Iterative Local Monte Carlo Technique for the Simulation of Si-MOSFETs

Jürgen Jakumeit, Thomas Sontowski, and Umberto Ravaioli[1]

II. Physikalisches Institut
Universität Köln, Zülpicher Str. 77, 50937 Köln, Germany
Phone: +49 221 4703588, Fax: +49 221 4702980
E-mail: jakumeit@uni-koeln.de

[1]Beckman Institute and
Department of Electrical and Computer Engineering
Unversity of Illinois at Urbana-Champaign
Urbana, IL 61801, USA

1. Introduction

Evolutionary algorithms can be used to implement transport simulations where the correct distribution function is approached by steering the evolution with the knowledge of a measured parameter constraint (e.g. a substrate current measurement). In order to facilitate the convergence of the mathematical optimization at the basis of evolutionary algorithms, we introduced a physical mutation operator (PMO)[1,2], which has the task to sort the most likely physical distribution out of the many unphysical ones which could be a solution of the same optimization problem. The PMO itself, without the complete evolutionary optimization process, was found to be very useful, providing a new level in the hierarchy of transport models based on particle kinetics for the approximate solution of Boltzmann equation. For this purpose, we formulated the Mutation Operator Monte Carlo method (MOMC) [3] and we tested this approach both in the context of evolutionary algorithm optimization and as a stand alone transport simulator. Here, we investigate thie features of this transport operator in detail, showing that the MOMC is a local Monte Carlo technique which combines features and advantages of the Monte Carlo approach with the stability of iterative algorithms.

2. Theory

The main part of the MOMC is the specific PMO employed, which mutates an initial energy distribution of electrons by Monte-Carlo-like simulation steps (see Fig. 1). A test Monte Carlo (MC) electron is placed at a randomly choosen energy E_{start} and position x_{start} in the device. The drift and scattering processes of this electron are simulated in the same way as in a conventional MC-simulation. The final energy E_{final} and position x_{final} reached by the MC-electron after the simulation are memorized in a drift table. Several such MC-steps are performed for each initial state E_{start} and x_{start} of a discretization of the device region and energy space, and are memorized in the drift table. Each MC-simulation step follows the electron movement only for several femtoseconds and stops after an isotropic scattering process takes place. The MOMC technique we present here represents a transport hierarchy level below the standard Monte Carlo approach because only the energy information of the MC-electron is memorized. Since momentum information is not utilized, one cannot interrupt the electron trajectory in the middle of a free flight or after a nonisotropic scattering, since this would obviously introduce an artificial elastic scattering process. The time involved with each MC simulation step is fairly short, therefore the test electrons do not travel very far and only a local area of the device is simulated at each step.

The second part of the MOMC uses the information collected in the drift table, that provides a connection between initial and final states, to calculate the electron distribution in the whole device. At each iteration, a fraction of the electron density, in each cell of the discretized device area and enery space, is transfered to the energy distribution at one of the final positions memorized in the drift table. After application of many such iterations, any initial energy distribution of electrons converges to a steady state distribution which reflects the physical model in the Monte Carlo steps. In this way a local Monte Carlo simulation is combined with an iterative algorithm.

The MC-simulation of the electron motion is based on a full band structure model based on the ensemble constant time technique for the time of flight generation [4,5]. Scatterings due to phonon absorption and emission, impact ionization and ionized impurities as well as surface scattering are included in the model. The inputs for the Monte Carlo simulation of the PMO are the initial

0-7803-4369-7/98 $10.00

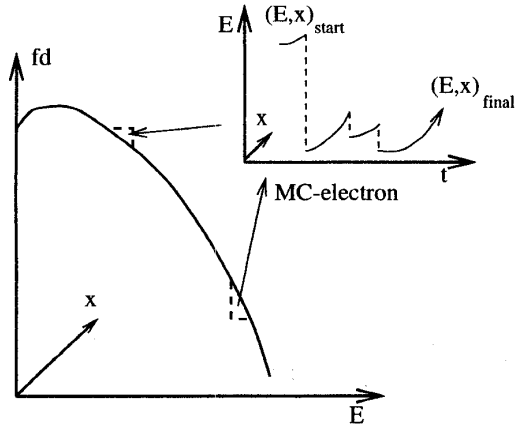

Figure 1: A schematic illustration of the of physical mutation procedure used in the MOMC

