Comparison of Numerical Quantum Device Models

Hans Kosina*, Gerhard Klimeck†, Mihail Nedjalkov*, and Siegfried Selberherr*

*Institute for Microelectronics, TU Vienna, Gusshausstrasse 27-29, Vienna, Austria
Telephone: +43-1-58801/36013, Fax: +43-1-58801/36099, Email: Kosina@iue.tuwien.ac.at

†Jet Propulsion Laboratory, California Institute of Technology
MS 169-315, 4800 Oak Grove Dr., Pasadena CA 91109, USA

***Abstract*— The Wigner equation and non-equilibrium Green's functions are two formalisms widely used in quantum device simulation. The Wigner equation, commonly solved by finite difference methods, is solved in this work by a recently developed Monte Carlo method. This method resolves both quantum interference and dissipation effects due to scattering with equal accuracy. Both limits, namely the pure quantum ballistic case and the scattering-dominated classical case are treated properly. A comparison of the Wigner MC solver and NEMO-1D is presented. Resonant tunneling diodes from the literature are chosen as benchmark devices. Current/voltage characteristics are compared for different temperatures and the effect of scattering on the current and the charge distribution is shown. Practical device simulation limitations of the Wigner MC method are discussed. Provided that numerical parameters of the Wigner MC method such as the coherence length and the grid size are chosen properly, results are obtained in good quantitative agreement with NEMO-1D.**

I. Introduction

At room temperature the electrical characteristics of nano-electronic and highly down-scaled microelectronic devices are influenced simultaneously by classical and quantum transport effects. Physical models capable of describing this mixed transport regime are given by the non-equilibrium Green's function (NEGF) formalism and the Wigner transport equation.

Based on the NEGF formalism, NEMO-1D [1] has served as a quantitatively predictive design and analysis tool for resonant tunneling diodes (RTDs). Such devices have been studied at room temperature including the dominant effects of band-structure [2] and at low temperatures including dominant scattering effects[1][3].

On the other hand, the Monte Carlo (MC) method is nowadays a well established, reliable and accurate numerical method for solving the Boltzmann equation. Because of the similarity of the Boltzmann equation and the Wigner equation it appears very promising to develop a MC method also for the solution of the latter equation. A particle-based method has the advantage that scattering processes can be included straightforwardly. MC approaches to solve the Wigner equation have been reported recently [4][5][6][7]. The major problem to overcome is that the Wigner potential does not represent a positive definite function. This so-called negative sign problem generally calls for the introduction of particles of negative statistical weight. A consequence of the negative sign problem is that even for a system of non-interacting particles the MC method has to include inter-particle interactions, allowing a transfer of, for example, the negative weight of one particle to the positive weight of another particle in order to achieve weight cancellation. Otherwise, if such a mechanism is not included, the MC method can be shown to be instable [7]. The particle weights of either sign grow exponentially at a very high rate, and because of the large degree of cancellation in the estimators the variance would also grow exponentially.

II. The Particle Model

The space-dependent Wigner equation for electrons including the Boltzmann collision operator $Q[f_w]$ reads

$$\left(\frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla_r + q\mathbf{E}\cdot\nabla_k\right) f_w = Q[f_w] + \Theta_w[f_w]\,. \qquad (1)$$

The classical force term $q\mathbf{E}$ is separated from the Wigner potential,

$$V_w(\mathbf{r},\mathbf{k}) = \int \frac{d\mathbf{s}}{2\pi i\hbar}\, e^{-i\mathbf{k}\cdot\mathbf{s}} \left(V(\mathbf{r}+\tfrac{\mathbf{s}}{2}) - V(\mathbf{r}-\tfrac{\mathbf{s}}{2}) + q\mathbf{s}\cdot\mathbf{E}\right)$$

and thus appears on the left hand side of (1) [5]. Because the Wigner potential assumes positive and negative values, it cannot directly be used as a probability density. However, the antisymmetry of V_w with respect to $\mathbf{k}$ allows the potential operator to be expressed solely in terms of the truncated Wigner potential [7], $V_w^+(\mathbf{k}) = \mathrm{Max}(0, V_w(\mathbf{k}))$, which is positive definite and thus amenable to a probabilistic interpretation. Expressing the Liouville operator in (1) as a total time derivative and writing the operators on the right hand side explicitly gives

