# Computational Electronics on GRID: A Mixed Mode Carrier Transport Model

P. Schwaha*, J. Cervenka*, M. Nedjalkov*, T. Gurov†, G. Arsov**, A. Misev**, A. Zoric‡ and S. Ilic‡

*Institute for Microelectronics, TU Wien, Gusshausstrasse 27-29, A-1040 Wien, Austria,
E-mails: (schwaha,cervenka,mixi)@iue.tuwien.ac.at
†IPP-BAS, Acad. G. Bonchev St., bl. 25A,
Sofia 1113, Bulgaria, E-mails: gurov@bas.bg
**FEEIT, Ss Cyril and Methodius University - Skopje, Macedonia,
E-mails: g.arsov@ieee.org, anastas@ii.edu.mk
‡FTS, University of Pristina, K. Mitrovica, Kosovo,
E-mails: aleksandar.zoric@pr.ac.rs, sinisa.ilic@pr.ac.rs

**Abstract.** The nano-era of semiconductor electronics introduces the necessity of simulation methods which describe the electron transport in ultra-small devices in a mixed mode where quantum-coherent processes are considered along with the de-coherence processes of scattering. The latter can be conveniently described in the Wigner picture of quantum mechanics, however the coherent counterpart gives rise to heavy numerical problems. We propose a scheme which combines the advantages of the Wigner function with the Green's function picture which is numerically efficient in coherent cases. An equation accounting for the scattering corrections to the coherent Wigner function is derived theoretically and a Monte Carlo algorithm for calculating these corrections is developed and implemented. The implementation is deployed on the SEE-GRID to facilitate the swift acquisition of results. Simulation results are presented as a final point.

**Keywords:** Wigner function, Monte Carlo method, Grid computing

## INTRODUCTION

Modeling and simulation of electronic devices is a part of a field of science which combines mathematics, physics, and electrical engineering methods and uses this union to design, analyze, and optimize the core components of integrated circuits. To the semiconductor industry it appears as the only alternative to an enormously expensive trial-and-error manufacturing approach. Using device modeling and simulation the physical characteristics of semiconductor devices are explored in terms of charge transport and electrical behavior [1]. The increase of complexity of the physical models describing the electron transport, as is required for progress in this field, needs to be accompanied by the application of efficient numerical algorithms, as the computational costs otherwise outgrow the available computational resources.

The nano-era of the semiconductor electronics raises the necessity of simulation methods which describe the electron transport in ultra-small devices as a mixed mode quantum process. The latter accounts for both, quantum-coherent processes of interaction with the device potential, and phase-breaking processes of de-coherence due to scattering with phonons and other crystal lattice imperfections. The Wigner formulation of quantum mechanics is particularly convenient for mixed mode transport description,

since it utilizes a phase space, where many classical notions are retained. In this picture the scattering can be accounted for in a straightforward way by using the Boltzmann collision models. While these models can be applied by using the well-developed classical simulation algorithms, the coherent counterpart gives rise to a heavy numerical burden. In contrast to this the alternative formulation of quantum transport in terms of Green's functions, is numerically efficient in the cases of coherent transport and problematic when considering phase-breaking processes.

We propose and approach which combines the advantages of the two pictures: Green's function calculations of the of coherent transport determined by the boundary conditions in the semiconductor device provide the coherent Wigner function. They determine the initial condition for an equation of the correction obtained by subtracting the coherent Wigner equation from the general coherent/de-coherent counterpart. Furthermore, this equation is approximated by its classical limit.

For very small devices, where the carrier dwelling time (the time needed for a carrier to abandon the device through the contacts/boundaries) is accordingly short, the initial condition may already be considered a sufficient correction accounting for the de-coherence effects.

A particle model based on numerical Monte Carlo (MC) theory is derived for the evaluation of the multidimensional integrals involved in the initial condition. An efficient implementation on a grid environment is considered, aiming to cope with the computational requirements of the developed Monte Carlo approach.

