VISTA Status Report
December 2009

O. Baumgartner, O. Ertl, R. Orio, P.J. Wagner, T. Windbacher, S. Selberherr

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

Contents

1 Numerical Quadrature of the Subband Distribution Functions in Strained Silicon UTB Devices

In this work, the $\mathbf{k}\cdot\mathbf{p}$ method is used to calculate the electronic subband structure. To reduce the computational cost of the carrier concentration calculation and henceforth the required number of numerical solutions of the Schrödinger equation, an efficient 2D $\mathbf{k}$-space integration by means of the Clenshaw-Curtis method is proposed. The suitability of our approach is demonstrated by simulation results of Si UTB double gate nMOS and pMOS devices.

1.1 Introduction

Strained silicon ultra-thin body MOSFETs are considered to be good candidates for CMOS integration in the post 22 nm technology nodes. An accurate description of such devices relies on the modeling of the subband structure. An efficient self-consistent Schrödinger-Poisson model for the calculation of the electronic subband structure is presented, taking into account band nonparabolicity and arbitrary strain [1]. A two-band $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian has been used for electrons and a six-band $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian for holes.

1.2 Calculation of the Subband Structure

The numerical modeling of the subband structure in ultra thin body SOI MOS structures relies on an accurate model of the bulk Hamiltonian. We applied a two-band $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian [2, 3] to describe the silicon conduction band around the X points.

$$\mathbf{H} = \begin{bmatrix} \mathrm{H}_- & \mathrm{H}_{\mathrm{bc}} \\ \mathrm{H}_{\mathrm{bc}} & \mathrm{H}_+ \end{bmatrix} \qquad \text{with}$$

$$\mathrm{H}_\mp = \mathcal{E}_\mathrm{c}(z) + \frac{\hbar^2 k_\mathrm{z}^2}{2m_\mathrm{l}} + \frac{\hbar^2\left(k_\mathrm{x}^2+k_\mathrm{y}^2\right)}{2m_\mathrm{t}} \mp \frac{\hbar^2 k_0 k_\mathrm{z}}{m_\mathrm{l}},$$

$$\mathrm{H}_{\mathrm{bc}} = D\epsilon_{\mathrm{xy}} - \frac{\hbar^2 k_\mathrm{x} k_\mathrm{y}}{M}.$$

$\mathcal{E}_\mathrm{c}$ denotes the conduction band edge energy, m_l and m_t are the longitudinal and transversal electron masses, respectively, and $\frac{1}{M} \approx \frac{1}{m_\mathrm{t}} - \frac{1}{m_\mathrm{e}}$. The shear strain deformation potential $\mathrm{D} = 14\,\mathrm{eV}$ and the off-diagonal strain component ϵ_{xy} describe the effects of shear strain on the bandstructure. $k_0 = 0.15\frac{2\pi}{a_0}$ corresponds to the distance of the valley to the X point.

To model the silicon valence band structure a $6\times6-\mathbf{k}\cdot\mathbf{p}$ Hamiltonian [4] has been implemented. Following the notation of Manku it is written as

$$\mathbf{H} = \mathcal{E}_\mathrm{v}\mathbf{I}_{6\times6} + \begin{bmatrix} \mathbf{S}+\mathbf{D} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{S}+\mathbf{D} \end{bmatrix} + \mathbf{H}_{\mathbf{so}},$$

where $\mathcal{E}_\mathrm{v}$ is the valence band edge and the perturbation matrix $\mathbf{S}$ and the deformation potential matrix $\mathbf{D}$ are given by

$$\mathbf{S} = \begin{bmatrix} Lk_\mathrm{x}^2+M(k_\mathrm{y}^2+k_\mathrm{z}^2) & Nk_\mathrm{x}k_\mathrm{y} & Nk_\mathrm{x}k_\mathrm{z} \\ Nk_\mathrm{x}k_\mathrm{y} & Lk_\mathrm{y}^2+M(k_\mathrm{x}^2+k_\mathrm{z}^2) & Nk_\mathrm{y}k_\mathrm{z} \\ Nk_\mathrm{x}k_\mathrm{z} & Nk_\mathrm{y}k_\mathrm{z} & Lk_\mathrm{z}^2+M(k_\mathrm{x}^2+k_\mathrm{y}^2) \end{bmatrix}$$

$$\mathbf{D} = \begin{bmatrix} l\varepsilon_{\mathrm{xx}}+m(\varepsilon_{\mathrm{yy}}+\varepsilon_{\mathrm{zz}}) & n\varepsilon_{\mathrm{xy}} & n\varepsilon_{\mathrm{xz}} \\ n\varepsilon_{\mathrm{xy}} & l\varepsilon_{\mathrm{yy}}+m(\varepsilon_{\mathrm{xx}}+\varepsilon_{\mathrm{zz}}) & n\varepsilon_{\mathrm{yz}} \\ n\varepsilon_{\mathrm{xz}} & n\varepsilon_{\mathrm{yz}} & l\varepsilon_{\mathrm{zz}}+m(\varepsilon_{\mathrm{xx}}+\varepsilon_{\mathrm{yy}}) \end{bmatrix}$$

As parameters for the silicon valence band structure without strain $L = -6.53$, $M = -4.64$, and $N = -8.75$ in units of $\frac{\hbar^2}{2m_\mathrm{e}}$ have been used [5]. l, m, and n are the strain deformation potentials for the valence band.

The spin orbit coupling is described by the Hamiltonian

$$\mathbf{H}_{\mathbf{so}} = -\frac{\mathcal{E}_{\mathrm{so}}}{3}\begin{bmatrix} 0 & \mathrm{i} & 0 & 0 & 0 & -1 \\ -\mathrm{i} & 0 & 0 & 0 & 0 & \mathrm{i} \\ 0 & 0 & 0 & 1 & -\mathrm{i} & 0 \\ 0 & 0 & 1 & 0 & -\mathrm{i} & 0 \\ 0 & 0 & \mathrm{i} & \mathrm{i} & 0 & 0 \\ -1 & -\mathrm{i} & 0 & 0 & 0 & 0 \end{bmatrix},$$

with the split off energy of silicon $\mathcal{E}_{\mathrm{so}} = 44\,\mathrm{meV}$.

Quantization is introduced in the bulk Hamiltonian by the substitution $k_\mathrm{z} \longrightarrow -\mathrm{i}\,\partial_\mathrm{z}$, where the z-axis is the quantization direction and corresponds to the normal of the

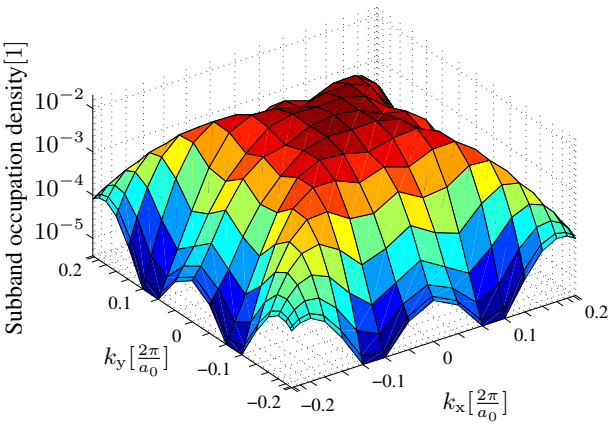


Figure 1: Occupation of the heavy hole band of Si in a 3 nm wide quantum well. The grid shows the nodes of the numerical quadrature.

(001) silicon crystal surface throughout this work. A finite difference scheme with hard wall boundary conditions has been used to discretize the Schrödinger equation. The resulting eigenvalue problem gives rise to discrete energies describing the subband structure.

1.3 Numerical Quadrature of the Subband Distribution Functions

The contribution of subband i and valley j to the equilibrium electron concentration is given by

$$n_{i,j}(z) = \int_{\mathrm{BZ}} \mathrm{d}^2\mathbf{k}\, |\psi_{i,j}(z)|^2 \, \frac{1}{(2\pi)^2} f_0\left(\mathcal{E}_{i,j}\left(k_\mathrm{x}, k_\mathrm{y}\right) - \mathcal{E}_\mathrm{F}\right),$$

where $\psi_{i,j}$ is the wave function and f_0 is the Fermi distribution and $\mathcal{E}_\mathrm{F}$ the Fermi level with a similar relationship holding for the hole concentration in a pMOS device. Therefore, to calculate the occupation of a subband a numerical, two-dimensional $\mathbf{k}$-space integration is required. This necessitates to solve the Schrödinger equation for every discrete point $(k_\mathrm{x}, k_\mathrm{y})$. Hence, one seeks after a numerical quadrature scheme that gives good accuracy on a coarse grid. In contrast to previous work [6] which made use of harmonic and cubic spline interpolation for $\mathbf{k}$-space integration, in this work the Clenshaw-Curtis method [7] has been applied. As nodes in the integration interval $[-1, 1]$ the zeros of the Chebyshev polynomial are used: $x_k := \cos(k\frac{\pi}{N})$ with $k = 0, 1, \ldots, N$. Following [8], the weights are written explicitly as

$$w_k = \frac{c_k}{N}\left(1 - \sum_{j=1}^{\lfloor N/2 \rfloor} \frac{b_j}{4j^2 - 1} \cos\left(2jk\tfrac{\pi}{N}\right)\right)$$

with $b_j = 1$ if $j = N/2$, or $b_j = 2$, if $j < N/2$, and $c_k = 1$ if $k \mod N = 0$, or $c_k = 2$ otherwise. For the $\mathbf{k}$-space integration of the subbands provided by the $\mathbf{k} \cdot \mathbf{p}$ Hamiltonian excellent accuracy has been achieved with only 19 nodes per k direction.

1.4 Results and Discussion

A (001) silicon UTB DG-MOSFET with 3 nm film and 1 nm oxide thickness has been simulated. For the nMOS device the donor doping of the polysilicon gates was $\mathrm{N_D} = 1.0 \times 10^{20}\,\mathrm{cm}^{-3}$ and the Si film was lightly p-doped at $\mathrm{N_A} = 2.0 \times 10^{16}\,\mathrm{cm}^{-3}$, while the complementary doping has been used for the pMOS device. The occupation function of the heavy hole band is depicted in Fig. 1. Equivalently the lowest unprimed subband of the nMOS device with and without shear strain is depicted in Fig. 2. The grid as shown in the figures corresponds to the nodes of the Clenshaw-Curtis quadrature. The zeros of the Chebyshev polynomial give an accumulation of grid points at the boundary of the integration domain. The integration intervals for the nMOS have been chosen as ten percent of the width of the Brillouin zone in each positive and negative direction around the valley. For the pMOS device the boundaries have been set at $k_{\mathrm{x,y}} = \pm 0.2\frac{2\pi}{a_0}$. Therefore, the domain has to be normalized accordingly to the interval $[-1, 1]$ of the Clenshaw-Curtis rule.

Fig. 3 shows the self-consistent conduction band edge and the electron concentration for the nMOS and Fig. 4 the corresponding result for the pMOS device. Within the well, the squared wave functions for the four lowest, twofold degenerate unprimed subbands are displayed at their corresponding energy levels. For each subband the electron density is calculated by $\mathbf{k}$-space integration. For the (001) Si-nMOS device the unprimed and primed subband ladder are taken into account to obtain the self-consistent solution.

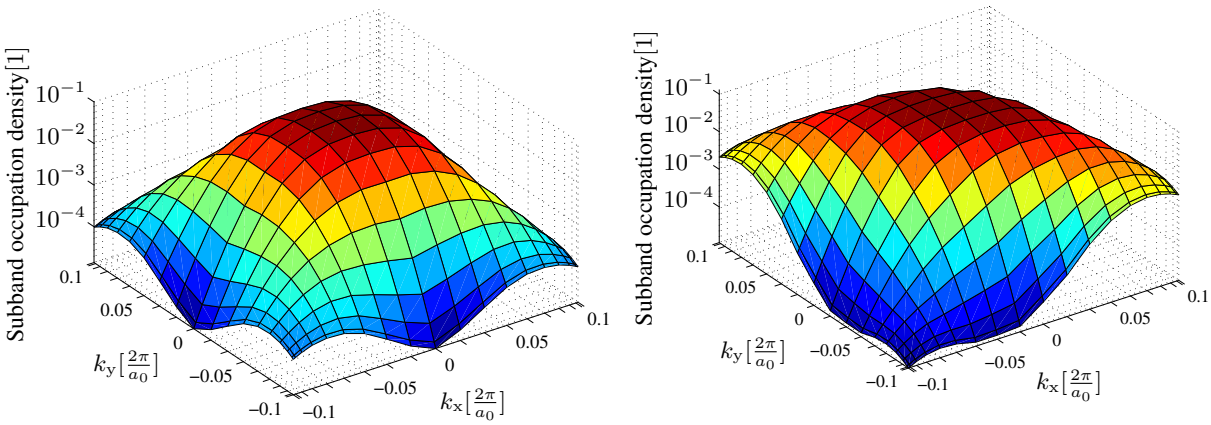


Figure 2: Occupation of the lowest unprimed subband of a 3 nm (001) silicon conduction band quantum well without strain (left) and with $\varepsilon_\mathrm{xy} = 0.5\%$ shear strain (right).

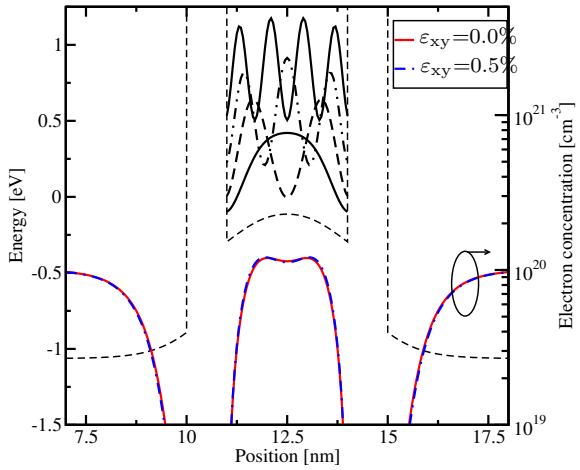


Figure 3: Self-consistent calculation of the conduction band edge and the electron concentration of a (001) Si-DG-nMOS with 3 nm well width and 1 nm oxide thickness. The normalized wave functions $[\mathrm{nm}^{-1}]$ are overlayed at their respective energy levels. The electron concentration is plotted for the unstrained case and for ε_{xy}=0.5% shear strain.

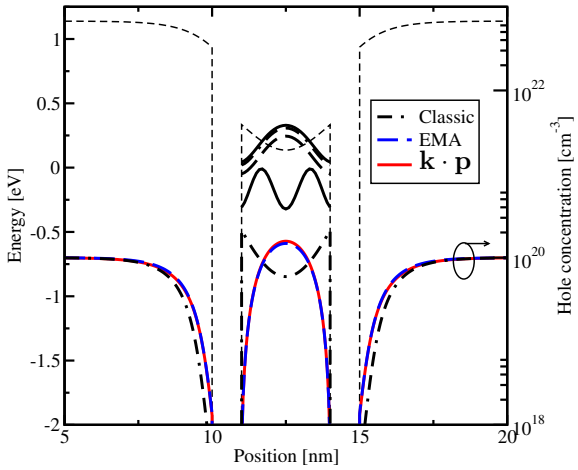


Figure 4: Same as Fig. 3 but for Si-DG-pMOS device.

The convergence behavior of the self-consistent Schrödinger/Poisson loop is shown in Fig. 5. The quadratic norm of the potential update after an iteration evolves similarly for a different number of nodes per k-direction. As depicted in the figure, the convergence behavior is good and hence the iteration scheme proves stable.

To give an impression of the accuracy of the numerical quadrature method a test with parabolic subbands has been conducted. Therefore, a (001) silicon DG-nMOS device has been simulated using the two-band $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian with k_0 and $\frac{1}{M}$ set to zero which corresponds to the parabolic effective mass approximation (EMA). Again, the unprimed and primed valleys are taken into account. This way, the self-consistent carrier concentration has been calculated and compared to the results of the EMA, where the 2D subband density is calculated analytically. The maximum relative difference of the electron concentration for the 3 nm silicon well has been used as measure of accuracy for the numerical quadrature. The results are depicted in Fig. 6. Furthermore, the CPU time for the calculation of the electron concentration on a single core of an Intel Core 2 Quad Q6600 machine is given. This includes the time for solving the Schrödinger equation for all points in $\mathbf{k}$-space and the following numerical quadrature. The curve shows the expected $\mathcal{O}(\mathrm{N}^2)$ behavior of the algorithm.

In Table 1 the minima of the unprimed (U) and primed (P) subbands are shown in units of eV. The five lowest eigenvalues of the subband ladders are summarized. The eigenvalues of the emulated "parabolic" $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian with k_0 and $\frac{1}{M}$ set to zero and numerically integrated subbands are compared to the energy levels resulting of the effective mass Schrödinger equation. The relative difference of the eigenvalues is given to show the accuracy of the self-consistent result using our proposed calculation scheme.

In Table 2 the effects of nonparabolicity and strain on the bound states is summed up. As in Table 1 the unprimed and primed subband ladder is shown in units of eV. The nonparabolic two-band $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian applied to a (001) silicon UTB device gives a two-fold degenerate unprimed valley which is located at the X-point. The four-fold degenerate primed valleys have their minimum at $k = \pm 0.15\frac{2\pi}{a_0}$.

By applying shear strain, the unprimed subbands at the X-point are split and shifted downwards with respect to the primed subband ladder, therefore, favoring the occupation of the unprimed valleys with lower transport mass. Whereas the occupation of the individual subbands is changed fundamentally, the effect on the total electron concentration is marginal as shown in Fig. 3.