$$\begin{aligned}\left[\frac{\mathrm{d}f_w}{\mathrm{d}t} + \underbrace{(\lambda+\alpha)}_{\text{out-scatt.}} f_w\right](\mathbf{k},\mathbf{r},t) = \\ \int f_w(\mathbf{k}',\mathbf{r},t)\,\underbrace{\left[S(\mathbf{k}',\mathbf{k}) + \alpha\,\delta(\mathbf{k}'-\mathbf{k})\right]}_{\text{in-scatt.}}\, d\mathbf{k}' \\ + \int \underbrace{V_w^+(\mathbf{q},\mathbf{r})\, f_w(\mathbf{k}-\mathbf{q},\mathbf{r},t)}_{\text{gen. pos. part.}}\, d\mathbf{q} \\ - \int \underbrace{V_w^+(\mathbf{q},\mathbf{r})\, f_w(\mathbf{k}+\mathbf{q},\mathbf{r},t)}_{\text{gen. neg. part.}}\, d\mathbf{q}\,. \qquad (2)\end{aligned}$$

0-7803-7826-1/03/$17.00

The three characteristic rates in this equation are the semiclassical scattering rate, $\lambda(\mathbf{k}) = \int S(\mathbf{k},\mathbf{k}')d\mathbf{k}'$, a self-scattering rate to be determined later, $\alpha(\mathbf{k},\mathbf{r}) \geq 0$, and a rate associated with the Wigner potential, $\gamma(\mathbf{r}) = \int V_w^+(\mathbf{r},\mathbf{k})\, d\mathbf{k}$. Equation (2) is now interpreted as a Boltzmann equation, where in- and out-scattering processes are exactly balanced, augmented by a generation term for positive particles and one for negative particles. Note that an interpretation of the very last term in (2) as an out-scattering term is ruled out by its non-locality in momentum space.

III. The Monte Carlo Method

In the same way as for the Boltzmann equation, a formal integration in time gives a path-integral equation for the Wigner function f_w. From the adjoint integral equation one can then derive forward MC algorithms. Various probabilities and probability densities employed in the MC algorithm can be directly identified from the integral-differential form (2). Introducing the rate $\mu = \lambda + \alpha$, which will determine the free-flight duration, and the normalized distributions $S(\mathbf{k}',\mathbf{k})/\lambda(\mathbf{k}')$ and $V^+(\mathbf{q},\mathbf{r})/\gamma(\mathbf{r})$, (2) is reformulated as:

$$\left[\frac{\mathrm{d}f_w}{\mathrm{d}t} + \mu\, f_w\right](\mathbf{k},\mathbf{r},t) = \int d\mathbf{k}'\, \mu(\mathbf{k}') f_w(\mathbf{k}',\mathbf{r},t) \times \left[\left\{\frac{\lambda}{\mu}\right\}\left\{\frac{S(\mathbf{k}',\mathbf{k})}{\lambda(\mathbf{k}')}\right\} + \left\{\frac{\alpha}{\mu}\right\}\left\{\delta(\mathbf{k}'-\mathbf{k})\right\} + \left\{\frac{\gamma}{\mu}\right\}\left(\left\{\frac{V_w^+(\mathbf{k}-\mathbf{k}')}{\gamma(\mathbf{r})}\right\} - \left\{\frac{V_w^+(\mathbf{k}'-\mathbf{k})}{\gamma(\mathbf{r})}\right\}\right)\right]. \quad (3)$$