# THEORETICAL MODEL

## Scattering-induced Wigner function correction

We propose an approach [2] which combines the advantages of the two pictures describing quantum phenomena, namely the Wigner function and the Green's function approaches. Green's function calculations of the coherent transport determined by the boundary conditions deliver the coherent Wigner function $f_w^c$.

$$\rho(x,x') = -2i\int G^<(x,x',E)\frac{dE}{2\pi}; \qquad f_w^c(x,k_x) = \frac{1}{2\pi}\int ds e^{-ik_x s}\rho(x+\frac{s}{2},x-\frac{s}{2}) \quad (1)$$

The lesser Green's function $G^<$ depends on the coordinates $x,x'$ and energy $E$. Using a center-of-mass transformation $x = \frac{x_1+x_2}{2}, \quad s = x_1 - x_2$ the coherent Wigner function $f_w^c(x,k_x)$ is obtained from the density matrix $\rho(x_1,x_2)$. Furthermore, $f_w^c$ is a solution of the coherent part of the Wigner-Boltzmann equation.

$$\begin{aligned} \frac{\hbar k_x}{m}\frac{\partial}{\partial x} f_w(x,\mathbf{k}) &= \int dk_x' V_w(x,k_x'-k_x) f_w(x,k_x',\mathbf{k}_{yz}) \\ &+ \int d\mathbf{k}' f_w(x,\mathbf{k}') S(\mathbf{k}',\mathbf{k}) - f_w(x,\mathbf{k})\lambda(\mathbf{k}) \end{aligned} \quad (2)$$

Here $V_w$ is the Wigner potential, while the phase-breaking processes are accounted for by the Boltzmann scattering operator by $S(\mathbf{k},\mathbf{k}')$, the scattering rate for a transition from $\mathbf{k}$ to $\mathbf{k}'$. The total out-scattering rate $\lambda(\mathbf{k})$ is linked to the scattering by

$\lambda(\mathbf{k}) = \int d\mathbf{k}' S(\mathbf{k}.\mathbf{k}')$. By setting the scattering rate $S$ (and thus $\lambda$) to zero the coherent problem is obtained from (2). The $\mathbf{k}_{yz}$ dependence remains arbitrary in this case and may be specified via the boundary conditions. Formally, the extrapolation must be such that $f_w^c(x,k_x')$ is recovered by an integration over $\mathbf{k}_{yz}$. Moreover, the Boltzmann scattering operator should cancel at the boundaries where standard equilibrium conditions are assumed. Hence, a Maxwell-Boltzmann distribution $f_{MB}(k_{yz}')$ is assumed in the $yz$ directions, and the functions

$$f_w^c(x,\mathbf{k}') = f_w^c(x,k_x')\frac{\hbar^2}{2\pi mkT}e^{-\frac{\hbar^2(k_y'^2+k_z'^2)}{2mkT}}; \qquad f_w^\Delta(x,\mathbf{k}) = f_w(x,\mathbf{k}) - f_w^c(x,\mathbf{k})$$

are introduced. By subtracting the coherent counterpart from (2) an equation for the correction $f_w^\Delta$ is obtained.

$$\begin{aligned}\frac{\hbar k_x}{m}\frac{\partial}{\partial x}f_w^\Delta(x,\mathbf{k}) = &\int dk_x' V_w(x,k_x'-k_x)\, f_w^\Delta(x,k_x',\mathbf{k}_{yz}) \\ &+ \int d\mathbf{k}'\, f_w^\Delta(x,\mathbf{k}')S(\mathbf{k}',\mathbf{k}) - f_w^\Delta(x,\mathbf{k})\lambda(\mathbf{k}) \\ &+ \int d\mathbf{k}'\, f_w^c(x,\mathbf{k}')S(\mathbf{k}',\mathbf{k}) - f_w^c(x,\mathbf{k})\lambda(\mathbf{k})\end{aligned} \tag{3}$$

The term in the last row may be determined from $f_w^c$ and may thus be considered known. Furthermore, the correction is zero at the device boundaries, since the same boundary conditions are assumed in both cases.

## Classical limit

The obtained equation is approximated using the classical limit.

$$\int \mathrm{d}k_x' V_w(x,k_x'-k_x)\, f_w^\Delta(x,k_x',\mathbf{k}_{yz}) = -\frac{eE(x)}{\hbar}\frac{\partial f_w^\Delta(x,k_x,\mathbf{k}_{yz})}{\partial k_x} \tag{4}$$

This approximation is valid for slowly varying potentials, so that the force $F(x) = eE(x)$ is described by a linear function within the spatial support of $f_w^\Delta$. This force induces Newton's trajectories for particles under its influence, which may be described by

$$X(t) = x - \int_t^0 \frac{\hbar K_x(\tau)}{m}d\tau \qquad K_x(t) = k_x - \int_t^0 \frac{F(X(\tau))}{\hbar}d\tau \qquad \mathbf{k}(t) = K_x(t),\mathbf{k}_{yz} \tag{5}$$