Similar simulations were carried out for the DG-pMOS device. In the effective mass approximation three types of holes have been considered. The heavy hole band with $m_{\mathrm{hh}} = 0.39\,m_{\mathrm{e}}$, the light hole band with $m_{\mathrm{lh}} = 0.19\,m_{\mathrm{e}}$ and the split off band with $m_{\mathrm{so}} = 0.24\,m_{\mathrm{e}}$ and a shift of 44 meV down from the valence band edge are included in the calculations. As illustrated in Fig. 4 this gives a good agreement of the EMA hole concentration with the self-consistent six-band $\mathbf{k}\cdot\mathbf{p}$ results. The calculated bound states in the UTB are summarized in Table 3. Furthermore, compressive stress of 1 GPa in [110] direction was applied. This gives an additional splitting of the heavy hole and light hole band. Under these conditions the transport mass in [110] of the highest band extracted from the $\mathbf{k}\cdot\mathbf{p}$ dispersion relation was $m = 0.17\,m_{\mathrm{e}}$ as compared to $m = 0.32\,m_{\mathrm{e}}$ in the unstrained case.

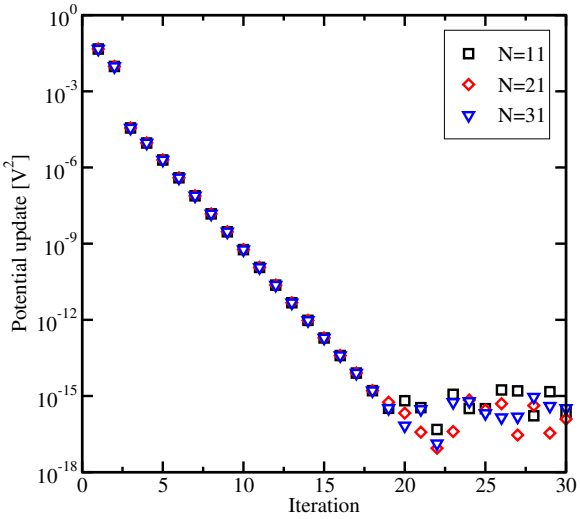


Figure 5: Potential update after each Schrödinger/Poisson iteration for different numbers of nodes per k-direction of the Clenshaw-Curtis subband integration. Starting from the classical solution all simulations give similar convergence behavior.

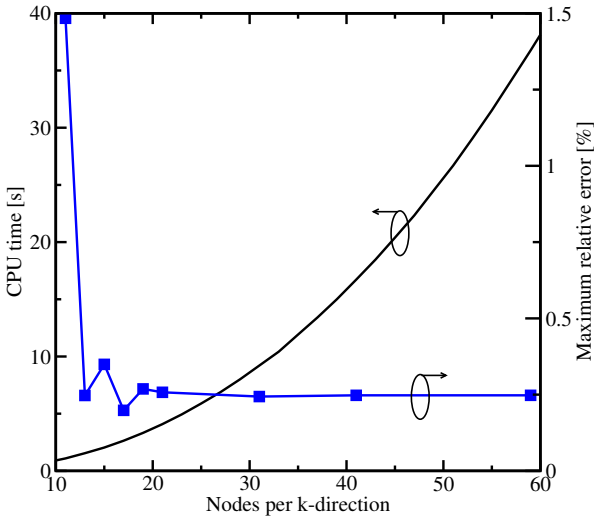


Figure 6: CPU time for a single Schrödinger/Poisson iteration for different number of nodes per k direction of the Clenshaw-Curtis subband integration. The maximum relative difference of the self-consistent carrier concentration within the well for numerically integrated parabolic subbands with respect to the effective mass approximation with analytically integrated subbands is given to show the good accuracy of the quadrature method.

1.5 Conclusion

Contrary to numerical solutions based on the one-band effective mass Schrödinger equation, this work considers a nonparabolic dispersion relation based on a $\mathbf{k} \cdot \mathbf{p}$ Hamiltonian. Furthermore, shear strain effects leading to a warping of the bandstructure are accounted for. The proposed numerical quadrature of the subbands has been successfully applied to electron and hole states in unstrained and strained Si. The self-consistent solutions for the band edges and carrier concentrations of a UTB Si nMOS and pMOS device are presented. The numerical quadrature proves as simple and yet robust method.

Table 1: The minima of the unprimed (U) and primed (P) subbands are shown in units of eV. To test the numerical quadrature the two-band $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian has been used with k_0 and $\frac{1}{M}$ set to zero which corresponds to parabolic bands. The relative difference of the eigenvalues is given to show the accuracy of the self-consistent result using numerical integration.

EMA		$\mathbf{k}\cdot\mathbf{p}$ parabolic		Relative difference	
U	P	U	P	U	P
-0.11210	0.04745	-0.11214	0.04741	3.6×10^{-4}	8.4×10^{-4}
-0.01494	0.61437	-0.01498	0.61434	2.7×10^{-3}	4.9×10^{-5}
0.19655	1.60525	0.19651	1.60521	2.0×10^{-4}	2.5×10^{-5}
0.49224	2.99374	0.49220	2.99371	8.1×10^{-5}	1.0×10^{-5}
0.87241	4.77407	0.87237	4.77404	4.6×10^{-5}	6.3×10^{-6}

Table 2: As in Table 1 the unprimed and primed subband ladder is shown in units of eV. By applying shear strain, the unprimed subbands are split and shifted downwards with respect to the primed ladder, therefore, favoring the occupation of the unprimed valleys with lower transport mass.

	$\mathbf{k}\cdot\mathbf{p}$ nonparabolic		$\mathbf{k}\cdot\mathbf{p}$ with $\varepsilon_{\mathrm{xy}}=0.5\%$	
	Unprimed	Primed	Unprimed	Primed
1	-0.10148	0.05773	-0.10336	0.07147
	-0.10148		-0.10071	
2	-0.00417	0.62431	-0.02814	0.63826
	-0.00417		0.04673	
3	0.20812	1.61512	0.22144	1.62886
	0.20812		0.23429	
4	0.50491	3.00361	0.51010	3.01732
	0.50491		0.53152	
5	0.88650	4.78393	0.89072	4.79764
	0.88650		0.91224	

Table 3: The bound states in a (001) Si DG-pMOS in units of eV. For the effective mass approximation the heavy hole, light hole and split off states have been considered. The results are compared to $\mathbf{k}\cdot\mathbf{p}$ simulations for unstrained and strained Si with compressive stress of 1 GPa in [110] direction.

	EMA			$\mathbf{k}\cdot\mathbf{p}$	
	HH	LH	SO	Unstrained	Strained
1	0.11094	0.00446	0.04758	0.02275	0.04029
2	-0.15444	-0.57724	-0.40559	-0.00009	-0.00842
3	-0.65224	-1.60061	-1.21528	-0.07281	-0.09870
4	-1.34922	-3.03304	-2.34895	-0.36674	-0.34967
5	-2.24353	-4.86962	-3.80274	-0.39456	-0.40065

2 Three-Dimensional Level Set Based Bosch Process Simulations Using Ray Tracing for Flux Calculation

This paper presents three-dimensional simulations of deep reactive ion etching processes, also known as Bosch processes. A Monte Carlo method, accelerated by ray tracing algorithms, is used to solve the transport equation, while advanced level set techniques are applied to describe the movement of the surface. With multiple level sets it is possible to describe accurately the different material layers which are involved in the process. All used algorithms are optimized in such a way, that the costs of computation time and memory scale more like with the surface size rather than with the size of the simulation domain. Finally the presented simulation techniques are used to simulate the etching of holes, whereas the influence of passivation/etching cycle times and hole diameters on the final profile is investigated.

2.1 Introduction

The invention of the Bosch process [9] enabled high aspect ratio etching by alternation of passivation and etching cycles and is used in semiconductor devices and microelectromechanical systems (MEMS) fabrication. In each cycle a chemically inert polymer layer is uniformly deposited using fluorocarbon gases. This passivation layer prevents the sidewalls from being attacked in the subsequent etching step (Fig. 7). By feeding a high frequency plasma with etch gases like SF_6, CF_4, or NF_3, a superposition of physical (directional) and chemical (isotropic) etching is obtained. This leads to a faster removal of the passivation layer at the bottom of the trench compared to the sidewalls due to the additional sputtering of the directional ions. After uncovering the substrate at the bottom chemical etching is dominant. Hence, in simple terms, in each cycle an isotropic etching process is started at the bottom of the trench. After many iterations profiles with high aspect ratios can be obtained. For optimal processing the passivation and etching cycle times have to be balanced. If the deposited passivation layer is too thin, the process time for the etching cycle has to be smaller to avoid the corrosion of the sidewalls increasing the number of required iterations. If the layer is too thick, the etching duration has to be increased resulting in a longer total process time. The choice of the process times has also an influence on the undulation of the sidewalls caused by the two-phase procedure. Computer simulations help to study parameter variations in order to optimize the process. Several simulators have been developed and applied to the Bosch process in the past [10]- [11].

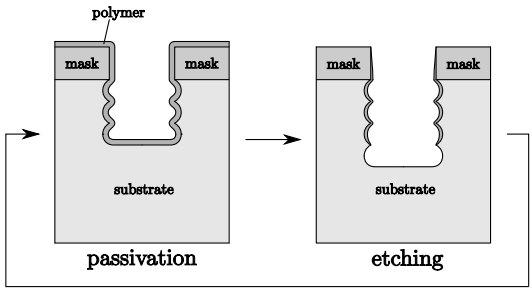


Figure 7: A schematic illustration of the Bosch process. The deposition of a passivation layer protects the sidewalls during the subsequent etching cycle.

A two-dimensional simulator using a string-cell hybrid method for surface evolution was presented in [10, 12]. Therein a simplified model for the particle transport is used. Etching is modeled by a constant etching rate superposed by a directional etching term which is proportional to the incident ion flux. For the passivation cycle a perfect conformal deposition is assumed, which is equivalent to a constant surface velocity. However, this model is not able to describe the lag effect [13] appropriately. Therefore, a geometric shape factor was introduced [14], accounting for different trench widths.

A simulation with a more sophisticated transport model is presented in [15], where different sticking probabilities and higher order re-emissions of neutral particles are incorporated using the ballistic transport-reaction model (BTRM) [16, 17]. Since the transport of neutrals to the surface is taken into account, the lag effect is inherently incorporated. The surface evolution is calculated using the level set method [18], which allows easy handling of topographic changes, while sub-grid resolution of the surface can be achieved. The transport equations, which result in surface integral equations, are solved by conventional integration techniques [19].

Another approach to calculate the particle transport is based on the Monte Carlo method which was first applied to Bosch process simulation in combination with a string-cell method for surface evolution in [20]. Many particle trajectories and their surface reactions are calculated to determine the surface rates.

Three-dimensional simulations of the Bosch process were recently reported in [11, 21]. Both use simplified transport models and do not incorporate higher order reemissions of neutrals. Instead, a uniform surface rate is assumed. The particle transport is calculated using conventional integral methods. For surface evolution a

voxel-based method and the level set method are used, respectively.

In the following we describe a new approach for three-dimensional Bosch process simulations. We use advanced level set techniques to represent the geometry and also the different material regions. To determine the reaction rates on the surface we apply a Monte Carlo method, accelerated by ray tracing algorithms and parallelization.

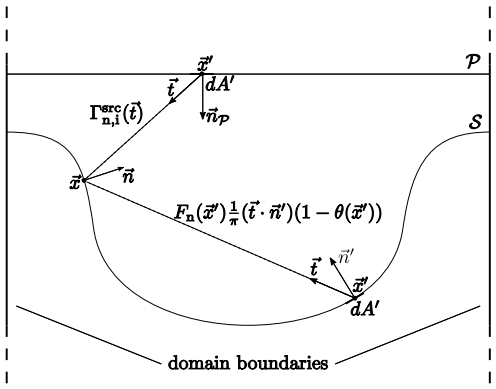


Figure 8: The arriving flux at point $\vec{x}$ on the surface $\mathcal{S}$ is the sum of the flux coming directly from the source plane $\mathcal{P}$ and the re-emitted flux originating from the surface itself. In case of reflective or periodic domain boundaries the regions of integration $\mathcal{P}$ and $\mathcal{S}$ are finite.

2.2 Model

The scope of this paper is the demonstration of three-dimensional Bosch process simulations by means of fast computation techniques. For this purpose we use the model as given in [15], where a Bosch process with alternating flows of SF_6 and fluorocarbon gases is described, including a full set of parameters. In the following we summarize the model and discuss the solution of the governing equations.

2.2.1 Particle Transport

The model is based on the BTRM [16, 17], where the mean free path of particles is assumed to be much larger than the typical structure sizes of the geometry. Hence, particle–particle interactions can be neglected at feature scale. The arrival angle distributions of all particles are given at a certain plane $\mathcal{P}$, called source plane, just above the surface $\mathcal{S}$ (Fig. 8). For neutral particles a cosine-like arrival angle distribution (flux per solid angle)

$$\Gamma^{\text{src}}_{\text{n}}(\vec{t}) = F^{\text{src}}_{\text{n}}\frac{1}{\pi}(\vec{t}\cdot\vec{n}_{\mathcal{P}}). \tag{1}$$

is assumed, while a more directional distribution is used for ions

$$\Gamma^{\text{src}}_{\text{i}}(\vec{t}) = F^{\text{src}}_{\text{i}}\frac{\kappa+1}{2\pi}(\vec{t}\cdot\vec{n}_{\mathcal{P}})^{\kappa}. \tag{2}$$

Here $\vec{t}$ denotes the direction and $\vec{n}_{\mathcal{P}}$ is the normal on the source plane pointing to the surface. $F^{\text{src}}_{\text{n}}$ and $F^{\text{src}}_{\text{i}}$ are the total incoming fluxes of neutrals and ions, respectively. The parameter κ is used to model the narrow angular distribution of ions [22]. For $\kappa \gg 1$ this distribution is equivalent to a normal distribution for the arrival angles with a standard deviation of $\sigma = \frac{1}{\sqrt{\kappa}}$.

The arriving flux for neutrals can be obtained by solving the surface integral equation

$$F_{\text{n}}(\vec{x}) = \int_{\mathcal{P}} \Gamma^{\text{src}}_{\text{n}}(\vec{t})\frac{\text{vis}(\vec{x},\vec{x}')(-\vec{t}\cdot\vec{n})}{\|\vec{x}-\vec{x}'\|^2}dA' + \int_{\mathcal{S}} F_{\text{n}}(\vec{x}')\frac{1}{\pi}(\vec{t}\cdot\vec{n}')(1-\theta(\vec{x}'))\frac{\text{vis}(\vec{x},\vec{x}')(-\vec{t}\cdot\vec{n})}{\|\vec{x}-\vec{x}'\|^2}dA', \tag{3}$$

with $\vec{t} = \frac{(\vec{x}-\vec{x}')}{\|\vec{x}-\vec{x}'\|}$. The first term describes the direct flux from the source. The second term is the flux which origins from the surface itself due to re-emission. $\text{vis}(\vec{x},\vec{x}')$ is the visibility function which returns 1 or 0, if the surface points $\vec{x}$ and $\vec{x}'$ are in line of sight or not, respectively. For neutrals diffusive re-emission with a sticking probability $\theta(\vec{x})$ is assumed. During the passivation cycle the sticking probability is uniform, because the whole surface gets covered with the same type of material. During the etching cycle the sticking probability depends on the local material on the surface. For ions a constant sticking probability of 1 is assumed. Therefore, the ion flux can be written as

$$F_{\text{i}}(\vec{x}) = \int_{\mathcal{P}} \Gamma^{\text{src}}_{\text{i}}(\vec{t})\frac{\text{vis}(\vec{x},\vec{x}')(-\vec{t}\cdot\vec{n})}{\|\vec{x}-\vec{x}'\|^2}dA'. \tag{4}$$

2.2.2 Surface Kinetics

The deposition and etching rates in both cycles of the Bosch process are simply modeled by linear combinations of neutral and ion fluxes

$$R(\vec{x}) = \alpha F_{\text{i}}(\vec{x}) + \beta F_{\text{n}}(\vec{x}). \tag{5}$$

The coefficients α and β are model parameters, and in case of etching they depend on the exposed material. The model assumes that three different types of material are involved in the Bosch process: the mask, the substrate, and the passivation (polymer) layer.

The numeric values of all parameters for the passivation and the etching cycle, which we used for all simulations, are listed in Table 4 and Table 5, respectively. Contrary to [15] we also consider mask etching by assuming a mask/substrate etch selectivity of 1:20. The coefficients α and β for the mask are adjusted accordingly. Furthermore, a spread of the arrival angles of ions is assumed ($\sigma = 2^\circ$).

Table 4: The numeric values of the parameters used for the simulation of the passivation cycle.

Parameter	Value
σ	2°
$F_{\text{n}}^{\text{src}}$	$2 \cdot 10^{18}$ atoms/(cm^{2}s)
$F_{\text{i}}^{\text{src}}$	$3.125 \cdot 10^{15}$ atoms/(cm^{2}s)
α	10 Å^3/atom
β	0.5 Å^3/atom
θ	0.1

Table 5: The numeric values of the parameters used for the simulation of the etching cycle.

Parameter	Value
σ	2°
$F_{\text{n}}^{\text{src}}$	10^{19} atoms/(cm^{2}s)
$F_{\text{i}}^{\text{src}}$	$4.375 \cdot 10^{15}$ atoms/(cm^{2}s)
α_{polymer}	125 Å^3/atom
$\alpha_{\text{substrate}}$	270 Å^3/atom
α_{mask}	13.5Å^3/atom
β_{polymer}	0.03 Å^3/atom
$\beta_{\text{substrate}}$	0.9 Å^3/atom
β_{mask}	0.045 Å^3/atom
θ_{polymer}	0.1
$\theta_{\text{substrate}}$	0.2
θ_{mask}	0.2

To solve the above-described equations two different methods are necessary. One for tracking the surface and the different material regions over time and a second to determine the fluxes on the surface. In the following two sections the numerical framework is presented to accomplish these tasks.

2.3 Surface Evolution

This section addresses the description of the geometry and of its evolution over time. For Bosch process simulation it is important that the profile evolution algorithm can handle different material regions. We use the level set method, since it allows a sub-cell accurate representation of the surface, while topographic changes are handled inherently.