In this equation all probabilities and probability distributions are enclosed in curly brackets. Now we chose α such that $\mu \geq \gamma$. Typical choices are $\mu = \lambda + \gamma$ or $\mu = \mathrm{Max}(\lambda,\gamma)$. Free flights are interrupted at a rate μ. At the end of a free flight one selects from the complementary probabilities λ/μ and α/μ either semiclassical or self scattering, and selects the final state $\mathbf{k}$ for a given initial state $\mathbf{k}'$ from S/λ or the δ-function, respectively. Additionally, with probability γ/μ one generates a pair of particles, whose signs are opposite. From the adjoint integral equation (not presented here) it can be seen, that the two generated states are $\mathbf{k}'+\mathbf{q}$ and $\mathbf{k}'-\mathbf{q}$, where $\mathbf{q}$ is selected from the density $V_w^+(\mathbf{q},\mathbf{r})/\gamma(\mathbf{r})$. Apparently, for $\gamma = 0$ no particles need to be generated and because of $\gamma/\mu = 0$, the self-scattering rate α can be chosen freely. In particular, α can be set to zero, such that the classical MC method is regained. The particle model and the associated probability distributions describe the general, time and space-dependent case. In this work we restrict ourselves to the stationary transport problem, for which a single-particle MC method is obtained.

Assuming that extended contact regions with high doping concentration are included in the simulation domain, one can safely neglect quantum effects in these regions and apply a classical distribution at the metal/semiconductor contact. As in the classical MC method, a particle is injected at a contact from a classical distribution and undergoes a sequence of accelerated free flights and scattering. The free-flight time is determined by the rate μ. In regions where the Wigner potential and hence the pair generation rate γ are non-zero, pairs of numerical particles are generated according to the generation terms in (2). Therefore, after each generation event one has to deal with three states, namely the generated ones, $\mathbf{k}'+\mathbf{q}$, $\mathbf{k}'-\mathbf{q}$, and the after-scattering state $\mathbf{k}$ generated from the semi-classical scattering operator (the term "in-scatt" in (2)). Because repeating this step will lead to an exponential increase in particle number, an additional measure has to be introduced to control the number of particles.

In the stationary MC algorithm developed such a measure is applied after each generation event, removing two in three particle states and continuing trajectory construction from the remaining state. The idea is that two particles of opposite weight and a sufficiently small distance in phase space can be assumed to annihilate each other. In a stationary method a phase space grid can be utilized, on which particle states are stored temporarily. Due to stationarity a particle stored in a cell at some time can be annihilated by a particle of opposite sign visiting the same cell at any other time. In the present algorithm this idea is realized as follows. First the weights of all three states are stored on the grid. Then the weights in the three cells are compared. The cell with the largest absolute weight is selected and the associated particle state is used to continue the trajectory. The sign of the particle weight is chosen such that the weight in the cell gets reduced. In this way, particle trajectories are constructed sequentially in the simulation. Furthermore, the method of selecting the continuing particle aims at compensating the local weight stored on the grid as much as possible. The residual weight on the grid has to be minimized as it is an indicator for the numerical error of the method. Simulations show that the weights on the grid cancel to a large extent.

IV. Results

Two benchmark devices have been chosen, where Device 1 is a resonant tunneling diode (RTD) described in [8]. Assuming a lattice constant of $0.565\ \mathrm{nm}$, the tunnel barriers are 5 mono-layers thick and 0.27 eV high, and the quantum well is 8 mono-layers wide. In the Wigner MC simulation, the length of the simulation domain was 178 mono-layers or 100.6 nm. The doping concentration in the contact regions is $2 \cdot 10^{18}\ \mathrm{cm}^{-3}$. The potential changes linearly only in the barrier/well/barrier regions and is constant in the contact regions.

In the Wigner MC simulation of Device 1 the coherence length is chosen to be $62.15\,\mathrm{nm}$, which corresponds to 110 lattice constants. Because of the rather large coherent off-resonant valley current of this RTD phonon scattering has only little effect on the I/V characteristics. Both simulators predict a slight increase in valley current due to inelastic scattering (Fig. 1). For the coherent case I/V characteristics at 77K and 300K are shown in Fig. 2. Compared with NEMO–1D, somewhat higher peak and valley currents are obtained by Wigner MC. The resonance voltages predicted by the two solvers agree very well. The coherence length has to be selected carefully when solving the Wigner equation numerically. The comparison of I/V characteristics shown in

