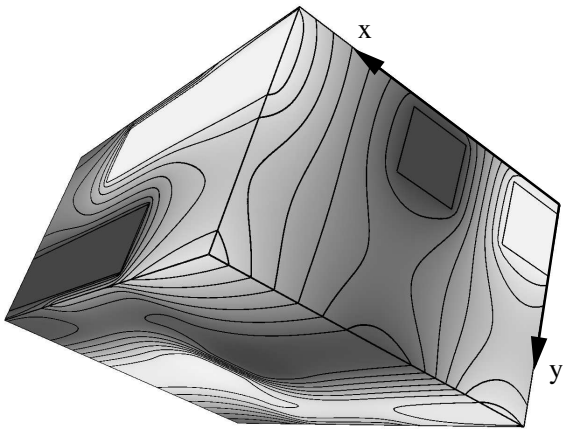

Figure 20: The Potential Distribution with xy Periodicity.

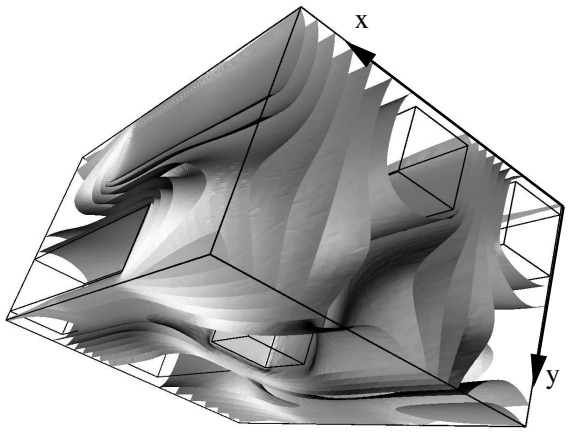

Figure 21: The Iso Faces of the Potential Distribution with xy Periodicity.

4 Three-Dimensional Analysis of Leakage Current in Non-Planar Oxides

We demonstrate the applicability of fully three-dimensional device simulation with the investigation of tunneling currents through oxides and show its benefit for the understanding of physical phenomena especially in the nanometre regime. We compare leakage current measurements from three oxides with different thicknesses (7 nm, 15 nm, and 50 nm), measured by an atomic force microscope (AFM), with simulated Fowler-Nordheim (FN) current distributions and show the necessity of including surface roughness as an essential part of three-dimensional simulation.

4.1 Introduction

During the investigation of gate leakage measurements of oxides with different thicknesses, as shown in Figure 22, it became apparent that taking only the flatband voltage of each of the measured devices into account is insufficient for understanding the measured data. This is shown in Figure 22 where the tunneling currents do not overlap with the theoretical Fowler-Nordheim (FN) curve.

Although the regions indicated in the figure exhibit the characteristics of FN tunneling the curves should overlap for this tunneling mechanism. It was suspected that three-dimensional effects due to surface roughness are at least partially responsible for the observed discrepancy. This spawned interest on how to describe these three-dimensional effects.

To investigate the influence of surface roughness on the electrical characteristics of oxides, height data sets obtained from AFM measurements were used as input parameters for modeling. The measured samples provided by *austriamicrosystems* corresponded to the ones subject to the leakage current investigation and were measured at the Institute for Solid State Electronics at the Technical University of Vienna.

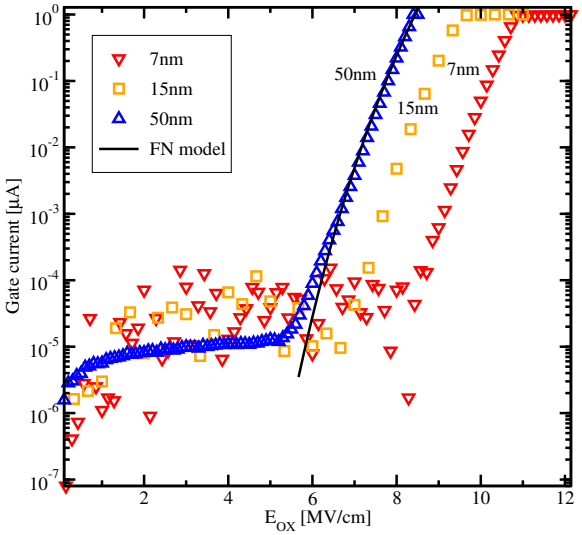

Figure 22: Comparison of the measured oxide tunneling currents. The measurements were performed at *austriamicrosystems*.

4.2 AFM Measurements

To overcome the difficulties with raw data sets from AFM measurement a pre-processing module (AFMStructureBuilder: ASBuilder) has been developed to correct the raw data sets and to perform the three-dimensional meshing step and contact building.

4.2.1 Raw Data Sets

The raw data set from AFMs measurement are used although some post-processing steps could be done within the measurement software. This is because these steps must be done very accurately and in correlation with the following device simulation steps for a detailed investigation. ASBuilder was developed with these considerations in mind. Figure 23 shows an output of the measurement software.

To enhance the raw data set ASBuilder can filter the data set with different options designed to compensate for different effects encountered during the measurement [41].

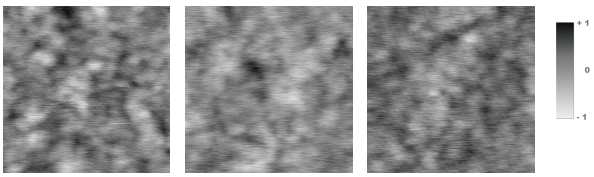

Figure 23: Two-dimensional height distribution of the AFM measurement (left: 7 nm, middle: 15 nm, right: 50 nm).

- Piezo drift
  Due to heating of the AFM tip during the measurement period the piezo crystal drift results in an z-offset of the measured data set. Within ASBuilder this piezo drift can be recomputed and compensated.
- Fast-scan-line noise
  To reduce fast-scan-line noise a discrete Fourier filter is used to suppress this kind of noise.
- Spike filtering
  To filter noise spikes gauss filters with different kernels can be applied within ASBuilder.

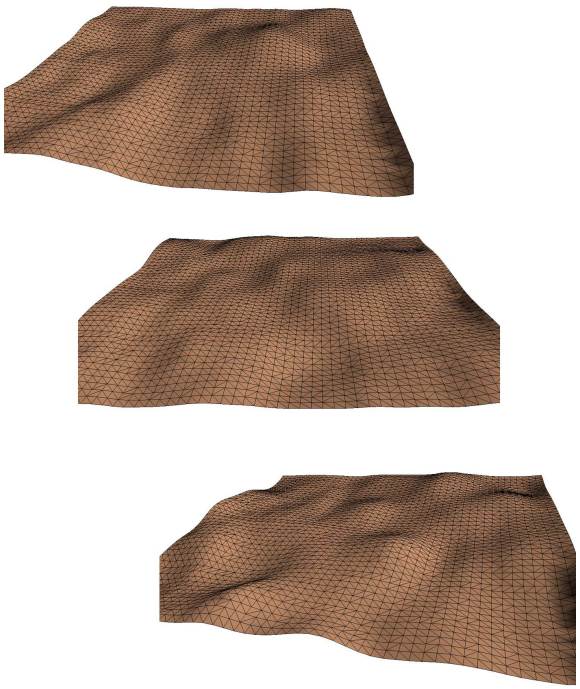

Figure 24: A detailed view of the height distribution (top: 7 nm, middle: 15 nm, bottom: 50 nm).

