Monte Carlo Grid Application for Electron Transport⋆

Emanouil Atanassov, Todor Gurov, Aneta Karaivanova, and Mihail Nedjalkov

IPP - Bulgarian Academy of Sciences,
Acad. G. Bonchev St., Bl.25A, 1113 Sofia, Bulgaria
emanouil@parallel.bas.bg, gurov@parallel.bas.bg, anet@parallel.bas.bg,
mixi@parallel.bas.bg

Abstract. In this paper we present a Grid application developed for electron transport problems called **SALUTE** (**S**tochastic **AL**gorithms for **U**ltra-fast **T**ransport in s**E**miconductors). We consider a physical model of a femtosecond relaxation of optically excited electrons which interact with phonons in an one-band semicondoctor. The electron-phonon interaction is switched on after a laser pulse creates an initial electron distribution. The Barker-Ferry equation is utilized as a quantum-kinetic model of the process under consideration. Two cases of this process are investigated - with and without an applied electric field. The electric field causes shift in the replicas, population of the semiclassically forbidden regions and influences the broadening and retardation of the electron distribution. The paper describes Grid implementation of these CPU-intensive algorithms. Using this application innovative results for different materials can be obtained. Here we present the first version of SALUTE which is used to obtain innovative results for GaAs materials. The results from a number of tests on MPI-enabled Grid are shown and disscussed.

Keywords: Monte Carlo methods, electron transport, Markov chains, parallel computing, Grid applications.

1   Introduction

Monte **C**arlo **M**ethods (MCMs) are based on the simulation of stochastic processes whose expected values are equal to computationally interesting quantities. MCMs offer simplicity of construction, and are often designed to mirror some process whose behavior is only understood in a statistical sense. However, there is a wide class of problems where MCMs are the only known computational methods of solution. It is well known that MCMs form the computational foundation for many fields including transport theory, quantum physics, computational chemistry, finance, etc.

In this paper we describe a Grid application, which solves a computationally intensive problem arising from semiconductor physics (see Section 2). The underlying stochastic algorithms are of Monte Carlo type. The application is called SALUTE for simplicity.

When using a MC method, a set of computational random numbers has to be generated and used to statistically estimate a quantity of interest. The probabilistic convergence rate is known to be approximately $O(N^{-1/2})$, where N is the number of underlying random samples. A serious drawback of the MC methods is the slow rate of convergence. In order to accelerate the convergence rate of the MC methods several techniques have been developed. Variance reduction methods, like antithetic varieties, stratification and importance sampling [1], reduce the variance which is a quantity to measure the probabilistic uncertainty. Another way to improve the convergence is to use quasi-random sequences instead of pseudorandom. While pseudorandom numbers are constructed to imitate the behavior of the truly random numbers, quasi-random sequences are deterministic and constructed to be as uniform as possible at the cost of correlation.

Parallelism is an alternative way to accelerate the convergence of a MC computation. If n processors execute n independent copies of a MC computation, the accumulated result has a variance n time smaller than that of a single copy. Due to the nature of MC methods, it is sometimes appropriate to have as many executions as possible to reduce the statistical error. On the other hand, for a distributed MC application: once a distributed task starts, it can usually be executed independently with almost no inter-process communications. Therefore, MC applications are widely perceived as computationally intensive but naturally parallel. The subsequent growth of computer power, especially that of the parallel computers and distributed systems, made distributed MC applications possible to carry out more and more ambitious calculations [2, 3]. Actually, many MC applications and software packages in science and engineering, for example, CHARMM [4] for macromolecular dynamics simulation implemented using MPI, PMC [5] for nuclear physics simulation on Livermore Message Passing System, and MESYST for simulation of 3D tracer dispersion in atmosphere running on a CRAY T3E parallel machine, have already taken advantage of the power of parallel systems to achieve a more accurate understanding of the problem or better performance of the computation.

Compared to the parallel computing environment, a large-scale distributed computing environment or a Computational Grid [6, 7, 8] has tremendous amount of computational power. Let us mention the EGEE Grid which today consists of over 18900 CPU in 171 Grid sites[1].

---

[1] The Enabling Grid for E-sciencE (EGEE) project is funded by the European Commission and aims to build on recent advances in grid technology and develop a secure grid infrastructure which is available to scientists 24 hours-a-day. The project aims to provide researchers in both academia and industry with access to major computing resources, independent of their geographic location. the EGEE project identifies a wide-range of scientific disciplines and their applications and supports a number of them for deployment. For more information see http://public.eu-egee.org/.

