Operator-Split Method for Variance Reduction in Stochastic Solutions of the Wigner Equation

M. Nedjalkov[1], E. Atanassov[2], H. Kosina[1] and S. Selberherr[1]

[1]Institute for Microelectronics, TU-Vienna, Austria
[2]CLPP, Bulgarian Academy of Sciences, Sofia, Bulgaria

Abstracts — The standard stochastic approach for simulation of carrier transport described by the Wigner equation introduces hard computational requirements. Averaged values of physical quantities are evaluated by means of numerical trajectories which accumulate statistical weight. The weight can take large positive and negative values which introduce large variance of the calculations. Aiming at variance reduction, we utilize the idea to split the weight so that a part is assigned to the trajectory and a part is left on a phase space grid for future processing. Formally this corresponds to a splitting of the kernel of the Wigner equation into two components. An operator equation is derived which couples the two kernel components and gives an answer how to further process the stored weight. The obtained Monte Carlo algorithm resembles a physical process of generation or annihilation of positive and negative particles. Variance reduction is achieved due to the partial annihilation of particles with opposite sign. Simulation results for a resonant-tunneling diode are presented.

1 Introduction

As described by the Wigner equation, the transport in nanoelectronic devices consists of coherent and dissipative parts. The coherent part accounts for the quantum interaction of the carriers with the device potential. Processes of dissipation due to phonons are treated as classical events via the Boltzmann collision operator.

We consider the case of stationary transport in one-dimensional devices. The four-dimensional phase space $(x, \mathbf{k})$ is composed by the real space coordinate x and the three-dimensional space of wave vectors $\mathbf{k} = (k_x, k_y, k_z)$. The device is bounded between two contact points x_1 and x_2. Open system boundary conditions are provided by classical equilibrium functions $f_b(x_b, \mathbf{k})$, $b = 1, 2$ [1]. A physical average $\langle A \rangle$ is expressed via the boundary conditions as follows:

$$< A >= \sum_{b=1,2} \int\limits_{K_+(x_b)} d\mathbf{k}_b \int\limits_0^\infty dt_0 \; |v_x(\mathbf{k}_b)| f_b(x_b, \mathbf{k}_b) e^{-\int\limits_0^{t_0} \mu(x_b(y), \mathbf{k}_b) dy} g(x_b(t_0), \mathbf{k}_b) \qquad (1)$$

$K_+(x_b)$ is the part of the wave vector space with inward directed x-velocities. $x_i(t) = x_i + v_x(\mathbf{k}_i)t$, $\mathbf{k}_i(t) = \mathbf{k}_i$ denote classical trajectories initialized by $(x_i, \mathbf{k}_i)$ at time 0, where $v_x(\mathbf{k}) = \frac{\hbar}{m} k_x$ is the x component of the carrier velocity. The function g is the solution of the adjoint integral form of the Wigner equation:

$$g(x, \mathbf{k}) = \int d\mathbf{k}' \int\limits_0^\infty dt \Gamma(x, \mathbf{k}', \mathbf{k}) e^{-\int\limits_0^t \mu(x(y), \mathbf{k}') dy} \theta_D(x) g(x(t), \mathbf{k}') + A(x, \mathbf{k}) \qquad (2)$$

$$
\begin{aligned}
\Gamma(x,\mathbf{k}',\mathbf{k}) \;=\;& S(\mathbf{k},\mathbf{k}') + \gamma(x)\delta(\mathbf{k}'-\mathbf{k}) \\
+\;& \left(V_w^+(x,k_x-k_x{}') - V_w^+(x,k_x{}'-k_x)\right)\delta(\mathbf{k}'_{yz}-\mathbf{k}_{yz})
\end{aligned}
\tag{3}
$$

S denotes the phonon scattering rate and $\lambda(\mathbf{k}) = \int S(\mathbf{k},\mathbf{k}')d\mathbf{k}'$ is the total out-scattering rate. V_w^+ is a positive function defined as $V_w^+ = \max(V_w,0)$. The antisymmetric Wigner potential V_w is defined by the device potential $V(x)$:

$$
V_w(x,k_x) = \int ds\, e^{-ik_x s}\frac{1}{i2\pi\hbar}\left(V(x-\frac{s}{2}) - V(x+\frac{s}{2})\right)
\tag{4}
$$