4.3 Further Processing of Data Sets

For further processing ASBuilder creates a surface triangulation of the corrected and adjusted height distribution. In order to accomplish this ASBuilder reads in the height distribution data set, corrects the data set and assembles a two-dimensional height distribution matrix. From this matrix an unstructured two-dimensional mesh is generated where the data set is meshed (ASBuilder) with a so called height-map meshing step where the height distribution is triangulated and elevated into three dimensions. The result can be seen in Figure 24.

The histograms of the height distributions presented in Figure 25 show the characteristics of the three different oxides. Here the 7 nm oxide has the flattest distribution which means that the surface roughness is equally distributed between the complete range of 6 nm to 8 nm. The presumably high fluctuations of the 15 and 50 nm oxide are insignificant when compared to the overall thickness of the oxide, while the same does not hold for the 7 nm oxide.

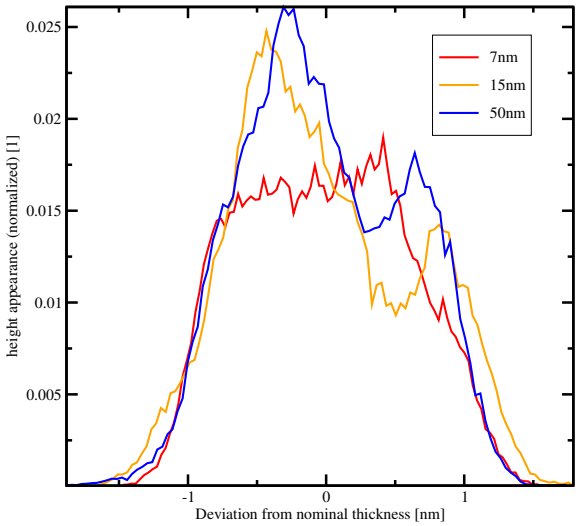

Figure 25: Histograms of the height distributions.

4.3.1 Building the Three-Dimensional Simulation Structures

To investigate oxide reliability in detail the prepared and triangulated surfaces of the

oxides are meshed by ASBuilder [42] into a three-dimensional object with a bottom and top metallic contact.

With this pre-processing steps of ASBuilder a completely three-dimensional object with the non-planar oxide element and two planar contacts is created which can be used as input data for any existing device simulation software such as Minimos-NT [31] to calculate the electric field distribution required for the modeling of FN tunneling.

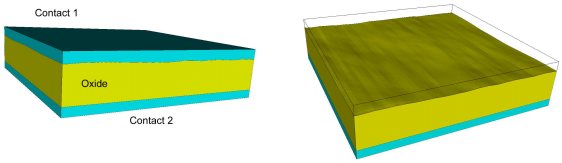

Figure 26: Three-dimensional oxide structure.

4.4 Simulation Methodology

After the extensive preparations outlined above several simulation steps need to be performed. First, the electric field distribution is calculated. The results of the three-dimensional electric field calculation are shown in Figure 27 and 28. Figure 27 depicts the absolute values of the electric field, while Figure 28 shows a cut through the three-dimensional simulation domain. Both figures illustrate the influence of the encountered surface roughness on the electric field. The field clearly shows peaks in the regions of thinner oxide inducing heightened electrical stress in these regions.

Due to the thickness of the oxides and the strong electric fields the leakage current is modeled as FN tunneling current. It is evaluated using the previously determined electric field distribution. The FN tunneling current is modeled by the well known expression [43, 44]

$$J = a\,|E|^2 \exp\left(-\frac{b}{|E|}\right). \qquad (30)$$

The parameters $a = 994.63 \times 10^{-9}$ A/V^2 and $b = 2.64 \times 10^{10}$ V/m were calibrated for the non-planar case of each oxide thickness and then used in the subsequent simulations.

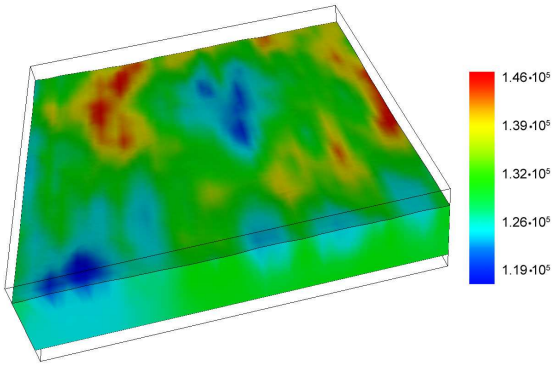

Figure 27: Results of a three-dimensional simulation of the electric field distribution for a $50\times50\,\mathrm{nm}^2$ region of a 7 nm oxide. The values of the electric field are in V/cm.

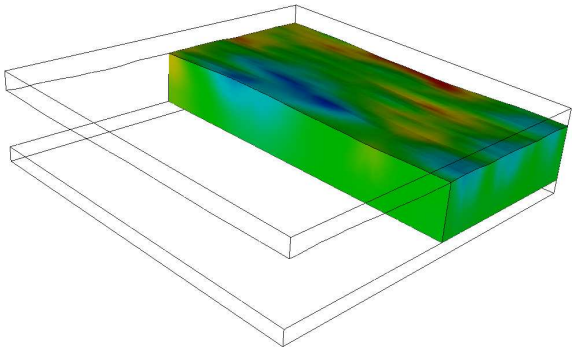

Figure 28: Cut through the simulation domain of a $50\times50\,\mathrm{nm}^2$ region of a 7 nm oxide.

4.5 Comparison of the Measured Leakage Currents

The area below the 50 nm oxide is ohmically connected to the bulk of the wafer, while the areas corresponding to the thinner oxides are insulated by pn-junctions. The measured structures are schematically presented in Figure 29 for the 50 nm oxide and in Figure 30 for the 7 and 15 nm oxides. This explains the differing noise levels visible in the measurement data (Figure 22) as the measurements of the 7 and 15 nm oxides also include noise from this junction.