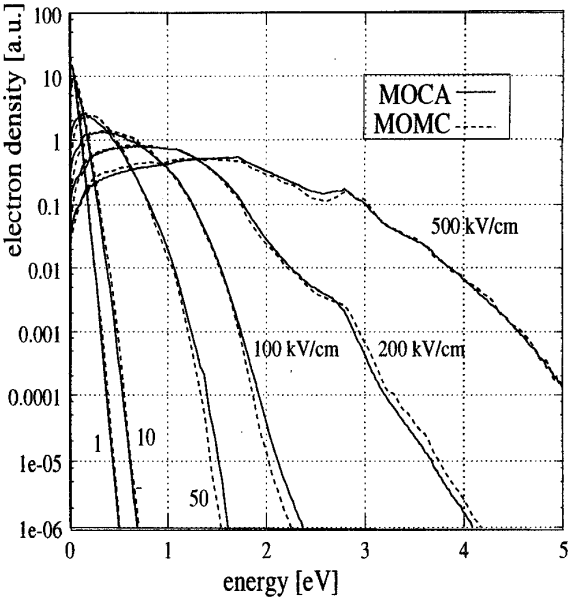

Figure 2: Results for bulk silicon (100-direction) from MOCA and the MOMC

carrier density and electric field or potential profile in the simulation region. Since our main interest is to use the MOMC to investigate hot carriers and calculate the high energy tail of the distribution, which has only a relatively small impact on the total carrier distribution, we employ here the technique as a post-processor on given field and carrier density profiles.

Theoretically, the MOMC can be seen as a conventional MC-simulation which is broken down into small separate segments, later recombined in different order by the iterative redistribution of electron densities. A main advantage of this approach is that the MC-electron is not associated with a certain electron density as in a conventional MC. Only later in the iterative step an electron density is associated with each drift process. This electron density is related to the density given by the electron distribution before the iteration. In this way the algorithm naturally assigns a high weight to particles in regions of high electron density and a low weight to particles in regions where the electron density is significantly lower, as is the case for high energy tails. Such a weighting of particles is essential for the calculation of hot electron behaviour and must be included in conventional MC-simulations by a special variance reduction technique. In the MOMC this statistical enhancement is automatically included. In more general terms, this piecewise MC-simulation allows a very efficient distribution of the computation time over the simulation area, making the MOMC significantly faster than standard Monte Carlo algorithms for the simulation of hot electron effects.

Parallelization of the MOMC is largely simplified by the fact that the local Monte Carlo steps need no information from the second part of the MOMC consisting of the iterative redistribution of charge. Therefore, the whole algorithm can be separated into: (a) computationally expensive but local Monte Carlo steps which are therefore memory inexpensive; (b)a computationally inexpensive iteration process which needs a large memory. The algorithm can be effectively parallelized on parallel computers and workstation cluster by a simple client server setup. The server forces as many clients as possible to calculate the MC-electron drift and create the corresponding drift tables. The server uses these tables for the iterative electron density redistribution.

3. Results

Figure 2 compares energy distributions of electrons calculated by the MOMC for bulk silicon (100-direction) with results obtained by MOCA, a full band Monte Carlo program developed at the University of Illinois [6]. The physical model in the two programs is essentially identical. For this example, no ionized impurity scattering is included. For fields between 1 kV/cm and 500kV/cm an excellent agreement could be achieved, demonstrating that the MOMC calculates energy distributions of electrons which are very comparable to the results of conventional MC-programs. In this simple case the MOMC is about ten times faster then MOCA in a purely scalar mode.