There are already MC applications running on the Grid (see some of the EGEE applications), and also publications like [9, 10, 11], for example, considering MCMs and large distributed computing paradigms.

Effectively exploring the power of distributed MC application requires that the underlying random number streams in each subtask are independent in a statistical sense. The main techniques used in parallel random number generators to distribute sequentially generated random number sequence among different processors include sequence splitting and leapfrog. One problem with the sequence splitting and leapfrog techniques is that we must either assume that the number of nodes is fixed or at least bounded which will restrict the scalability of the distributed MC computations. Another way to generate parallel random number sequences is to produce independent sequences by properly parameterizing pseudorandom generators [12]. The SPRNG (Scalable Parallel Random Number Generators) library [13] was designed to use parameterized pseudorandom generators to provide independent random streams. Some generators in SPRNG can generate up to $2^{31}-1$ independent random number streams with sufficiently long period and good quality [13]. These generators meet the requirements of the most grid-based MC applications.

The intrinsically parallel aspect of MC applications makes them an ideal fit for the grid-computing paradigm. In general, grid-based MC applications can divide the MC task into a number of subtasks (called jobs in EGEE Grid) by the *task-split service* and utilize the grid's *Workload Management Service* (also called *Resource Broker* in EGEE) to dispatch these independent subtasks to different nodes [14]. The *connectivity services* are provided using the *gridftp* protocol. The execution of a subtask takes advantage of the Storage Elements to store the executable, intermediate results, and to store each subtask's final (partial) result. When the subtasks are done, the EGEE *Workload Management Service* and the *Logging and Bookkeeping Service* are used to collect the information about the results and then the *task-gathering service* gets the results of the successfully executed subtasks. We have prepared a task-split service and a task-gathering service using the SQL language. The paper is organized as follows: After an overview (the Introduction) we describe the application (section 2), then we discuss the parallel and Grid implementation of SALUTE, and, finally, we give some conclusions and directions for future work.

2   Description of the Application

Many problems of interest in transport theory and related areas can be cast as a Fredholm integral equation of the second kind [1]. It is instructive to rewrite such equation as:

$$f = Kf + \varphi, \tag{1}$$

where K is an integral operator, and f, φ are functions in a Banach space of integrable functions.

The equation (1) shows that the kernel, K, appears as a linear operator. MCMs are merely probabilistic ways of applying linear operators. The formal
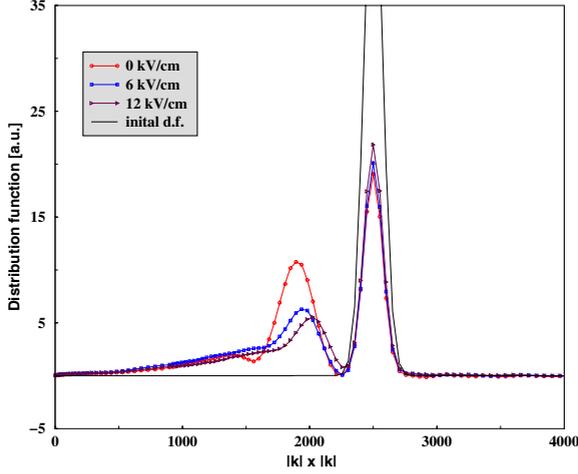
Fig. 1. MC solution at positive direction on the z axe (t=200 femtoseconds)

solution of the equation is presented as a Neumann series with terms obtained by consecutive application of K. The general MCM problem is to calculate linear functionals of the solution:

$$J(f) = \int_D h(x)f(x)dx = (h, f), \qquad (2)$$

where h a given function. In particular, the MCM approach consists of the following steps (1) to define a suitable random variable whose mathematical expectation is equal to the solution, (2) to determine appropriate initial and transition probability densities by using the kernel of the integral operator after some necessary transformations for the chosen random process, (3) to apply various variance reduction techniques to reduce the statistical error, and, finally, to formulate the appropriate statistical estimator.

MCMs for quantum transport in semiconductors and semiconductor devices have been actively developed during the last decade [15, 16]. If temporal or spatial scales become short, the evolution of the semiconductor carriers cannot be described in terms of the Boltzmann transport [17] and therefore a quantum description is needed. We note the importance of active investigations in this field: nowadays nanotechnology provides devices and structures where the carrier transport occurs at nanometer and femtosecond scales. As a rule quantum problems are very computationally intensive and require parallel and Grid implementations.

