IEICE TRANS. ELECTRON., VOL.E83–C, NO.8 AUGUST 2000

PAPER *Special Issue on SISPAD'99*

Full-Band Monte Carlo Device Simulation of a Si/SiGe-HBT with a Realistic Ge Profile

Christoph JUNGEMANN[†a)], **Stefan KEITH**[†], *and* **Bernd MEINERZHAGEN**[†], *Nonmembers*

SUMMARY This work presents the first comprehensive full-band Monte Carlo model for the simulation of silicon/silicon-germanium devices with arbitrary germanium profiles. The model includes a new CPU and memory efficient method for the discretization of the Brillouin zone based on adaptive nonuniform tetrahedral grids and a very efficient method for transfers through heterointerfaces in the case of irregular $\vec{k}$-space grids. The feasibility of the FB-MC simulation is demonstrated by application to an industrial HBT with a graded germanium profile and different aspects of the microscopic carrier transport are discussed. Internal distributions of the transistor are calculated with a very low noise level and high efficiency allowing a detailed investigation of the device behavior.

key words: *silicon-germanium, full-band Monte Carlo, HBT*

1. Introduction

Silicon/silicon-germanium heterojunction bipolar transistors (Si/SiGe HBTs) are promising devices for low-cost high frequency applications in mobile communication, because most of the standard Si production processes can be used for SiGe device fabrication and the bandgap can be engineered by varying the Ge concentration [1], [2]. Base widths of less than 20 nm have been realized allowing cut-off frequencies of over 100 GHz [3]. On the other hand, the modeling of these nano-scale devices is more complex than that of Si devices and the reliability of the standard simulation tools like the drift-diffusion (DD) or hydrodynamic (HD) model is questionable. Hence, there is an increasing demand for a more physics based modeling of SiGe heterodevices. The solution of the Boltzmann transport equation by means of the full-band Monte Carlo method (FB-MC) only relies on few simplifications of carrier transport physics [4]. Thus, this method is expected to describe the device behavior of heterodevices in a more reliable way than the standard tools. But this simulation technique has its own problems. The bottleneck is the required CPU time. Efficient algorithms for the band structure description and the particle propagation are necessary to minimize the computational effort.

In this work we present the first FB-MC device simulator for SiGe devices. In Sect. 2 a detailed description of the newly developed efficient FB-MC methods for SiGe is given. In Sect. 3 the details of the device simulation of an HBT test structure of an industrial pilot line process with a realistic Ge profile are described [3], [5]. In Sect. 4 the feasibility of an FB-MC simulation of this realistic HBT is demonstrated and details of the quasiballistic carrier transport are discussed. Of course, the FB-MC simulator is not restricted to HBTs and an example of a SiGe PMOSFET simulation can be found in Ref. [6].

Manuscript received November 15, 1999.
Manuscript revised February 21, 2000.

[†]The authors are with the University of Bremen, Postfach 330 440, Kufsteiner Straße, 28334 Bremen, Germany.
a) E-mail: junge@item.uni-bremen.de

2. Simulation Model

The SiGe device simulator used in this work is based on the Si FB-MC program described in Ref. [7]. The extension of the efficient particle scattering algorithms of Ref. [7] for Si to the case of SiGe is straight forward. The model of Ref. [8] for phonon and alloy scattering in SiGe is used. Scattering by ionized impurities is taken into account including a doping dependent heuristic adjustment to match experimental minority and majority mobility data. The new methods for phase space discretization and transfers through heterointerfaces are described below.

In a SiGe device the band structure depends on the Ge content and doping concentration and its spatial variation is splitted into two parts. One part is the absolute value of the band edge being the minimum band energy. The band shifts of the conduction and valence bands relative to relaxed Si due to the Ge content are given by [1], [9]

$$\Delta E_{\mathrm{c}}^{\mathrm{Ge}} = -0.196\mathrm{eV}\, x_{\mathrm{Ge}} + 0.396\mathrm{eV}\, x_{\mathrm{Ge}}^2 \quad (1)$$

$$\Delta E_{\mathrm{v}}^{\mathrm{Ge}} = -0.7\mathrm{eV}\, x_{\mathrm{Ge}} \quad (2)$$

