Wigner Ensemble Monte Carlo: Challenges of 2D Nano-Device Simulation

M. Nedjalkov[1], H. Kosina[1], and D. Vasileska[2]

[1] Institute for Microelectronics, Technical University of Vienna
Gusshausstrasse 27-29/E360, A-1040 Vienna, Austria
[2] Department of Electrical Engineering,
Arizona State University, Tempe, AZ 85287-5706, USA

Abstract. We announce a two dimensional WIgner ENSemble (WIENS) approach for simulation of carrier transport in nanometer semiconductor devices. The approach is based on a stochastic model, where the quantum character of the carrier transport is taken into account by generation and recombination of positive and negative particles. The first applications of the approach are discussed with an emphasis on the variety of raised computational challenges. The latter are large scale problems, introduced by the temporal and momentum variables involved in the task.

1 Introduction

The Wigner formulation of the quantum statistical mechanics provides a convenient kinetic description of carrier transport processes on the nanometer scale, characteristic of novel nanoelectronic devices. The approach, based on the concept of phase space considers rigorously the spatially-quantum coherence and can account for processes of de-coherence due to phonons and other scattering mechanisms using the models developed for the Boltzmann transport. Almost two decades ago the coherent Wigner equation has been utilized in a deterministic 1D device simulators [3,4,1]. The latter have been refined towards self-consistent schemes which take into account the Poisson equation, and dissipation processes have been included by using the relaxation time approximation. At that time it has been recognized that an extension of the deterministic approaches to two dimensions is prohibited by the enormous increase of the memory requirements, a fact which remains true even for todays computers. Indeed, despite the progress of the deterministic Boltzmann simulators which nowadays can consider even 3D problems, the situation with Wigner model remains unchanged. The reason is that, in contrast to the sparse Boltzmann scattering matrix, the counterpart provided by the Wigner potential operator is dense. A basic property of the stochastic methods is that they turn the memory requirements of the deterministic counterparts into computation time requirements. Recently two Monte Carlo methods for Wigner transport have been proposed [7,5]. The first one has been derived by an operator splitting approach. The Wigner function is presented by an ensemble of particles which are advanced in the phase space and carry

I. Lirkov, S. Margenov, and J. Waśniewski (Eds.): LSSC 2007, LNCS 4818, pp. 139–147, 2008.
© Springer-Verlag Berlin Heidelberg 2008

the quantum information via a quantity called affinity. The latter is updated at consecutive time steps and actually originates from the Wigner potential, whose values are distributed between particles according their phase space position. This ensemble method has been applied in a self-consistent scheme to resonant-tunneling diodes (RTD's), the scattering with phonons is accounted in a rigorous way [7]. Recently it has been successfully extended to quasi two dimensional simulations of double gate MOSFET's [6]. The second method is based on a formal application of the Monte Carlo theory on the integral form of the Wigner equation. The action of the Wigner potential is interpreted as generation of couples of positive and negative particles. The quantum information is carried by their sign, all other aspects of their evolution including the scattering by phonons are of usual Boltzmann particles. The avalanche of generated particles is controlled by the inverse process: two particles with opposite sign entering given phase space cell annihilate. The approach offers a seamless transition between classical and quantum regions, a property not yet exploited for practical applications.

WIENS is envisaged as an union of theoretical and numerical approaches, algorithms and experimental code for 2D Wigner simulation of nanostructures. In contrast to device simulators which, being tools for investigation of novel structures and materials rely on well established algorithms, WIENS is comprised by mutually related elements which must be developed and tested for relevance and viability. Many open problems need to be addressed such as the choice of the driving force in the Wigner equation, pure quantum versus mixed classical-quantum approaches, the correct formulation of the boundary conditions, appropriate values for the parameters and a variety of possible algorithms. We present the first results in this direction. In the next section a semi-discrete formulation of the Wigner equation for a typical MOSFET structure is derived. An Ensemble particle algorithm is devised in the framework of the second approach. Next, simulation experiments are presented and discussed. It is shown that, despite the nanometer dimensions, the temporal and momentum scales introduce a large scale computational problem.