The measured data includes several effects that complicate the analysis. The flatband voltage is one such interference within this measurement. The pn-junctions included in the 7 and 15 nm

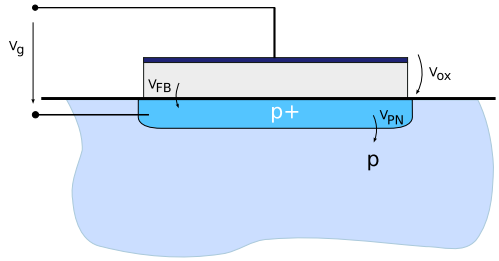

Figure 29: Structure of measurement arrangement (50 nm).

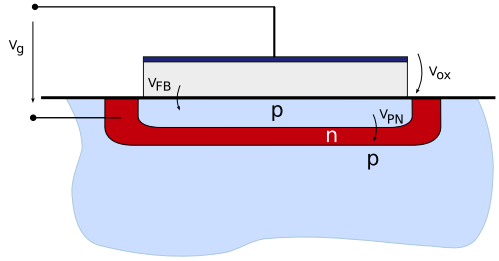

Figure 30: Structure of measurement arrangement (7 nm and 15 nm).

structures contribute another parasitic effect within the measurement.

Both of these effects need to be taken into account in order to enable a correct modeling of the leakage current. Figure 31 shows the influence of the flatband voltage. While the influence on the 50 nm oxide is marginal, there is a significant impact for the thinner oxides. The compensation of the influence of the pn-junctions on the thinner oxides is shown in Figure 32. Again the influence is larger for smaller oxide thicknesses.

$$E'_{\mathrm{ox}} = \frac{V - V_{\mathrm{pn}}}{t_{\mathrm{ox}}} \tag{31}$$

After amending for these effects the measurement curves almost overlap as can be observed in Figure 34. This is an indication of a common mechanism of the leakage current which is readily found in FN tunneling.

4.6 Simulation Results

After considering the flatband voltage, the pn-junction voltage and the previously determined correction voltages the regions depicted in Figure 22 overlap, as can be seen in Figure 34, and can then be simulated with the FN tunneling model. The result obtained from this simulation is also depicted in Figure 34.

The agreement between the measured leakage current and the simulation result is excellent. Using the parameters obtained from the non-planar case, a simulation with planar surfaces is performed as well. This is done by calculating the average height of the oxide from the distribution and assuming a parallel plate capacitor. This corresponds to an effective thickness extracted from CV measurements. The results of this computation is shown in Figure 33. As expected the non-planar curves overlap. The discrepancy between the planar and non-planar case increases with decreasing oxide thickness.

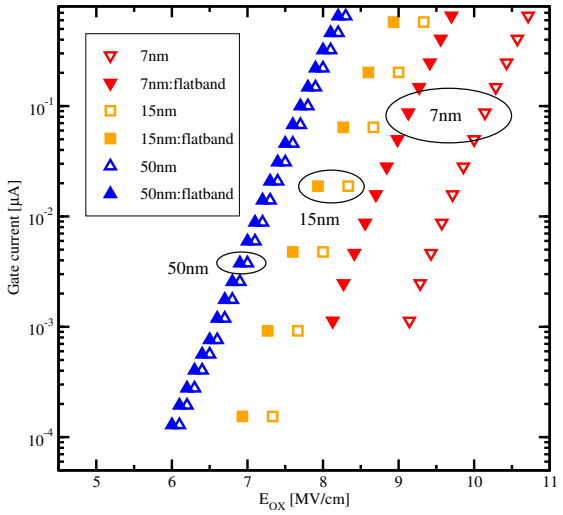

Figure 31: Comparison of the original data set and the corrected set obtained by inclusion of the flatband voltage.

This indicates that the relative roughness is responsible for this deviation which, as already stated above, increases as the oxide thickness is reduced. From this it is evident that non-planar effects are increasingly important as oxide thicknesses shrink. From the comparison of the fully three-dimensional and the planar simulations correction voltages can be derived.

The observed tunneling current is not only important for the overall power consumption of devices but also for the reliability of the devices [43], as the tunneling charge carriers are responsible for damaging the oxide and deteriorating the performance of the device.

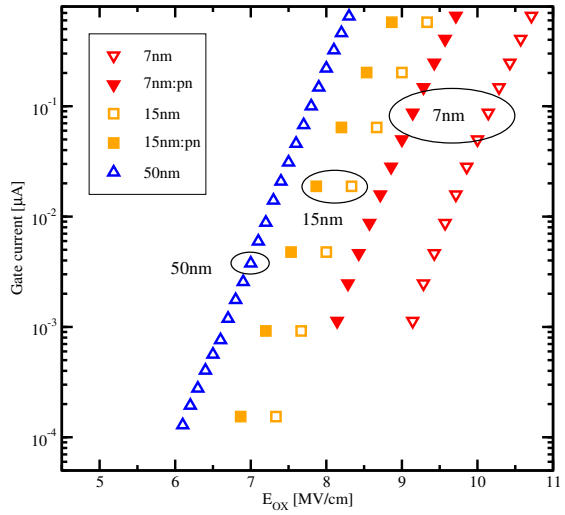

Figure 32: Comparison of the original data set and the data corrected by the pn-junction.

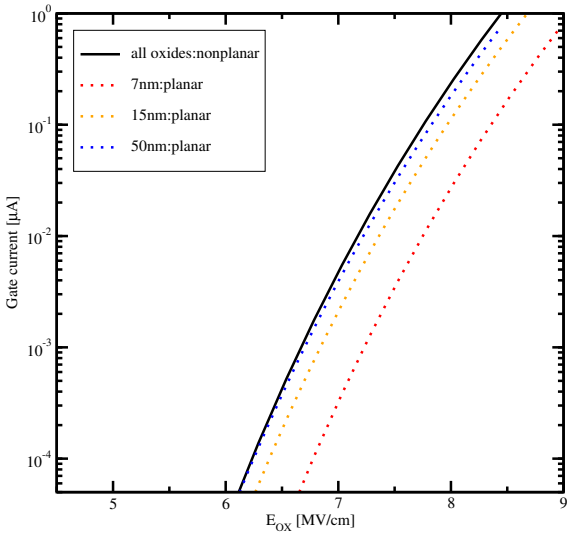

Figure 33: Comparison of the influence of three-dimensional surface roughness effects.

4.7 Conclusion

Due to the growing complexity of the structures of modern semiconductor devices and the ongoing shrinking to smaller dimensions, device simulations in two dimensions are no longer sufficient because of dominant three-dimensional effects. This is especially true for oxide properties due to the reduction of oxide thickness to only a few atomic layers.