These trajectories are initialized by $x, k_x, 0$. With these settings a trajectory is called forward in the case of $t > 0$ and backward otherwise. A backward trajectory crosses the boundary of the device at a certain time $t_b$, so that $f_w^\Delta(X(t_b),\mathbf{k}(t_b)) = 0$. Using Equation (5), the approximated equation can be transformed into a Fredholm integral equation of the second kind:

$$f_w^\Delta(x,\mathbf{k}) = \int_{t_b}^0 dt \int \mathrm{d}\mathbf{k}'\, f_w^\Delta(X(t),\mathbf{k}')S(\mathbf{k}',\mathbf{k}(t))e^{-\int_t^0 \lambda(\mathbf{k}(\tau))d\tau} \tag{6}$$

$$+\int_{t_b}^{0} dt \left\{ \int \mathrm{d}\mathbf{k}' f_w^c(X(t),\mathbf{k}') S(\mathbf{k}',\mathbf{k}(t)) - f_w^c(X(t),\mathbf{k}(t))\lambda(\mathbf{k}(t)) \right\} e^{-\int_t^0 \lambda(\mathbf{k}(\tau))d\tau}$$

The free term is given by the second row of (6) and determined by $f_w^c$. The solution to the Fredholm integral equation may be represented as a Neumann series with terms obtained by iterative application of the kernel to the free term. The series corresponds to a Boltzmann kind of evolution process, where the free term corresponds the initial condition. Thereby, the full mixed mode problem posed by the boundary conditions is transformed into a classical evolution of the quantum-coherent solution $f_w^c$. However, the latter may accommodate negative values and can thus not be interpreted as the initial distribution of classical electrons. But it is possible to model this situation by using positive and negative particles during the evolution process. In this fashion, the quantum information is retained in Equation (6) by accounting for the sign of the evolving particles. The boundary is still represented using $t_b$, however, its physical meaning is changed: the boundary is purely absorbing, as all trajectories with evolution time $t < t_b < 0$ do not contribute to the solution. In very small devices the carrier dwelling time can be so short that the probability for multiple scattering events tends to zero. In such cases the initial condition itself presents the correction $f_w^\Delta$. Otherwise the evaluation of the initial condition is a necessary to obtain an appropriate solution. Next, the particle approach developed to this end using numerical Monte Carlo theory is introduced.

## MONTE CARLO ALGORITHM

The computational task is specified as the calculation of the value of the two components $f_{0A}$ and $f_{0B}$ of the initial conditions at the given points $(x^i,k_x^j)$: $f_{0A}^{i,j} = f_{0A}(x^i,k_x^j)$ and similarly for $f_{0B}$. Particle approaches are suitable for the calculation of the inner product of two functions: in our case it is the averaged value $I_{A,B}(\Omega)$ of $f_{0A,B}$ in a given domain $\Omega$ of the phase space.

$$I_{A,B} = \int dx \int dk_x \theta_\Omega(x,k_x) f_{0A,B}(x,k_x) = \int dx \int dk_x \int dk_y \int dk_z \theta_\Omega(x,k_x) f_{0A,B}(x,\mathbf{k})$$

The domain indicator $\theta_\Omega(x,k_x)$ is unity, if its arguments fall within $\Omega$, and 0 otherwise. Then $\Omega = \Omega^{i,j}$ can be determined by the phase space area using a small volume $\Delta = \Delta k_x \Delta x$ around $(x^i,k_x^j)$ so that $f_{0A,B}^{i,j} = I_{A,B}(\Omega^{i,j})/\Delta$. Another peculiarity is the point wise evaluation of $f_w^c$ which gives rise to the approximation:

$$\int dx^t \int dk_x^t f_w^c(x^t,k_x^t) \simeq \sum_{i,j} f_w^c(i,j)\Delta \tag{7}$$

Here we focus on the contribution of the first component $f_{0A}(x,k_x)$. The same approach is employed to treat the second component $f_{0B}(x,k_x)$.

$$I_A = \int_{-\infty}^{0} dt \int dx \int dk_x \int dk_y \int dk_z \int \mathrm{d}\mathbf{k}' \frac{\hbar^2}{2\pi mkT} e^{-\hbar^2(k_y'^2+k_z'^2)/2mkT} \tag{8}$$
$$f_w^c(X(t),k_x') S(\mathbf{k}',\mathbf{k}(t)) e^{-\int_t^0 \lambda(K_x(\tau),\cdot)d\tau} \theta_\Omega(x,k_x)\theta_D(X(t))$$