2 Semi-discrete Wigner Equation

A system of carriers is considered in a typical 2D structure, for example of a MOSFET shown in Fig. 1. The device shape is a perfect rectangle with the highly doped Source and Drain regions at the left and right bottom ends respectively. At the very bottom, to the left and right of the Gate are shown parts of the leads. These supply carriers and thus specify the boundary conditions in the two tinny strips marked in black. It is assumed that the current flows between the device and the leads only, that is, at the rest of the device boundary including the region under the gate carriers are reflected. The state of the carriers is characterized by a wave function which becomes zero at and outside these boundaries. Accordingly the density matrix $\rho(\mathbf{r}_1, \mathbf{r}_2)$, will vanish if some of the arguments $\mathbf{r} = (x, y)$ is outside of the rectangle $(0, \mathbf{L}) = (0, L_x); (0, L_y)$ determined by the device dimensions. This condition holds everywhere but on the leads. We postpone the

discussion of the leads and first introduce the Wigner function as obtained from the density matrix by the continuous Wigner-Weyl transform:

$$f_w(\mathbf{r},\mathbf{k},t) = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty} d\mathbf{s} e^{-i\mathbf{k}\mathbf{s}} \rho(\mathbf{r}+\frac{\mathbf{s}}{2},\mathbf{r}-\frac{\mathbf{s}}{2},t); \tag{1}$$

where $\mathbf{r} = \frac{\mathbf{r}_1+\mathbf{r}_2}{2}$, $\quad \mathbf{s} = \mathbf{r}_1 - \mathbf{r}_2$.

The condition $0 < \mathbf{r}_1, \mathbf{r}_2 < \mathbf{L}$, (which holds everywhere but on the leads) gives rise to the following condition for $\mathbf{s}$:

$$-\mathbf{L}_c < \mathbf{s} < \mathbf{L}_c \qquad \mathbf{L}_c = 2\min\left(\mathbf{r}, (\mathbf{L}-\mathbf{r})\right). \tag{2}$$

The confinement of $\mathbf{s}$ allows to utilize a discrete Fourier transformation in the definition (1). Two details must be adjusted. $\mathbf{L}_c$, which is actually the coherence length, depends on $\mathbf{r}$. Thus the discretization of the wave vector space will change with the position which is inconvenient. Fortunately, since any partially continuous function of $\mathbf{s}$ defined in given interval $(\mathbf{a},\mathbf{b})$ can be presented by the Fourier states of a larger interval, we can conveniently fix the coherence length to the maximal value of the minimum in (2): $\mathbf{L}_c = \mathbf{L}$. As the values of the function outside $(\mathbf{a},\mathbf{b})$ are extrapolated to 0, we must formally assign to ρ a domain indicator $\theta_D(\mathbf{s})$, which becomes zero if (2) is violated. However, according the above considerations, θ_D is implicitly included in ρ. From a physical point of view the problem whether to consider the leads or not is still open: Usually leads are included in 1D simulations and the integral in (1) is truncated somewhere deep in them for numerical reasons. In a recent study [2] it is argued that in the leads there are processes which entirely destroy the quantum interference between the point $\mathbf{r}-\mathbf{s}$, Fig. 1, and the corresponding counterpart $\mathbf{r}+\mathbf{s}$. In this way the integral in (1) has to be truncated already at the point where the line along $\mathbf{s}$ enters into the gate (G) region. Alternatively, if leads are considered, the segment of this line lying in the gate region (and the corresponding counterpart in the device) are excluded from the integral. The correlation is enabled after the point of entering into the leads. However, even in this case, $\mathbf{s}$ remains bounded.