2.3.1 Level Set Method

The basic idea of the level set method is to describe a boundary by means of a continuous function [18]. For a given surface $\mathcal{S}$ a level set function Φ is initialized in such a way that $\mathcal{S}$ can be obtained as its zero level set

$$\mathcal{S} = \{\vec{x} : \Phi(\vec{x}) = 0\}. \tag{6}$$

The advantage of this representation is that the propagation of a boundary driven by a given velocity field $V(\vec{x})$ can be easily determined by solving the level set equation

$$\frac{\partial \Phi}{\partial t} + V(\vec{x})\|\nabla \Phi\| = 0. \tag{7}$$

If discretized on a Cartesian grid, this equation can be easily solved by means of simple finite difference schemes. To guarantee a stable time integration a Courant–Friedrichs–Lewy (CFL) condition has to be fulfilled, which restricts the maximum advancement of the surface per time step. For our calculations we used a maximum step size of 0.1 grid spacings.

In topography simulations the surface velocities are only defined on the surface. Therefore, to get the required surface velocity field an extension technique has to be applied [23]. To keep the level set function a signed distance function, it was proposed to take for each grid point the surface velocity of its closest surface point [24]. Later, we will discuss this mapping and more generally, how to couple the transport equation solver with the level set method.

2.3.2 Sparse Field Level Set Method

The original level set method shows a non-linear scaling of computation time and memory consumption with the surface size, since the level set function is stored and integrated over time for all grid points of the simulation domain. A linear scaling law for the computation time was achieved by the narrow band level set method which only considers active grid points close to the surface for time integration. The approach makes use of the fact that the level set values of grid points far away do not influence the actual position of the surface. A further enhancement of this method is the sparse field level set method [25], which has the advantage that only a single layer of active grid points, namely those with an absolute level set value less than 0.5, must be considered for time

integration. Only for those grid points the surface velocity field has to be known, making the mapping from the surface very easy. Another advantage of the sparse field level set method is that periodic re-initializations of the level set function as needed for conventional level set methods are not necessary. The level set values at neighboring grid points, which are required to determine the derivatives of the level set function, are obtained by a simple and fast update scheme.

2.3.3 Run-Length-Encoding

To reduce the memory consumption for storing a level set function the recently developed hierarchical run-length-encoding (HRLE) data structure [26] was implemented. Only for grid points close to the surface the explicit level set values are stored. For all other grid points only the signs of the level set function are stored using run-length compression. The availability of the signs of all grid points is very useful, since the sign of the level set reveals on which side of the level set a grid point is located.

The HRLE data structure enables fast sequential and random access to grid points with constant and sublogarithmic complexity, respectively. In combination with the sparse field level set method a perfect linearly scaling level set algorithm in terms of surface size can be realized.

Boolean operations like union or intersection of two regions play a role in multi-level-set methods, where different material regions are represented by more than one level set. These operations can be expressed as minimum or maximum of the corresponding level set functions [27], which can be computed with an optimal linear complexity using the HRLE data structure.

2.3.4 Multiple Materials

The simulation of the Bosch process requires accurate handling of all three involved material regions: The substrate, the mask, and the passivation layer, labeled by $\Omega_{\text{substrate}}$, Ω_{mask}, and Ω_{polymer}, respectively. Three level sets are used to represent the whole geometry. They are defined as follows

$$\begin{aligned} \Phi_1(\vec{x}) \leq 0 \quad &\Leftrightarrow \quad \vec{x} \in \Omega_{\text{substrate}}, \\ \Phi_2(\vec{x}) \leq 0 \quad &\Leftrightarrow \quad \vec{x} \in \Omega_{\text{substrate}} \cup \Omega_{\text{mask}}, \\ \Phi(\vec{x}) \leq 0 \quad &\Leftrightarrow \quad \vec{x} \in \Omega_{\text{substrate}} \cup \Omega_{\text{mask}} \cup \Omega_{\text{polymer}}. \end{aligned} \tag{8}$$

Here the zero level set of Φ is equal to the surface, while those of Φ_1 and Φ_2 can be assigned to interfaces. The representation of a structure consisting of three different material regions is illustrated in Fig. 9. If the level set functions are initialized using a metric function, the inequation

$$\Phi_1(\vec{x}) \geq \Phi_2(\vec{x}) \geq \Phi(\vec{x}) \tag{9}$$

holds.

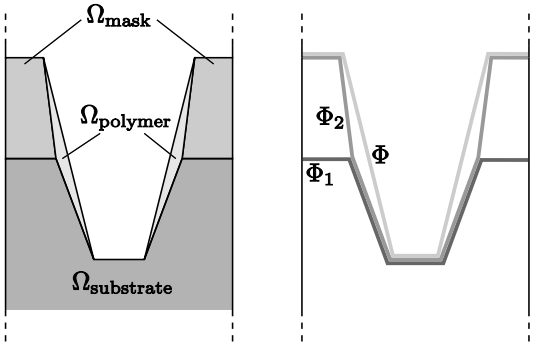


Figure 9: The different material regions which have to be considered during a Bosch process simulation (left) and their representation using level sets (right).

Obviously there are other alternatives to choose the level sets to represent this structure. For example, it is possible to describe each material region by one enclosing level set. However, by nature, if the level set functions are discretized on a Cartesian grid, it is not possible to resolve layers accurately which are thinner than one grid spacing. Therefore, it is possible that the passivation layer suddenly disappears, if a certain thinness is reached during the etching cycle. Consequently, the etching of the underlying substrate starts too early, leading to wrong profiles. This effect is intensified due to the etch rate ratio and due to the multiple repetitions during the Bosch process. Therefore, it is very important to resolve the passivation layer accurately. With the level set configuration as defined in (8) also very thin layers can be resolved.

A time integration step consists of solving the level set equation for the surface level set function Φ and subsequent adapting the interface level sets Φ_1 and Φ_2 using the boolean operation

$$\Phi_i^{(t+\Delta t)}(\vec{x}) = \max\left(\Phi_i^{(t)}(\vec{x}), \Phi^{(t+\Delta t)}(\vec{x})\right). \tag{10}$$

It should be noted that this adaption rule maintains inequation (9). As mentioned previously, the maximum of two level set functions can be constructed very efficiently using the HRLE data structure.

The type of material at a certain surface point $\vec{x}$ can be obtained from the level set functions as follows:

$$\Phi_1(\vec{x}) = \Phi_2(\vec{x}) = \Phi(\vec{x}) \quad \Rightarrow \quad \text{substrate}, \tag{11}$$
$$\Phi_1(\vec{x}) > \Phi_2(\vec{x}) = \Phi(\vec{x}) \quad \Rightarrow \quad \text{mask}, \tag{12}$$
$$\Phi_1(\vec{x}) \geq \Phi_2(\vec{x}) > \Phi(\vec{x}) \quad \Rightarrow \quad \text{polymer}. \tag{13}$$

The surface velocities of different materials are taken into account during time integration. If the surface front reaches another material within a time step (during the etching cycle), the different surface rates are incorporated adequately. A detailed description of this methodology can be found in [28].

2.4 Flux Calculation

Every time step the surface rates have to be determined to enable the profile evolution calculation using the level set method. For this purpose the flux equation (3) has to be solved. Especially in three dimensions it is crucial to use fast techniques and algorithms to speed up the whole topography simulation.

Conventionally, this surface integral equation is solved by discretization of the surface using triangle [29] or voxel elements [30], resulting in a system of linear equations. The system matrix contains the visibility factors which have to determined for each pair of elements. If they are visible from each other, the corresponding system matrix entry is non-zero. Generally the system matrix is dense, which to set up and to solve is computational intensive, since at least a quadratic scaling law with surface size can be expected. The visibility check can lead to an even worse scaling [31].

The particle fluxes are often calculated using an explicit representation of the surface. However, surface extraction algorithms like the marching cubes algorithm [32] result in a huge number of surface elements, revealing the importance of a well scaling algorithm. A way to reduce the number of elements is coarsening of the resulting surface mesh [33]. However, this approach does not only reduce the number of elements and hence the computation time, it also reduces the resolution of the flux. This is a problem, since even on plane regions of the surface the flux can change abruptly due to shadowing. Therefore, coarsening is limited and the unfavorable scaling law is maintained.

2.4.1 Ray Tracing

Since ballistic transport of particles is assumed, the flux calculation is quite analogous to rendering a scene in computer graphics. Due to the ballistic transport of particles the propagation is linear like that of light rays. A widely applied technique to get a realistic picture of a three-dimensional scene is ray tracing [34], a Monte Carlo technique, where a huge number of light rays is simulated. Applied correspondingly to our problem, many particle trajectories are calculated. Whenever, a particle hits the surface it contributes locally to the surface. Thus, the main task is to find the first intersections of rays with the surface. Spatial subdivision can reduce the complexity of finding the first intersection to an expected logarithmic scaling $\mathcal{O}(\log N)$ [35], where N is the number of surface discretization elements, or in our case the number of surface grid cells. Grid cells having corners with different signed level set values contain parts of the surface, which consequently have to be checked for intersection. To optimize the data structure for fast traversals we use binary subdivision along grid planes with simultaneous consideration of a cost function based on surface area heuristics (SAH) as described in [36]. As exemplified in Fig. 10 only a small number of boxes have to be traversed to find the intersection with the surface. Ray tracing can be directly applied to the level set surface representation. The intersection can be found by tri-linear interpolation of the level set function within grid cells and finding the zero-crossing along the particle ray [37].

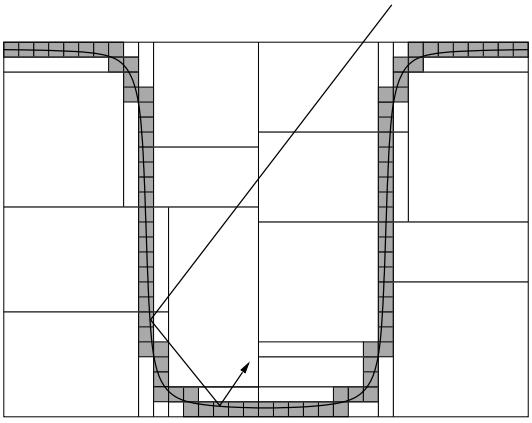


Figure 10: Spatial subdivision accelerates the calculation of particle trajectories. Within the surface cells (gray) tri-linear interpolation is used to find the intersection with the surface.

Since ray tracing is a statistical method, its accuracy strongly depends on the number of simulated rays. To obtain a desired accuracy the number of simulated particle trajectories has to scale with the surface size, to keep the statistical events per unit area constant. Therefore an overall complexity of $\mathcal{O}(N \log N)$ can be achieved using ray tracing, which is a much better scaling law than that for solving the flux balance equation directly.

To be able to determine the incident flux for a certain surface point a reference area has to be defined to relate the number of incidences to the local fluxes. Each particle hitting a reference area of size A^{ref} contributes to the local flux of the corresponding surface point following

$$\Delta F = \frac{F^{\text{src}}}{n \cdot A^{\text{ref}}}. \tag{14}$$

Here n denotes the number of simulated particles which are launched per unit area from the source plane. In principle, these reference areas can be arbitrarily shaped plain areas. For example, the triangles of a surface mesh, or, as we will describe in the next section, tangential disks can be used as reference areas. It is only important that they are localized around the surface point for which the flux has to be determined. However, it is not necessary that the sum of all reference areas equals the real physical area of the surface. In particular, it is even possible that they overlap. In this case an incident particle can contribute to the fluxes of various reference areas following (14).

According to our model neutral particles have a sticking probability much less than 1. Hence, also higher order re-emissions have to be incorporated. This can be performed by continuing the particle trajectory calculation in compliance with the applied re-emission law. The particle trajectory is stopped with a probability equal to the sticking probability. Elsewise, a new direction is randomly chosen in accordance to the used re-emission model, and the particle is re-emitted. In contrast to reality where a particle only contributes to the surface velocity at the point where it finally remains sticking, a particle trajectory contributes to the flux each time it reaches the surface, independent from being re-emitted or not. Hence, more statistical events are generated and a better accuracy is obtained.

Alternatively, instead of re-emitting a particle following the complementary sticking probability, it is also possible to assign a weight factor w to the particle as described in [38]. Starting with an initial value $w^{(0)} = 1$ the particle contributes to the local flux according to its weight factor

$$\Delta F = w \cdot \frac{F^{\text{src}}}{n \cdot A^{\text{ref}}}. \tag{15}$$

In contrast to the first method the particle is always re-emitted, however, with a reduced weight factor

$$w^{(k+1)} = w^{(k)} \cdot (1 - \theta_{\text{stick}}). \tag{16}$$

This method is equivalent with the first one, because the expected contribution to the local flux of a particle which is re-emitted k times is the same in both cases

$$\langle \Delta F \rangle = \rho_k \cdot \frac{F^{\text{src}}}{n \cdot A^{\text{ref}}} = w^{(k)} \cdot \frac{F^{\text{src}}}{n \cdot A^{\text{ref}}}. \tag{17}$$

Here $\rho_k = (1 - \theta_{\text{stick}})^k$ denotes the probability that a particle is re-emitted k times. The trajectory calculation is stopped, if the weight factor falls under a certain fraction $w < w_{\text{limit}}$, or, if the particle leaves the simulation domain upwards. In our simulations we used $w_{\text{limit}} = 10^{-3}$. The error introduced by aborting the particle trajectory is given by w_{limit}. Usually, the error is smaller, because the particle leaves the simulation domain after a couple of re-emissions before reaching this critical weight. For the latter method a better accuracy can be expected especially at regions which are unlikely reached by lower order particles.

2.4.2 Coupling with Surface Evolution

In the following we describe how to link the ray tracing algorithm for flux calculation with the level set method. On the one hand side the surface velocities at all active grid points have to be determined as needed for the sparse field level set method. On the other hand side reference areas for the flux calculation using ray tracing have to be defined. In [39] it was proposed to choose for each active grid point an environment around its closest surface point. However, this approach requires a triangulation of the surface.

As already mentioned ray tracing can be performed directly using the implicit level set surface representation. To avoid an explicit surface representation at all, which increases not only the memory requirements but also the calculation time due to the surface extraction algorithm, a disk with predefined radius ρ is set up for each active grid point. These disks serve as reference areas ($A^{\text{ref}} = \pi\rho^2$) for the calculation of the fluxes for the corresponding active grid points. Their positions are chosen in such a way that they are tangential to the surface at the closest surface point of the corresponding grid point. The closest surface point of a grid point $\vec{p}$ can be approximated by

$$\vec{p}' = \vec{p} + d \cdot \vec{n} = \vec{p} + \frac{\Phi(\vec{p})}{\|\nabla\Phi(\vec{p})\|} \cdot \frac{\nabla\Phi(\vec{p})}{\|\nabla\Phi(\vec{p})\|}. \tag{18}$$

d denotes the distance to the closest surface point and $\vec{n}$ is the normal vector. As applied, both expressions can be estimated from the surface level set function Φ [25]. Fig. 11 shows the tangential disk for an active grid point. Whenever a particle hits the disk, it contributes to the flux of the corresponding grid point according to (15). As shown it might be necessary to continue the trajectory calculation after finding the intersection with the surface to ensure a proper calculation of the fluxes. In our simulations the particle rays are extended for 3 grid spacings from the intersection point. Then, in case of a neutral particle, for which diffusive re-emission is assumed, the trajectory is continued from the memorized intersection point. The direction is randomized in accordance with diffusive re-emission. The surface normal is obtained from the tri-linear interpolated level set function in the corresponding grid cell.

Since for all active grid points $\vec{p}$, $|\Phi(\vec{p})| \leq \frac{1}{2}$ and for the gradient $\|\nabla\Phi(\vec{p})\| \geq 1$ holds except for some special cases, the distance to the surface is always within a half grid spacing $|d| \leq \frac{1}{2}$. Thus, if the radius is chosen in

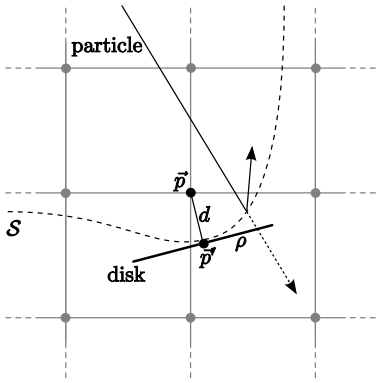


Figure 11: The tangential disk for an active grid point $\vec{p}$. All particles hitting the disk contribute to the local flux of the grid point. Due to the curvature of the surface it can be necessary to continue the trajectory calculation for a couple of grid spacings (dashed) to calculate the flux correctly. However, re-emission takes place at the surface intersection point.

such a manner that

$$\rho \leq \sqrt{1 - \left(\tfrac{1}{2}\right)^2} \approx 0.866, \tag{19}$$

the disk is almost always within the 8 grid cells which are adjacent to the corresponding active grid point. In very rare cases the distance d has to be reduced to fit the disk into these cells. Hence, the same data structure can be used as for the tri-linear interpolation, which requires for each surface cell links to all its corners in order to access the corresponding level set values. Therefore, it is sufficient within a grid cell to check its 8 corners, if they are active and if their corresponding tangential disks are hit by the particle.

The choice of the disk size is a compromise between statistical and spatial accuracy. If the disk size is too large, the calculated fluxes are spatially averaged. In case of disk radii much larger than the grid spacing the spatial resolution of the flux, and consequently that of the surface velocity, might be not sufficient for an accurate time evolution of the surface. Larger disks also intensify the previously mentioned problem at surface regions with larger curvature, resulting in additional errors. Furthermore, if (19) is not satisfied, the disks of much more grid points have to be checked for intersections, which slows down the ray tracing algorithm and also requires additional data structures. Otherwise, if the disk size is too small, only a few particle rays hit the disk leading to a poor statistical accuracy of the fluxes. The statistical errors are inversely proportional to the chosen disk radius. A good choice is a value close to the upper limit in (19). We compared simulations with $\rho = 0.4$ and $\rho = 0.8$, where 4 times more particles are used for the first case to obtain the same statistical accuracy. However, we could not observe an improvement of the simulation result for the smaller radius. Consequently, it does not make sense to further decrease the radius. In our simulations $\rho = 0.8$ is used which seems to be a good choice.