and due to the doping dependent apparent bandgap narrowing by

$$\Delta E^{\mathrm{Bgn}} = E_0 \left(\ln \frac{N_{\mathrm{T}}}{N_{\mathrm{Ref}}} + \sqrt{\left(\ln \frac{N_{\mathrm{T}}}{N_{\mathrm{Ref}}} \right)^2 + 0.5} \right) \quad (3)$$

with $E_0 = 6.92\,\mathrm{meV}$ and $N_{\mathrm{Ref}} = 1.3 \cdot 10^{17}/\mathrm{cm}^3$ and where N_{T} is the total doping concentration [10], [11]. In accordance with Ref. [12] the above doping dependent apparent bandgap narrowing developed for Si is also used without any modifications in the case of SiGe. The total band offsets read

$$\Delta E_{\mathrm{c}}^{\mathrm{T}} = \Delta E_{\mathrm{c}}^{\mathrm{Ge}} - \frac{1}{2} \Delta E^{\mathrm{Bgn}} \quad (4)$$

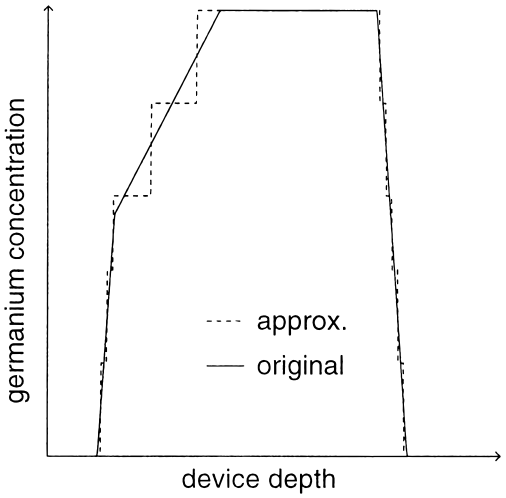

Fig. 1 Original germanium profile and approximation for the selection of the band structures.

$$\Delta E_{\mathrm{v}}^{\mathrm{T}} = \Delta E_{\mathrm{v}}^{\mathrm{Ge}} - \frac{1}{2}\Delta E^{\mathrm{Bgn}} \ . \tag{5}$$

These band offsets are calculated for each node of the real space grid. Within a cell of the grid they are linearly interpolated.

The second part of the band structure, the energy bands measured relative to the absolute band edge only depend on the Ge concentration. They are calculated with the nonlocal empirical pseudopotential method [13]. In contrast to the band edge the energy bands are assumed to be constant within a cell of the real space grid. Because of the high number of grid cells this might result for a general Ge profile in a large number of Ge concentrations for which the energy bands have to be stored. Since our simulator needs about 50 MBytes per Ge value to store all the necessary data, the maximum number of energy bands has to be limited. This is done by approximating the Ge profile with a staircase-like function with the appropriate number of steps (Fig. 1). Please note that this approximation is only applied to the energy bands and not to the band edges.

Concerning energy bands the simulated device consists due to the above approximation of layers with different Ge concentrations. Since the Ge content produces strain in the pseudomorphicly grown device structure, the lattice constant is anisotropic and varies in growth direction (here x-direction), whereas it is given perpendicular to the growth direction by the lattice constant of the relaxed Si bulk. The anisotropy distorts the Brillouin zone compared to the case of relaxed Si. This distortion is removed by mapping the original $\vec{k}$-space into the $\vec{k}^*$-space with