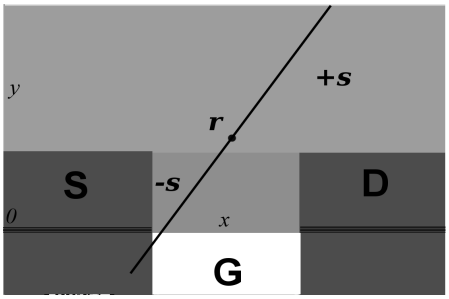

Fig. 1. Typical 2D structure

In this case $\mathbf{L}_c$ must be augmented to $(0, L_x); (-L_y, L_y)$. By changing $\mathbf{s}$ to $2\mathbf{s}$ and applying the discrete Fourier transform:

$$f(\mathbf{r}, \mathbf{n}) = \frac{1}{\mathbf{L}_c} \int_{\mathbf{L}_c/2}^{\mathbf{L}_c/2} d\mathbf{s} e^{i2\mathbf{n}\Delta\mathbf{k}\mathbf{s}} \rho(\mathbf{r}, \mathbf{s})$$

$$\rho(\mathbf{r}, \mathbf{s}) = \sum_{\mathbf{n}=-\infty}^{\infty} e^{i2\mathbf{n}\Delta\mathbf{k}\mathbf{s}} f(\mathbf{r}, \mathbf{n})$$

$$\Delta\mathbf{k} = \pi/\mathbf{L}_c$$

to the von-Neumann equation for the density matrix we obtain a Wigner equation which is continuous in space and discrete in momentum:

$$\frac{\partial f(\mathbf{r}, \mathbf{M}, t)}{\partial t} + \frac{\hbar}{m} \mathbf{M}\Delta\mathbf{k} \frac{\partial f}{\partial \mathbf{r}}(\mathbf{r}, \mathbf{M}, t) = \sum_{\mathbf{m}=-\infty}^{\infty} V_w(\mathbf{r}, \mathbf{m}) f_w(\mathbf{r}, (\mathbf{M} - \mathbf{m}), t).$$

Here the Wigner potential is defined as:

$$V_w(\mathbf{r}, \mathbf{M}) = \frac{1}{i\hbar} \frac{1}{\mathbf{L}_c} \int_{-\mathbf{L}_c/2}^{\mathbf{L}_c/2} d\mathbf{s} e^{-i2\mathbf{M}\Delta\mathbf{k}\mathbf{s}} (V(\mathbf{r} + \mathbf{s}) - V(\mathbf{r} - \mathbf{s})) \theta_D(\mathbf{s}) \tag{3}$$

We note the presence of the domain indicator in this definition. According the particle sign approach [5], the Wigner potential generates particles with a frequency given by the Wigner out-scattering rate γ obtained from (3):

$$\gamma(\mathbf{r}) = \sum_{\mathbf{M}=0}^{\infty} |V_w(\mathbf{r}, \mathbf{M})| \tag{4}$$

The shape and magnitude of γ strongly depend on the treatment of the leads, as it will be shown in what follows.

3 Numerical Aspects and Simulations

The above theoretical considerations are presented for a coherent transport, where the interaction with phonons is switched off. As the corresponding coherent algorithm is the core module of WIENS, it must be carefully developed and tested. We furthermore focus on the stationary case, where the boundary conditions control the carrier transport. The initial picture of the algorithm under development is of an ensemble of particles which is evolved in the phase space at consecutive time steps. The boundary conditions are updated after each step in the usual for device simulations way.

The magnitude of γ is of order of $[10^{15}/s]$ so that the avalanche of particles does not allow an individual treatment of each particle as in the classical Ensemble Monte Carlo algorithm. Particles must be stored on grid points of a mesh in the phase space, where the inverse process of annihilation occurs. Thus along

with the wave vector spacing $\Delta\mathbf{k}$, also the real space must be divided into cells of size $\Delta\mathbf{r}$. Actually two arrays are needed to store the particles.