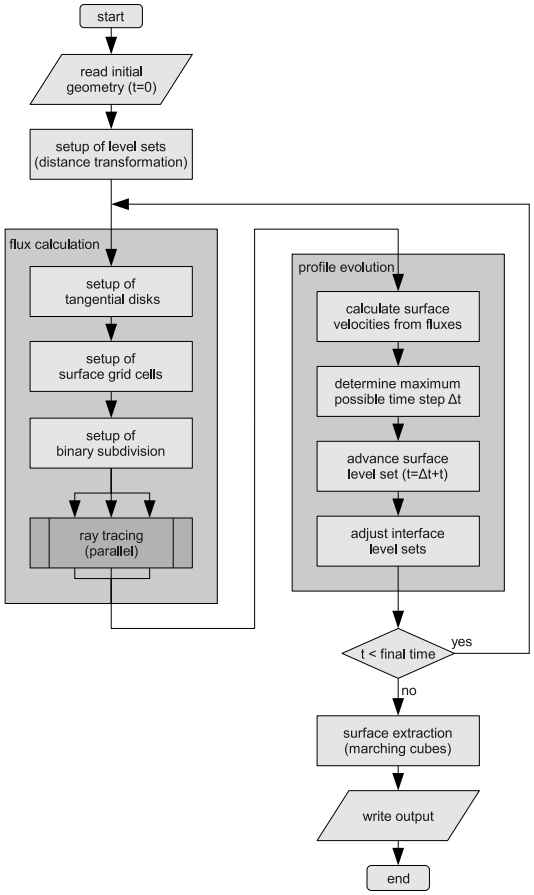


Figure 12: The simulation algorithm.

2.5 Simulation

Assuming that small changes in geometry have only a small impact on the local fluxes, which is also known as pseudo-steady state assumption [16], the flux can be considered constant during the whole time step. Therefore a simulator can simply pass the surface velocities obtained from the calculated fluxes to the profile evolution algorithm.

2.5.1 Algorithm

An overview of the whole algorithm is shown in Fig. 12. After reading the initial geometry a distance transformation initializes the level set functions. Then a loop over the flux calculation and the profile evolution modules is started, until the final time is reached.

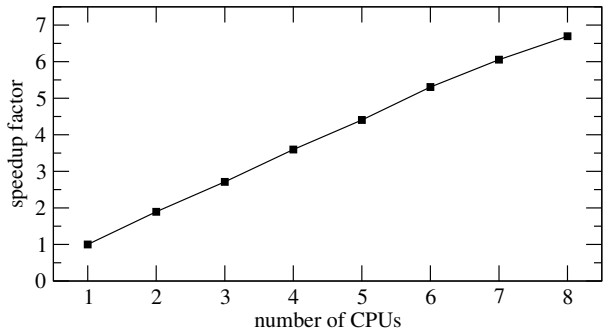


Figure 13: The speedup of ray tracing versus the number of used CPUs.

Within the flux calculation part the tangential disks are set up first. Then all cells are determined which contain parts of the surface or parts of the tangential disks. Links to their corner grid points are stored, since they are necessary for the tri-linear interpolation and for the ray-disk intersection tests. Subsequently, the simulation domain is subdivided into boxes in such a manner that all surface cells represent individual boxes. This additional data structure speeds up the ray tracer which calculates the particle fluxes for all active grid points. Within the profile evolution module the surface velocities are computed from the fluxes. Then the maximum possible time step according to the CFL-condition is determined and used for integrating the level set equation (7) over time. Afterwards the interface level sets Φ_1 and Φ_2 are adjusted accordingly (10).

After the final time is reached, the marching cubes algorithm is applied to extract explicit representations of the surface and the interface level sets, which are used for visualization.

2.5.2 Parallelization

For good statistical accuracy a huge number of particles has to be simulated each time step. Despite the application of fast algorithms, the simulator spends most of the time with ray tracing. To resolve this bottle neck we use parallelization. Since individual trajectories are independent from each other due to ballistic transport, their calculation can be simply distributed among multiple cores. Especially on shared memory architectures, which are getting more and more popular due to the increasing number of processor cores, the parallelization is straightforward using OpenMP [40]. To get for all threads independent streams of random numbers, which are required for ray tracing, we used the Scalable Parallel Random Number Generators Library (SPRNG) [41]. The ray tracing speedup shows a very good scaling with the number of applied CPUs (Fig. 13).

2.6 Results and Discussion

For all in the following presented simulations we use the same parameters, as described in Section Section 2.2 for the passivation and the etching cycle. For all calculations reflective boundary conditions are used for both lateral directions. If not mentioned differently, the grid spacing is 25nm. The radii of the tangential disks are set to 0.8 grid spacings.

2.6.1 Process Time Variations

The effect of different passivation and etching cycle durations is studied on a structure existing of a substrate and a 1μm thick mask, which has a cylindrical hole with diameter 2.5μm. Despite the rotational symmetry this problem can not be straightforwardly reduced to two dimensions. The introduction of cylindrical coordinates leads to non-linear particle trajectories, which makes the determination of the visibilities in the particle transport equation (3) much more difficult. For convex holes, where all points are visible from each other, the solution of the transport equation using cylindrical coordinates was demonstrated in [19]. However, due to the rippled, non-convex side walls of the hole, which evolve during the Bosch process, this method can not be applied.

In three dimensions the simulation domain can be reduced to a quarter due to the reflective boundary conditions and the twofold reflection symmetry of the hole. However, to proof the symmetry of the solution and to avoid reflections to generate our final visualizations the process is simulated on half of the domain, which was discretized using a grid with lateral extensions 140×70. 100 particles for each involved species are launched per grid unit area each time step from the open boundary ($n = 100$). Hence, in total 1.96 million particle trajectories are calculated.

The final profiles after 20 cycles with different process times for deposition (5s and 8s) and etching (11s and 13s) are given in Fig. 14. The results show the influence of the process time on the depths of the holes, tilt angles of the side walls, and the resulting polymer layers. Since also mask etching is incorporated, its final thickness can also be studied. Such simulations can help to find the optimal process parameters.

2.6.2 Lag Effect

Next the influence of the hole diameter on the final profile is investigated. A Bosch process with 6s passivation followed by 12s etching cycles is applied on a 1μm thick

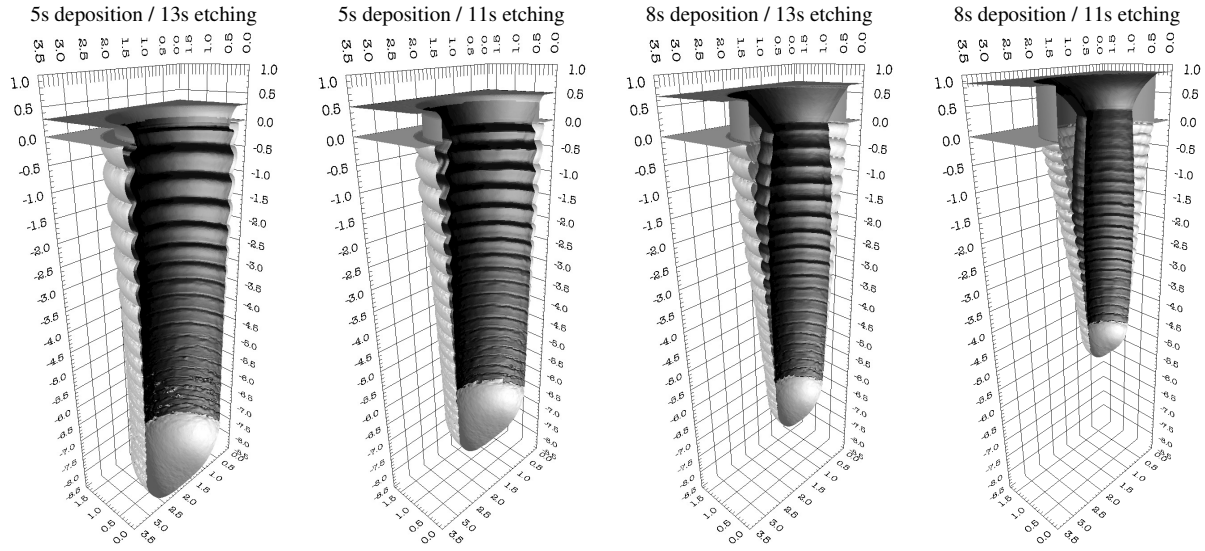


Figure 14: The final profiles after 20 cycles for different combinations of deposition and etching process times. The zero level sets of the functions Φ_1 (light gray), Φ_2 (dark gray), and Φ (black) are visualized. Lengths are given in μm. The grid spacing is 25nm, which corresponds to a grid with lateral extensions 140×70.

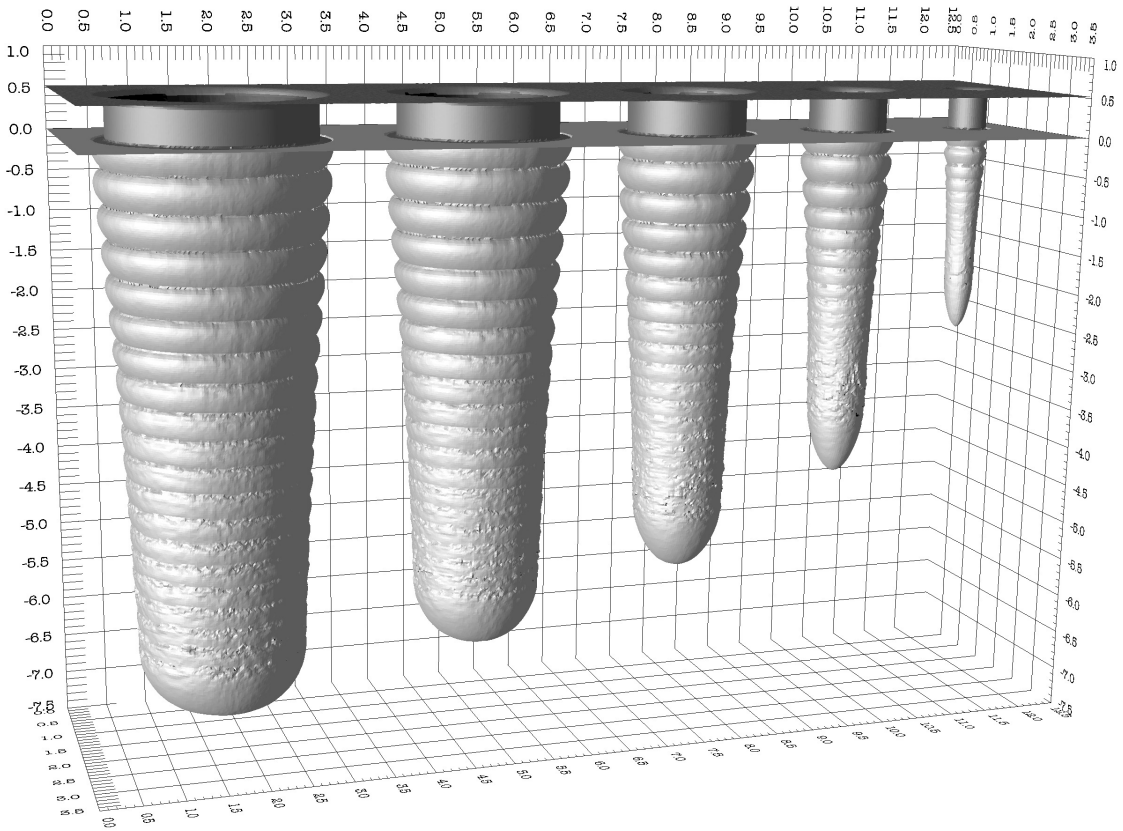


Figure 15: Deep reactive ion etching of holes with varying diameters (2.5μm, 2μm, 1.5μm, 1μm, and 0.5μm). The different depths are a result of the lag effect. The structure is resolved on a grid with lateral extensions 500×140.

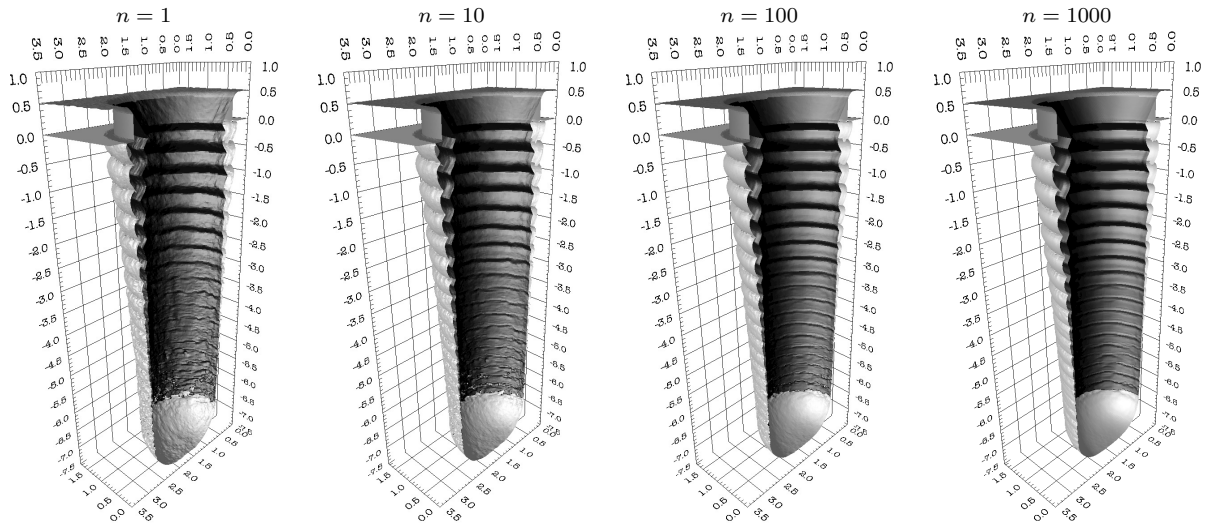


Figure 17: The final profiles for different n. Apart from the roughness of the surface the results are very similar, es over 3 orders of magnitude.

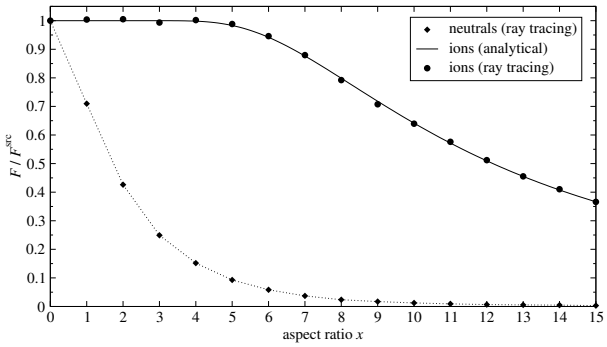


Figure 16: The characteristic dependence of the neutral and ion fluxes at the bottom center on the aspect ratio.

perforated mask with cylindrical holes of varying diameters (2.5μm, 2μm, 1.5μm, 1μm, and 0.5μm).

The simulation domain is resolved on a grid with extensions 500×140 proving the practicability of the applied techniques on larger geometries. Despite this large simulation domain the total memory consumption does not exceed 300MB during the whole simulation thanks to the applied adaptive memory saving data structures.

For each of both species 100 particle trajectories are calculated per grid unit area ($n = 100$), which gives 12.5 millions in total for each time step. Using 8 cores of AMD Opteron 2222 processors (3GHz) the total computation time is about 2 days. 6480 time steps are necessary to simulate all 20 cycles of the Bosch process. The sequential part of the algorithm takes 3.4s and the parallelized ray tracing takes 24s in average. The runtimes increase continuously during the whole simulation due to the increasing depths of the holes and the increasing surface area. However, the runtime of these simulation can be reduced drastically by lowering the accuracy, which is described in the next section.

Fig. 15 shows the final profile after 20 cycles. The different etching depths due to the lag effect can be clearly seen. With increasing aspect ratio the effective etching rate decreases.

To analyze the reason of the lag effect in more detail, the ion and neutral fluxes are calculated at the bottom center of cylindrical holes for various aspect ratios $x = d/2r$. Here d denotes the depth and r the radius of the hole. The ion fluxes obtained by ray tracing are in very good agreement with those calculated analytically (Fig. 16). The analytical expression

$$F_\mathrm{i} = F_\mathrm{i}^\mathrm{src} \left(1 - \left(\frac{2x}{\sqrt{1+4x^2}} \right)^{\kappa+1} \right) \qquad (20)$$

can be derived from (2) by integration over the open solid angle. For the calculation of the neutral flux the sticking coefficient is set to 0.1, which corresponds to the sticking probability of neutrals on the passivation layer as used in our model. The results show that the neutral flux is much more affected by the aspect ratio than the directional ion flux ($\sigma = 2°$). With increasing depth the hole surface area increases, leading to a smaller fraction of particles, which remain sticking at the bottom and not at the sidewalls.

According to our Bosch process model (5) and Table 4 the neutral flux is the main contribution to the deposition rate of the passivation layer. Hence, with increasing aspect ratio the thickness of the deposited passivation layer decreases due to the smaller neutral fluxes. However, the ion flux is not alike reduced. As consequence, the

passivation layer is faster etched through. The ion flux countervails the lag effect, because the substrate is attacked earlier in the etching cycle for larger aspect ratios. However, this head start is more than compensated by the larger substrate etch rate for smaller aspect ratios due to the larger neutral fluxes.

2.6.3 Accuracy vs. Runtime

Inaccuracies of the final profiles are mainly caused by statistical and spatial discretization errors. Both can be reduced at the expense of the runtime.

Statistical errors can be reduced by simulating more particles. Principally, the statistical accuracy of the final profiles strongly depends on the total number of particles, which are calculated during the whole simulation. Therefore, larger time steps also require the simulation of more particles each time step to obtain final profiles of similar quality. Time steps can only be increased by choosing a weaker CFL-criterion, hence allowing larger advancements of the surface, at the expense of the time resolution. However, if the total number of simulated particles is kept constant, the runtime is only marginally reduced, because for good accuracies the simulator spends most of the time with ray tracing anyway. As trade-off we limit the maximum advancement of the surface by 0.1 grid spacings, as already mentioned earlier.