In particular we have shown that by considering only the effective oxide thickness obtained for

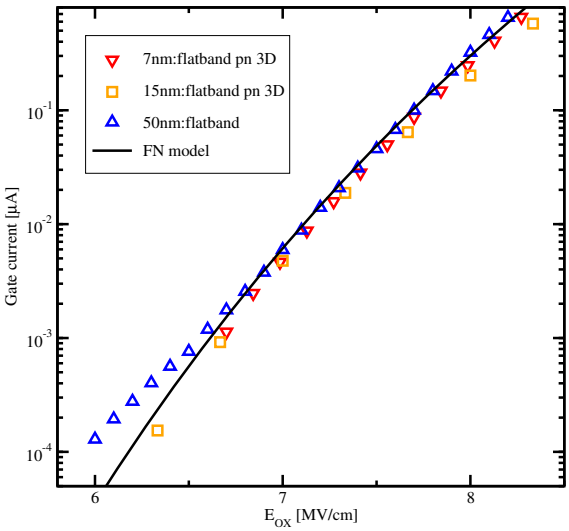

Figure 34: Final simulation compared to corrected measurement data sets.

instance from CV measurements the estimated FN currents are significantly underestimated due to the non-planarity of the oxide. This effect increases for decreasing oxide thicknesses and has to be considered for oxide reliability considerations.

5 The Effect of Degeneracy on Electron Transport in Strained Silicon Inversion Layers

The effect of degeneracy both on the phonon-limited mobility and the effective mobility including surface-roughness scattering in unstrained and biaxially tensile strained Si inversion layers is analyzed. We introduce a new method for the inclusion of the Pauli principle in a Monte Carlo algorithm. We show that incidentally degeneracy has a minor effect on the bulk effective mobility, despite non-degenerate statistics yields unphysical subband populations and an underestimation of the mean electron energy. The effective mobility of strained inversion layers slightly increases at high inversion layer concentrations when taking into account degenerate statistics.

5.1 Introduction

Strained Si has emerged as a promising material for the improvement of Si MOS technology because of its enhanced carrier mobility [45] [46] [47]. Strain can be induced by epitaxially growing thin Si layers on relaxed $Si_{1-y}Ge_y$ substrates, or alternatively, by processing additional cap layers over the transistors. The latter method is especially suitable for mass production because it requires only a slight modification of the process flow [48].

Surprisingly, from a theoretical viewpoint the mobility enhancement caused by strain is still an issue of discussion. The reason for this is manifold: It was claimed that using the well established models for scattering in the two dimensional electron gas (2DEG) the mobility gain of strained Si (SSi) at low effective fields should be compensated by more pronounced surface roughness scattering at large effective fields [49].

Due to this fact, only with the assumption of much smoother strained Si-SiO_2 interfaces one should be able to get qualitative agreement with experimental data. Even though it seemed to be unphysical to change the smoothness of the strained Si-SiO_2 interface, the Monte Carlo (MC) community has adopted this assumption [50] or simply not responded to the troubling fact.

When trying to clarify this dissatisfactory status two main difficulties arise: First, there exists a variety of surface roughness scattering models, and it is not clear which approximations to the general expression given by Ando [51] are allowed. Second, there is a discordance whether and how degeneracy effects should be included in transport calculations of inversion layers.

In this paper, the ways to include the Pauli principle in a MC algorithm are revised and critically compared to each other. The usual method, where the Pauli blocking factor $1-f(\mathbf{k})$ is approximated using the equilibrium distribution function $f_{\mathrm{FD}}(\mathbf{k})$, can be shown to lead to unphysical subband populations, kinetic energies, and mobilities. The reason being that at high degeneracy the error $\varepsilon(\mathbf{k}) = f(\mathbf{k}) - f_{\mathrm{FD}}(\mathbf{k})$ is dominant. A new MC algorithm accounting for the Pauli exclusion principle is proposed which is less sensitive to the error $\varepsilon(\mathbf{k})$.

The paper is organized as follows: Section II describes the new approach to implement the Pauli exclusion principle in the MC method. It is shown that in the low field limit the proposed algorithm yields the same mobility as the Kubo Greenwood formula, while other algorithms do not. We use the new method to extract velocity profiles and illustrate the large effect of degeneracy on the electron system. Finally, the simulated effective mobility curves for unstrained and biaxially tensile strained Si on relaxed $Si_{1-y}Ge_y$ substrates are presented in Section III.

5.2 Inclusion of the Pauli Principle in Monte Carlo Simulations

In transport calculations of the 2DEG forming in the channel of MOSFETs the inclusion of the Pauli principle is expected to be important since the lowest subband may lie well below the Fermi level in the regime of moderate and high effective fields (high inversion layer concentrations). This leads

to modified subband populations and an elevated mean kinetic energy of electrons as compared to the nondegenerate case. A change in the mobility is therefore to be expected.

The proposed algorithm is based on the following reformulation of the degenerate scattering operator.

$$\begin{aligned} Q[f]_{\mathbf{k}} = & \\ = \int f_{\mathbf{k}'}(1-f_{\mathbf{k}})S_{\mathbf{k}',\mathbf{k}}\,d\mathbf{k}' - \int f_{\mathbf{k}}(1-f_{\mathbf{k}'})S_{\mathbf{k},\mathbf{k}'}\,d\mathbf{k}' & \\ = \int f_{\mathbf{k}'}S_{\mathbf{k}',\mathbf{k}} - f_{\mathbf{k}}S_{\mathbf{k},\mathbf{k}'} + \underbrace{f_{\mathbf{k}}f_{\mathbf{k}'}(S_{\mathbf{k},\mathbf{k}'} - S_{\mathbf{k}',\mathbf{k}})}_{\text{additional term}}\,d\mathbf{k}'. & \end{aligned} \tag{32}$$

The last term represents a nonlinear correction to the non-degenerate scattering operator. To linearize the scattering operator it is common to keep one factor of the product $f_{\mathbf{k}}f_{\mathbf{k}'}$ constant and to treat the other as the unknown.