Finally

$$
\mu(x,\mathbf{k}) = \lambda(\mathbf{k}) + \gamma(x); \qquad \gamma(x) = \int V_w^+(x,k_x)dk_x
\tag{5}
$$

The function γ is interpreted as the out-scattering rate of the Wigner potential in strict analogy with the phonon out-scattering rate λ. The indicator of the device domain $\theta_D(x)$ is unity if $x_1 \le x \le x_2$ and zero otherwise.

$\langle A\rangle$ can be expanded into a series, which is convenient for application of the numerical Monte Carlo (MC) theory. By following the same steps applied in the case of classical transport ([2]), equation (2) can be formally written into a compact form as:

$$
g(Q) = \int dQ'\, K(Q,Q')g(Q') + A(Q); \qquad Q = (x,\mathbf{k})
\tag{6}
$$

The solution is expanded into a Neumann series:

$$
g(Q) = \int dQ'(\delta(Q-Q') + \sum_{n=1}^{\infty} K^n(Q,Q'))A(Q') = (I-K)^{-1}A,
$$

where I is the identity operator and $K^n(Q,Q') = \int dQ_1 K(Q,Q_1)K^{n-1}(Q_1,Q')$ A replacement in (1) gives rise to a series which is written as inner product:

$$
\langle A\rangle = (b,(I-K)^{-1}A)
\tag{7}
$$

2   Basic Monte Carlo Method

The consecutive terms of the series are evaluated by means of numerical trajectories. Trajectories are constructed by using an initial density $P(Q)$ and a transition density $PP(Q,Q')$. Each trajectory consists of an initial point Q_0, selected by P, followed by points $Q_1 \dots Q_n$ selected by consecutive applications of the transition density PP. The following quantity

$$
\xi_n = \frac{b(Q_0)}{P(Q_0)}\frac{K(Q_0,Q_1)}{PP(Q_0,Q_1)}\dots\frac{K(Q_{n-1},Q_n)}{PP(Q_{n-1},Q_n)}\tilde{A}_n
\tag{8}
$$

estimates the nth term of (7). $\tilde{A}_n$ is a functional of A determined by the synchronous ensemble or time-integration method ([3]). The initial density can be taken from the classical single-particle MC algorithm: $P = b/\Phi$, where $\Phi > 0$ is a normalization constant. It is shown that trajectories which begin at a device boundary and end at a device boundary provide independent realizations of the random variable associated to $\langle A\rangle$.

That is, $\sum_n \xi_n^{(j)}$ calculated on the jth trajectory is the jth sample value for $\langle A \rangle$. $\langle A \rangle$ is evaluated by the sample mean. The term in front of $\tilde{A}_n$ in (8) is the accumulated statistical weight. It depends on the choice of PP by the ratio $w = K/PP$, called weight factor. In the case of Boltzmann transport PP can be chosen such that the weight factor is always unity. This is not possible for the quantum case, where the accumulated weight rapidly grows with the iteration number n.

The kernel K will be reformulated as a product of a transition density PP and a weight factor w. PP is composed by conditional probability densities which are introduced by following the order of their application.

$$K(x, \mathbf{k}', \mathbf{k}, t) = p_t(t, x, \mathbf{k}) \frac{\Gamma(x(t), \mathbf{k}', \mathbf{k})}{\mu(x, \mathbf{k})} \theta_D(x(t)); \quad p_t = \mu(x(t), \mathbf{k}) e^{-\int_0^t \mu(x(y), \mathbf{k}) dy}$$

p_t generates a value of t associated with a free flight time of a particle which drifts over a piece of a Newton trajectory between the initial state $(x, \mathbf{k})$ at time 0 and the before-scattering state $(x(t), \mathbf{k})$. This state is used in the remaining term $\Gamma(x(t), \mathbf{k}', \mathbf{k})/\mu(x(t), \mathbf{k})$ to select the value of $\mathbf{k}'$,

