Monte Carlo Algorithm for Mobility Calculations in Thin Body Field Effect Transistors: Role of Degeneracy and Intersubband Scattering

V. Sverdlov, E. Ungersboeck, and H. Kosina

Institute for Microelectronics, Technical University of Vienna,
Gusshausstrasse 27–29, A 1040 Vienna, Austria
sverdlov@iue.tuwien.ac.at

Abstract. We generalize the Monte Carlo algorithm originally designed for small signal analysis of the three-dimensional electron gas to quasi-two-dimensional electron systems. The method allows inclusion of arbitrary scattering mechanisms and general band structure. Contrary to standard Monte Carlo methods to simulate transport, this algorithm takes naturally into account the fermionic nature of electrons via the Pauli exclusion principle. The method is based on the solution of the linearized Boltzmann equation and is exact in the limit of negligible driving fields. The theoretically derived Monte Carlo algorithm has a clear physical interpretation. The diffusion tensor is calculated as an integral of the velocity autocorrelation function. The mobility tensor is related to the diffusion tensor via the Einstein relation for degenerate statistics. We demonstrate the importance of degeneracy effects by evaluating the low-field mobility in contemporary field-effect transistors with a thin silicon body. We show that degeneracy effects are essential for the correct interpretation of experimental mobility data for field effect transistors in single- and double-gate operation mode. In double-gate structures with (100) crystal orientation of the silicon film degeneracy effects lead to an increased occupation of the higher subbands. This opens an additional channel for elastic scattering. Increased intersubband scattering compensates the volume inversion induced effect on the mobility enhancement and leads to an overall decrease in the mobility per channel in double-gate structures.

1 Introduction

Monte Carlo is a well-established numerical method to solve the Boltzmann transport equation. Traditionally, the so called forward Monte Carlo technique [5] is used to find the distribution function. Within this approach, particles are moving on classical trajectories determined by Newton's law. The motion along the trajectory is interrupted by scattering processes with phonons and impurities. Scattering is modeled as a random process. The duration of a free flight, the scattering mechanism and the state after scattering are selected randomly from

I. Lirkov, S. Margenov, and J. Waśniewski (Eds.): LSSC 2007, LNCS 4818, pp. 157–164, 2008.

a given probability distribution which is characteristic of the scattering process. This technique of generation sequences of free flights and scattering events appears to be so intuitively transparent that it is frequently interpreted as a direct emulation of transport process. Due to the Pauli exclusion principle, scattering into an occupied state is prohibited. Therefore, scattering rates depend on the probability that the final state is occupied, given by the distribution function, which is the solution of the Boltzmann equation. Dependence of scattering rates on the solution makes the Boltzmann transport equation nonlinear. In many cases the occupation numbers are small and can be safely neglected in transport simulations for practically used devices. With downscaling of semiconductor devices continuing, the introduction of double-gate (DG) silicon-on-insulator field-effect transistors (FETs) with ultra-thin (UTB) silicon body seems increasingly likely [6]. Excellent electrostatic channel control makes them perfect candidates for the far-end of ITRS scaling [1]. However, in UTB FETs degeneracy effects are more pronounced, and their proper incorporation becomes an important issue for accurate transport calculations.

Different approaches are known to include degeneracy effects into Monte Carlo algorithms. One method is to compute the occupation numbers self-consis- tently [2,8]. This approach is applicable not only to mobility simulations at equilibrium but also for higher driving fields [7]. When the distribution function is close to the equilibrium solution, the blocking factor can be approximated with the Fermi-Dirac distribution function [3]. A similar technique to account for degeneracy effects was recently reported in [14].

In this work we use a Monte Carlo algorithm originally developed for three-dimensional simulations [10] which was recently generalized to a quasi-two-dimen- sional electron gas [11,4]. This method incorporates degeneracy effects *exactly* in the limit of vanishing driving fields and is valid for arbitrary scattering mechanisms and for general band structure. We demonstrate that in UTB DG FETs degeneracy effects lead to a qualitatively different mobility behavior than in the classical simulations. Degeneracy results in higher occupation of upper subbands which substantially increases intersubband scattering in (100) UTB DG FETs, resulting in a mobility decrease.