$$\vec{k}^* = \underline{\underline{T}}\,\vec{k} \ , \tag{6}$$

where the transformation matrix is given by

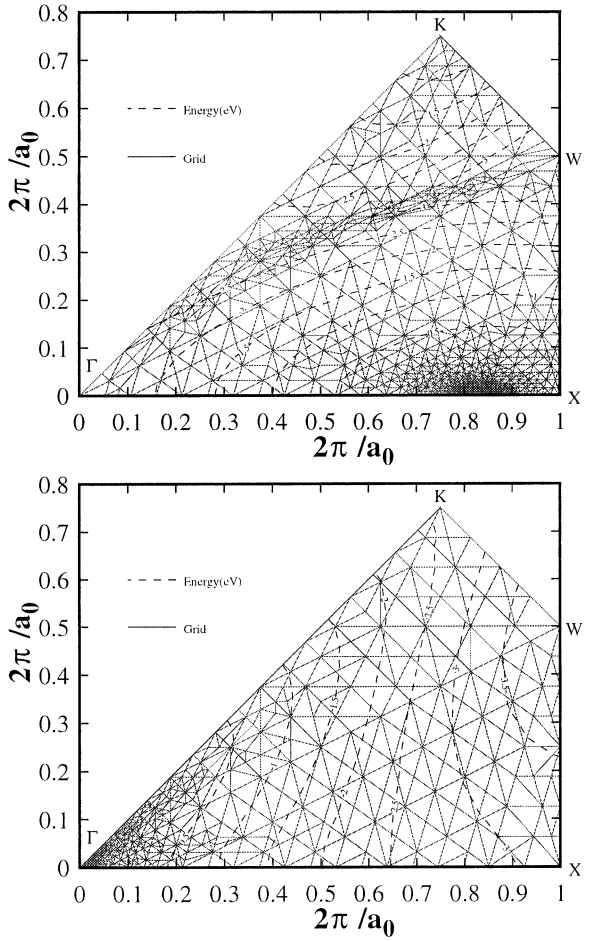

Fig. 2 Nonuniform adaptive tetrahedral mesh of the first conduction band (upper graph) and of the first valence band (lower graph) in the k_x, k_y-plane. Equienergy lines are for relaxed Si.

$$\underline{\underline{T}} = \begin{pmatrix} \frac{a_\parallel}{a_0} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \ . \tag{7}$$

$a_\parallel$ is the lattice constants parallel to the growth direction and a_0 is the lattice constant of relaxed Si.

The irreducible wedge of the strained material comprises in the $\vec{k}^*$-space three irreducible wedges of relaxed Si. One of these wedges of relaxed Si is discretized with a nonuniform tetrahedral grid, which is then mapped into the other two wedges to discretize the full irreducible wedge of the strained material. This grid is generated with the adaptive method described in Ref. [7] by minimizing the discretization error of the band energy [14]. Since the individual energy bands strongly differ from each other, grids are produced for each energy band. On the other hand, for a given energy band the same grid is used for all Ge concentrations to save memory. In order to obtain a good grid for all relevant Ge concentrations the discretization error of the energy is simultaneously minimized in the entire Brillouin zone for relaxed Si, strained $Si_{0.6}Ge_{0.4}$

Table 1 Number of grid nodes N^{nod} and tetrahedra N^{tet} per conduction or valence band.

Band	N^{nod}	N^{tet}
4c	2763	12439
3c	2761	12499
2c	3051	13494
1c	6572	29735
1v	3082	13111
2v	2209	9354
3v	1284	5226

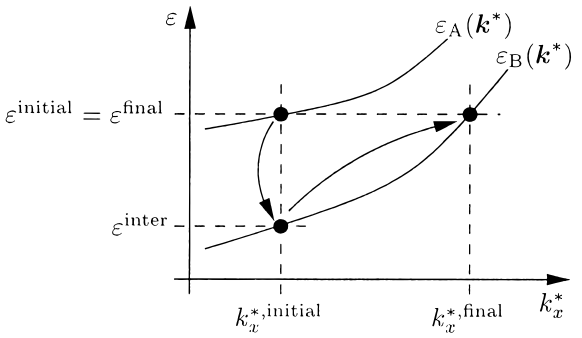

Fig. 3 Selection of the final state after a heterointerface transition from band structure A to B.

on relaxed Si, and strained Si on relaxed $Si_{0.6}Ge_{0.4}$. In Fig. 2 the grids of the first conduction band and valence band are shown and in Table 1 the number of grid nodes and tetrahedra is given. Compared to the original grids for relaxed Si of Ref. [7] these grids contain about 30% more tetrahedra and grid nodes.

When a particle moves from one cell of the real space grid into another one the band structure might change and the new particle state in the $\vec{k}^*$-space is calculated under conservation of the total energy, the band index and the component of the wave vector parallel to the interface between both cells (Fig. 3). In order to reduce the CPU time the search for the new particle state starts with the $\vec{k}^*$-vector just before the transition. This particle state is a valid state in the new band structure, because the grid of the energy band is the same as before the transition, and the CPU intensive search for a $\vec{k}^*$-vector in an irregular grid (Fig. 2) is avoided. But the new state does not satisfy energy conservation (Fig. 3). This is corrected by moving the particle in k_x^*-direction from tetrahedron to tetrahedron until energy conservation holds. (Please note, that the interface is always parallel to the y, z-plane and that k_x^* is the component normal to the interface.) The maximum search distance for k_x^* in the $\vec{k}^*$-space is $4\pi/a_0$. The new particle state must have a velocity component perpendicular to the interface which points into the new cell. In the case that no valid new state can be found, the transition is rejected and the particle is reflected at the interface. Since in most cases the energy mismatch of the transition is small, the new state is rapidly found.