$$\frac{\Gamma(x, \mathbf{k}', \mathbf{k})}{\mu(x, \mathbf{k})} = \frac{\lambda(\mathbf{k})}{\mu(x, \mathbf{k})} p_{ph}(\mathbf{k}, \mathbf{k}') + \frac{\gamma(x)}{\mu(x, \mathbf{k})} \times \tag{9}$$

$$\left(p_2 p_w^+(x, k_x - k_x') - p_2 p_w^-(x, k_x - k_x') + p_2 p_\delta(k_x - k_x') \right) \delta(\mathbf{k}_{yz} - \mathbf{k}'_{yz}) \frac{1}{p_2}$$

$$p_{ph}(\mathbf{k}, \mathbf{k}') = \frac{S(\mathbf{k}, \mathbf{k}')}{\lambda(\mathbf{k})}; \quad p_w^\pm(x, k_x) = \frac{V_w^+(x, \pm k_x)}{\gamma(x)}; \quad p_\delta(k_x) = \delta(k_x); \quad p_2 = \frac{1}{3}$$

Γ/μ is interpreted as follows. According to (5) λ/μ and γ/μ are two complementary probabilities, which determine the type of interaction. Classical scattering by phonons occurs with a probability λ/μ. In this case the after-scattering state $\mathbf{k}'$ is chosen with probability density p_{ph}. The alternative interaction is due to the Wigner potential. It is comprised of the three terms, enclosed in the brackets. p_2 has been introduced with the purpose to select which one of the three densities p_w^+, p_w^- and p_δ generates the after-scattering state $(k_x', \mathbf{k}_{yz})$.

The consecutive application of the described conditional probabilities generates a transition between $(x, \mathbf{k})$ and $(x(t), \mathbf{k}')$. The transition corresponds to one realization of PP. The remaining term is the weight factor $w = \pm 1/p_2 = \pm 3$ where the minus sign applies if p_w^- is selected.

Obviously only events of quantum interaction change the statistical weight. It can be shown that the mean statistical weight is evaluated by of $\pm e^{\gamma n_w}$ where n_w is the mean number of quantum interaction events per trajectory [4]. The argument of the exponent depends on V_w and the length of the considered device. Their magnitudes in nanoelectronic devices are such that the weight and thus the variance of the simulation results become huge. This precludes the practical implementation of the method.

3 The Operator-Split Method

Particle splitting is an established approach [5] for statistical enhancement in classical Monte Carlo simulations. A classical particle entering a rarely visited region of the

phase space is split into sub-particles with corresponding fraction of the particle weight. The sub-particles are consecutively removed by simulation of any particular sub-particle trajectory until it exits from the device. The process continues until all stored particles are removed.

This idea can not be applied directly in quantum simulations. The reason is in the increase of the magnitude of the weight on quantum trajectories. Suppose a trajectory which initiates from the boundary with a unit weight. The magnitude of the weight increases so that a fraction of the weight is periodically stored in some points of the phase space. The next step this primary weight is removed by trajectories which begin from these phase space points. During this simulation step a secondary weight is stored etc. The process can be infinitely continued: the stored weight cannot be removed from the simulation domain. This problem is solved with the help of an integral equation based on the splitting of the integral operator in (6). Two operators $L = mK$, $M = (1 - m)K$, $m < 1$ are introduced, which obey the operator equality:

$$(I - K)^{-1} = (I - L)^{-1} + (I - K)^{-1}M(I - L)^{-1} \tag{10}$$

(10) is proved by multiplication by $(I - L)$ and $(I - K)$. Substituting (10) into (7) one obtains the following expansion:

$$\langle A \rangle = (b, (I - K)^{-1}A) = \sum_{k=0}^{\infty}(b, \left((I - L)^{-1}M\right)^k (I - L)^{-1}A) \tag{11}$$