Near thermodynamic equilibrium, f can be approximated by the Fermi-Dirac distribution function f_{FD}. The key point of the new method is that a symmetric approximation with respect to $\mathbf{k}$ and $\mathbf{k}'$ is employed.

$$f(\mathbf{k})f(\mathbf{k}') \approx \frac{1}{2}\big(f(\mathbf{k})f_{FD}(\mathbf{k}') + f_{FD}(\mathbf{k})f(\mathbf{k}')\big) \tag{33}$$

Using this approximation the scattering operator can be expressed in terms of a modified transition rate $\widehat{S}_{\mathbf{k},\mathbf{k}'}$ and scattering rate $\widehat{\lambda}_{\mathbf{k}}$ as

$$Q[f]_{\mathbf{k}} = \int f(\mathbf{k}')\widehat{S}_{\mathbf{k}',\mathbf{k}}\,d\mathbf{k}' - f(\mathbf{k})\widehat{\lambda}_{\mathbf{k}}, \tag{34}$$

with

$$\begin{aligned} \widehat{S}_{\mathbf{k},\mathbf{k}'} &= S_{\mathbf{k},\mathbf{k}'}\left(1-\frac{1}{2}f_{FD}(\mathbf{k}')\right) + S_{\mathbf{k}',\mathbf{k}}\frac{1}{2}f_{FD}(\mathbf{k}') \\ \widehat{\lambda}_{\mathbf{k}} &= \int \widehat{S}_{\mathbf{k},\mathbf{k}'}\,d\mathbf{k}'. \end{aligned} \tag{35}$$

A simple error analysis shows the advantage of this formulation. Consider a highly degenerate state $\mathbf{k}$, characterized by $f(\mathbf{k}) \approx 1$. A direct approximation of the blocking factor $(1-f(\mathbf{k}))$ can give completely wrong results, because the approximation of the blocking factor is determined by the error, $1-(f_{FD}+\varepsilon) \approx \varepsilon$. In the formulation e:mod$_t$*ransrate*, *however*, *becauseof*$\varepsilon \ll 1$ the effect of the error will be negligible, $1-(f_{FD}+\varepsilon)/2 \approx 1/2$.

The modified transition rate e:mod$_t$*ransrateisgivenbyalinearcomb* and backward rate $S_{\mathbf{k}',\mathbf{k}}$. The latter can be expressed in terms of the forward rate by means of the principle of detailed balance. The modified scattering rates for phonon emission and absorption become,

$$\begin{aligned} \widehat{\lambda}_{\text{em}} &= \lambda_{\text{em}} \cdot \left(1 - \frac{1}{2}\frac{f_{\text{FD}}(\varepsilon_f)}{N_0+1}\right) \\ \widehat{\lambda}_{\text{ab}} &= \lambda_{\text{ab}} \cdot \left(1 + \frac{1}{2}\frac{f_{FD}(\varepsilon_f)}{N_0}\right), \end{aligned} \tag{36}$$

where ε_f denotes the final energy and N_0 the equilibrium phonon distribution function,

$$N_0 = \frac{1}{\exp\left(\frac{\hbar\omega_0}{k_B T}\right) - 1}. \tag{37}$$

For elastic scattering mechanisms the modified scattering rates do not change from the classical ones, $\widehat{\lambda}_{\mathbf{k}} = \lambda_{\mathbf{k}}$. In the simulation of the 2DEG to a good approximation one can assume scattering with surface roughness, impurities, and acoustical phonons to be elastic.

To implemenent the Pauli principle in a conventional MC program for non-degenerate statistics the only modifications necessary are the replacement of the classical scattering rates by the modified ones.

5.2.1 Comparison of Algorithms

The new algorithm has been compared to two other methods found in the literature [52, 53]. The first algorithm to include the Pauli principle in the MC technique [52] is based on a self-consistent iterative algorithm that uses a rejection technique to account for the occupation probability of the final state at each scattering event. Since this auxiliary self-scattering mechanism is proportional to the occupation of the final states, the algorithm prevents a large number of classically allowed transitions.

A different approach to include degeneracy in MC simulations was given in [53]. Inelastic scattering rates are multiplied with a factor of $(1-f_{\text{FD}}(\varepsilon_f))/(1-f_{\text{FD}}(\varepsilon_i))$, where ε_i (ε_f) denotes the initial (final) electron energy. This additional

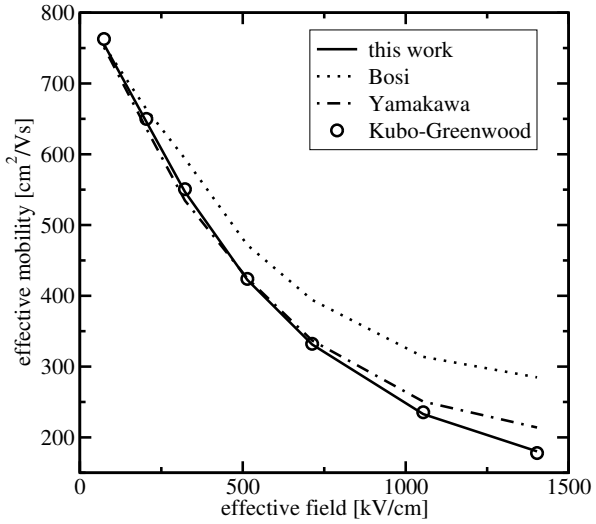

Figure 35: The simulated effective mobility using the new algorithm (solid line) is compared to results of a non-selfconsistent version of Bosi's algorithm [52] (dotted line), the algorithm proposed by [53], and to the mobility calculated with the Kubo-Greenwood formalism (open circles).

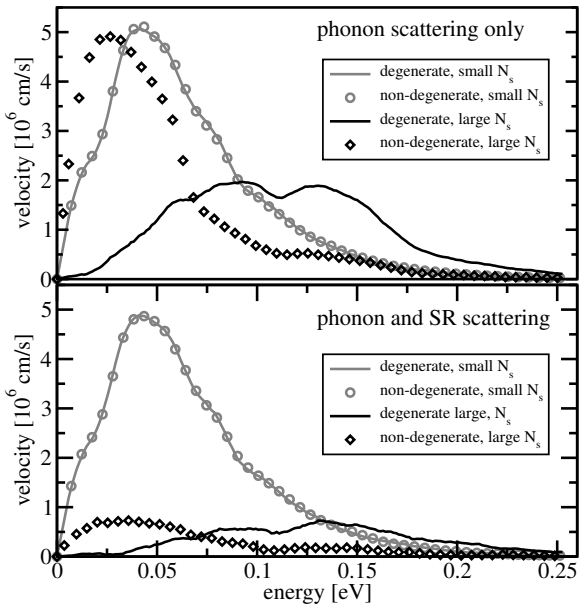

Figure 36: Mean velocities resulting from simulations with (lines) and without (symbols) inclusion of the Pauli principle for an inversion layer density $N_s \approx 10^{11}$ cm^{-2} (grey) and $N_s \approx 1.5^{13}$ cm^{-2} (black).

factor stems from the use of Fermi-Dirac statistics within the relaxation time approximation [54].

In the limit of vanishing field the mobility can also be calculated using the relaxation time approximation. Thus we compare our MC algorithm and a modified method from [52] and [53] to the mobility calculated from the Kubo-Greenwood expression [55]. From Figure 35 it can be seen that the new method yields the closest agreement, whereas a non-selfconsistent implementation of the algorithm proposed in [52], where $f(\mathbf{k})$ has been approximated by the equilibrium distribution function $f_{\mathrm{FD}}(\mathbf{k})$, and the algorithm proposed by [53] overestimate the effective mobility.

5.3 Results

We calculate the mean velocities as a function of total energy in the small-field limit. Note that the overall mobility will be proportional to the integral of this function.