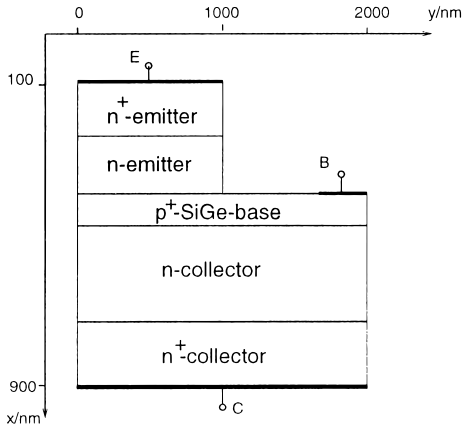

Fig. 4 The 2D SiGe-HBT test structure.

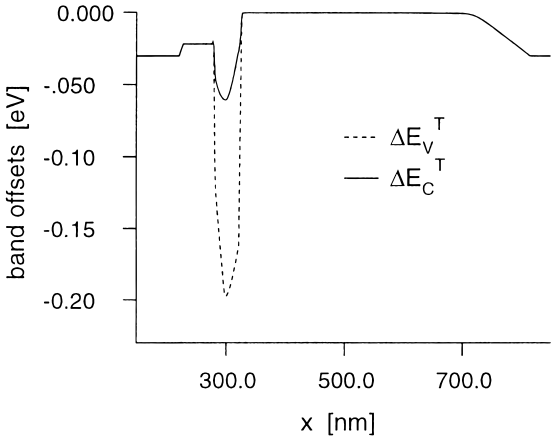

Fig. 5 Conduction and valence band offsets.

3. Device Simulation

In Fig. 4 the 2D outline of the HBT test structure is shown [5], [15]. The maximum acceptor doping is $3 \cdot 10^{19}/\mathrm{cm}^3$. The lowly doped emitter and the collector have a donor doping of about $3 \cdot 10^{18}/\mathrm{cm}^3$ and $2 \cdot 10^{16}/\mathrm{cm}^3$, respectively. In the emitter and collector contact region the maximum donor concentration is limited to $10^{19}/\mathrm{cm}^3$ in order to reduce the effect of carrier self-heating during the MC simulations. It has been checked by HD simulations that this reduction of the maximum donor doping does not obscure the simulation results for the collector current in comparison with simulations based on the original doping profile.

The Ge profile is graded with a maximum Ge content of about 20%. It is similar to the one shown in Fig. 1. The resultant band offsets (Fig. 5) lead to a blocking energy barrier for electrons passing the base/collector junction and for holes passing the base/emitter junction. Here, the Ge profile is approximated for the selection of the energy bands by six different Ge concentrations. Although within our FB-MC code any desired number of band structures is

possible, the number is limited by the available CPU memory. Less than 50 MBytes are required for one band structure and the total for this simulation is about 300 MBytes including the particle data and statistics. In addition, simulations with ten different Ge concentrations were performed and no statistically significant differences compared to the case of six Ge concentrations were found. The FB-MC model is self-consistently solved with Poisson's equation and the CPU efficiency of the FB-MC model is similar to the one of the analytical MC model described in Ref. [16].

In order to reduce the CPU time 1D simulations were performed with the FB-MC model. Holes were considered by solving a nonlinear Poisson equation with a constant quasifermi level for the holes [17] which was computed with a consistent DD model [18]. 1D and 2D simulations with and without a constant quasifermi level for the holes by the DD, HD, and FB-MC model showed that these 1D simulations yield accurate results for the collector current as long as the device is not operated too far in the high injection regime. Besides being more CPU efficient the assumption of a constant quasifermi level for the holes has the advantages of avoiding spurious heating of the electrons in the base region by holes. This spurious heating is due to the small statistical weight assigned to the electrons in the base by statistical enhancement compared to the holes [19]. Of course, evaluation of a base current is not possible with this approximation. The real space grid has 195 nodes with a nonequidistant spacing. In the base and around it a spacing of 1 nm is used to resolve the details of the carrier transport. The number of particles per cell of the real space grid is 50 being maintained by the Multiple refresh method [20]. The total number of simulated electrons is about 13600 including extra particles for the highly doped regions. The high number of particles ensures that the simulated plasma is collisionless [19].