One way to improve the total runtime considerably apart from using more and faster CPUs is to simulate less particles at the expense of the statistical accuracy. The influence of the number of particles n, which are launched per unit area from the top of the simulation domain, on accuracy and runtime is studied on the basis of a 6s/12s Bosch process simulation. Fig. 17 shows the final profiles for $n = 1$, $n = 10$, $n = 100$, and $n = 1000$ after 20 cycles. Interestingly, even for the least accurate case quite good results are obtained. Although the final surface is very rough the qualitative characteristics are maintained. The etched hole is only 3.8% deeper than for $n = 1000$.

From measurements of the total runtime T for different n and different number of CPUs we obtain the relation

$$T \approx N_{\text{steps}} \cdot (n/p \cdot 0.31\text{s} + 0.52\text{s}) \,. \qquad (21)$$

Here p denotes the speedup due to parallelization as given in Fig. 13. Fig. 18 shows plots of T and the number of time steps N_{steps}, which are necessary for the whole simulation. The times are referred to 4.8h, which is the runtime for $n = 1$ on a single CPU. For small n the number of time steps increases due to the poorer statistics. The expected maximum surface velocity of all

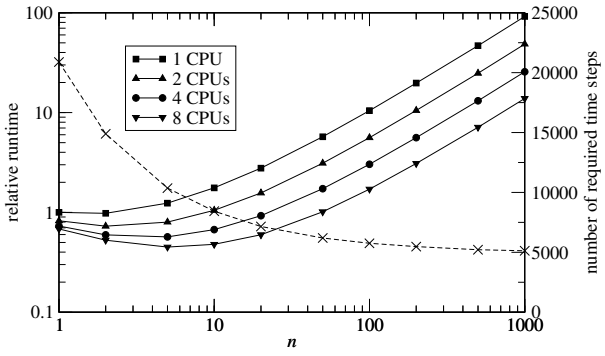


Figure 18: The relative runtimes for different accuracies n and different numbers of used CPUs. In addition the number of required time steps is also given (dashed).

grid points is larger for smaller n due to the larger variation of the velocity distribution. Hence, according to the CFL-criterion the maximum allowed time step must be smaller, leading to more time steps. This on the other hand side implies that the total number of simulated particles during the whole calculation increases, which improves the accuracy again. Consequently, a minimum exists for the total runtime as can be seen in Fig. 18. Choosing a value for n smaller than that of the minimum is not favorable.

Another way to speed up the calculation is the usage of coarser spatial discretizations. In Fig. 19 the simulation results are compared for $n = 100$ and grid spacings of 70nm and 14nm, which correspond to grids with lateral extensions 50×25 and 250×125, respectively. The runtime between both calculations differs by a factor more than $5^3 = 125$. This comes from the fact that the number of surface discretization elements N scales inversely quadratically with the grid spacing. Furthermore, due to the CFL-condition, the number of required time steps scales inversely with the grid spacing.

Despite the large difference in computation time the results look again very similar. Due to the sub-time step resolution of the material dependent etch rates within our multi-level-set framework [28] the error introduced by the coarse grid is kept small. The final depths differ only by 2.1%. However, the coarse grid is not able to represent the rippled sidewalls accurately.

2.7 Conclusion

We applied modern techniques like the sparse field level set method and the HRLE data structure for the profile evolution as well as ray tracing algorithms for the flux calculation to three-dimensional simulations of the Bosch process. The presented multi-level-set approach

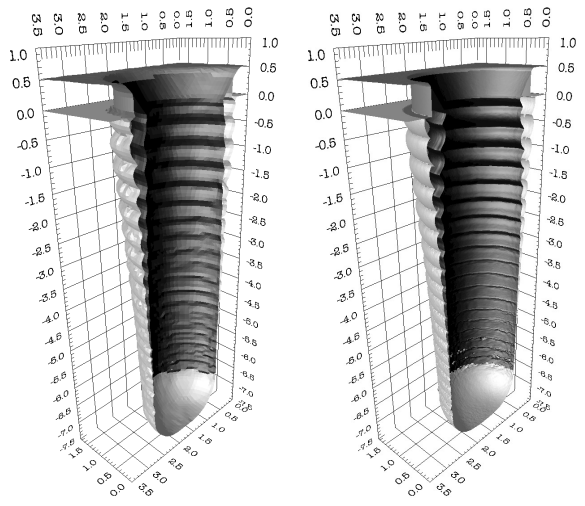


Figure 19: The final results of calculations on grids with grid spacing 70nm (left) and 14nm (right).

allows an accurate, robust, and memory efficient handling of different material regions, including thin layers.

Due to its simple parallelization and due to the increasing number of processor cores, ray tracing becomes an alternative to common direct integration methods for particle transport calculation.

Although the Monte Carlo method was only demonstrated with a relative simple model, it is capable to solve more complex ones, where for example energy dependent sputter rates or specular reflexions of ions are incorporated. In contrast to direct integration methods such effects can be straightforwardly implemented without increasing the algorithmic complexity. A three-dimensional simulation of a more complex reactive ion etching model was already demonstrated in [42].

At the expense of accuracy the whole calculation can be drastically accelerated by reducing the number of simulated particles or the grid resolution. The results still reflect the qualitative characteristics. Therefore, the Monte Carlo method is useful for the fast examination of parameter variations. After finding the optimized set of parameters a final simulation can be carried out to get an appropriately accurate profile.

For future work it may be interesting to incorporate mask charging effects [43], where the emerging electric field leads to non-linear trajectories of ions, which complicates ray tracing.

3 The Effect of Copper Grain Size Statistics on the Electromigration Lifetime Distribution

We investigate the influence of the statistical distribution of copper grain sizes on the electromigration time-to-failure distribution. We have applied a continuum multi-physics electromigration model which incorporates the effects of grain boundaries for stress build-up. The peak of tensile stress develops at the intersection of copper grain boundaries with the capping layer. It is shown that the electromigration lifetimes follow lognormal distributions. Moreover, the increase of the standard deviation of the grain size distribution results in an increase of the electromigration lifetimes standard deviation. The results strongly imply that the lognormal distribution of the grain sizes is a primary cause for the lognormal distribution of electromigration lifetimes.

3.1 Introduction

The continuous shrinking of the dimensions of on-chip interconnects and the introduction of advanced backend-of-line (BEoL) manufacturing process steps increases the complexity of physical phenomena behind electromigration failure. The total wiring length amounts to kilometers arranged in several levels of metallization with millions of interlevel connections. The tendency of modern technologies to increase the interconnect length and, at the same time, to reduce the cross section, makes the interconnect structures more and more susceptible to electromigration. Currently, integrated circuits are often designed using simple and conservative design rules to ensure that the resulting circuits meet reliability goals. However, this precaution leads to reduced performance for a given circuit and metallization technology.

Electromigration data have been described by lognormal distributions [44]. Although the origin of the lognormal distribution of electromigration lifetimes is not entirely clear, it has been argued that the diffusion process in connection with the effect of microstructure on electromigration provides the basis for the lognormal distribution [45]. In copper dual-damascene interconnects the main diffusivity path is along the copper/capping layer interface. This interfacial diffusion is affected by the orientation of the grains. As the copper grain sizes seem to follow lognormal distributions in typical dual-damascene process technology [45] and due to the influence of microstructure on the electromigration process, the lognormal distribution has been used as the underlying statistics for electromigration lifetimes. However, it has been discussed whether this choice is the most appropriate [46, 47]. The understanding of the electromigration lifetime distribution is crucial for the extrapolation of the times to failure obtained empirically from accelerated tests to real operating conditions, as performed by a modified form of the Black equation [44].

Also, it has been shown that the microstructure plays a key role regarding the failure mechanisms in copper dual-damascene interconnects [48]. It affects electromigration in different ways. Grain boundaries are natural locations of atomic flux divergence, they act as fast diffusivity paths for vacancy diffusion [49], and they act as sites of annihilation and production of vacancies [50].

The main challenge in electromigration modeling is the diversity of physical phenomena which have to be taken into account for an adequate description of the problem. Electromigration transport is also accompanied by material transport driven by the gradients of material concentration, mechanical stress, and temperature. Furthermore, taking into account the effects of interfaces and grain boundaries as fast diffusivity paths imposes new challenges for electromigration modeling.

In this work we investigate the origin of the statistical distribution of electromigration times to failure as a function of the distribution of copper grain sizes. The effect of lognormal grain size distributions on the distribution of electromigration lifetimes of fully three-dimensional copper dual damascene interconnect structures is studied based on numerical simulations. We have applied a continuum multi-physics electromigration model which incorporates the effects of grain boundaries for stress build-up. Moreover, we have developed a tool to include the microstructure into the simulations based on a given statistical distribution of grains sizes.

3.2 Electromigration Modeling

Several driving forces are responsible for the vacancy transport in a conductor line under electromigration. The combination of these driving forces leads to the total vacancy flux given by

$$\vec{J}_v = -D_v\left(\nabla C_v + \frac{|Z^*e|}{k_BT}C_v\nabla\varphi + \frac{f\Omega}{k_BT}C_v\nabla\sigma\right), \quad (22)$$

where D_v is the vacancy diffusion coefficient of the dominant transport path, C_v is the vacancy concentration, Z^*e is the effective charge, f is the vacancy relaxation ratio, Ω is the atomic volume, σ is the hydrostatic stress, k_B is Boltzmann's constant, and T is the temperature.

Vacancies accumulate or vanish in sites of flux divergence, and this dynamics is described by the continuity equation

$$\frac{\partial C_v}{\partial t} = -\nabla \cdot \vec{J}_v + G(C_v), \tag{23}$$

where $G(C_v)$ is the source function which models vacancy generation and annihilation processes [51]. The source term plays a major role for the mechanical stress buildup and is taken into account only at interfaces and grain boundaries. It comprises three processes, namely, the exchange of point defects between adjacent grains, the exchange of point defects between grains and grain boundaries, and the formation/annihilation of point defects at grain boundaries.

In our model grain boundaries are treated as separate regions which can trap or release vacancies [52], as shown in Fig. 20. We denote the vacancy concentration from both sides of the grain boundary as C_v^1 and C_v^2, respectively, and the concentration of immobile vacancies which are trapped inside the grain boundary as C_v^{im}.

The trapping rate of vacancies at the grain boundary, which corresponds to the generation/recombination rate, is controlled by the atomic fluxes J_v^1 and J_v^2, yielding [52]

$$\frac{\partial C_v^{im}}{\partial t} = G = \frac{1}{\tau}\Big[C_v^{eq} - C_v^{im}\Big(1 + \frac{2\,\omega_R}{\omega_T(C_v^1 + C_v^2)}\Big)\Big], \tag{24}$$

where ω_T is the trapping rate of vacancies from both neighboring grains, ω_R is the release rate, and C_v^{eq} is the equilibrium vacancy concentration inside the grain boundary, given by

$$C_v^{eq} = C_v^0 \exp\Big(\frac{\sigma_{nn}\Omega}{k_B\,T}\Big), \tag{25}$$

where C_v^0 is the equilibrium vacancy concentration in the absence of stress and σ_{nn} is the stress component normal to the grain boundary. In (24) τ represents the vacancy relaxation time which characterizes the efficiency of the grain boundary as vacancy sink/source [52]

$$\frac{1}{\tau} = \frac{\omega_T(C_v^1 + C_v^2)}{\delta}. \tag{26}$$

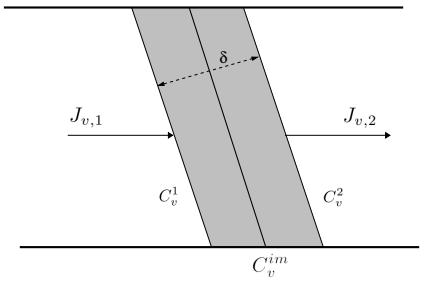


Figure 20: Grain boundary model.

Sarychev *et al.* [53] introduced the contribution of vacancy migration and generation/annihilation processes for stress build-up in a three-dimensional model of stress evolution during electromigration. Considering the grain boundary model we have proposed that the strain growth from both sides of the grain boundary is proportional to the growth rate of immobile vacancies

$$\frac{\partial \varepsilon_{kk}}{\partial t} = \Omega\Big[(1-f)\nabla \cdot \vec{J}_v + f\frac{\partial C_v^{im}}{\partial t}\Big], \tag{27}$$

where ϵ_{kk} is the trace of the strain vector.

Equation (27) shows that vacancies trapped at the grain boundaries are responsible for build-up of tensile stress. When the grain boundaries are able to capture large amounts of vacancies, a high tensile stress develops.

The system of equations formed by (22)–(27) is solved until a stress threshold (σ_{th}) for void nucleation is reached at an intersection of grain boundaries with the capping layer. These intersections are considered sites of weak adhesion and, consequently, most susceptible to void nucleation [54].

Gleixner *et al.* [55] showed that the stress threshold is given by

$$\sigma_{th} = \frac{2\gamma_s \sin\theta_c}{R_p}, \tag{28}$$

where R_p is the radius of the adhesion-free patch, θ_c is the contact angle between the void and the surface, and γ_s is the surface energy.

3.3 Simulation Approach

Equations (22)–(27) are solved using the finite element method (FEM) until the stress threshold for void nucleation is reached at some weak adhesion point. We consider the intersection of grain boundaries with the copper/capping layer interface as natural places of weak adhesion [54]. As grain boundaries and interfaces act as fast diffusivity paths, the diffusion coefficient in (22) has to be adapted for these regions. We have used $D_v^{gb} = 10^4 D_v^{bulk}$ for grain boundaries and $D_v^{Cu-cap} = 10^5 D_v^{bulk}$ for the copper/capping layer interface [56]. It should be pointed out that all model parameters are equal for all grains and all simulated structures. Grain boundaries, and generally, material interfaces of the geometry have to be supplied with an appropriately fine FEM mesh. This is necessary in order to provide sufficient resolution for the local dynamics described by the proposed model.

In order to include the grain distribution into the numerical simulations, a microstructure generator tool has been

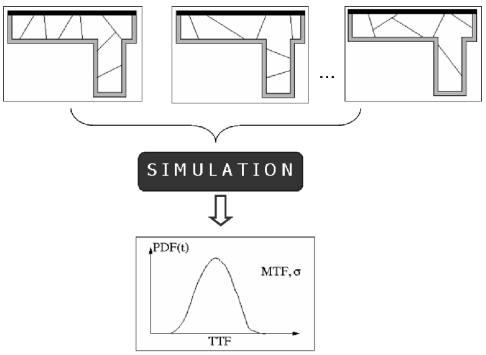


Figure 21: Schematic simulation procedure.

developed. Given a specific interconnect structure and providing the tool with a median grain size and corresponding standard deviation, it generates a lognormal distribution of grain sizes. Then, following this distribution, the interconnect line is cut along its length by the planes that form the grain boundaries. Furthermore, the angles between the grain boundaries' planes and the line surface follow a normal distribution with median value of 90 °C. The corresponding standard deviation can also be specified.

In Fig. 21 we present the schema of the simulation procedure. Three standard deviations for the distribution of grain sizes are considered, namely 0.1, 0.3 and 0.6. For each of them 20 dual-damascene interconnect structures were created with the microstructure generator. As the interconnect line is assumed to present a bamboo-like structure, the median grain size is equal to the line width, 0.10 μm. The barrier, capping and interlayer dielectric layers are Ta, SiN, and SiO_2, respectively. The corresponding interconnect structure is shown in Fig. 22.

The applied current density is 1.5 MA/cm^2, and the temperature is 300 °C. We have used a stress threshold value as failure criterion, which means that the electromigration time to failure represents the time for a void nucleation to occur. Thus, the time to failure is determined by the time for the stress to reach a given threshold value at some intersection between a grain boundary and the SiN layer.

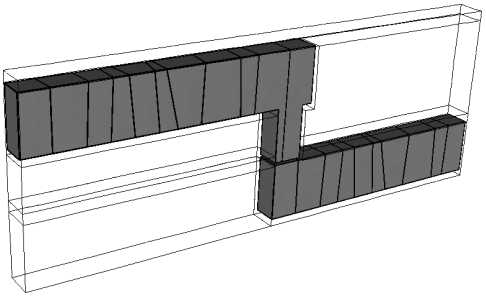


Figure 22: Dual-damscene interconnect structure.

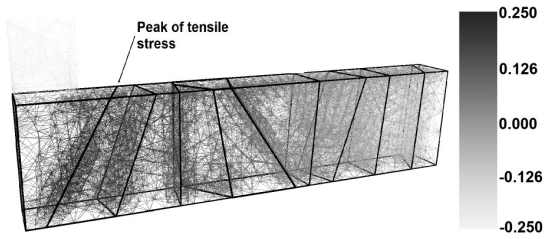


Figure 23: Hydrostatic stress distribution in a simulated interconnect (in MPa). The peak value is located at grain boundaries, where vacancies are trapped.

3.4 Discussion

Fig. 24 shows the hydrostatic tensile stress development for the structures with grain size standard deviation of 0.3. The stress peak value follows the peak of trapped vacancy concentration and is located at the intersection of grain boundaries with the capping layer, as shown by Fig. 23.