Results of a one-dimensional MOMC for a 75 nm Si-MOSFET are shown in Fig. 3 for 1.5 V a) or 3 V b) applied to gate and drain (source and substrate are grounded). The energy distribution of electrons along the channel calculated by the MOMC is compared with results from MOCA and Damocles (the IBM full band Monte Carlo device simulator [7]). For this test the MOMC starts with the potential profile and carrier density calculated self-consistently by MOCA as input. For

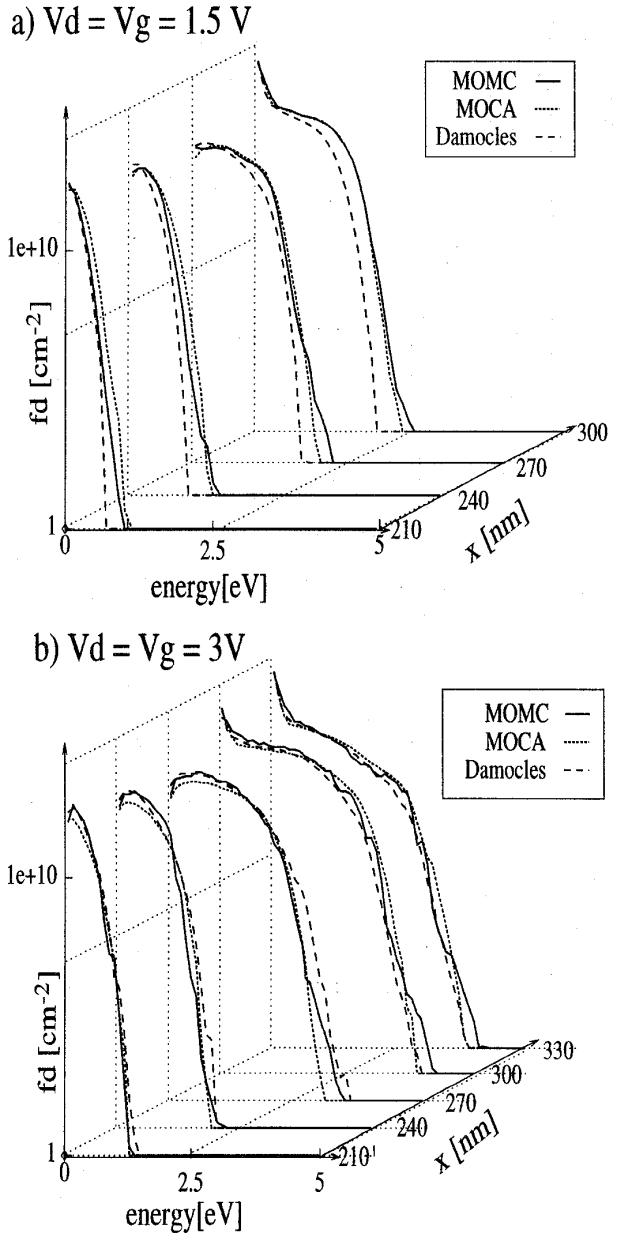

Figure 3: Comparison between results from the MC-programs Damocles and MOCA with the MOMC for an Si-MOSFET with 50 nm effective channel length. 1.5 V a) and 3 V are applied to source and drain. The transition from channel o drain is at 275 nm.

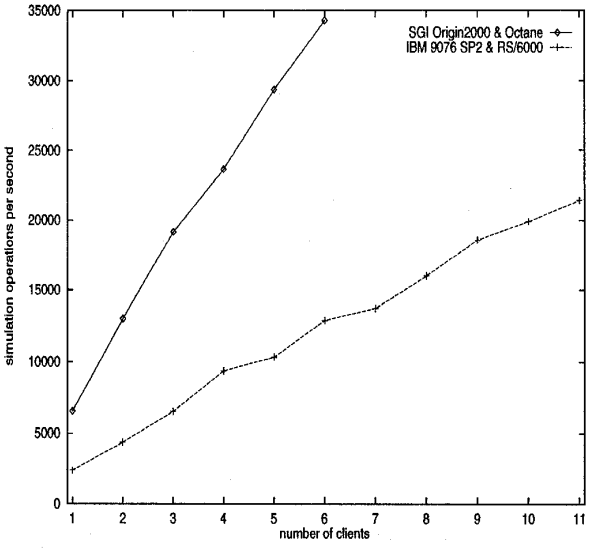

Figure 4: Almost linear speed up of the MOMC on different parallel architectures using MPI for communication