2 Simulation Method

In order to obtain the low-field mobility in UTB FETs, we compute the response of a quasi-two dimensional electron system to the small electric field $\mathbf{E}(t)$. The system is described by a set of subband functions $\psi_n(z)$ in the confinement direction and dispersions $E_n(\mathbf{k})$ relating subband energies to the two-dimensional quasi-momentum $\mathbf{k} = (k_x, k_y)$ along Si/SiO_2 interfaces. Representing the distribution function as $f(E_n(\mathbf{k}, t)) = f_0(E_n(\mathbf{k})) + \delta f_n(\mathbf{k}, t)$, where $f_0(E_n(\mathbf{k}))$ is the equilibrium Fermi-Dirac distribution and $\delta f_n(\mathbf{k}, t)$ is a small perturbation, we arrive to the system of coupled linearized subband equations:

$$\frac{\partial \delta f_n(\mathbf{k}, t)}{\partial t} = -e\mathbf{E}(t)\nabla_{\mathbf{k}} f_0(E_n(\mathbf{k})) + Q_n[\delta f], \tag{1}$$

where Q_n is the scattering operator of the linearized Boltzmann equation

$$Q_n[\delta f] = \sum_m \int \frac{d^2\mathbf{k}'}{(2\pi)^2}(\Lambda_{nm}(\mathbf{k},\mathbf{k}')\delta f_m(\mathbf{k}',t) - \Lambda_{mn}(\mathbf{k}',\mathbf{k})\delta f_n(\mathbf{k},t)). \quad (2)$$

The scattering rates $\Lambda_{mn}(\mathbf{k},\mathbf{k}')$ in (2) are related to the rates $S_{mn}(\mathbf{k},\mathbf{k}')$ of the original Boltzmann equation via

$$\Lambda_{mn}(\mathbf{k}',\mathbf{k}) = (1 - f_0(E_m(\mathbf{k}')))S_{mn}(\mathbf{k}',\mathbf{k}) + f_0(E_m(\mathbf{k}'))S_{mn}(\mathbf{k},\mathbf{k}'), \quad (3)$$

where $f_0(E)$ is the Fermi-Dirac distribution function, and $E_n(\mathbf{k})$ is the total energy in the n-th subband. The equation for the perturbation has a form similar to the Boltzmann equation, with two important differences: (i) a source term which depends on the small driving field and is proportional to the derivative of the equilibrium function is present, and (ii) it has renormalized scattering rates which enforce the equilibrium solution of the homogeneous equation (1) to be $f_0(E_n(\mathbf{k}))(1 - f_0(E_n(\mathbf{k})))$, and not $f_0(E_n(\mathbf{k}))$.

In order to calculate the mobility, a subband Monte Carlo method is used to solve the system (1). Following the procedure outlined in [10], we assume the time dependence of the driving field to be a set of instantaneous delta-like pulses:

$$\mathbf{E}(t) = \mathbf{E}_0\tau\sum_i \delta(t - t_i), \quad (4)$$

In (4) $\mathbf{E}_0$ is the value of the field averaged over a long simulation time T

$$\mathbf{E}_0 = \frac{1}{T}\int_0^T dt\,\mathbf{E}(t).$$

Then τ is the average period between the delta-pulses.

We compute the current response $\mathbf{I}_i(t)$ produced by an electric field pulse at the moment t_i as

$$\mathbf{I}_i = eH(t - t_i)\sum_m \int \frac{d^2\mathbf{k}}{(2\pi)^2}\mathbf{v}_m(\mathbf{k})\delta f_m(\mathbf{k}, t - t_i), \quad (5)$$

where $\mathbf{v}$ is the velocity, and $H(t)$ is the Heaviside function. The instanteneous current density $\mathbf{J}(t) = \sum_i \mathbf{I}_i$ is calculated as the sum over current densities $\mathbf{I}_i$ produced by all pulses i. The current density value averaged over the long time T is expressed as $\mathbf{J} = \left(\sum_i \int_0^T dt\mathbf{I}_i(t)\right)/T$. The low field mobility is defined as $\mu_{\alpha\beta} = J_\alpha/(enE_\beta)$, where the direction of β-axis coincides with the direction of $\mathbf{E}_0$, and n is the carrier concentration. Now the mobility can be easily computed using a single-particle Monte Carlo technique.

The method can be illustrated as follows. The diffusion tensor $D_{\alpha\beta}$ is calculated as an integral of the velocity autocorrelation function [9]

$$D_{\alpha\beta} = \int_0^\infty d\tau\,\langle v_\alpha(t)v_\beta(t+\tau)\rangle\,, \quad (6)$$

where angular brackets denote the time averaging over the stochastic dynamics determined by the rates $\Lambda_{mn}(\mathbf{k}, \mathbf{k}')$ of the *linearized* multi-subband Boltzmann scattering integral in case of degenerate statistics. The mobility tensor $\tilde{\mu}_{\alpha\beta}$ is related to the diffusion tensor via the Einstein relation for degenerate statistics

$$\tilde{\mu}_{\alpha\beta} = eD_{\alpha\beta}\frac{1}{n}\frac{\delta n}{\delta E_F}, \tag{7}$$

where E_F is the Fermi level.