Collecting the times to failure from Fig. 24 and calculating the cumulative failure percentages resulted in the distributions of electromigration lifetimes shown in Fig. 25. The lifetimes are fitted by lognormal distributions. The obtained standard deviations are 0.0065, 0.0080, and 0.0085 for the grain size distributions with standard deviations of 0.1, 0.3, and 0.6, respectively. The standard deviation for a lognormal distribution is given by

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(ln\ TTF_i - ln\ MTF)^2}. \qquad (29)$$

where TTF_i is the time to failure of the i-th test structure, N is the number of test structures, and MTF is the mean time to failure of the lognormal distribution

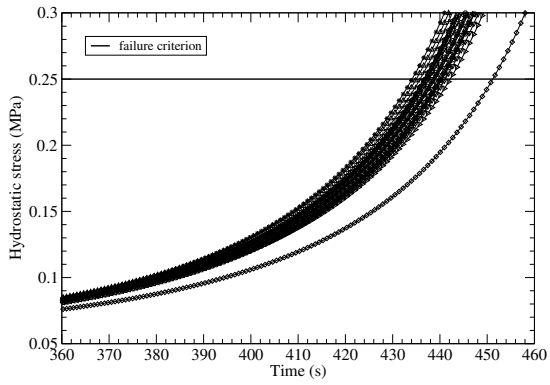
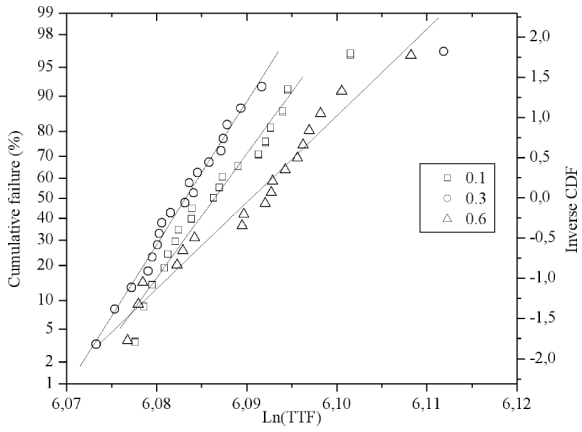


Figure 24: Peak of hydrostatic stress development for the set with grain size standard deviation of 0.3.

$$ln\ MTF = \frac{1}{N}\sum_{i=1}^{N} ln\ TTF_i, \tag{30}$$



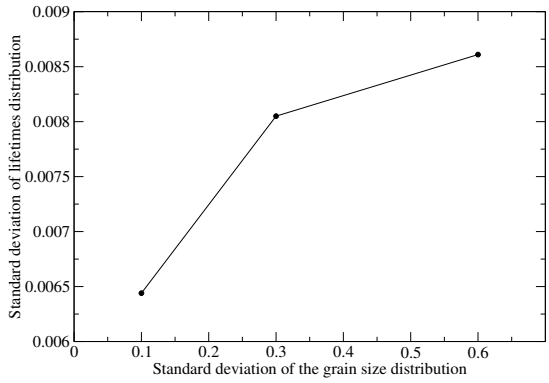


Figure 26: Electromigration lifetime standard deviation for different standard deviations of grain size.

The standard deviations for the electromigration lifetimes are very small compared to those frequently observed in experiments [44]. Several factors can explain this behavior. First, for convenience, we have used a small value of stress threshold as failure criterion to determine the interconnect lifetime. As can be seen from Fig. 24, the variation of the lifetimes can be more pronounced for higher stress thresholds. Second, the simulation parameters and material properties are independent of the grain distribution. This means that mechanical properties and diffusivities, for example, are equal and constant for all grains in an interconnect line, for all simulated structures. This is clearly not the case in real experiments, as it is well known that material properties vary according to the grain orientation. It is expected that atomic diffusion along the copper/capping layer interface changes from grain to grain, inducing a flux divergence at the corresponding grain boundary. Moreover, the diffusivities are different from line to line as the grain distribution varies. Therefore, given the simplifications we have made, the small standard deviations obtained from our simulations should be expected.

Nevertheless, our results show that the grain distribution still affects the electromigration lifetime distribution.When the grain size distribution exhibits a smaller standard deviation the corresponding interconnect lines have a more uniform distribution of the grains. As a consequence, the stress build-up has smaller variations yielding a smaller standard deviation of the electromigration lifetimes. On the other hand, increasing the grain size standard deviation, the lines exhibit significant differences in the grain structures. This leads to increased variations for the stress development. Thus, a bigger standard deviation of electromigration lifetimes is expected. This behavior is presented in Fig. 26. It shows that the increase of the standard deviation of the distribution of grains sizes increases the standard deviation of the electromigration lifetime distribution.

3.5 Conclusion

We analyzed the electromigration failure development in typical copper dual-damascene interconnect structures based on numerical simulations. A continuum electromigration model which describes mechanical stress build-up in connection with the microstructure effect was applied. We observed that the peak of tensile stress is located at the intersection of grain boundaries with the capping layer, following the peak of trapped vacancy concentration. This shows that the microstructure has a decisive impact on the determination of void nucleation sites. The simulation results indicate that the lognormal distribution of the copper grain sizes is a primary cause for the lognormal distributions of the electromigration lifetimes. Moreover, an increase of the standard deviation of the grain size distribution leads to an increase of the electromigration time-to-failure distribution.

4 Possible Correlation Between Flicker Noise and Bias Temperature Stress

A link between Bias Temperature Stress (BTS, NBTI) and flicker noise ($1/f$-noise) is explored by comparing flicker noise data to charge pumping data. Large-area devices are shown to initially have very low, bias independent normalized flicker noise. After BTS the normalized noise increases considerably and becomes gate bias dependent. Small-area devices are shown to exhibit bias dependent burst noise (RTS) in addition to flicker noise, regardless of BTS.

4.1 Introduction

When subjected to strong-inversion bias and high temperatures, the drain current of MOSFETs degrades, a phenomenon known as *Bias Temperature stress (BTS)*. The drain current degradation is often described as an increase of the threshold voltage, but other parameters, foremost the carrier mobility and the sub-threshold slope, degrade as well. The exact physical mechanism responsible for BTS are still controversial, but there is ample evidence that both interface states, possibly created by breaking the bonds of passivating hydrogen [57], and oxide traps play a role [58].

Since flicker noise has been used as a diagnostic tool in various places before [59, 60], we conducted a series of flicker noise measurements on MOSFETs that previously experienced BTS degradation. To assess the amount of degradation, the increase in interface state density was monitored using charge pumping measurements [61].

4.2 Methodology

The devices measured were pMOSFETs with $W/L = 50\,[\mu\mathrm{m}]/10\,[\mu\mathrm{m}]$ and $t_{ox} = 30\,[\mathrm{nm}]$. We studied three wafers that differed only in the back-end-of-line processing, described in [62]. One process variant resulted in a high initial interface trap density N_{it0}, but showed a comparatively low increase ΔN_{it} after BTS; this wafer is referred to as wafer A. Wafer B had both medium initial interface traps and medium increase of traps after BTS. The third wafer (C) exhibited a high ΔN_{it}, resulting in the highest post-stress interface state density, despite the fact that this wafer's initial interface state density was lowest. On every wafer at least two neighbouring devices were measured using constant-baselevel charge pumping at $2\,[\mathrm{MHz}]$ [63]. Next, on every wafer one device was stressed for $e3[\mathrm{s}]$ at $V_{gs} = -17.5\,[\mathrm{V}]$ and $175\,[^\circ\mathrm{C}]$, and charge pumping was done again immediately upon release of stress.

Then, noise measurements were performed in the linear region of the MOSFETs ($|V_{ds}| < 0.3[\mathrm{V}]$) at gate voltages $V_{gs} = -1.54\,[\mathrm{V}], -3.07\,[\mathrm{V}], -4.59\,[\mathrm{V}]$. Fig. 27 depicts the spectra for a fresh and a stressed device at weak inversion. Although the bias point is approximately the same, the noise power density is tenfold for the stressed device. Prior to the noise measurements, the $I_d(V_{gs})$-characteristic of a fresh device was measured, and the SPICE level-1 model was fitted yielding the parameters $V_t = -0.95\,[\mathrm{V}]$, $\beta = 1.23e-4\,[\mathrm{A/V^2}]$, and $\theta = 0.128\,[\mathrm{V^{-1}}]$.

In addition to the large-area transistors, small-area transistors with $W/L = 2.4\,[\mu\mathrm{m}]/2.6[\mu\mathrm{m}]$ were examined. Fig. 28 shows that with these devices the noise is not conveniently described by a pure $1/f$-dependence. Because of the smaller number of free carriers the Lorentzians of distinct traps may be visible, and their superposition yields $1/f^\gamma$-noise with γ appreciably deviating from unity, as predicted by the criterion $N < 1/(4\pi\alpha)$ in [64].

Using the empirical relation [65]

$$\left.\frac{S_{V_{ds}}}{V_{ds}^2}\right|_{I_d=\mathrm{const}} = \left.\frac{S_{I_d}}{I_d^2}\right|_{V_{ds}=\mathrm{const}} = \frac{S_{r_{ds}}}{r_{ds}^2} = \frac{\alpha}{Nf}\,, \tag{31}$$

an α was calculated for every device at every bias by taking $f_0 S_{V_{ds}}(f_0)/V_{ds}^2$ at $f_0 = 10\,[\mathrm{Hz}]$. Care was taken to verify a $1/f$-dependence of $S_{V_{ds}}$ around f_0, which was the case for all large-area transistors. Assuming a homogeneous channel, the number of carriers in the channel N was obtained via $N = L^2/(\mu q r_{ds})$, where r_{ds} is the (measured) channel resistance at $V_{ds} \approx 0$, q the elementary charge, and the carrier mobility μ was calculated from the SPICE model parameter β.

4.3 Results

The fresh devices showed very low α values around 10^{-6} that were only weakly dependent on the gate voltage. The values were quite similar for all three wafers. The stressed devices exhibited considerably higher noise power, corresponding to higher α values, that moreover turned out to be bias dependent: For weak inversion, α of the stressed devices was up to ten times the value of the fresh ones, where at strong inversion fresh and stressed devices had comparable α values, cf. Fig. 29.

Since the conductivity is proportional to the product of carrier number N and carrier mobility μ, assuming that both N and μ fluctuate *independently* allows to split

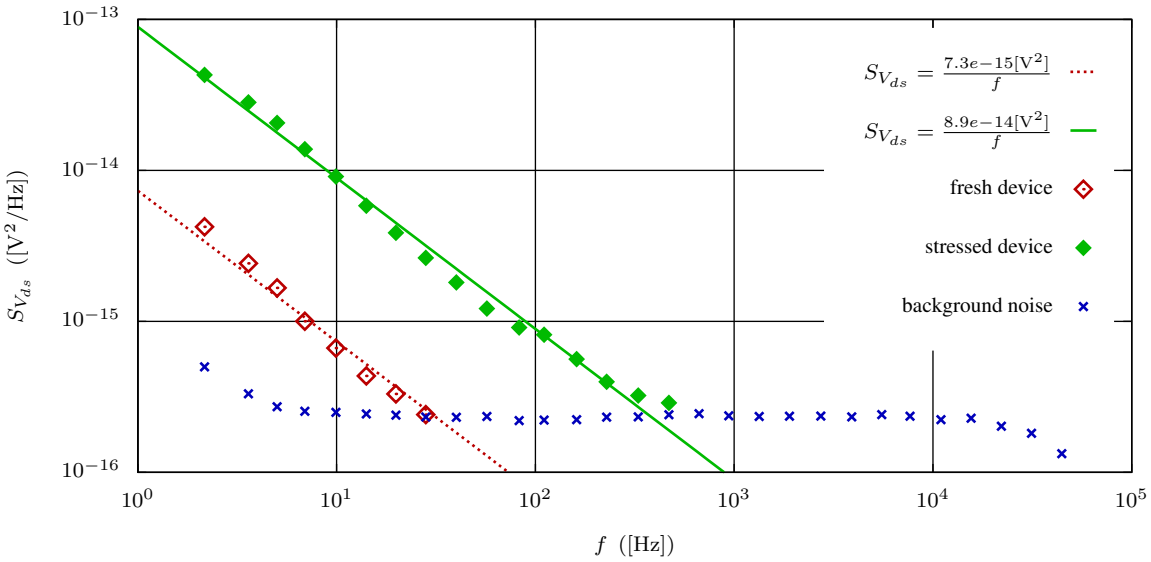


Figure 27: Noise spectrum of a large-area MOSFET (wafer C) biased at $V_{gs} = -1.54\,[\mathrm{V}]$ before and after BTS and respective least-squares fits. The crosses show the background noise ($I_d = 0$).

$\alpha = \alpha_\mu + \alpha_N$. In a first order approximation, mobility reduction and parasitic resistances are negligible, hence α_μ is *independent of gate bias*. It seems likely that the unstressed (fresh) devices just show this kind of flicker noise, i.e. $\alpha_f = \alpha_\mu$, which is a bulk noise effect. Continuing with this interpretation, for stressed devices $\alpha_N = \alpha - \alpha_\mu = \alpha_s - \alpha_f = \Delta\alpha$ is the flicker noise component due to carrier number fluctuations. According to the McWhorter theory, $\alpha_N \propto 1/V_g^*$, as nicely confirmed by Fig. 29.

4.4 Conclusion

The low-frequency noise behaviour of large-area pMOSFETs subjected to bias temperature stress was investigated. Unstressed devices showed very low and gate bias independent α values around origin for the flicker noise in these devices. After BTS, the devices showed considerably increased, gate bias dependent α values, indicating that in competition with the bulk mobility noise, a surface-provoked noise component of the McWhorter type emerges.

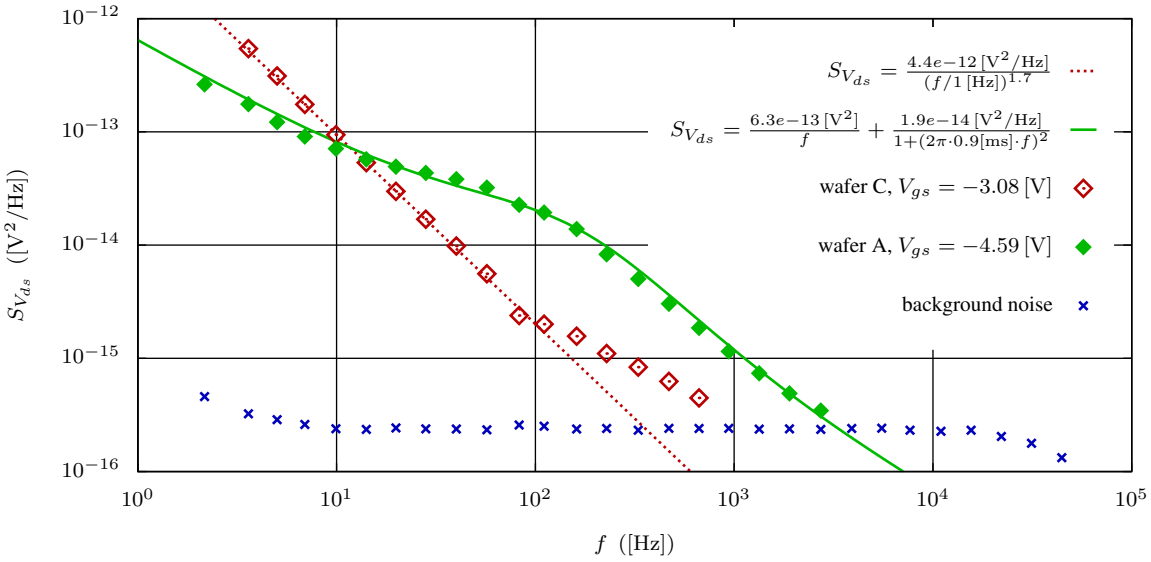


Figure 28: Noise spectra of small-area MOSFETs (unstressed) for different biases. The spectra clearly deviate from the $1/f$-form: One can partly be fitted with a much higher slope, thus more resembling a $1/f^2$-spectrum. The other one can be fitted by a superposition of a 'true' $1/f$-component and a Lorentzian. This situation is characteristic for the presence of a single dominant trap. These traps were also visible in the time domain.

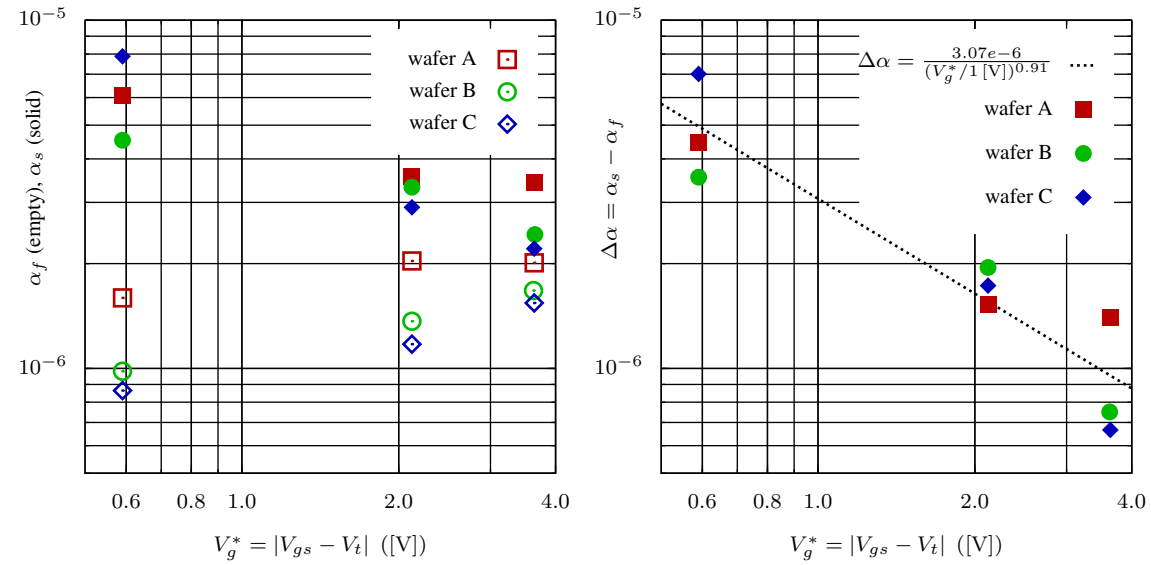


Figure 29: *Left:* Dependence of calculated α values on the effective gate voltage (empty symbols: fresh devices; solid symbols: stressed devices). *Right:* Increase of α due to BTS, and least squares fit to the data, indicating $\Delta\alpha \propto 1/V_g^*$.