At the beginning of an evolution step the initial one, f_1, is occupied by particles, while the second one, f_2, is empty. Particles are consecutively taken from f_1, initiating from randomly chosen phase space coordinates around the corresponding grid point. A selected particle evolves untill the end of the time step and then is stored in f_2. The stored wave vector corresponds to the initial value since there is no accelerating field. The particle gives rise to secondary, ternary etc. particles which are evolved in the phase space for the rest of the time and then stored in f_2. As they are generated on the same grid in the wave vector space, the assignment in straightforward. As a rule the particles injected by the boundary conditions are slow (low wave vector), while these generated by the Wigner potential are fast. The wave vector ranges over several orders of magnitude so that the task for the position assignment becomes a large scale problem: A straightforward approach is to assign the particle position to the nearest grid point. For fast particles which cross several cells during the time step this introduces small error in the spatial evolution. More dramatic is the situation with the slow particles: if during the time step a particle crosses a distance less than a half of the mesh step it can remain around a grid point for a long time. This is an example for artificial diffusion which is treated in the following way:

(i) Slow particles, e.g., these which belong to the ground cell around the origin of the wave vector are treated in a standard ensemble approach: the phase space evolution is followed continuously throughout the device.
(ii) The grid assignment is chosen stochastically, according a probability proportional to the distance to the neighborhood grid points.

Another large-scale aspect is related to the range of the time constants involved in the problem. The existence of very fast particles imposes an evolution step of few hundreds of femtosecond. The time step t_a between successive assignments to the grid is of order of femtosecond, while the total evolution must be above picosecond in order to reach the stationary conditions. Accordingly large computational times are expected. Thus the first objective of the computations is to investigate the convergence, to optimize where possible the algorithm and to find appropriate values for the parameters leading to stable results. The latter are a necessary condition for solving the physical aspects of the problem: finding a proper normalization, a choice of the boundary conditions and investigation of the quantum phenomena will be focused on a next step.

Two structures, A and B of the type shown on Fig. 1 are considered in the experiments. A Γ valley semiconductor with a small effective mass (0.036) has been chosen to enhance the effects of tunneling. The potential and dimensions of device A are twice as small as compared to device B.

The potential of B is shown in Fig. 2 and Fig. 3 as a contour plot respectively. The potential is obtained from the self-consistent Boltzmann-Poisson solution: In the quantum case the entire potential is used to calculate $\gamma(\mathbf{r})$ and the scattering probability table. The driving force is zero so that the carriers perform a free motion throughout the device until the reflecting boundaries.

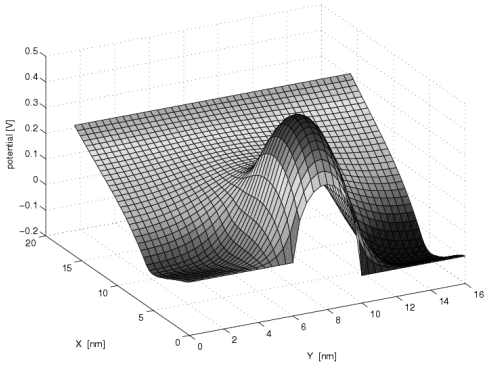

Fig. 2. Device potential

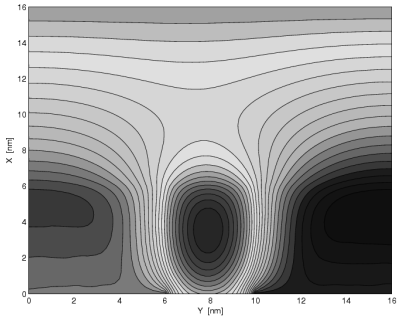

Fig. 3. Contour plot of the potential

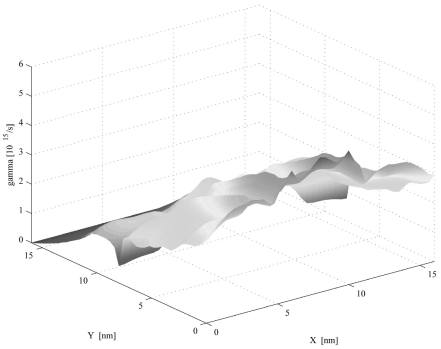

Fig. 4. $\gamma(\mathbf{r})$, device B

Fig. 4 shows $\gamma(\mathbf{r})$ computed for the case including the leads. At the injecting boundaries in the source and drain regions (compare Fig. 1) the generation rate is very high so that an injected particle feels the Wigner potential already at the boundary. On contrary, if leads are excluded the generation rate is zero at this boundary.