4. Results

Simulations have been performed for a collector bias of 2 V, a base bias of 0 V and three emitter voltages (-0.65 V, -0.70 V, and -0.75 V) at an ambient temperature of 305 K. Due to the constant base/collector voltage the potential (Fig. 6) does not vary much in the base and collector region as a function of the emitter bias. The small differences in the lowly doped collector between the results for the two lowest emitter voltages and the highest one are caused by the strong increase of the space charge in the collector with the base/emitter voltage (onset of high injection, Fig. 7). These small differences in the potential have only a minor impact on the electron velocity (Fig. 8) and temperature (Fig. 9) in the base and collector. In the case of the latter quantity the peak value is slightly reduced for the highest base/emitter voltage due to the reduced maximum electric field in the lowly doped collector. A strong velocity overshoot (Fig. 8) is found at the base/collector junction just before the maximum of the electron temperature (Fig. 9) demonstrating the quasiballistic nature of the transport. In the case of the lowest base/emitter voltage the electron density (Fig. 7) varies over more

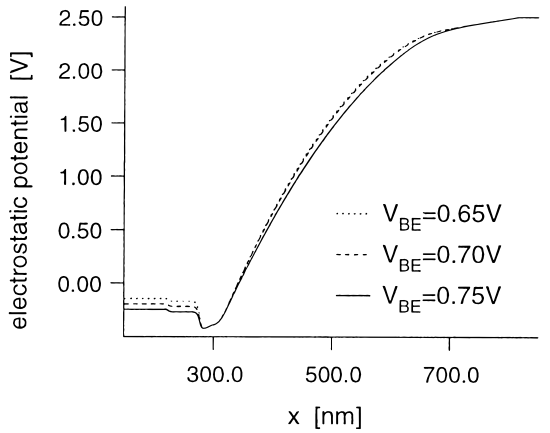

Fig. 6 Electrostatic potential for $V_{\mathrm{CB}} = 2.0$ V.

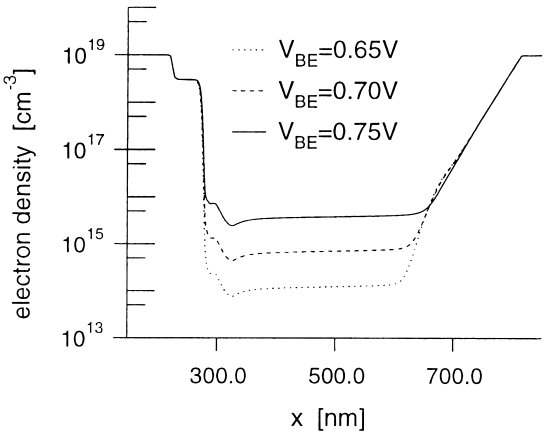

Fig. 7 Electron density for $V_{\mathrm{CB}} = 2.0$ V.

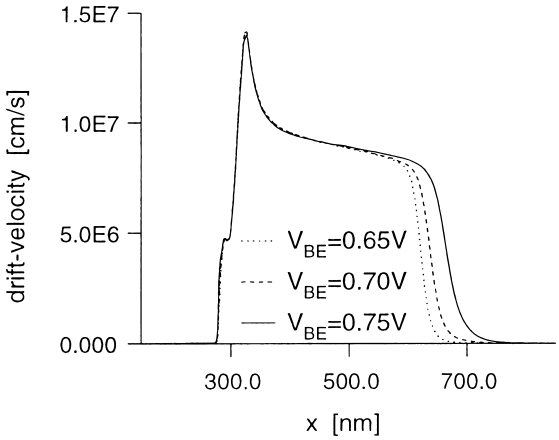

Fig. 8 Electron drift velocity for $V_{\mathrm{CB}} = 2.0$ V.