In order to compute the mobility, we accumulate three temporary estimators t, w_β, and $\nu_{\alpha\beta}$ during the Monte Carlo simulations:

(i) initialize $t = 0$, $w_\beta = 0$, $\nu_{\alpha\beta} = 0$, and start the particle trajectory with the stochastic dynamics determined by the scattering rates $\Lambda_{mn}(\mathbf{k}, \mathbf{k}')$ from (3) of the *linearized* multi-subband Boltzmann equations;

(ii) before each scattering event update $\nu_{\alpha\beta}$, w_β, and t:

$$\begin{aligned} t &= t + \frac{\tau(j)}{1 - f(E(j))}, \\ w_\beta &= w_\beta + v_\beta(j)\tau(j), \\ \nu_{\alpha\beta} &= \nu_{\alpha\beta} + \tau(j)v_\alpha(j)w_\beta(j); \end{aligned}$$

(iii) When t is sufficiently large, compute the mobility tensor as

$$\tilde{\mu}_{\alpha\beta} = \frac{e}{k_B T}\frac{\nu_{\alpha\beta}}{t},$$

where $v_\alpha(j)$ denotes the α-component of the velocity, $E(j)$ is the particle energy, $f(E)$ is the Fermi-Dirac function, and $\tau(j)$ is the time of j-th free flight. The convergence of the method is improved by resetting $w_\beta = 0$ each time a velocity randomizing scattering event occurs.

3 Degeneracy Effects and Intersubband Scattering

We demonstrate the importance of degeneracy effects by evaluating the low-field mobility in inversion layers and in UTB FETs. The phonon-limited mobility in inversion layers shows a different behavior if the Pauli exclusion principle is taken into account. However, if surface roughness scattering is included, the relative difference decreases, and the universal mobility curve can be reproduced equally well using both degenerate and nondegenerate statistics [14], as shown in Fig. 1.

In UTB FETs degeneracy effects are expected to be more pronounced. We consider as example a 3 nm thick (100) UTB FET. The nondegenerate statistics is assured by using the rates $S_{mn}(\mathbf{k}', \mathbf{k})$ of the original Boltzmann equation in the Monte Carlo algorithm described above. Results of mobility calculations for single-gate (SG) and DG structures, with and without degeneracy effects taken into account in the Monte Carlo simulations are summarized in Fig. 2. Mobility in a DG FET is plotted as function of the carrier concentration per channel, or $N_{DG}/2$. When degeneracy effects are neglected, the DG mobility is superior as

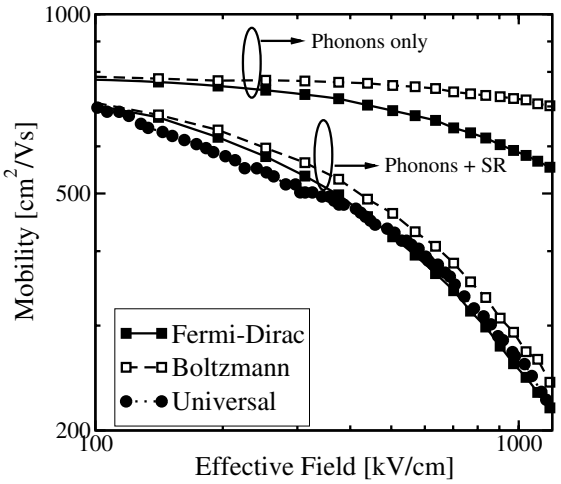

Fig. 1. Effective mobility of a Si inversion layer at (100) interface computed with Boltzmann (open symbols) and Fermi-Dirac (filled symbols) statistics reproduces well the universal mobility curve [12] (circles). Phonon-limited mobility for degenerate and nondegenerate statistics is also shown.

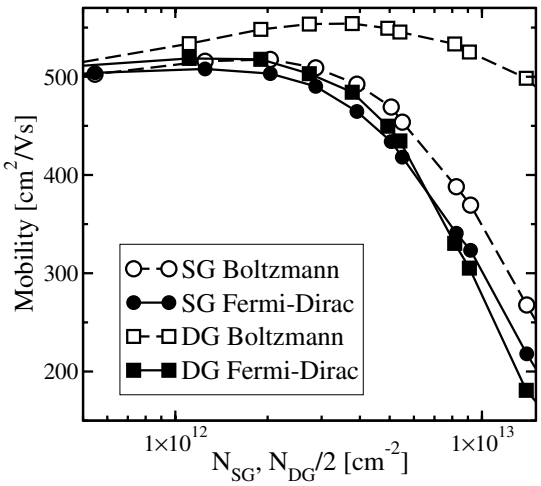

Fig. 2. Mobility in 3 nm thick (100) SG (circles) and DG (squares) structures computed with Boltzmann (open symbols) and Fermi-Dirac (filled symbols) statistics

compared to the SG mobility. When the degeneracy effects are included, behavior of the DG mobility is qualitatively different. Namely, the DG mobility becomes lower than the SG mobility at high carrier concentrations, in agreement with experimental data [13].