Figures 5 and 6 show γ in device A, computed for either of the two cases. There is a profound difference of the shape and magnitude of this quantity on the two pictures. The contour plots of the classical and quantum densities in device A, no leads considered, are compared in Fig. 7 and 8. Both densities have the same shape, however the quantum counterpart is more spread inside the device which can be related to tunneling effects. We note that the quantum density is due to effects of generation and recombination only, so that the basic similarity was the first encouraging result. In particular, since the small generation rate $\simeq 10^{14}/s$ the time $t_a$between successive assignments is a femtosecond, so that the effect of artificial diffusion is negligible. Here the algorithms treating this effect have

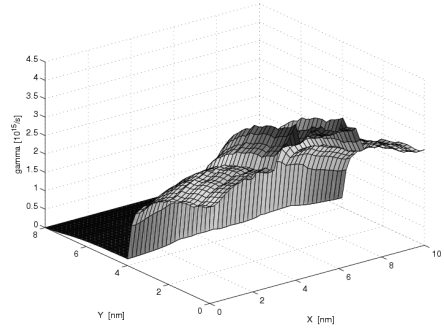

Fig. 5. $\gamma(\mathbf{r})$, device A

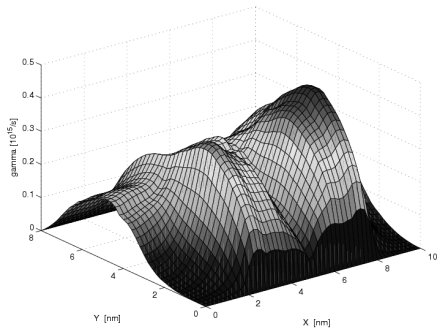

Fig. 6. $\gamma(\mathbf{r})$, device A, leads are excluded

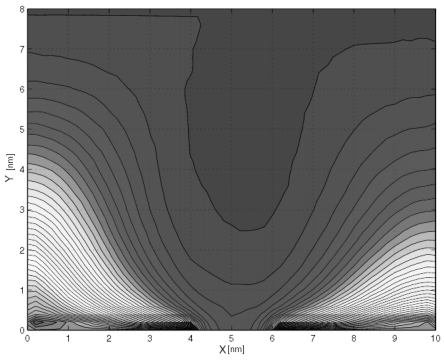

Fig. 7. Boltzmann carrier density

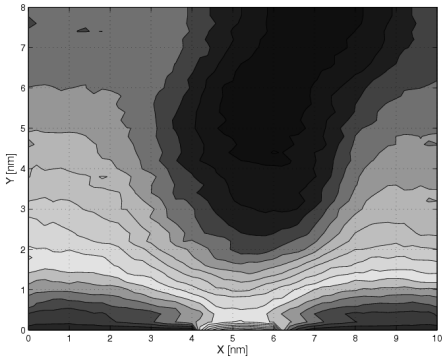

Fig. 8. Wigner carrier density

been introduced and tested by reducing t_a. Fig. 9 demonstrates the much slower convergence of the quantum current as a function of the evolution time.

The potential of device B, Fig. 3, has a shape of two valleys separated by a high barrier with a maximum between the valleys. Between this maximum and the high potential at the bottom of the base there is a saddle point. The carrier density is expected to follow this pattern: carriers fill the valleys while their number should decrease with the raise of the potential. This is essentially the behavior of the quantum densities at Fig. 10 and Fig. 11. The former is obtained for a spatial step of 0.2nm giving rise to a Wigner function with 41.10^6 elements. Since the great number of particles the evolution time reached after 30 days of CPU time of a regular 2.5 GHz PC is one picosecond only. Fig. 11 corresponds to a 0.4nm step: the dimension of the Wigner function is around one order of magnitude smaller and the convergence is as much faster. The use of such step became possible due to the algorithms avoiding the artificial diffusion. A difference in the normalization factor along with some nuances in the shape exist, e.g., in the density around the saddle point and the position of the left peak.