A very interesting behavior can be observed when comparing the mean velocities resulting from simulations with classical and Fermi-Dirac statistics. The upper plot of Figure 36 shows that when considering only phonon scattering the mean velocities (grey open symbols) coincide for both simulation modes in the non-degenerate regime ($E_f - E_0 \approx -0.13$ eV). At high inversion layer concentrations, where the 2DEG is highly degenerate ($E_f - E_0 \approx 0.8$ eV), a shift of the mean velocity distribution (black closed circles) toward higher energies and a decrease of its peak can be observed as compared to the non-degenerate mean velocity (black closed diamonds).

The coincidence of the mean velocities in the non-degenerate regime is merely a test that the degenerate-algorithm converges to the classical algorithm for the non-degenerate 2DEG. At high inversion layer concentrations the different mean velocities can be interpreted as follows: In non-degenerate simulations electrons will have an equilibrium energy of $\mathrm{k_B}T$ whereas the mean energy in the degenerate case can be more than twice as much. Since phonon scattering is merely proportional to the density of states, which is an increasing step-like function in the

2DEG, electrons being at higher energies – as it is the case in degenerate simulations – experience more scattering and thus the phonon-limited mobility is strongly decreased when including the Pauli exclusion principle in the simulation (see Figure 37).

The lower plot of Figure 36 shows the mean velocities when surface roughness scattering has been included in simulations. In the non-degenerate regime (grey symbols) surface roughness scattering does not play an important role, and the mean velocities compare well with the simulation results with only phonon scattering included. Interestingly now even in the degenerate 2DEG the mean velocities (black closed symbols) do not differ as much and the large peak in the mean velocity at low energies that could be observed before when the Pauli principle was not included is here suppressed. This stems from the influence of surface roughness scattering which is more effective at small energies in contrast to phonon scattering.

Finally from Figure 36 one can also observe that due to degeneracy electrons at energies below the Fermi level have smaller velocities which corresponds to the general picture that these electrons have little contribution to transport.

The new algorithm is used to extract the effective mobility in unstrained and biaxially tensile strained Si inversion layers. For the simulations a one-dimensional Schrödinger-Poisson solver [56] is used with modifications to account for the energy splitting between the twofold and the fourfold conduction-band minima and for the change of the band-gap [49]. From these results the matrix elements for phonon and surface roughness scattering and the form factors are calculated following [57].

Screening of the surface roughness scattering was included according to [57] while impurity scattering has been ignored as we mainly focus on the high-density (high effective field) region. A non-parabolic bandstructure ($\alpha = 0.5$ eV^{-1}) was used leading to a 20% reduction of the phonon-limited mobility at 300 K in agreement with [49]. For all simulations presented here a uniform doping concentration of 2×10^{16} cm^{-3} has been assumed.

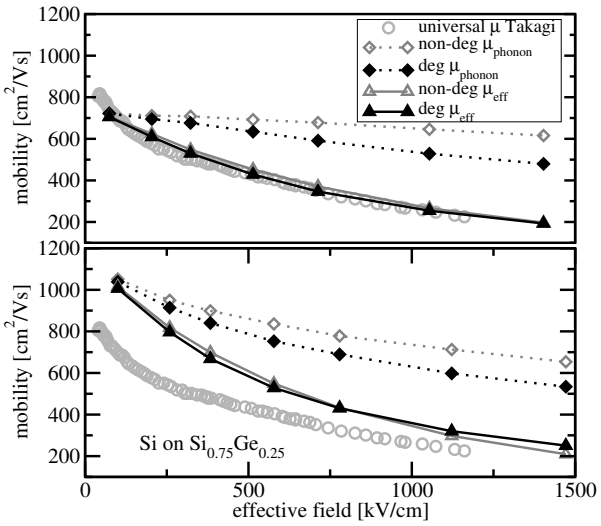

Figure 37: Calculated phonon-limited and effective mobility compared to the universal mobility curve with and without degeneracy effects for unstrained Si (upper curve) and biaxially strained Si (lower curve).

The simulated mobility curve for unstrained Si in the upper plot of Figure 37 shows good agreement with the universal mobility curve by Takagi [58]. Surprisingly, the effective mobility resulting from simulations with degenerate statistics are in close agreement to those using classical statistics even though the phonon-limited mobility experiences a noticeable reduction when using degenerate statistics.

As previously discussed, this close agreement can only be understood from the cancellation of two effects: Degeneracy leads to an increase of the mean kinetic energy. This leads to an increase in phonon scattering and a decrease in the mobility. At the same time electrons with larger kinetic energies experience less effective surface-roughness scattering, thus the surface roughness limited mobility is increased. In unstrained Si by chance these two effects cancel each other at all effective fields, and the difference between a simulation with non-degenerate and degenerate statistics is very small.

Having calibrated our model against the unstrained universal mobility curve, a simulation of biaxially strained Si on relaxed $Si_{1-y}Ge_y$

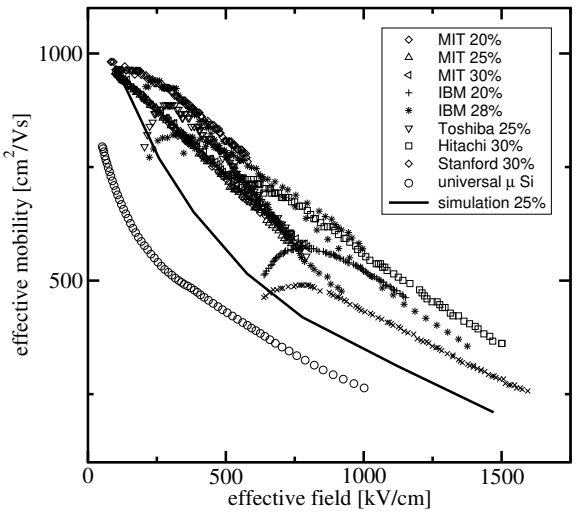

Figure 38: Comparison of simulations results for the effective mobility of biaxially strained Si on $Si_{1-y}Ge_y$ with $y = 0.25$ with published data from different experimental groups for various substrate compositions of Ge.

substrates with $y = 0.25$ was performed. From Figure 37 it can be observed that in SSi inversion layers, where the ratio between phonon and surface roughness scattering is different – due to suppressed intervalley transitions – simulations with degeneracy effects yield higher mobilities $\mu_{\text{sr,deg}} > \mu_{\text{sr,nondeg}}$.

Our simulation results for the inversion layer mobility in biaxially strained Si on $Si_{1-y}Ge_y$ for various Ge contents suggest a saturation of the mobility enhancement at $y \approx 25\%$. As can be seen from Figure 37 the anomalous intersection of the strained and unstrained mobility curve from [49] was not observed, however Figure 38 indicates that the predicted mobility for SSi is still underestimated.