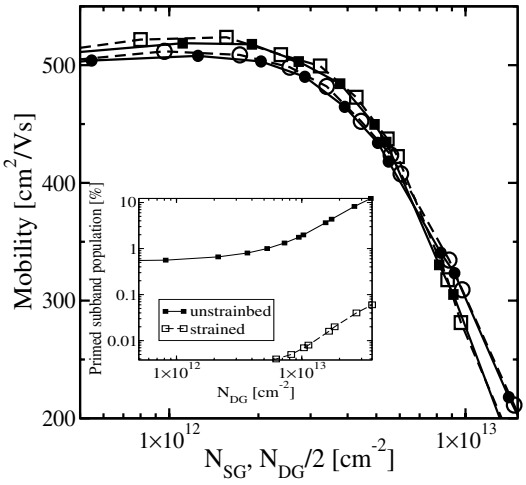

Fig. 3. Mobility in (100) 3 nm thick DG (squares) and SG (circles) structures computed with (open symbols) and without (filled symbols) in-plane biaxial stress of 1.6 GPa. Inset: occupation of primed subbands in relaxed (filled symbols) and biaxially stressed (open symbols) DG structure.

The difference between the mobility values for degenerate and nondegenerate statistics shown in Fig. 2 looks surprising. Indeed, at high carrier concentrations the principal scattering mechanism limiting the low-field mobility is elastic surface roughness scattering. For elastic scattering the forward and inverse scattering rates are equal: $S^{el}_{mn}(\mathbf{k}', \mathbf{k}) = S^{el}_{nm}(\mathbf{k}, \mathbf{k}')$, so that the Pauli blocking factor cancels out from the equations for the elastic scattering rates (3), and degeneracy effects seem to be irrelevant. This is not correct, however, since the Pauli blocking factor is also present in the inelastic electron-phonon part of the total scattering integral and ensures the equilibrium solution to be the Fermi-Dirac distribution function. In case of Fermi-Dirac statistics the Fermi level in a DG FET is higher than in a SG FET, due to twice as high carrier concentration for the same gate voltage [11]. This results in a higher occupation of upper subbands. To study the influence of the occupation of *primed* subbands on the mobility lowering in (100) DG FETs we apply a biaxial stress of 1.6 GPa. This level of stress provides an additional splitting between the primed and unprimed subbands high enough to depopulate the primed ladder completely. Results of the mobility simulation in 3 nm DG and SG structures, with biaxial stress applied, are shown in Fig. 3 together with the results for the unstrained structure. Both mobilities in strained and unstrained structures are similar in the whole range of concentrations. The inset displays the population of primed subbands in a 3 nm DG structure, showing that the primed ladder in a strained FET is practically depopulated. Since the mobilities of strained and unstrained UTB FETs are almost equivalent for both SG and DG structures, it then follows that the higher occupation of primed subbands is not the reason for the DG mobility

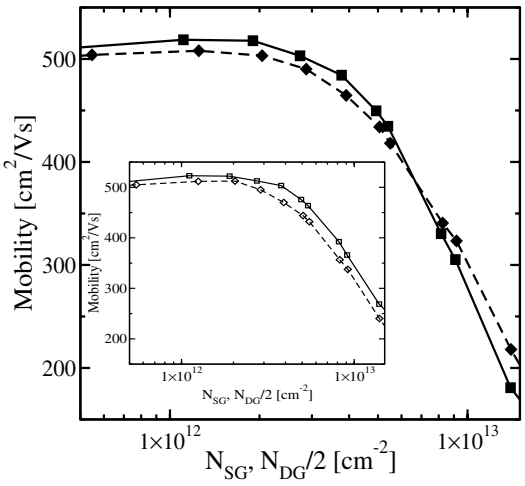

Fig. 4. Mobility in 3 nm thick (100) structures, computed with and without (Inset) inter-subband scattering. Higher carrier concentration in a UTB DG structure (squares) at the same gate voltage pushes the Fermi-level up and opens additional inter-subband scattering channels between unprimed subbands. It decreases the mobility in (100) UTB DG FETs below its SG values (diamonds) at high carrier concentrations.

lowering in (100) DG UTB structures. Another consequence of twice as high carrier concentration in a DG UTB FET is the higher occupation of upper *unprimed* subbands. When the carrier energy is above the bottom E_1 of the next unprimed subband, and intensive elastic intersubband scattering occurs. This additional scattering channel leads to a step-like increase in the density of after-scattering states, which results in higher scattering rates.