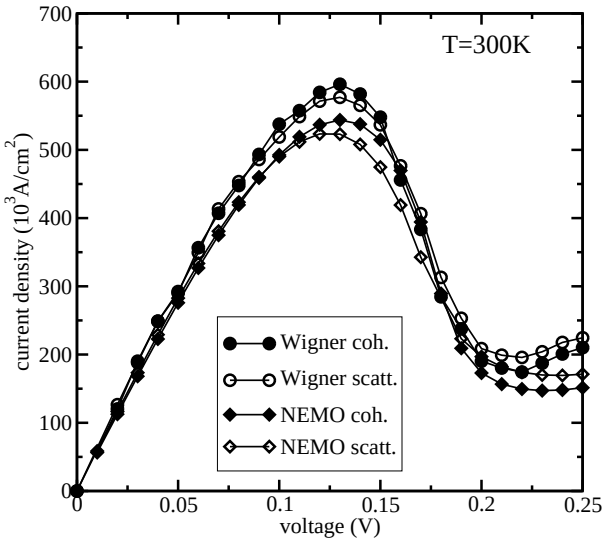

Fig. 1. I/V characteristics of Device 1 at 300K obtained from Wigner MC and NEMO-1D. Transport is coherent (coh.) or dissipative (scatt). Both simulators predict a small effect of phonon scattering on the current.

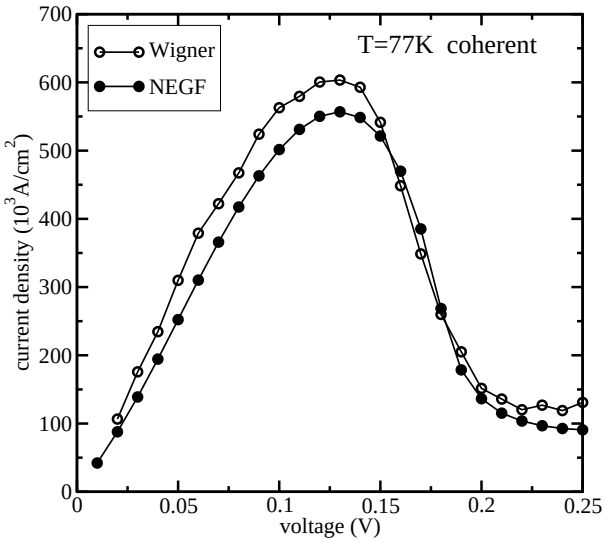

Fig. 2. I/V characteristics of Device 1 at 77K obtained from Wigner MC and NEMO–1D.

Fig. 3 demonstrates that only a sufficiently large coherence length gives a realistic result. A too short coherence length results in an overestimation of the valley current.

The layer structure of Device 2 is taken from [4], assuming a barrier height of 0.3 eV, a barrier width of 3 nm, and a well width of 5nm. The potential changes linearly in a region starting 10 nm before the emitter barrier and extending 19 nm after the collector barrier. In the contact regions the doping is 10^{16} cm^{-3} and a constant potential is assumed. In both devices the effective mass is $0.067m_0$ independent of position. The electron concentration in Device 2 is plotted in Fig. 4. At $0.3V$ this device is close to off-resonance. Both simulators predict a significant increase in electron concentration in front of the first barrier and in the quantum well when phonon

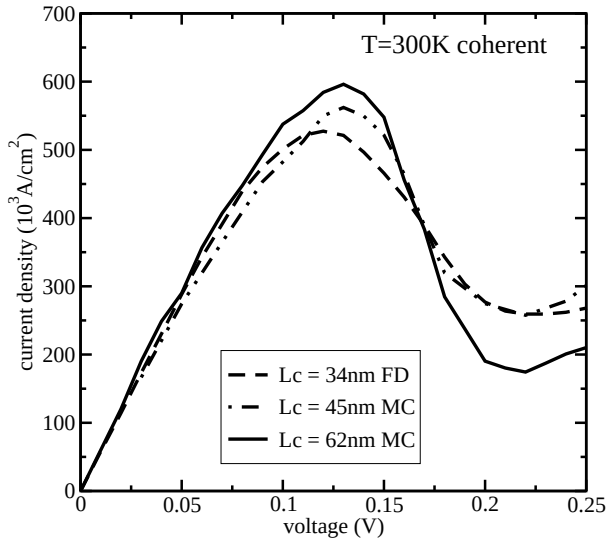

Fig. 3. Effect of the coherence length on the I/V characteristics in Wigner simulations. The finite difference (FD) result is taken from [8].