5 Modeling of Low Concentrated Buffer DNA Detection with Suspend Gate Field-Effect Transistors (SGFET)

The experimental data of a suspend gate field-effect transistor (SGFET) have been analyzed with three different models. A SGFET is a MOSFET with an elevated gate and an empty space below it. The exposed gate-oxide layer is biofunctionalized with single stranded DNA, which is able to hybridize with a complementary strand. Due to the intrinsic charge of the phosphate groups (minus one elementary charge per group) of the DNA, large shifts in the transfer characteristics are induced. Thus label-free, time-resolved, and in-situ detection of DNA is possible. It can be shown that for buffer concentrations below mmol/l the Poisson-Boltzmann description it is not valid anymore. Because of the low number of counter ions at small buffer concentrations, the screening of the oligo-deoxynucleotides/DNA is more appropriately described with the Debye-Hückel model. Additionally we propose an extended Poisson-Boltzmann model which takes the closest possible ion distance to the oxide surface into account, and we compare the analytical soultion of this model with the Poisson-Boltzmann and the Debye-Hückel model.

5.1 Introduction

The need for fast, cheap, reliable, and in-situ detection of DNA, antibody, protein and tumor markers, also known as “point of care” applications, requires new technological approaches. Today the detection of DNA needs several complex and time consuming steps, like amplification of the DNA by polymerase chain reaction (PCR) or reverse transcription (RT), followed by a procedure to add certain molecules which are able to fluroescent or radiate (called labeling), and at last an optical read out of the experimental data with a microarray reader [66, 67].

One promissing approach is to exchange the optical detection mechanism by an electrically working principle [68–74]. The field-effect based approach has several advantages over the optical method. The application of a field-effect transistor eases the integration of amplifying and analyzing circuits on the same chip, thus reducing the costs for the read-out equipment. Additionally, the use of semiconductor process technology enables mass production and a corresponding huge decrease in price per device. In this work the experimental data of a biosensor for detecting DNA are studied via three different models. The biosensor is a suspend gate field-effect transistor (SGFET). This device is a MOSFET with a raised gate and an empty space beneath it (see Fig. 30). Within this empty gap the gate-oxide layer is chemically modified with single stranded DNA which is able to hybridize with a complementary strand. Due to the intrinsic charge of the phosphate groups (minus one elementary charge per group) of the DNA large shifts in the transfer characteristics are induced. Thus label-free, time-resolved, and in situ detection of DNA is possible. Interestingly the commonly used Poisson-Boltzmann models are not able to reproduce the experimental data, while the Debye-Hückel model [75] works, although its validity in the used regime is questionable.

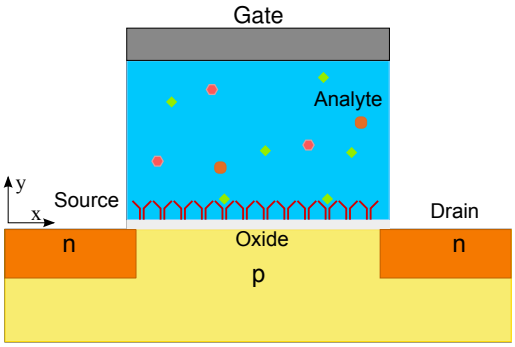


Figure 30: Scheme of suspend gate field-effect transistor.

Finally we introduce an extended Poisson-Boltzmann formulation which takes the closest possible approach between ions into account. In an aqueous solution the salt ions are covered with water molecules. Due to the thereby increased effective ion radius there is a minimum distance between the ions and the oxide surface, called outer Helmholtz plane (OHP). Within this OHP there is no screening.

5.2 Experimental Data

In the work of Harnois [76] 60 oligo-deoxynucleotides (ODN), also known as single stranded DNA, were attached onto a glutaraldehyd coated nitride layer. Then one test run with mismatched ODNs and one test run with matching ODNs were carried out. The runs with the mismatching DNA show no relevant change in the output curves, while for the matching single stranded DNA a big shift in the threshold voltage becomes visible. The results show two interesting properties. Firstly, a threshold voltage shift of about $800\,\mathrm{mV}$ between the probe curve and the target transfer curve and, secondly, the probe transfer curve lies in the middle between target and reference. The average threshold voltage shift is in a range from several mV to $100\,\mathrm{mV}$ [77], depending on the buffer concentration, the $800\mathrm{mV}$ shift is quite big and the Poisson-Boltzmannregime shows a big shift between ref-

erence and probe/target ($\sim 100\,\mathrm{mV}$), but a much smaller shift between probe and target ($10 - 20\,\mathrm{mV}$) [67].

5.3 Simulation

First a Poisson-Boltzmannmodel was utilized which treats the buffer as continuous ion concentrations weighted with Boltzmann type terms ($e^{-\frac{qV}{k_B T}}$) (Fig. 32), combined with a space charge density that corresponds to 60 base pairs (probe) and 120 base pairs (target).

$$\epsilon_0 \nabla \cdot (\epsilon_{Ana} \nabla \psi(x,y)) = -\sum_{\xi \in S} \xi\, q\, c_\xi^\infty e^{-\xi \frac{q}{k_B T}(\psi(x,y)-\psi_\mu)} + \rho_{Space}(x,y) \tag{32}$$

k_B denotes Boltzmann's constant, T the temperature in Kelvin, and $\xi \in S$, where S contains the valences of the ions in the electrolyte. ϵ_0 describes the permittivity of vacuum, and q denotes the elementary charge. ψ_μ is the chemical potential. c_σ^∞ is the ion concentration in equilibrium, while $\epsilon_{Ana} \approx 80$ is the relative permittivity of water. The second model also uses the Poisson-Boltzmann description but assumes a sheet charge density at the oxide-analyte interface (Fig. 33).

$$\epsilon_0 \nabla \cdot (\epsilon_{Ana} \nabla \psi(x,y)) = -\sum_{\xi \in S} \xi\, q\, c_\xi^\infty e^{-\xi \frac{q}{k_B T}(\psi(x,y)-\psi_\mu)} + \sigma_{Sheet}(x)\, \delta(y - y_0) \tag{33}$$

The third model uses the Debye-Hückel formulation which can be derived by linearizing the Poisson-Boltzmann model (Fig. 34).

$$\epsilon_0 \nabla \cdot (\epsilon_{Ana} \nabla \psi(x,y)) = \frac{2q^2}{k_B T}(\psi(x,y) - \psi_\mu) \sum_{\xi \in S} \xi^2 c_\xi^\infty + \rho_{Sace}(x,y) \tag{34}$$

5.4 Discussion

Fig. 32, Fig. 33, and Fig. 34 show the transfer characteristics for the unprepared SGFET (reference), the prepared but unbound (probe), and when the DNA has bound to functionalized surface (target), respectively. For better comparison between experimental data and our simulation, the curves of the experiment are in discrete grey tones included. As can be seen for Fig. 32 and Fig. 33, even for the low salt concentration of $0.6\,\mathrm{mmol}$, the shift between the reference curve and the probe/target is bigger than between the probe and target curves. This behavior complies with the observations by [67] and is attributed to the nonlinear screening of the used models. Looking at Fig. 35 and Fig. 36 shows that doubling the charge at the interface does not lead to a doubled potential shift. Nevertheless there is a bigger shift for the sheet charge model due to the description of the DNA charge as sheet with infinite small height. Therefore less screening compared to the space charge model that distributes the same amount of charge over $20\,\mathrm{nm}$ takes place.

However, just by decreasing the salt concentration it is impossible to fit the experimental data. Nevertheless the Debye-Hückel model shows acceptable agreement with the experimental data for the same parameters as in the Poisson-Boltzmann description (Fig. 34). Here, doubling the amount of charge leads to twice the potential shift Fig. 37 because of the linear screening term in the model (34).

In order to understand why the Poisson-Boltzmann model fails and the Debye-Hückel model works, one hast to look for the validity constraints of the used models. For instance, assuming a volume of $10 \cdot 10 \cdot 20\,\mathrm{nm}^3$ for a single 60 bases DNA strand and one mmol sodium-chloride bulk concentration leads to an average concentration of about one sodium/chlorine atom within this given volume. So there will be no strong nonlinear screening in this regime. The Poisson-Boltzmann model treats the salt concentration as continuous quantity, so it is overestimating the screening and therefore is not valid for small salt concentrations.

The Debye-Hückel model can be derived by expanding the exponential terms into a Taylor series and neglecting all terms higher than second order [75]. According to the laws of series expansion $\frac{q\Psi}{k_B T} \ll 1$ and thus the potential has to be small compared to the thermal energy. By treating the ions as infinite small point charges, the mean distance between the ions in the solution must be big and therefore the bulk salt concentration low. However, even though only one of the constraints is fullfilled, the Debye-Hückelmodel is able to fit the data.

Additionally we investigated a modified Poisson-Boltzmann model. This modified model takes the average closest possible approach of two ions within the liquid into account and is able to reproduce the Stern layer, where no screening takes place [67]. For better comparison to the other two models we study the one-dimensional analytical solutions for the Debye-Hückel, the Poisson-Boltzmann, and the extended Poisson-Boltzmann model.

Reformulating the Laplace term to

$$\frac{d\varphi^2}{dx^2} = -\frac{dE}{dx} = E \cdot \frac{dE}{d\varphi} \tag{35}$$

and transforming the equations with

$$\varphi = \frac{q\psi}{k_B T} \quad \text{and} \tag{36}$$

$$\frac{1}{\lambda^2} = \frac{2qc_0}{k_B T \epsilon_0 \epsilon_{\text{Ana}}} , \tag{37}$$

leads to the following differential equations:

$$E \cdot \frac{dE}{d\varphi} = \frac{1}{\lambda^2} \sinh(\varphi) \tag{38}$$

for the Poisson-Boltzmannmodel [78],

$$E \cdot \frac{dE}{d\varphi} = \frac{1}{\lambda^2} \varphi , \tag{39}$$

for the Debye-Hückelmodel. Integrating these equations twice gives the following solutions:

$$\varphi(x) = 2 \ln\left(\frac{1 + e^{-x/\lambda} \tanh(\varphi_0/4)}{1 - e^{-x/\lambda} \tanh(\varphi_0/4)}\right) \tag{40}$$

$$E(x) = \frac{4}{\lambda} \frac{e^{-x/\lambda} \tanh(\varphi_0/4)}{1 + e^{-x/\lambda} \tanh(\varphi_0/4)} , \tag{41}$$

for the Poisson-Boltzmannmodel and

$$\varphi(x) = \varphi_0 \, e^{-x/\lambda} \tag{42}$$

$$E(x) = \varphi_0/\lambda \, e^{-x/\lambda} \tag{43}$$

for the Debye-Hücke model. Our proposed extended Poisson-Boltzmann model is formulated as

$$E \cdot \frac{dE}{d\varphi} = \frac{2}{\lambda^2} \frac{(a - (a-1)\cosh(\varphi/2))\sinh(\varphi/2)}{((1-a) + a\cosh(\varphi/2))^3} \tag{44}$$

or simplified,

$$E(\varphi) = \frac{2}{\lambda} \frac{\sinh(\varphi/2)}{1 - a + a\cosh(\varphi/2)}, \tag{45}$$

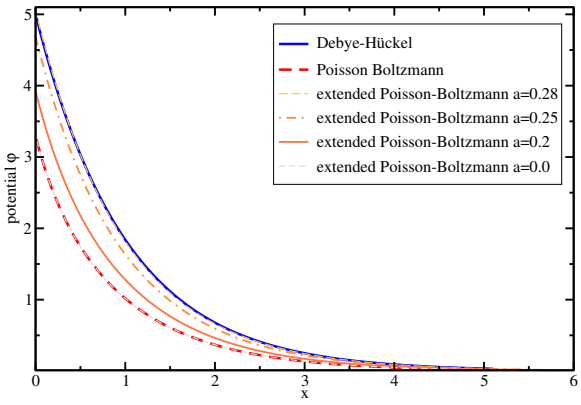


Figure 31: Analytic solution of different models at same interface charge.

where a is the closest average distance between ions. For the limit $a \to 0$ the initial Poisson-Boltzmann formulation is obtained. Fig. 31 shows the behavior of the extended Poisson-Boltzmann model. Close to the surface the extended model shows no screening, also known as the Stern layer [79]. The stern layer arises from the salt ions which are covered in a shell of water molecules. This water shell causes a minimum distance to the oxide surface (OHP) and generates a region without screening. While when one gets outside the OHP strong nonlinear screening takes place (Gouy-Chapman diffusive layer). Fig. 31 confirms that for $\mathrm{a} = 0$ the potential of the Poisson-Boltzmann model is recovered. Increasing a leads to reduced screening and generates for $\mathrm{a} = 0.28$ a similar behavior like the Debye-Hückel model. For better comparability the calculations were carried out in dimensionles units and with the same surface charge.

5.5 Conclusion

Decreasing the salt concentration does not improve the result of the Poisson-Boltzmann model. The reason is that due to nonlinear screening doubling the charge density does not lead to twice the potential shift (shown in Fig. 35). The Debye-Hückel formulation produces the best fit. Two conditions for this model must be met [75]. Firstly, the salt concentration has to be low and, secondly, the potential in the exponential terms has to be small compared to $\mathrm{k_B T}$. Despite the fact that the potential is not small enough to satisfy the linear model, it is able to reproduce the experimental data. A possible reason is that the Poisson-Boltzmann model overestimates screening. Indeed, for small salt concentrations the Poisson-Boltzmann model breaks down at high potential values, when there are not enough ions to cause screening. Therefore, the physical behavior is far more complex and requires further investigation.

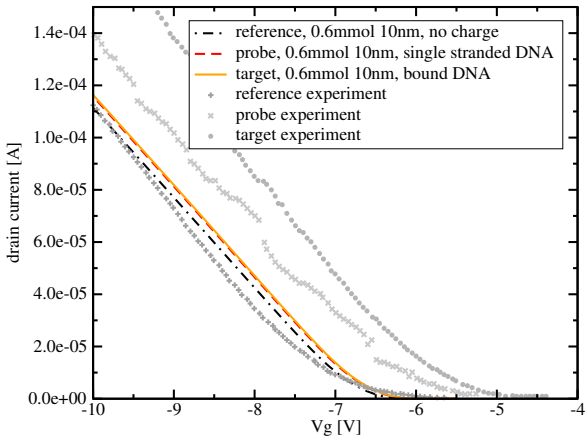


Figure 32: Transfer characteristics of a SGFET for Poisson-Boltzmannmodel and DNA charge modeled via space charge density.

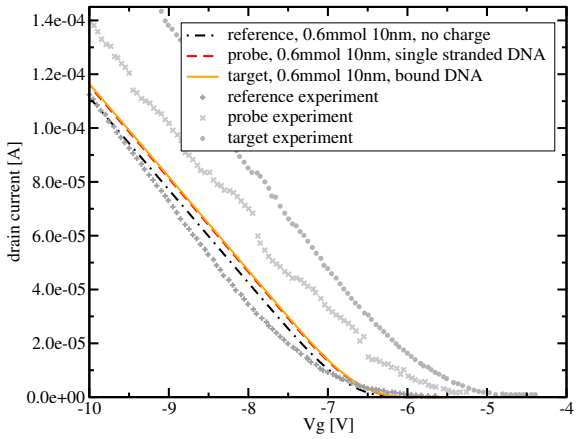


Figure 33: Transfer characteristics of a SGFET for Poisson-Boltzmannmodel and DNA charge modeled via sheet charge density.

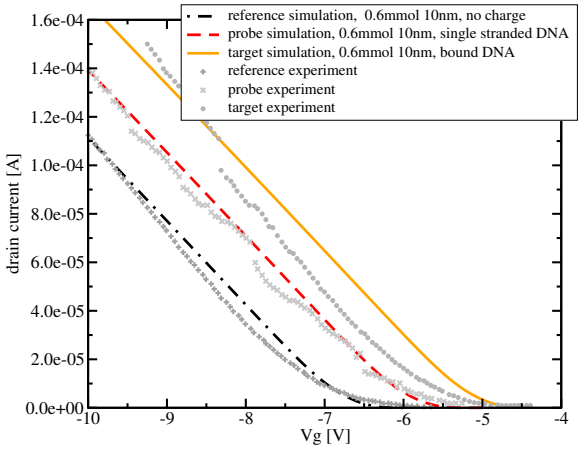


Figure 34: Transfer characteristics of a SGFET for Debye-Hückelmodel and DNA charge modeled via space charge density.

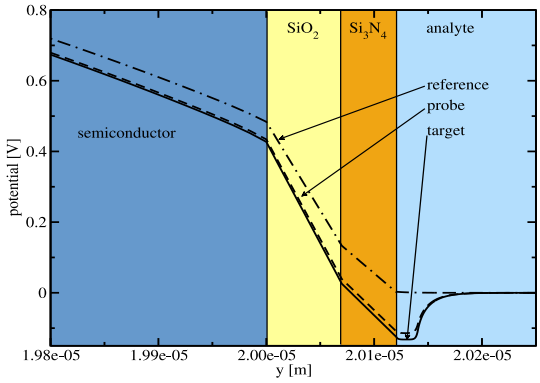


Figure 35: Potential for the Poisson-Boltzmannmodel with space charge, starting from the semiconductor (left) and ending in the analyte (right). It can be seen that doubling the charge does not lead to twice the potential shift due to nonlinear screening.

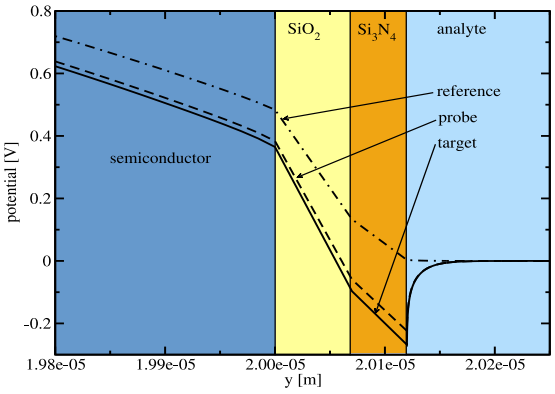


Figure 36: Potential for the Poisson-Boltzmannmodel with sheet charge, starting from the semiconductor (left) and ending in the analyte (right). Here the shift is a bit increased but far away from the values from the measurement. However, also here doubling the charge does not lead to twice the potential shift due to nonlinear screening.