Now we derive a MC algorithm which evaluates the consecutive terms in (11). For the sake of simplicity we consider the coherent case, $\lambda = 0$. A choice of $m = 1/3$ leads to the convenient relations $L = \pm PP$ and $M = 2L$. With the help of (8) the first term $\langle A \rangle_1 = (b, (I - L)^{-1}A)$ is evaluated by the sample mean

$$\langle A \rangle_1 \simeq \frac{\Phi}{J} \sum_{j,n} \text{sign}(L(Q_0^{(j)}, Q_1^{(j)})) \ldots \text{sign}(L(Q_{n-1}^{(j)}, Q_n^{(j)})\tilde{A}_n^{(j)} \tag{12}$$

where J is the number of the used trajectories. The second term $\langle A \rangle_2 = (b, (I - L)^{-1}M(I - L)^{-1}A)$ can be reformulated with the help of an operation of projection in the phase space. Formally we use a delta function to introduce the auxiliary function W_1 which has a meaning of a weight stored in point Q of the phase space

$$W_1(Q) = (b, (I - L)^{-1}M\delta)(Q).$$

This can be seen from the estimator of W_1, which is obtained by replacing $\tilde{A}_n^{(j)}$ in (12) by $2\delta(Q - Q_n^{(j)})$. Both W_1 and $\langle A \rangle_1$ can be evaluated within a single computational step. The second average $\langle A \rangle_2$ is expressed as an inner product with the stored weight: $\langle A \rangle_2 = (W_1, (I - L)^{-1}A)$. This term can be estimated by the sample mean computed by using of J_1 trajectories:

$$\langle A \rangle_2 \simeq \frac{1}{J_1} \sum_{j,n} \frac{W_1(Q_0^{(j)})}{P_1(Q_0^{(j)})} \text{sign}(L(Q_0^{(j)}, Q_1^{(j)})) \ldots \text{sign}(L(Q_{n-1}^{(j)}, Q_n^{(j)})\tilde{A}_n^{(j)} \tag{13}$$

The initial density P_1 for these trajectories can be conveniently selected as:

$$P_1(Q) = |W_1(Q)|/||W_1||; \qquad ||W_1|| = \int |W_1(Q)|dQ.$$

If J_1 is selected to be the nearest integer to $||W_1||$, (13) simplifies to:

$$\langle A \rangle_2 \simeq \sum_{j,n} \text{sign}(W_1(Q_0^{(j)}))\text{sign}(L(Q_0^{(j)}, Q_1^{(j)})) \ldots \text{sign}(L(Q_{n-1}^{(j)}, Q_n^{(j)})\tilde{A}_n^{(j)} \qquad (14)$$

The term composed by the product of the sign factors in (12), (13) and the corresponding estimator of W_1 can be associated with a particle which has unit weight and a sign which changes during the evolution. This picture can be maintained, if the pre-factor Φ/J in (12) is transferred to the left of the equation. Then the estimated quantities become $C\langle A \rangle_1$, CW_1 and $C\langle A \rangle_2$ where $C = J/\Phi$. Moreover CW_1, which determines the number of trajectories J_1 becomes integer.

The steps for evaluation of the first two terms in (11) can be summarized as follows. The stochastic process is described in terms of particles. A phase space mesh is introduced, where the indicators of the mesh cells play the role of projection operators. An estimator η and an array estimator ν with components associated to the cells are initialized to zero. During the first simulation step J positive particles are injected from the boundary with probability P. They evolve over trajectories constructed by successive realizations of PP until exiting the device. At each transition between points $(x, \mathbf{k})$ and $(x(t), \mathbf{k}')$

- $\tilde{A}_n$ is computed, multiplied by the sign of the particle and added to η;

- The sign s of the particle is changed, if $\mathbf{k}'$ is generated from p_w^-;

- $2s$ is added to the value stored in that component of ν, which corresponds to the cell of $(x(t), \mathbf{k}')$. We say that two particles with sign s are stored in that cell.

We note a very important property of this model. Particles in a given phase space point have a common probabilistic future. This peculiarity can be approximately generalized for sufficiently small phase space cells. Two particles with opposite signs, which initiate their evolution from the same cell have opposite contributions to the estimators of the averages. Such particles can be annihilated so that a cell can contain only particles with a given sign. This step computes the quantities $C\langle A \rangle_1 \simeq \eta$ and $CW_1 \simeq \nu$.