scattering is switched on. This effect can be understood as follows. With the assumed piece-wise linear potential profile a triangular potential well forms in front of the first barrier. In this emitter notch a quasi bound state forms. In the coherent case electrons reside in states above the emitter band edge and cannot occupy the lower notch state. With inelastic scattering, however, the notch state can be populated which increases the density in the emitter notch. As a consequence, the density in the quantum well is significantly higher and so is the valley current.

V. Discussion

These examples demonstrate that a numerical solver based on the Wigner equation can provide quantitatively correct results. One requirement is that the coherence length is chosen sufficiently large. The completeness relation of the discrete Fourier transform, which reflects Heisenberg's uncertainty principle, $\Delta k_x = \pi/L_c$ shows that a small coherence length L_c will result in a coarse grid in momentum space, and resonance peaks might not be resolved properly. In the past the Wigner equation has been solved most frequently by finite difference methods. Due to the non-locality of the potential operator all points in momentum space are coupled and the sparsity pattern of the matrix is very poor. Therefore, increasing the number of grid points in k-space, related to the coherence length by $N_k = L_c/\Delta x$, can easily lead to prohibitive memory and computation time requirements. This might be one reason why quantitatively correct solutions were difficult to obtain in the past.

The MC method allows the number of k-points to be increased. In this work the Wigner potential is discretized using approximately $N_k = 10^3$ points. However, high performance RTDs with very high peak/valley current ratio pose a severe problem for the MC method. In such a device the density can vary over several orders of magnitude, which often can not

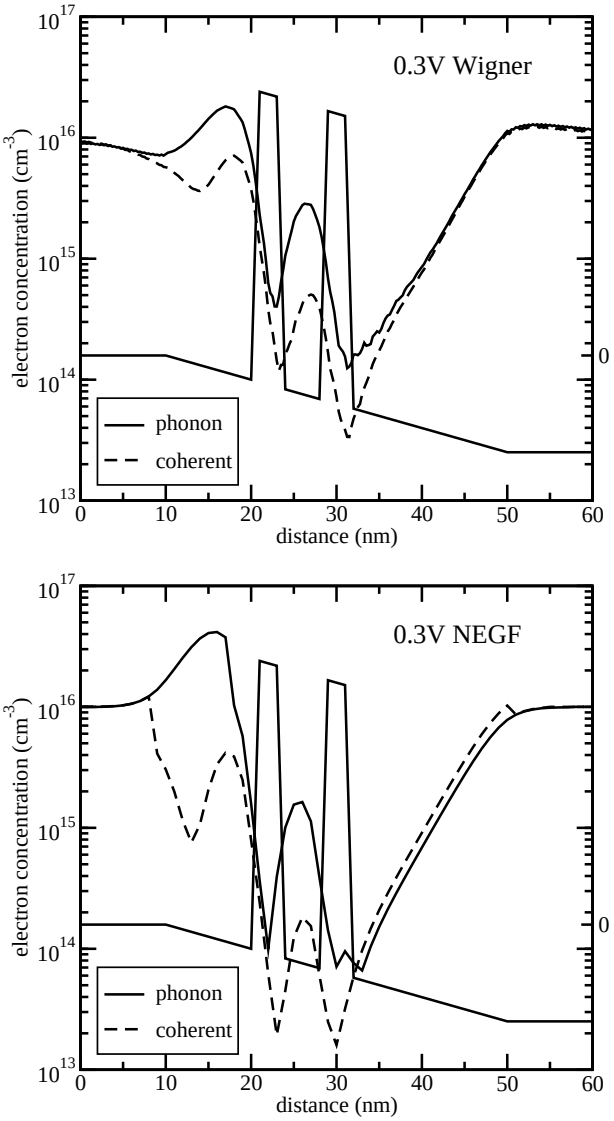

Fig. 4. Effect of scattering on the electron concentration in Device 2 obtained from Wigner MC and NEMO–1D. With scattering the the population of the emitter notch states increases.

be resolved by the MC method, a problem well known also from the classical MC method. Furthermore, the resonance peak might be so narrow that a resolution with an equi-distant k-grid is not well suited.