1.5 V at drain and gate the MOMC and MOCA agree very well. Damocles arrives at a slightly different solution for Poisson's equation leading to a somewhat smaller heating of the electron gas. In the case of 3 V at drain and gate (Fig. 3b) all three simulation approches give very similar distributions. Taking into account the different physical models and enhancement techniques of the three programs, the overall agreement between the three results is quite encouraging. A comparison of the simulation time is not too meaningful here, because the simulation parameters of MOCA and Damocles were not optimized towards a fast calculation of the energy distribution of electrons. In addition both programs calculate a two-dimensional selfconsistent solution and a significant amount of computation time is used for the solution of Poisson's equation. The one-dimensional results shown in Fig. 3 are calculated from the two-dimensional results by averaging over the depth of the MOSFET. In any event, the MOMC procedure is quite fast in comparison, needing less than 15 minutes on an HP-735 workstation to calculate the electron distribution shown in Fig. 3.

The parallel client server approach was tested with a two-dimensional version of the MOMC. This version is still in the testing phase, therefore a simpler physical model based on analytical bands rather than the full band structure was used. Because of this, a comparison of the electron distribution with results from full band Monte Carlo simulators like MOCA and Damocles would not be too useful at this stage, particularly for deeply scaled MOS devices. Nevertheless, computational tests are significant and give a good indication of parallel performance. The scalability of the parallel MOMC was tested on an SGI-Origin & SGI-Octane cluster and an IBM SP2 with 11 POWER2-nodes. The communication via an ATM network was performed using MPI. Figure 4

demonstrates that an almost linear speed up could be achieved on both systems. Tests on a larger SGI-Origin platform available at NCSA in Illinois, are planned for the near future. The time involved with data communication is nearly negligible. A similarly good perfomance could be achieved by a communication scheme via NFS. In this case the clients write the drift table on the file system. Such a simple parallel approach might be interesting for computational environments where MPI is not installed.

4. Conclusions

We developed a full band MOMC scheme and compared results with energy distributions of electrons obtained from conventional MC-simulators, MOCA and Damocles. The agreement for bulk silicon simulations is excellent. A one-dimensional simulation for a short channel Si-MOSFETs also shows quite comparable results. This shows that the MOMC procedure can be a valuable basic level in the hierarchy of particle-based transport models. The MOMC provides a much reduced computational cost and sufficient accuracy, which make it very attractive for fast applications. Parallelization on different platforms using a simple client server approach showed an almost linear speed up. The MOMC breaks down the conventional Monte Carlo simulation into many local MC-steps which are then drawn upon by an iteration process to build up the simulation. This piece-wise MC-approach enables one to distribute the computation effort very efficiently across the simulation area and simplifies the parallelization with respect to a traditional Monte Carlo procedure. Also, statistical enhancement of the high enery regions is automatically achieved. Complex physical models, necessary for the simulation of hot electron phenomena, like the full band structure can be included in the MOMC. We conclude that the MOMC is a very suitable approach to supplement the hierarchy of particle simulations, and in view of its efficiency the MOMC can be used as a quick evaluation tool of hot carrier trends with which to explore large variationis in the input parameter space, while leaving more physically detailed calculations to complete full band Monte Carlo models when necessary.

5. Acknowledgements

This work was partially supported by the Semiconductor Research Corporation, contract 96-CJ-816, and by the National Center for Computational Electronics (NCCE), National Science Foundation grant ECS 95-26127. The authors would like to thank IBM Corp. for making available the simulator Damocles by special arrangement with NCCE.

References

1. J. Jakumeit, U. Ravaioli, K. Hess, J. Appl. Phys. **80**, p.5061 (1996)
2. J. Jakumeit, U. Ravaioli, K. Hess, Proceedings of IWCE-4, to be published on VLSI Design
3. J. Jakumeit, A. Duncan, U. Ravaioli, K. Hess, Proceedings of IWCE-5, to be published on VLSI Design
4. B. Shapo, C. Ball, I. Kizilyalli, U. Ravaioli, Superlattices and Microstructures **4**, p.39 (1988)
5. U. Ravaioli, in *Monte Carlo Device Simulation: Full Band and Beyond*, Ed. K. Hess, Kluwer, p.267 (1991)
6. S. E. Laux, M. V. Fischetti, in *Monte Carlo Device Simulation: Full Band and Beyond*, Ed. K. Hess, Kluwer, p.1 (1991)
7. A. Duncan, U. Ravaioli, J. Jakumeit, IEEE Trans. ED, p.867 (1998)