Due to the use of the device domain indicator $\theta_D$, the lower bound of the time integral can been extended to $-\infty$, as $\theta_D$ takes care of providing the correct value. The backward parametrization of the trajectories is later changed to a forward one in order to achieve a more heuristic picture of the evolution of the real carriers. Two important properties of the trajectories will be utilized: Since a trajectory obeys a system of first order differential equations, any phase space point reached by the trajectory at any given time can be used for initialization. A full notation of a trajectory $X(t), K_x(t)$ contains the initialization point: $X(t) = X(t;x,k_x,0) = x^t, \quad K_x(t) = K_x(t;x,k_x,0) = k_x^t$. Furthermore, by reparametrizeing a trajectory the initialization point can be changed from $x,k_x,0$ to $x^t,k_x^t,t$ so that $x = X(0,x^t,k_x^t,t), \quad k_x = K_x(0,x^t,k_x^t,t)$. For stationary transport the absolute clock is replaceable by a relative one: trajectories are invariant with respect to a shift of both, initialization and parametrization time.

Applying this procedure to (8), with using (7) results in:

$$I_A(\Omega^{n,m}) = \sum_{i,j} \int_0^\infty dt \int dk_x^t \int dk_y \int dk_z \int \mathrm{d}\mathbf{k}'_{yz} \theta_D(x_i^t) \left\{ \frac{\bar{h}^2}{2\pi mkT} e^{-\frac{\bar{h}^2(k_y'^2+k_z'^2)}{2mkT}} \right\} f_w^c(i,j)\Delta$$
$$\left\{ \frac{S(\mathbf{k}',k_x^t,k_y,k_z)}{\lambda(\mathbf{k}')} \right\} \left\{ \lambda(K_x^t(t),\cdot) e^{-\int_0^t \lambda(K_x^t(\tau),\cdot)d\tau} \right\} \frac{\lambda(\mathbf{k}')}{\lambda(K_x^t(t),\cdot)} \theta_{\Omega^{n,m}}(X^t(t),K_x^t(t))$$

where $\mathbf{k}'$ now denotes $(k_{xj},\mathbf{k}'_{yz})$. This integral has been reformulated in such a fashion that the terms in the curly brackets represent probability densities. The first bracket is the normalized Gaussian distribution, while the next two brackets contain the well known probabilities for scattering and free flight. From this integral expression, the following algorithm is obtained:

1. Associate an estimator $\xi^{n,m}$ to each node $n,m$.
2. Loop over $i,j$ nodes corresponding to $x^t$ and $k_x'$ integrals, and initiate $l = 1,2,\ldots N$ trajectories from each node:
3. select the $k'_{yl},k'_{zl}$ values according to the term in the first curly brackets, thus accounting for the $k_y,k_z$ integrals.
4. Select a wave vector according to the term in the second curly-brackets. Input parameters are $k'_{xj},k'_{yl},k'_{zl}$, the particular value of the after-scattering wave vector is denoted by $\mathbf{k} = k_{xl}^t,k_{yl},k_{zl}$.
5. The point $x_i^t,k_{xl}^t$ initializes the trajectory $K_{xl}^t(t),X_l^t(t)$ at time $t = 0$. Generate a free-flight time value $t_l$ according to the term in the last curly brackets.
6. Add the weight

$$w_l = f_w^c(i,j)\Delta\lambda(k'_{xj},k'_{yl},k'_{zl})/\lambda(K_{xl}^t(t_l),k_{yl},k_{zl})$$

   to the estimator $\xi^{n,m}$ associated to the mesh node $n,m$ nearest to the point $K_{xl}^t(t_l),X_l^t(t_l)$
7. At the end of the $i,j$ loop divide $\xi^{n,m}$ by $N$.