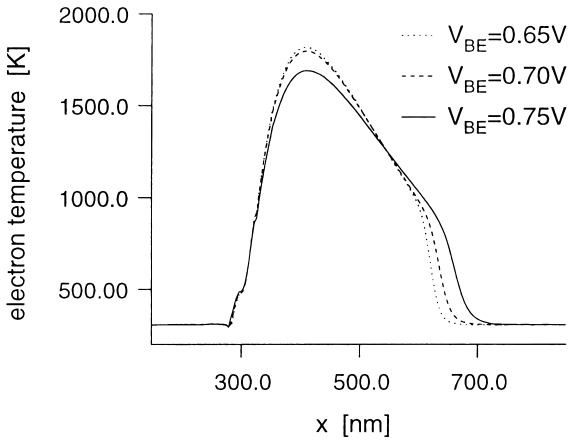

Fig. 9 Dynamic electron temperature for $V_{\mathrm{CB}} = 2.0\,\mathrm{V}$.

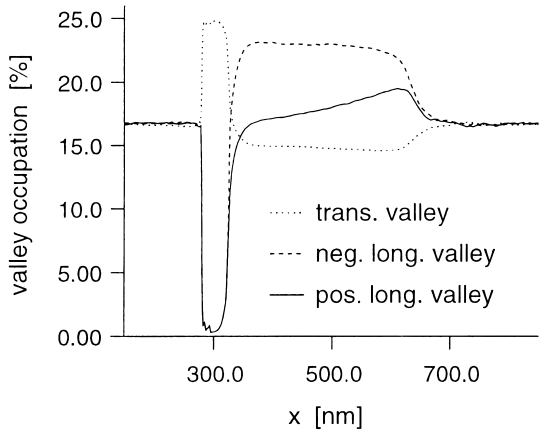

Fig. 10 Relative electron population for $V_{\mathrm{CB}} = 2.0\,\mathrm{V}$ and $V_{\mathrm{BE}} = 0.70\,\mathrm{V}$ for a transversal valley, the negative and positive longitudinal valley.

Table 2 Collector current.

V_{BE} [V]	I_C [mA/cm]	
	MC	Exp.
0.65	17.5	18.9
0.70	99.2	107.8
0.75	544.9	608.0

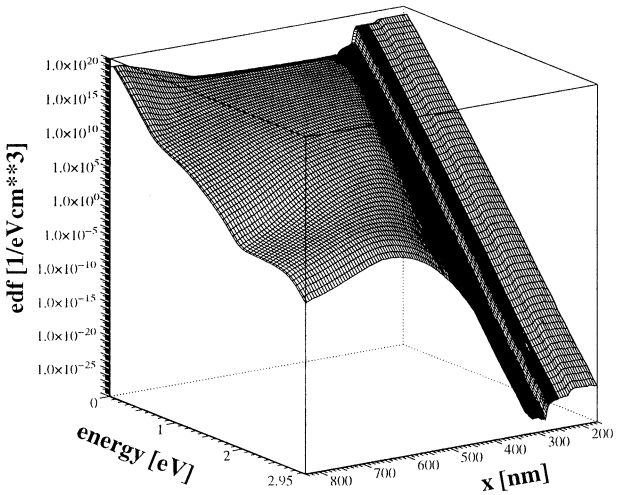

Fig. 11 Electron energy distribution function for $V_{\mathrm{CB}} = 2.0\,\mathrm{V}$ and $V_{\mathrm{BE}} = 0.75\,\mathrm{V}$.

than five orders of magnitude. Nevertheless, the statistical noise is efficiently kept small by the Multiple Refresh method.

The collector current was sampled with an estimator based on the Ramo-Shockley theorem [21]. The results are shown in Table 2 and good agreement of the simulation results with experiment is found for nearly two orders of magnitude [3], [5], [15].

It is one of the definite advantages of the FB-MC simulation that it facilitates detailed microscopic investigations, which go beyond the possibilities of the standard HD or DD device models. For example, in Fig. 10 the relative valley population of the electrons is depicted for the FB-MC simulation. In the emitter all valleys are more or less uniformly populated, because the electron ensemble is nearly in thermal equilibrium (Fig. 9). Due to the biaxial compressive strain in the base the two longitudinal valleys of the first conduction band are shifted upwards in energy, whereas the four transversal valleys are lowered [1]. Consequently, the longitudinal valleys are rarely occupied in the base region and the transversal-valleys are populated with nearly 25%. In the collector region all six valleys are again equivalent, but the strong electric field increases the population of the longitudinal valleys because of the higher longitudinal mass. In the collector contact the electron gas returns to a state near thermal equilibrium and all valleys are populated by nearly the same number of electrons.