An estimator η_1 is introduced in the second step, which computes the value of $C\langle A \rangle_2$ by following the same scheme utilized for $C\langle A \rangle_1$. The trajectories now start from the cells and are simulated until their exit from the device. As it follows from (13), the number of the trajectories which initiate from a given cell and the initial sign correspond to the number and sign of the particles stored in that cell. An estimator ν_1 can be computed in the same way as ν which corresponds to the particles stored at this step. It can be easily shown that ν_1 evaluates the quantity CW_2, where W_2 is the stored weight for the third term in (11):

$$W_2(Q) = (b, (I - L)^{-1}M(I - L)^{-1}M\delta)(Q)$$

$\langle A \rangle_3$ is obtained as an inner product with W_2: $\langle A \rangle_3 = (W_2, (I - L)^{-1}A)$. This term resembles the expression for $\langle A \rangle_2$ with W_2 replacing W_1. Then step two can be repeated for the estimators η_2 and ν_2 which evaluate $C\langle A \rangle_3$ and the stored particles for the next term in (11). An iterative algorithm is obtained, which computes the consecutive terms of the series for $C\langle A \rangle$.

Aiming at optimization, we propose several modifications in the algorithm.

Since $C\langle A\rangle$ is evaluated by $\sum_n \eta_n$, one estimator η is used in all steps.

The action of the Wigner potential can be interpreted in terms of particles as follows. One of the three components of the transition density PP is selected to create three particles in the final phase space point. One of them continues the trajectory, the remaining two are stored. In an alternative interpretation each kernel component can be used to create a particle. Formally this corresponds to a replacement of p_2 in (9) with unity. Two of the created particles carry the sign of the initial particle. The particle generated by p_w^- has the opposite sign. In this way three phase space cells are updated, which assists the annihilation process. The trajectory is continued by a particle from that one of the three cells, which contains the highest number of stored particles.

The constant Φ can be evaluated by using of the fact that the number of electrons N in given device domains is known. If θ is the indicator of that domain, then $\langle\theta\rangle = N$. The quantity $C\langle\theta\rangle = B$ can be calculated during the simulation, which determines the ratio $C = J/\Phi = B/N$. $\langle A\rangle$ is calculated by $\langle A\rangle \simeq \eta B/N$.

The number of the injected boundary particles J must be chosen sufficiently large in order to attain a reliable approximation of the first term in (11). The evaluation of all remaining terms depends on this choice. Alternatively a moderate value of J can be used in the following modification of the algorithm. Injection from the boundary alternates injection from the phase space cells. Particles stored from a boundary injection are added to the particles stored from the previous injection from the cells. After the Rth boundary injection, the stored weight approximates $C\sum_{k=1}^{R} W_k(Q)$, which can be proved by induction. The subsequent step of injection from the cells completes the estimation of the functional $C\sum_{k=1}^{R+1}\langle A\rangle_k$. We take the mean of these estimates at steps $0,1,\ldots,R$ as an approximation of the functional $C\langle A\rangle$. Since the functional $C\langle A\rangle_k$ is summed $R-k$ times and the ratios $(R-k)/R$ tend to 1 as R tends to infinity, our estimate of $C\langle A\rangle$ is asymptotically correct. We call this algorithm QMC-I.

A second algorithm, called QMC-II, is based on the notion that injection of negative particles from the boundary destroys the statistics obtained by the injected positive particles. A careful selection of the ratio of the injected positive and negative particles, along with the annihilation in the cells keeps the number of the stored particles low [6]. The steps of removing the stored particles can be ommited.

4 Simulation Experiments

A benchmark resonant-tunneling diode (RTD) has been chosen from the literature [7] for simulation experiments. Simulation results obtained by the two MC algoritms and by NEMO–1D based on non-equilibrium Green's functions [6] are compared. A common physical model is used for all simulations. The effective mass is uniform 0.067 and the barrier height is 0.27eV. The barriers have a thickness of 2.825nm and the well has a thickness of 4.52nm. The doping level in the contacts is $2\cdot 10^{18}\text{cm}^{-3}$, linear potential drop across the central device is assumed.