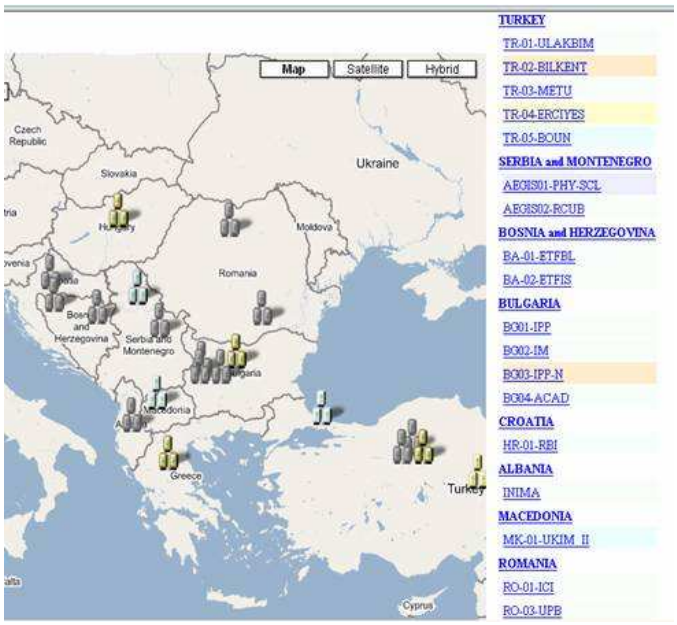


**FIGURE 1.** SEEGRID Infrastructure in South-East Europe

## IMPLEMENTATION OF THE ALGORITHM ON GRID

A computational grid is a computing environment which enables the unification of geographically widely distributed computing resources into one big (super)computer [3]. The individual computing resources commonly consist mostly of computer clusters or several individual computers, which are interconnected by a high-speed wide area network. The grid is a computer system which is, at this moment, primarily intended for supporting e-Science, however the technology itself is very adaptable for a very wide area of present and future computer use. The major goal of a grid is to enable the clustering and unification of distributed computing and data processing resources. This is done to accumulate and coordinate as much computing power as possible and make it available for use by applications, which have a particularly high demand for computing resources. Examples of scientific applications greatly benefiting or even necessitating a grid are from the fields of particle physics, climate analysis, biomedical research, meteorology etc.

The algorithm under consideration, which has been described in a previous Section, requires considerable computational power and time to obtain accurate results. Therefore, a hundred jobs using different input data have been run on grid clusters included in the SEE-GRID infrastructure.

The SEE-GRID infrastructure contains computational and storage resources, which are made available by more than fifteen partners from different countries in South East Europe, for scientific applications. An illustration of the distributed nature of the SEE-GRID is given in Figure 1. Currently, the SEE-GRID provides more than 35 clusters with a total number of CPUs greater than 2000 and a storage capacity exceeding 400 TB [4].

**TABLE 1.** Similarities and differences between a grid and a PC.

| PCs or workstations | Grid clusters |
|---|---|
| Login with a username and password ("Authentication") | Login with digital credentials; single sign-on ("Authentication") |
| Use rights given to user ("Authorisation") | Use rights given to user ("Authorisation") |
| Run jobs | Run jobs |
| Manage files: create them, read/write, list directories | Manage files: create them, read/write, list directories |
| Components are linked by a bus | Services are linked by the Internet |
| Operating system | Middle ware |
| One admin domain | Many admin domains |

This Grid infrastructure was built using the gLite middleware [5], which provides Web Service APIs for most of its services, and also provides new types of services, such as the gLite WMS, gLite FTS, AMGA, etc. The middle ware also improves the reliability and scalability of other services. The following lists services available in the SEE-GRID:

- Each of the SEE-GRID clusters provides the mandatory services:
  - Computing element (CE) - provide user access to grid resources
  - Worker nodes (WN) - execute jobs, perform calculations
  - Storage element (dCache, DPM or classic SE) - reliable data storage
  - MON box (Monitoring and accounting) - monitor the current grid status and report completed jobs and their resource use
- The worker nodes provide the computational resources, while the storage elements provide storage resources
- The set of services, which are not tied to a specific site are called core services. In the SEE-GRID the core services are distributed among partners. They include:
  - VOMS (Virtual organization management system)
  - MyProxy
  - R-GMA registry/schema server (distributed data-base)
  - BDII (provides comprehensive information about the resources)
  - WMS (distributes and manages the jobs among different grid sites)
  - FTS (file transfer service)
  - AMGA (meta data catalog)

The differences and similarities how users can make use of a grid on the one hand and workstations or PCs on the other hand are shown in Table 1.

Our tests are performed mainly on two main grid clusters - BG03-NGCCP and BG-4-ACAD which are situated in the Institute for Parallel Processing. The specifications of the grid clusters used in our experiments are:

- The BG03-NGCC grid cluster has 25 worker nodes, which contain 2xIntel Xeon E5430 2.66 GHz Quad Core CPU (total 200 Cores, $>$ 400 kSI2000) with 16 GB RAM on each node.