The electron energy distribution function was calculated with the Multiple Refresh technique applied to the phase space. To this end the energy between 0.0 eV and 3.0 eV was divided into bins of 50 meV width for each cell of the real space grid. Each bin contains 8 particles. Only in the bins with the lowest energy 50 particles were used. The total number of particles was about 100000 for this 1D simulation. The resultant energy distribution function is shown in Fig. 11. Over 50 orders of magnitude are obtained with a low noise level. In the emitter the typical Maxwell-Boltzmann distribution is found, whereas in the collector region strong nonequilibrium effects are present.

5. Conclusion

To the best knowledge of the authors the first FB-MC model for SiGe devices is presented in this work. The Brillouin zone is discretized with memory efficient and at the same time very accurate nonuniform tetrahedral grids. Arbitrary Ge profiles can be resolved without excessive memory requirements. The new algorithms

for particle transport are so efficient, that the FB-MC model is as fast as models based on analytical band structures. The use of sophisticated statistical enhancement methods results in a low noise level together with a high accuracy making the FB-MC model the ideal tool to study the reliability of the simpler but more CPU efficient HD and DD models.

Acknowledgement

The authors thank Dr. A. Schüppen of TEMIC Semiconductor GmbH for helpful discussions and providing the device structure and measured data. Partial funding by the Bundesministerium für Bildung, Forschung und Technologie under contract no. 01M2416C is gratefully acknowledged.

References

[1] F. Schäffler, "High–mobility Si and Ge structures," Semicond. Sci. Technol., vol.12, pp.1515–1549, 1997.

[2] A. Schüppen, "SiGe–HBTs for mobile communication," Solid-State Electron., vol.43, pp.1373–1381, 1999.

[3] A. Schüppen, U. Erben, A. Gruhle, H. Kibbel, and H. Schumacher, "Enhanced SiGe heterojunction bipolar transistor with 160 GHz–f_{max}," IEDM Tech. Dig., pp.743–746, 1995.

[4] K. Hess, ed., Monte Carlo Device Simulation: Full Band and Beyond, Kluwer, Boston, 1991.

[5] A. Schüppen, H. Dietrich, U. Seiler, H. von der Ropp, and U. Erben, "A SiGe RF technology for mobile communication systems," Microwave Engineering Europe, pp.39–46, June 1998.

[6] C. Jungemann, S. Keith, and B. Meinerzhagen, "Full-band Monte Carlo simulation of a 0.12 μm-Si-PMOSFET with and without a strained SiGe-channel," IEDM Tech. Dig., pp.897–900, San Francisco, USA, 1998.

[7] C. Jungemann, S. Keith, M. Bartels, and B. Meinerzhagen, "Efficient full-band Monte Carlo simulation of silicon devices," IEICE Trans. Electron., vol.E82-C, no.6, pp.870–879, June 1999.

[8] F.M. Bufler, Full-band Monte Carlo simulation of electrons and holes in strained Si and SiGe, Dissertation, Universität Bremen, Bremen, 1997, (H. Utz Verlag Wissenschaft, München: 1998).

[9] D. Nuernbergk, Simulation des transportverhaltens in $Si/Si_{1-x}Ge_x$/Si–heterobipolartransistoren, Dissertation, Technische Universität Ilmenau, Ilmenau, 1999, (H. Utz Verlag Wissenschaft, München: 1999).

[10] J.W. Slotboom and H.C. de Graaf, "Measurements of bandgap narrowing in Si bipolar transistors," Solid-State Electron., vol.19, pp.857–862, 1976.

[11] D.B. M. Klaassen, J.W. Slotboom, and H.C. de Graaf, "Unified apparent bandgap narrowing in n- and p-type silicon," Solid-State Electron., vol.35, pp.125–129, 1992.

[12] R.J. E. Hueting, J.W. Slotboom, A. Pruijmboom, W.B. de Boer, C.E. Timmering, and N.E. B. Cowern, "On the optimization of SiGe-base bipolar transistors," IEEE Trans. Electron Devices, vol.43, pp.1518–1524, 1996.