Fig. (1) compares the current-voltage characteristis of the RTD obtained by the three approaches. A typical (for stochastic methods) variance of the simulation results is demonstrated for moderate simulation times. Convergence is achieved after a very long simulation presented for QMC-II algorithm. The agreement with the deterministic curve
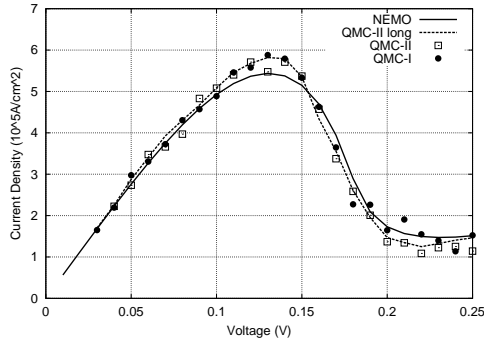
Figure 1: Current-Voltage characteristics obtained by simulations with NEMO–1D, QMC-I and QMC-II.
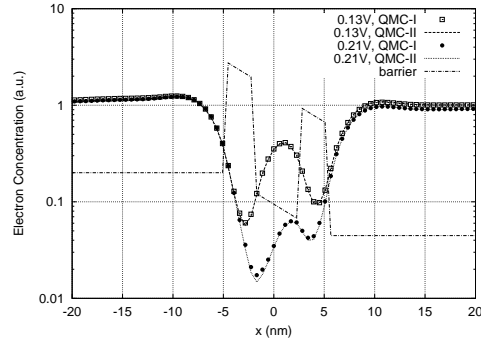
Figure 2: Comparison of the electron concentrations obtained by simulations with QMC-I and QMC-II.

of NEMO–1D solver is good already for the results obtained with moderate simulation times.

Fig. (2) shows the electron concentration in the central part of the simulated device as obtained by the two MC algorithms. The chosen voltage values correspond to the peak and valley points of the current-voltage characteristic. The corresponding results are in excellent agreement.

5 Conclusions

The proposed Monte Carlo method has been successfully applied for realistic nanoelectronic devices. Variance reduction has been achieved by annihilation of positive and negative particles into the phase space cells. The particle picture associated to the stochastic process has been used to interpret and explain quantum transport phenomena in Wigner representation.

References

[1] W. Frensley, "Boundary conditions for open quantum systems driven far from equilibrium," *Reviews of Modern Physics*, vol. 62, no. 3, pp. 745–789, 1990.

[2] H. Kosina, M. Nedjalkov, and S. Selberherr, "The Stationary Monte Carlo Method for Device Simulation - Part I: Theory," *J.Appl.Phys.*, vol. 93, no. 6, pp. 3553–3563, 2003.

[3] M. Nedjalkov, H. Kosina, and S. Selberherr, "Monte Carlo Algorithms for Stationary Device Simulation," *Mathematics and Computers in Simulation*, vol. 62, no. 3-6, pp. 453–461, 2003.

[4] M. Nedjalkov, H. Kosina, and S. Selberherr, "Stochastic Interpretation of the Wigner Transport in Nanostructures," *Microelectronics Journal*, vol. 34, pp. 443–445, 2003.

[5] C. Wordelman, T. Kwan, and C. Snell, "Comparison of Statistical Enhancement Methods for Monte Carlo Semiconductor Simulation," *IEEE Trans.Computer-Aided Design*, vol. 17, no. 12, pp. 1230–1235, 1998.

[6] H. Kosina, G. Klimeck, M. Nedjalkov, and S. Selberherr, "Comparison of Numerical Quantum Device Models," in *Proc. Simulation of Semiconductor Processes and Devices*, (Boston, USA), pp. 171–174, Sept. 2003.

[7] H. Tsuchiya, M. Ogawa, and T. Miyoshi, "Simulation of Quantum Transport in Quantum Devices with Spatially Varying Effective Mass," *IEEE Trans.Electron Devices*, vol. 38, no. 6, pp. 1246–1252, 1991.