To demonstrate the importance of intersubband scattering for mobility calculations, we artificially switch off the scattering between unprimed subbands. We consider degenerate statistics and restore screening. Results of the mobility calculations for a 3 nm thick UTB structure, with and without intersubband scattering are shown in Fig. 4. Without intersubband scattering, the DG FET mobility is higher than the corresponding SG mobility, in analogy to nondegenerate results. It confirms our finding that as soon as the additional intersubband scattering channel becomes activated, the DG mobility value sinks below the SG mobility.

4 Conclusion

The Monte Carlo algorithm originally designed for small signal analysis of the three-dimensional electron gas response is generalized to quasi-two-dimensional electron systems. In the limit of vanishing driving fields the method includes degeneracy effects exactly. The method is valid for arbitrary scattering rates and includes realistic band structure. The Monte Carlo method is applied to compute

the low-field mobility in UTB FETs. It is demonstrated that degeneracy effects play a significant role in compensating the volume inversion induced mobility enhancement in (100) DG structures. They lead to a significant occupation of higher subbands in the unprimed ladder, which results in increased intersubband scattering and mobility lowering.

References

1. International Technology Roadmap for Semiconductors — 2005 edn.(2005), http://www.itrs.net/Common/2005ITRS/Home2005.htm
2. Bosi, S., Jacoboni, C.: Monte Carlo High Field Transport in Degenerate GaAs. J. Phys. C: Solid State Phys. 9, 315–319 (1976)
3. Fischetti, M.V., Laux, S.E.: Monte Carlo Analysis of Electron Transport in Small Semiconductor Devices Including Band-Structure and Space-Charge Effects. Physical Review B 38, 9721–9745 (1988)
4. Jungemann, C., Pham, A.T., Meinerzhagen, B.: A Linear Response Monte Carlo Algorithm for Inversion Layers and Magnetotransport. In: Proc. Intl. Workshop Comput. Electronics, pp. 13–14 (May 2006)
5. Kosina, H., Nedjalkov, M., Selberherr, S.: Theory of the Monte Carlo Method for Semiconductor Device Simulation. IEEE Trans. Electron Devices 47, 1899–1908 (2000)
6. Likharev, K.K.: Sub-20-nm Electron Devices. In: Morkoc, H. (ed.) Advanced Semiconductor and Organic Nano-Techniques, pp. 239–302. Academic Press, New York (2003)
7. Lucci, L., et al.: Multi-Subband Monte-Carlo Modeling of Nano-MOSFETs with Strong Vertical Quantization and Electron Gas Degeneration. In: IEDM Techn. Dig., pp. 531–534 (2005)
8. Lugli, P., Ferry, D.K.: Degeneracy in the Ensemble Monte Carlo Method for High Field Transport in Semiconductors. IEEE Trans. Electron Devices 32, 2431–2437 (1985)
9. Reggiani, L., et al.: Diffusion and fluctuations in a nonequilibrium electron gas with electron-electron collisions. Phys. Rev. B 40, 12209–12214 (1989)
10. Smirnov, S., et al.: Monte Carlo Method for Modeling of Small Signal Response Including the Pauli Exclusion Principle. J. Appl. Phys. 94, 5791–5799 (2003)
11. Sverdlov, V., et al.: Mobility for High Effective Field in Double-Gate and Single-Gate SOI for Different Substrate Orientations. In: Proc. EUROSOI 2006, pp. 133–134 (March 2006)
12. Takagi, S.I., et al.: On the Universality of Inversion Layer Mobility in Si MOSFET's: Part I — Effects of Substrate Impurity Concentration. IEEE Trans. Electron Devices 41, 2357–2362 (1994)
13. Uchida, K., Koga, J., Takagi, S.: Experimental Study on Carrier Transport Mechanisms in Double- and Single-Gate Ultrathin-Body MOSFETs — Coulomb Scattering, Volume Inversion, and δT_{SOI}-induced Scattering. In: IEDM Techn. Dig., pp. 805–808 (2003)
14. Ungersboeck, E., Kosina, H.: The Effect of Degeneracy on Electron Transport in Strained Silicon Inversion Layer. In: Proc. Intl. Conf. on Simulation of Semiconductor Processes and Devices, pp. 311–314 (2005)