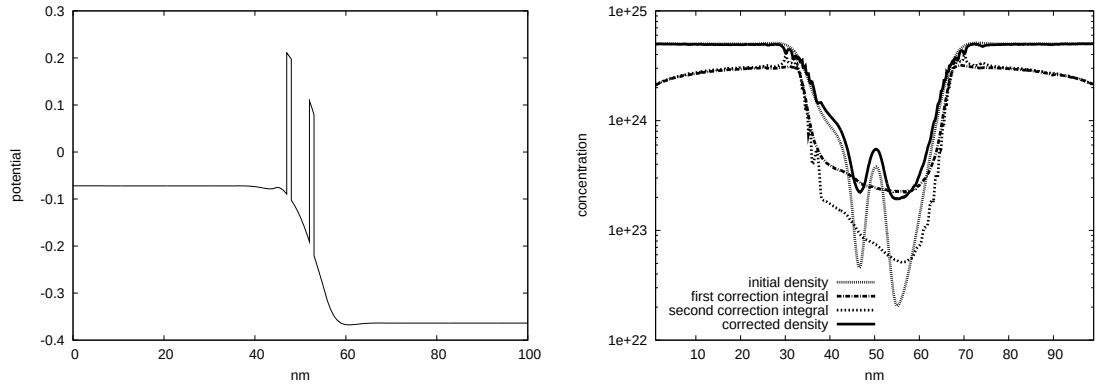


**FIGURE 2.** A potential profile of an RTD under investigation (left). The initial (coherent) density, the two computed correction integrals, and the correction of the density due to scattering (right).

- The BG04-ACAD has 40 worker nodes with 2xOpteron 2,4 GHz (total 80 cores $>$ 120 KSI2000), 4GB RAM on each node, and a low-latency Myrinet interconnect for MPI jobs.

Numerical results are presented in the next paragraph.

## NUMERICAL EXPERIMENTS

The presented algorithm has been applied to a resonant tunneling diodes (RTDs) as shown in the left part of Figure 2. As previously indicated, the initial density is obtained by using a Green's function calculation. This data is then fed to the implementation of the presented Monte Carlo algorithm. The formulated Monte Carlo algorithm is easily parallelized as the individual tasks do not share interdependencies and hence have no need for communication once calculation has been started. This low communication makes this application very suited for deployment on a grid. An almost linear speed is obtained by running several different instances of the Monte Carlo code and then simply averaging the all the results of the individual simulation runs. Care has only to be taken to account for this kind of parallelization in the random number generator to avoid distortions of the probabilities. Since the noise of the obtained results only decreases proportional to $\sqrt{n}$, with $n$ being the number of samples, the quality of the results greatly benefit from the opportunities of running multiple jobs simultaneously, which are provided by the grid. The overall simulation time may thereby be decreased from approximately one week to a single day.

The results from these calculations are depicted in the right part of Figure 2. An increase of the density within the quantum well regions can be observed, which may be attributed to scattering. Furthermore the correction tens to 0 in the contact regions, as has been predicted in the theoretical considerations.

## CONCLUSIONS

A scheme to treating quantum phenomena in a mixed mode, which accounts for both, a coherent description as well as the coherence breaking scattering mechanisms has been presented. After a theoretical derivation of the procedure and a presentation of the used run time environment of the grid, numerical results from an implementation have been presented.

## ACKNOWLEDGMENTS

This work is supported by ASO through the GRINKO project, grant no:K-04-2008 and the NSF of Bulgaria through grant no: DO02-146/2008, as well by the EC under the FP7 Research Infrastructure, SEE-GRID-SCI project, contract no. 211338.

## REFERENCES

1. S. M. Goodnick and D. Vasileska, "Computational electronics," in *Encyclopedia of Materials: Science and Technology*, K. H. J. Buschow, R. W. Cahn, M. C. Flemings, E. J. Kramer, and S. Mahajan, Eds., New York, 2001, vol. 2, pp. 1456–1471, Elsevier.
2. O. Baumgartner, P. Schwaha, M. Karner, M. Nedjalkov, and S. Selberherr, "Coupling of non-equilibrium Green's function and Wigner function approaches," in *Proc. Simulation of Semiconductor Processes and Devices*, Hakone, Japan, ISBN: 978-1-4244-1753-7, pp. 931–934, 2008.
3. I. Foster and C. Kesselman, "The Grid Blueprint for a New Computing Infrastructure," Morgan Kaufmann, San Francisco, 1999.
4. http://goc.grid.sinica.edu.tw/gstat/seegrid/
5. http://glite.web.cern.ch/glite/documentation/default.asp