[13] M.M. Rieger and P. Vogl, "Electronic-band parameters in strained $Si_{1-x}Ge_x$ alloys on $Si_{1-y}Ge_y$ substrates," Phys. Rev. B, vol.48, pp.14276–14287, 1993.

[14] R.K. Smith and J. Bude, "Highly efficient full band Monte Carlo simulations," Proc. Int. Workshop on Computational Electronics, Leeds, pp.224–230, Aug. 1993.

[15] M. Bartels, S. Decker, B. Neinhüs, K.H. Bach, A. Schüppen, and B. Meinerzhagen, "Comprehensive hydrodynamic simulation of an industrial SiGe heterobipolar transistor," Proc. BCTM, pp.105–108, Minneapolis, Minnesota, USA, 1999.

[16] P. Graf, F.M. Bufler, B. Meinerzhagen, and C. Jungemann, "A comprehensive SiGe Monte Carlo model for transient 2D simulations of HBTs," IEDM Tech. Dig., pp.881–884, 1997.

[17] M.V. Fischetti and S.E. Laux, "Monte Carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects," Phys. Rev. B, vol.38, pp.9721–9745, 1988.

[18] B. Neinhüs, S. Decker, P. Graf, F.M. Bufler, and B. Meinerzhagen, "Consistent hydrodynamic and Monte-Carlo simulation of SiGe HBTs based on table models for the relaxation times," VLSI Design, vol.8, pp.387–391, 1998.

[19] A. Pacelli and U. Ravaioli, "Analysis of variance-reduction schemes for ensemble Monte-Carlo simulation of semiconductor devices," Solid-State Electron., vol.41, pp.599–605, 1997.

[20] C. Jungemann, S. Decker, R. Thoma, W.-L. Engl, and H. Goto, "Phase space multiple refresh: A general purpose statistical enhancement technique for Monte Carlo device simulation," IEEE J. Tech. Comp. Aided Design, no.2, 1997.

[21] H. Kim, H.S. Min, T.W. Tang, and Y.J. Park, "An extended proof of the Ramo-Shockley theorem," Solid-State Electron., vol.34, pp.1251–1253, 1991.

Christoph Jungemann received the Dipl.-Ing. degree and the Dr.-Ing. degree in electrical engineering in 1990 and 1995 from the RWTH Aachen (Technical University of Aachen), Aachen, Germany. From 1990 to 1995 he was a research and teaching assistant at the "Institut für Theoretische Elektrotechnik," RWTH Aachen. From 1995 until 1997 he was employed at the Research and Development facility of Fujitsu Limited, Kawasaki, Japan. Since 1997 he is a chief engineer at the "Institut für Theoretische Elektrotechnik and Mikroelektronik," University of Bremen. His main research interests are full-band Monte Carlo simulation of Si and SiGe devices, numerical device modeling and transport in inversion layers.

Stefan Keith received the Dipl.-Ing. degree in electrical engineering in 1994 from the RWTH Aachen (Technical University of Aachen), Aachen, Germany. Since then he was a research Assistant at the "Institut für Theoretische Elektrotechnik," RWTH Aachen, and moved in 1995 as a research and teaching Assistant to the "Institut für Theoretische Elektrotechnik and Mikroelektronik," University of Bremen. His current research interests are full-band Monte Carlo device simulation of Si-MOSFETs and SiGe based heterojunction devices.

Bernd Meinerzhagen received the Dipl.-Ing. degree in electrical engineering in 1977, the Dipl.-Math. degree in mathematics in 1981, the Dr.-Ing. degree in electrical engineering in 1985 and the "venia legendi" in 1995 all from the RWTH Aachen (Technical University of Aachen), Aachen, Germany. From 1978 to 1986 as a research and teaching assistant at the RWTH Aachen he worked mainly on the development of numerical device modeling codes. In 1986 he joint AT&T Bell Laboratories in Allentown, PA, as a Member of Technical Staff, where he developed advanced numerical models for MOS substrate and gate currents. He went back to the RWTH Aachen in 1987 and became head of the research and development group for Silicon technology modeling and simulation (TCAD). In 1995 he was appointed Professor at the University of Bremen, where his current research interests include TCAD and the theory of electromagnetic fields and networks.