In a Wigner function based simulation of 1D heterostructures fundamental simulation parameters such as the coherence length are closely linked with physical device parameters such as the spacing from the contacts. This handicap stems from the choice of plane wave basis sets in a quantum mechanical regime of broken translational invariance. While analytically appealing, this basis set proves to be numerically difficult. The NEGF formalism used in NEMO-1D uses real space local orbitals and energy as a quantum mechanical basis. Devices extended over a micrometer can be simulated with an atomic resolution in real space and a resolution of arbitrarily sharp resonances on an average workstation without numerical instabilities.

It is our conclusion that the novel Wigner MC method is not an optimal method for RTD simulation. However, since the method describes quantum effects and scattering effects with equal accuracy it is considered a predictive tool especially whenever some kind of quasi-ballistic transport condition without energetically sharp resonances is present. One strength of the Wigner function approach is the treatment of contact regions. Non-equilibrium transport can be simulated in the whole device formed by a central quantum region embedded in extended classical regions. The presented Wigner MC method can bridge the gap between classical device simulation and pure quantum ballistic simulations.

VI. Conclusion

A MC method for the simulation of non-equilibrium transport in nanostructures has been presented. The method solves the Wigner equation including the Boltzmann scattering operator. The Wigner MC solver has been verified by comparative simulations with NEMO-1D. For simplified test structures the numerical solution of the Wigner equation is found in good agreement with the results of NEMO-1D. In the Wigner simulation the coherence length turned out to be a critical parameter that has to be chosen properly. The effect of phonon scattering on the device characteristics and the internal profile of the density is discussed.

Acknowledgment

The work at TU Vienna has been supported by the European Commission, project IST-10828, NANOTCAD. Part of this work was carried out at the Jet Propulsion Laboratory, California Institute of Technology under a contract with the National Aeronautics and Space Administration. Funding was provided under grants from ONR, ARDA/NSA, and JPL.

References

[1] R. Lake, G. Klimeck, R. Bowen, and D. Jovanovic, "Single and multiband modeling of quantum electron transport through layered semiconductor devices," *J.Appl.Phys.*, vol. 81, pp. 7845–7869, 1997.

[2] R. Bowen, G. Klimeck, R. Lake, W. Frensley, and T. Moise, "Quantitative Resonant Tunneling Diode Simulation," *J.Appl.Phys.*, vol. 81, pp. 3207–3213, 1997.

[3] G. Klimeck, R. Lake, and D. Blanks, "Role of interface roughness scattering in self-consistent resonant tunneling diode simulation," *Physical Review B*, vol. 58, pp. 7279–7285, 1998.

[4] L. Shifren and D. Ferry, "Wigner function quantum Monte Carlo," *Physica B*, vol. 314, pp. 72–75, 2002.

[5] P. Bordone, A. Bertoni, R. Brunetti, and C. Jacoboni, "Monte Carlo Simulation of Quantum Electron Transport Based on Wigner Paths," *Mathematics and Computers in Simulation*, vol. 62, pp. 307–314, 2003.

[6] M. Nedjalkov, R. Kosik, H. Kosina, and S. Selberherr, "Wigner Transport through Tunneling Structures - Scattering Interpretation of the Potential Operator," in *Proc. Simulation of Semiconductor Processes and Devices*, (Kobe, Japan), pp. 187–190, Business Center for Academic Societies Japan, 2002.

[7] H. Kosina, M. Nedjalkov, and S. Selberherr, "Quantum Monte Carlo Simulation of a Resonant Tunneling Diode Including Phonon Scattering," in *Nanotech*, (San Francisco), pp. 190–193, Computational Publications, Feb. 2003.

[8] H. Tsuchiya, M. Ogawa, and T. Miyoshi, "Simulation of Quantum Transport in Quantum Devices with Spatially Varying Effective Mass," *IEEE Trans.Electron Devices*, pp. 1246–1252, 1991.