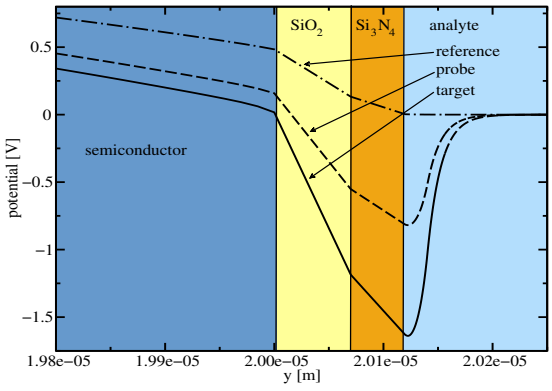


Figure 37: Potential for the Debye-Hückelmodel with space charge, starting from the semiconductor (left) and ending in the analyte (right). It can be seen that doubling the charge leads to twice the potential shift due to the weaker linear screening.

References

[1] O. Baumgartner, M. Karner, V. Sverdlov, and H. Kosina. Numerical Study of the Electron Subband Structure in Strained Silicon UTB Devices. In *EUROSOI*, 2009.

[2] J.C. Hensel, H. Hasegawa, and M. Nakayama. Cyclotron Resonance in Uniaxially Stressed Silicon. II. Nature of the Covalent Bond. *Phys. Rev.*, 138(1A):A225–A238, 1965.

[3] E. Ungersböck, 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 Transactions on Electron Devices*, 54(9):2183–2190, 2007.

[4] T. Manku and A. Nathan. Valence Energy-Band Structure for Strained Group-IV Semiconductors. *J. Appl. Phys.*, 73(3):1205–1213, 1993.

[5] F.L. Madarasz, J.E. Lang, and P.M. Hemeger. Effective Masses for Nonparabolic Bands in P-Type Silicon. *J. Appl. Phys.*, 52(7):4646–4648, 1981.

[6] A.T. Pham, B. Meinerzhagen, and C. Jungemann. A Fast $\mathbf{k}\cdot\mathbf{p}$ Solver for Hole Inversion Layers with an Efficient 2D k-Space Discretization. *J. Comp. Electron.*, 7(3):99–102, 2008.

[7] C.W. Clenshaw and A.R. Curtis. A Method for Numerical Integration on an Automatic Computer. *Num. Math.*, 2:197–205, 1960.

[8] J. Waldvogel. Fast Construction of the Fejér and Clenshaw-Curtis Quadrature Rules. *BIT*, 46:195–202, 2006.

[9] F. Lärmer and A. Schilp. Patent Nos. DE4241045 (Germany, issued 5 December 1992), US5,501,893 (U.S., issued 26 March 1996).

[10] R. Zhou, H. Zhang, Y. Hao, D. Zhang, and Y. Wang. Simulation of Profile Evolution in Etching-Polymerization Alternation in DRIE of Silicon with SF_6/C_4F_8. In *Proc. 16th IEEE Int. Micro Electro Mechanical Systems Conf. (MEMS)*, 2003.

[11] A. Hössinger, Z. Djurić, and A. Babayan. Modeling of Deep Reactive Ion Etching in a Three-Dimensional Simulation Environment. In *Proc. Int. Conf. on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2007.

[12] R. Zhou, H. Zhang, Y. Hao, and Y. Wang. Simulation of the Bosch Process with a String-Cell Hybrid Method. *J. Micromech. Microeng.*, 14(7):851–858, 2004.

[13] R.A. Gottscho, C.W. Jurgensen, and D.J. Vitkavage. Microscopic Uniformity in Plasma Etching. *J. Vac. Sci. Technol. B*, 10(5):2133–2147, 1992.

[14] Y. Tan, R. Zhou, H. Zhang, G. Lu, and Z. Li. Modeling and Simulation of the Lag Effect in a Deep Reactive Ion Etching Process. *J. Micromech. Microeng.*, 16(12):2570–2575, 2006.

[15] G. Kokkoris, A. Tserepi, A.G. Boudouvis, and E. Gogolides. Simulation of SiO_2 and Si Feature Etching for Microelectronics and Microelectromechanical Systems Fabrication: A Combined Simulator Coupling Modules of Surface Etching, Local Flux Calculation, and Profile Evolution. *J. Vac. Sci. Technol. A*, 22(4):1896–1902, 2004.

[16] T.S. Cale and G.B. Raupp. A Unified Line-Of-Sight Model of Deposition in Rectangular Trenches. *J. Vac. Sci. Technol., B*, 8(6):1242–1248, 1990.

[17] T.S. Cale, T.P. Merchant, L.J. Borucki, and Andrew H. Labun. Topography Simulation for the Virtual Wafer Fab. *Thin Solid Films*, 365(2):152–175, 2000.

[18] J.A. Sethian. *Level Set Methods and Fast Marching Methods*. Cambridge Univ. Press, 1999.

[19] G. Kokkoris, A.G. Boudouvis, and E. Gogolides. Integrated Framework for the Flux Calculation of Neutral Species Inside Trenches and Holes During Plasma Etching. *J. Vac. Sci. Technol. A*, 24(6):2008–2020, 2006.

[20] A. Shumilov and I. Amirov. Modeling of Deep Grooving of Silicon in the Process of Plasmochemical Cyclic Etching/Passivation. *Russian Microelectronics*, 36(4):241–250, 2007.

[21] G. Sun, X. Zhao, H. Zhang, L. Wang, and G. Lu. 3-D Simulation of Bosch Process with Voxel-Based Method. In *Proc. 2nd IEEE Int. Conf. on Nano/Micro Engineered and Molecular Systems (IEEE-NEMS)*, 2007.

[22] R.A. Gottscho. Ion Transport Anisotropy in Low Pressure, High Density Plasmas. *J. Vac. Sci. Technol. B*, 11(5):1884–1889, 1993.

[23] D. Adalsteinsson and J.A. Sethian. The Fast Construction of Extension Velocities in Level Set Methods. *J. Comput. Phys.*, 148(1):2–22, 1999.

[24] R. Malladi, J.A. Sethian, and B.C. Vemuri. Shape Modeling with Front Propagation: A Level Set Approach. *IEEE T. Pattern. Anal.*, 17(2):158–175, 1995.

[25] R.T. Whitaker. A Level-Set Approach to 3D Reconstruction from Range Data. *Int. J. Comput. Vision*, 29(3):203–231, 1998.

[26] B. Houston, M.B. Nielsen, C. Batty, O. Nilsson, and K. Museth. Hierarchical RLE Level Set: A Compact and Versatile Deformable Surface Representation. *ACM Trans. Graph.*, 25(1):151–175, 2006.

[27] A. Pasko, V. Adzhiev, A. Sourin, and V. Savchenko. Function Representation in Geometric Modeling: Concepts, Implementation and Applications. *The Visual Computer*, 11(8):429–446, 1995.

[28] O. Ertl and S. Selberherr. A Fast Level Set Framework for Large Three-Dimensional Topography Simulations. *Computer Physics Communications*, 180(8):1242–1250, 2009.

[29] H. Liao and T.S. Cale. Three-Dimensional Simulation of an Isolation Trench Refill Process. *Thin Solid Films*, 236(1-2):352–358, 1993.

[30] D. Adalsteinsson and J.A. Sethian. A Level Set Approach to a Unified Model for Etching, Deposition, and Lithography III: Redeposition, Reemission, Surface Diffusion, and Complex Simulations. *J. Comput. Phys.*, 138(1):193–223, 1997.

[31] P.L. O'Sullivan, F.H. Baumann, and G.H. Gilmer. Simulation of Physical Vapor Deposition Into Trenches and Vias: Validation and Comparison with Experiment. *J. Appl. Phys.*, 88(7):4061–4068, 2000.

[32] W.E. Lorensen and H.E. Cline. Marching Cubes: A High Resolution 3D Surface Construction Algorithm. *SIGGRAPH Comput. Graph.*, 21(4):163–169, 1987.

[33] C. Heitzinger, A. Sheikholeslami, F. Badrieh, H. Puchner, and S. Selberherr. Feature-Scale Process Simulation and Accurate Capacitance Extraction for the Backend of a 100-nm Aluminum/TEOS Process. *IEEE T. Electron. Dev.*, 51(7):1129–1134, 2004.

[34] J. Arvo and D. Kirk. A Survey of Ray Tracing Acceleration Techniques. In *An Introduction to Ray Tracing*. Academic Press Ltd., 1989.

[35] V. Havran. *Heuristic Ray Shooting Algorithms*. Dissertation, Department of Computer Science and Engineering, Faculty of Electrical Engineering, Czech Technical University in Prague, 2000.

[36] I. Wald and V. Havran. On Building Fast kd-Trees for Ray Tracing, and on Doing that in O(N log N). In *Proc. IEEE Symposium on Interactive Ray Tracing*, 2006.

[37] G. Marmitt, A. Kleer, I. Wald, H. Friedrich, and P. Slusallek. Fast and Accurate Ray-Voxel Intersection Techniques for Iso-Surface Ray Tracing. In *Proc. 9th Int. Fall Workshop Vision, Modeling, and Visualization (VMV)*, 2004.

[38] T. Smy, S.K. Dew, and R.V. Joshi. Efficient Modeling of Thin Film Deposition for Low Sticking Using a Three-Dimensional Microstructural Simulator. *J. Vac. Sci. Technol., A*, 19(1):251–261, 2001.

[39] O. Ertl, C. Heitzinger, and S. Selberherr. Efficient Coupling of Monte Carlo and Level Set Methods for Topography Simulation. In *Proc. Int. Conf. on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2007.

[40] OpenMP C and C++ Application Program Interface.

[41] Michael M. and Ashok S. Algorithm 806: SPRNG: A Scalable Library for Pseudorandom Number Generation. *ACM Trans. Math. Softw.*, 26(3):436–461, 2000.

[42] O. Ertl and S. Selberherr. Three-Dimensional Topography Simulation Using Advanced Level Set and Ray Tracing Methods. In *Proc. Int. Conf. on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2008.

[43] K.P. Giapis, G.S. Hwang, and O. Joubert. The Role of Mask Charging in Profile Evolution and Gate Oxide Degradation. *Microelectron. Eng.*, 61-62:835–847, 2002.

[44] M. Hauschildt, M. Gall, S. Thrasher, P. Justison, R. Hernandez, H. Kawasaki, and P. S. Ho. Statistical Analysis of Electromigration Lifetimes and Void Evolution. *J. Appl. Phys.*, 101:043523, 2007.

[45] M. Hauschildt. *Statistical Analysis of Electromigration Lifetimes and Void Evolution in Cu Interconnects*. Dissertation, The University of Texas at Austin, 2005.

[46] J.R. Lloyd and J. Kitchin. The Electromigration Failure Distribution: The Fine-Line Case. *J. Appl. Phys.*, 69(4):2117–2127, 1991.

[47] M. Gall, C. Capasso, D. Jawarani, R. Hernandez, H. Kawasaki, and P. S. Ho. Statistical Analysis of Early Failures in Electromigration. *J. Appl. Phys.*, 90(2):732–740, 2001.

[48] L. Arnaud, T. Berger, and G. Reimbold. Evidence of Grain-Boundary Versus Interface Diffusion in Electromigration Experiments in Copper Damascene Interconnects. *J. Appl. Phys.*, 93(1):192–204, 2003.

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

[50] R.W. Balluffi. Grain Boundary Diffusion Mechanisms in Metals. *Metall. Trans. A*, 13:2069–2095, 1982.

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

[52] 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.

[53] M. E. Sarychev, Yu. V. Zhitnikov, L. Borucki, C.-L. Liu, and T. M. Makhviladze. General Model for Mechanical Stress Evolution During Electromigration. *J. Appl. Phys.*, 86(6):3068–3075, 1999.

[54] A. V. Vairagar, S. G. Mhaisalkar, A. Krishnamoorthy, K. N. Tu, A. M. Gusak, M. A. Meyer, and E. Zschech. In Situ Observation of Electromigration-Induced Void Migration in Dual-Damascene Cu Interconnect Structures. *Appl. Phys. Lett.*, 85(13):2502–2504, 2004.

[55] 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. Mater. Res.*, 12:2081–2090, 1997.

[56] V. Sukharev, E. Zschech, and W. D. Nix. A Model for Electromigration-Induced Degradation Mechanisms in Dual-Inlaid Copper Interconnects: Effect of Microstructure. *J. Appl. Phys.*, 102:053505, 2007.

[57] K.O. Jeppson and C.M. Svensson. Negative Bias Stress of MOS Devices at High Electric Fields and Degradation of MNOS Devices. *J. Appl. Phys.*, 48(5):2004–2014, 1977.

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

[59] J.H. Scofield, T.P. Doerr, and D.M. Fleetwood. Correlation Between Preirradiation $1/f$ Noise and Postirradiation Oxide-Trapped Charge in MOS Transistors. *IEEE Transactions on Nuclear Science*, 36(6):1946–1953, 1989.

[60] Xiaosong Li and L. K. J. Vandamme. $1/f$ Noise in MOSFET as a Diagnostic Tool. *Solid-State Electronics*, 35(10):1477–1481, 1992.

[61] G. Groeseneken, H.E. Maes, N. Beltran, and R.F. de Keersmaecker. A Reliable Approach to Charge-Pumping Measurements in MOS Transistors. *IEEE Transactions on Electron Devices*, ED-31(1):42–53, 1984.

[62] M. Nelhiebel, J. Wissenwasser, Th. Detzel, A. Timmerer, and E. Bertagnolli. Hydrogen-Related Influence of the Metallization Stack on Characteristics and Reliability of a Trench Gate Oxide. *Microelectronics Reliability*, 45(9-11):1355 – 1359, 2005. Proceedings of the 16th European Symposium on Reliability of Electron Devices, Failure Physics and Analysis.

[63] T. Aichinger and M. Nelhiebel. Charge pumping Revisited - the Benefits of an Optimized Constant Base Level Charge Pumping Technique for MOSFET Analysis. *IEEE Integrated Reliability Workshop 2007 Final Report*, 2007.

[64] L. K. J. Vandamme and F. N. Hooge. What Do We Certainly Know About $1/f$ Noise in MOSTs? *IEEE Transactions on Electron Devices*, 55(11):3070–3085, 2008.

[65] F N Hooge, T G M Kleinpenning, and L K J Vandamme. Experimental studies on $1/f$ noise. *Reports on Progress in Physics*, 44(5):479–532, 1981.

[66] Michael C. Pirrung. How to Make a DNA Chip. *Angew. Chem. Int. Ed.*, 41:1276–1289, 2002.

[67] M. W. Shinwari, M. J. Deen, and D. Landheer. Study of the Electrolyte-Insulator-Semiconductor Field-Effect Transistor (EISFET) with Applications in Biosensor Design. *Microelectronics Reliability*, 47(12):2025–2057, 2007.

[68] K. Y. Park, Y.S. Sohn, C.K. Kim, H.S. Kim, Y.S. Bae, and S.Y. Choi. Development of FET-Type Albumin Sensor for Diagnosing Nephritis. *Biosensors and Bioelectronics*, 23:1904–1907, 2008.

[69] K.M. Park, S.K. Lee, Y.S. Sohn, and Choi S.Y. BioFET Sensor for Detection of Albumin in Urine. *Electronic Letters*, 44(3), 2008.

[70] S. Gupta, M. Elias, X. Wen, J. Shapiro, and L. Brillson. Detection of Clinical Relevant Levels of Protein Analyte Under Physiologic Buffer Using Planar Field Effect Transistors. *Biosensors and Bioelectronics*, 24:505–511, 2008.

[71] Z. Gao, A. Agarwal, A.D. Trigg, N. Singh, C. Fang, C. Tung, Y. Fan, K.D. Buddharaju, and J. Kong. Silicon Nanowire Arrays for Label-Free Detection of DNA. *Analytical Chemistry*, 79(9):3291–3297, 2007.

[72] H. Im, X. . Huang, B. Gu, and Y. . Choi. A Dielectric-Modulated Field-Effect Transistor for Biosensing. *Nature Nanotechnology*, 2(7):430–434, 2007.

[73] E. Stern, J.F. Klemic, D.A. Routenberg, P.N. Wyrembak, D.B. Turner-Evans, A.D. Hamilton, D.A LaVan, T.M. Fahmy, and M.A. Reed. Lable-free Immunodetection with CMOS-compatible Semiconducting Nanowires. *Nature Letters*, 445(1):519–522, 2007.

[74] A. Girard, F. Bendria, O. De Sagazan, M. Harnois, F. Le Bihan, A.C. Salaün, T. Mohammed-Brahim, P. Brissot, and O. Loréal. Transferrin Electronic Detector for Iron Disease Diagnostics. *IEEE Sensors*, 2006.

[75] P. Debye and E. Hückel. Zur Theorie der Elecrolyte: I. Gefrierpunktserniedrigung und verwandte Erscheinungen. *Physikalische Zeitschrift*, 24(9):185–206, 1923.

[76] M. Harnois, O.De Sagazan, A. Girard, A-C. Salaün, and T. Mohammed-Brahim. Low Concentrated DNA Detection by SGFET. In *Transducers & Eurosensors*, Lyon, France, 2007.

[77] A. Poghossian, A. Cherstvy, S. Ingebrandt, A. Offenhäusser, and M. J. Schöning. Possibilities and Limitations of Label-Free Detection of DNA Hybridization with field-effect-based devices. *Sensors and Actuators, B: Chemical*, 111-112:470–480, 2005.

[78] B.V. Derjaguin and L.D. Landau. Theory of the Stability of Strongly Charged Lyophobic Sols and the Adhesion of Strongly Charged Particles in Solutions Electrolytes. *Russian Journal of Experimental and Theoretical Physics (ZhETF)*, 11(15):663, 1945.

[79] O. Stern. Zur Theorie der elektrolytischen Doppelschicht. *Zeitschrift für Elektrochemie und angewandte physikalische Chemie*, 30:508, 1924.