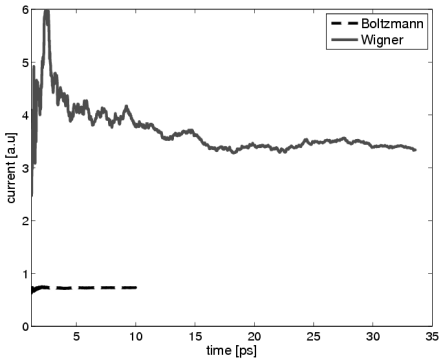

Fig. 9. Convergence of the current

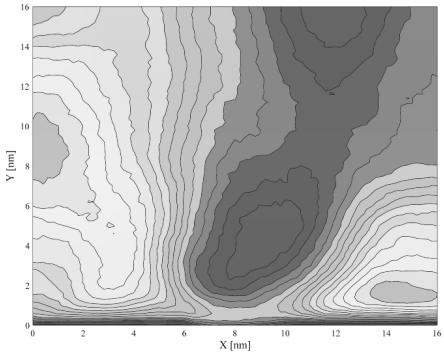

Fig. 10. Carrier density for a 0.2 nm mesh

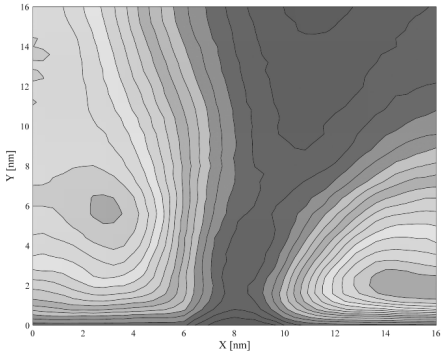

Fig. 11. Carrier density for a 0.4 nm mesh

Several factors can be responsible for this difference: the annihilation mesh is different, the number of the used **k** states, the option that Fig. 10 is not yet in a stationary state. These problems can not be answered without implementation of MPI and GRID technologies which is currently underway.

4 Conclusions

A stochastic approach is developed within a semi-discrete Wigner-Weyl transform, for which the problem of 2D Wigner transport is not an impossible numerical task. The obtained first results are qualitative and mainly demonstrate the convergence, which, furhtermore characterizes a large-scale computational problem. MPI and GRID technologies must be implemented to address the physical aspects such as proper boundary conditions, normalization and resolution of the incorporated quantum phenomena.

Acknowledgment

This work has been partially supported by the Österreichische Forschungsgemeinschaft (ÖFG), Project MOEL239.

References

1. Biegel, B., Plummer, J.: Comparison of self-consistency iteration options for the Wigner function method of quantum device simulation. Phys. Rev. B 54, 8070 (1996)
2. Ferrari, G., Bordone, P., Jacoboni, C.: Electron Dynamics Inside Schort-Coherence Systems. Physics Letters A 356, 371 (2006)
3. Frensley, W.R.: Boundary conditions for open quantum systems driven far from equilibrium. Rev. of Modern Physics 62, 745 (1990)
4. Kluksdahl, N.C., et al.: Self-consistent study of resonant-tunneling diode. Phys. Rev B 39, 7720 (1989)
5. Nedjalkov, M., et al.: Unified particle approach to Wigner-Boltzmann transport in small semiconductor devices. Phys. Rev. B 70, 115319 (2004)
6. Querlioz, D., et al.: Fully Quantum Self-Consistent Study of Ultimate DG-MOSFETs Including Realistic Scattering Using Wigner Monte Carlo Approach, IEDM — Int.Electron Devices Meeting (2006)
7. Shifren, L., Ringhofer, C., Ferry, D.K.: A Wigner Function-Based Quantum Ensemble Monte Carlo Study of a Resonant Tunneling Diode. IEEE Trans.Electron Devices 50, 769 (2003)