5.4 Conclusion

By means of MC simulations we are able to deduce the effect of degeneracy both on the phonon-limited mobility and the effective mobility including surface-roughness scattering. It is shown that in the unstrained case the inclusion of the Pauli principle leads to a noticeable reduction of the phonon-limited mobility, but has almost no impact on the effective mobility. The effective mobility of strained inversion layers increases slightly at high inversion layer concentrations when taking into account degenerate statistics. Thus a correct treatment of degenerate carrier statistics of the 2DEG of strained Si inversion layers is important.

However, this study cannot explain the experimental mobility enhancement for SSi, which is still underestimated at large effective fields, where surface roughness scattering dominates. Thus a careful revision of surface roughness scattering might be needed to achieve the correct mobility enhancements.

References

[1] M. Yan, K. N. Tu, A. Vairagar, S. G. Mhaisalkar, and A. Krishnamoorthy. The Effect of Immersion Sn Coating on the Electromigration Failure Mechanism and Lifetimes of Cu Dual Damascene Interconnects. *Mater. Res. Soc. Symp. Proc.*, 863:B9.9.1–B9.9.6, 2005.

[2] H. Ceric, V. Deshpande, Ch. Hollauer, S. Holzer, T. Grasser, and S. Selberherr. Comprehensive Analysis of Vacancy Dynamics Due to Electromigration. In *Proc. of the 12th International Symposium on the Physical and Failure Analysis of Integrated Circuits*, pp 100–103, 2005.

[3] H. Ceric and S. Selberherr. An Adaptive Grid Approach for the Simulation of Electromigration Induced Void Migration. In *Proc. SISPAD*, pp 253–257, 2002.

[4] E. Zschech, M. A. Meyer, S. G. Mhaisalkar, and A. V. Vairagar. Effect of Interface Modification on Em-Induced Degradation Mechanisms in Copper Interconnects. In *Proc. of ICMAT*, 2005. in press.

[5] P. H. Dederichs and K. Schroeder. Anisotropic Diffusion in Stress Fields. *Physical Review B*, 17(6):2524–2536, 1978.

[6] W. W. Mullins. Mass Transport at Interfaces in Single Component Systems. *Metall. Mater. Trans. A (USA)*, 26(8):1917–1929, 1995.

[7] M. A. Korhonen, P. Borgesen, K. N. Tu, and C. Y. Li. Stress Evolution due to Electromigration in Confined Metal Lines. *J.Appl.Phys.*, 73(8):3790–3799, 1993.

[8] R. Kirchheim. Stress and Electromigration in Al-Lines of Integrated Circuits. *Acta Metallurg. Mater.*, 40(2):309–323, 1992.

[9] F. C. Larche and J. W. Cahn. The Interactions of Composition and Stress in Cystalline Solids. *Acta Metall.*, 33(3):331–357, 1985.

[10] D. N. Bhate, A. F. Bower, and A. Kumar. A Phase Field Model for Failure in Interconnect Lines Due to Coupled Diffusion Mechanisms. *J. Mech. Phys. Solids*, 50:2057–2083, 2002.

[11] J. Betten. *Kontinuumsmechanik*. Springer, Berlin, 2001.

[12] M. E. Sarychev and Yu. V. Zhitnikov. General Model for Mechanical Stress Evolution During Electromigration. *J.Appl.Phys.*, 86(6):3068 – 3075, 1999.

[13] C. P. Flynn. *Point Defects and Diffusion*. Clarendon Press, Oxford, 1972.

[14] M. R. Sorensen, Y. Mishin, and A. F. Voter. Diffusion mechanisms in Cu grain boundaries. *Physical Review B*, 62(6):3658–3673, 2000.

[15] J. J. Clement. Electromigration Modeling for Integrated Circuit Interconnect Reliability Analysis. *IEEE Trans. on Dev. and Mat. Rel.*, 1(1):33–42, 2001.

[16] R. Rosenberg and M. Ohring. Void Formation and Growth During Electromigration in Thin Films. *J.Appl.Phys.*, 42(13):5671–5679, 1971.

[17] V. Sukharev. Simulation of Microstructure Influence on EM-Induced Degradation in Cu Interconnects. *Stress-Induced Phenomena in Metallization: 8th International Workshop, AIP Conference Proceedings*, 2005. in press.

[18] E. Zschech and V. Sukharev. Microstructure Effect on EM-Induced Copper Interconnect Degradation. *Microelectronic Engineering*, 82:629–638, 2005.

[19] E. M. Vogel, K. Z. Ahmed, B. Hornung, K. Henson, P. K. McLarty, G. Lucovsky, J. R. Hauser, and J. J. Wortman. Modeled Tunnel Currents for High Dielectric Constant Dielectrics. *IEEE Trans.Electron Devices*, 45(6):1350–1355, 1998.

[20] J. Zhang, J. S. Yuan, Y. Ma, and A. S. Oates. Design Optimization of Stacked Layer Dielectrics for Minimum Gate Leakage Currents. *Solid-State Electron.*, 44(12):2165–2170, 2000.

[21] H. S. Wong. Beyond the Conventional Transistor. *IBM J.Res.Dev.*, 46(2/3):133–168, 2002.

[22] R. Tsu and L. Esaki. Tunneling in a Finite Superlattice. *Appl.Phys.Lett.*, 22(11):562–564, 1973.

[23] B. Khairurrijal, W. Mizubayashi, S. Miyazaki, and M. Hirose. Analytic Model of Direct Tunnel Current Through Ultrathin Gate Oxides. *J.Appl.Phys.*, 87(6):3000–3005, 2000.

[24] A. Gehring and S. Selberherr. On the Calculation of Quasi-Bound States and Their Impact on Direct Tunneling in CMOS Devices. In *Proc. SISPAD*, pp 25–28, 2004.

[25] E. Cassan. On the Reduction of Direct Tunneling Leakage through Ultrathin Gate Oxides by a One-Dimensional Schrdinger-Poisson Solver. *J.Appl.Phys.*, 87(11):7931–7939, 2000.

[26] A. Dalla Serra, A. Abramo, P. Palestri, L. Selmi, and F. Widdershoven. Closed- and Open-Boundary Models for Gate-Current Calculation in n-MOSFETs. *IEEE Trans.Electron Devices*, 48(8):1811–1815, 2001.

[27] C. L. Fernando and W. R. Frensley. An Efficient Method for the Numerical Evaluation of Resonant States. *J.Appl.Phys.*, 76(5):2881–2886, 1994.

[28] R. Clerc, A. Spinelli, G. Ghibaudo, and G. Pananakakis. Theory of Direct Tunneling Current in Metal-Oxide-Semiconductor Structures. *J.Appl.Phys.*, 91(3):1400–1409, 2002.

[29] S. Odermatt AND M. Luisier AND B. Witzigmann. Bandstructure Calculation using the k-p-Method for Arbitrary Potentials with Open Boundary Conditions. *J.Appl.Phys.*, 97(4):046101–046103, 2005.

[30] M. Karner, A. Gehring, H. Kosina, and S. Selberherr. Ecient Calculation of Quasi-Bound State Tunneling in CMOS Devices. In *Proc. SISPAD*, pp 35–38, 2004.

[31] Institut für Mikroelektronik, Technische Universität Wien, Austria. *MINIMOS-NT 2.1 User's Guide*, 2005. http://www.iue.tuwien.ac.at/software.

[32] R. Sabelka and S. Selberherr. A Finite Element Simulator for Three-Dimensional Analysis of Interconnect Structures. *Microelectronics Journal*, 32:163–171, 2001.

[33] J. L. Volakis, A. Chatterjee, and L. C. Kempel. *Finite Element Method for Electromagnetics*. IEEE Press, New York, 1998.

[34] R. Bauer and S. Selberherr. Preconditioned CG-Solvers and Finite Element Grid. In *Proc. Colorado Conference on Iterative Methods (CCIM)*, Breckenridge, USA, 1994.

[35] I. Koren, Z. Koren, and D. Pradhan. Designing Interconnection Buses in VLSI and WSI for Maximum Yield and Minimum Delay. *IEEE Journal of Solid State Circuits*, 23(3):859–866, 1988.

[36] N. Thepayasuwan and A. Doboli. Layout Conscious Bus Architecture Synthesis for Deep Submicron Systems on Chip. In *Proc. of the Design, Automation and Test in Europe Conference and Exhibition (DATE 04)*. IEEE Computer Society, 2004.

[37] P. Fleischmann, R. Sabelka, A. Stach, R. Strasser, and S. Selberherr. Grid Generation for Three-dimensional Process and Device Simulation. In *Proc. SISPAD*, pp 161–166, 1996.

[38] P. Fleischmann and S. Selberherr. A New Approach to Fully Unstructured Three-dimensional Delaunay Mesh Generation with Improved Element Quality. In *Proc. SISPAD*, pp 129–130, 1996.

[39] J. R. Shewchuk. Triangle: Engineering a 2D Quality Mesh Generator and Delaunay Triangulator. In *Proc. First Workshop on Applied Computational Geometry*, pp 124–133. Association for Computing Machinery, 1996.

[40] W. Schroeder, K. Martin, and B. Lorensen. *The Visualization Toolkit: An Object-*

Oriented Approach to 3D Graphics. Prentice-Hall, 1996.

[41] C. Chao-Jung. *Development of a Traceable Atomic Force Microscope with Interferometer and Compensation Flexure Stage*. Dissertation, Technische Universität Ilmenau, 2003.

[42] P. Fleischmann and S. Selberherr. Enhanced Advancing Front Delaunay Meshing in TCAD. In *SISPAD*, pp 99–102, 2002.

[43] A. Ghetti. *Gate Oxide Reliability: Physical and Computational Models*, pp 201–258. Springer, 2004.

[44] M. Lenzlinger and E. H. Snow. Fowler-Nordheim Tunneling into Thermally Grown SiO_2. *J.Appl.Phys.*, 40(1):278–283, 1969.

[45] L. J. Huang, J. Chu, S. Goma, C. Emic, S. Koester, D.Canaperi, P. Mooney, S. Cordes, J. Speidell, R. Anderson, and H. Wong. Carrier mobility enhancement in strained Si-on-Insulator fabricated by wafer bonding. *VLSI Symp. Tech.Dig.*, pp 57–58, 2001.

[46] K. Rim, J. Chu, H. Chen, K. Jenkins, T. Kanarsky, K. Lee, A. Mocuta, H. Zhu, R. Roy, J. Newbury, J. Ott, K. Petrarca, P. Mooney, D. Lacey, S. Koester, K. Chan, D. Boyd, M. Ieong, and H. Wong. Characteristics and device design of sub-100 nm strained Si N- and PMOSFETs. *VLSI Symp. Tech.Dig.*, pp 98–99, 2002.

[47] N. Sugii, D. Hisamoto, K. Washio, N. Yokoyama, and S. Kimura. Enhanced performance of strained-Si MOSFETs on CMP SiGe virtual substrate. In *Proc. Intl. Electron Devices Meeting*, pp 737–740, 2001.

[48] S. E. Thompson, M. Armstrong, C. Auth, M. Alavi, and M. Buehler. A 90-nm Logic Technology Featuring Strained-Silicon. *IEEE Trans.Electron Devices*, 51(11):1790–1797, 2004.

[49] M. V. Fischetti, F. Gamiz, and W. Hänsch. On the enhanced electron mobility in strained-silicon inversion layers. *J.Appl.Phys.*, 92(12):7320–7324, 2002.

[50] J. R. Watling, L. Yang, M. Borici, R. C. Wilkins, A. Asenov, J. R. Barker, and S. Roy. The impact of interface roughness scattering and degeneracy in relaxed and strained Si n-channel MOSFETs. *Solid-State Electron.*, 48:1337–1346, 2004.

[51] T. Ando, A. B. Fowler, and F. Stern. Electronic properties of two-dimensional systems. *Review of Modern Physics*, 54(2):437–672, 1982.

[52] S. Bosi and C. Jacoboni. Monte Carlo High-Field Transport in Degenerate GaAs. *Journal of Physics C*, 9(2):315–319, 1976.

[53] S. Yamakawa, H. Ueno, K. Taniguchi, C. Hamaguchi, K. Miyatsuji, K. Masaki, and U. Ravaioli. Study of Interface Roughness Dependence of Electron Mobility in Si Inversion Layers Using the Monte Carlo Method. *J.Appl.Phys.*, 79(2):911–916, 1996.

[54] D. Roychoudhury and P. K. Basu. Mobility of electrons in a quantized inversion layer due to phonon scattering. *Physical Review B*, 22(12):6325 – 6329, 1980.

[55] M. V. Fischetti, Z. Ren, P. M. Solomon, M. Yang, and K. Rim. Six-Band K.P Calculation of the Hole Mobility in Silicon Inversion Layers: Dependence on Surface Orientation, Strain, and Silicon Thickness. *J.Appl.Phys.*, 94(2):1079–1095, 2003.

[56] D. Vasileska and Z. Ren. *SCHRED 2.0 User's Manual*. http://www.nanohub.org, 2000.

[57] C. Jungemann, A. Edmunds, and W. L. Engl. Simulation of Linear and Nonlinear Electron Transport in Homogenous Silicon Inversion Layers. *Solid-State Electron.*, 36(11):1529–1540, 1993.

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