A Wigner equation for the nanometer and femtosecond transport regime has been derived from a three equations set model based on the generalized Wigner function [18]. The full version of the equation poses serious numerical challenges. A simplified version of the equation (in the homogeneous case - the Levinson or Barker-Ferry equation, [19]) is analyzed in the framework of SALUTE. The physical model describes a femtosecond relaxation process of optically excited
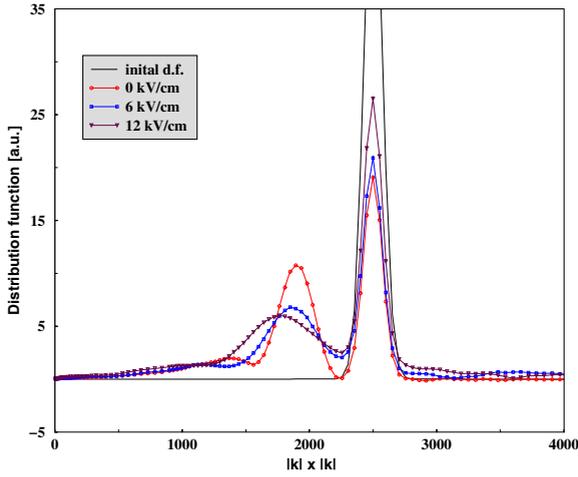
Fig. 2. MC solution at negative direction on the z axe (t=200 femtoseconds)

electrons which interact with phonons in single-band semiconductor. The interaction with phonons is switched on after a laser pulse creates an initial electron distribution.

Experimentally, such processes can be investigated by using ultra-fast spectroscopy, where the relaxation of electrons is explored during the first hundred femtoseconds after the optical excitation. In our model we consider a low-density regime, where the interaction with phonons dominates the carrier-carrier interaction. Two cases are studied using SALUTE: electron evolution in presence and in absence of electric field. The electric field causes shift in the replicas population of the semiclassically forbidden regions and influences the broadening and
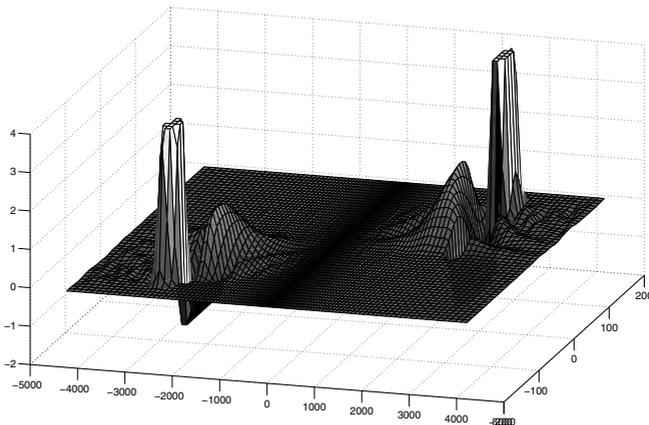
Fig. 3. Distribution of optically generated electrons in a quantum wire

retardation of the electron distribution (see Figures 1-3). The intracollisional field effect is clearly demonstrated as an effective change of the phonon energy, which depends on the field direction and the evolution time.

Another formulation of the Wigner equation considers inhomogeneous case when the electron evolution depends on the energy and space coordinates. The problem is relevant e.g. for description of the ultrafast dynamics of confined carriers. Particularly we consider a quantum wire, where the carriers are confined in the plane normal to the wire by infinite potentials. The initial condition is assumed both in energy and space coordinates. Grid implementations have been used to obtain the space and energy dependence of the evolution in the zero field case (see Figure 3).

SALUTE integrates a set of Monte Carlo algorithms for simulation of ultra-fast carrier transport in semiconductors considered in above mentioned versions of the Wigner equation. A detailed description of the algorithms can be found in [20, 21, 22].

3   Grid Implementation

SALUTE solves an NP-hard problem concerning the evolution time using Monte Carlo algorithms which are inherently parallel. Thus, SALUTE is a very good candidate for implementations on MPI-enabled Grid sites. It is proved [20, 22] that stochastic error has order $O(N^{-1/2})$, where N is the number of samples of the MC estimator for a given evolution time t. When t is fixed and $N \to \infty$ the error decreases, but for N fixed and t large, the factor for the error looks ominous. That is why the problem of estimating the electron energy distribution function for long evolution times with small stochastic error requires combining both MC variance reduction techniques and distributed or parallel computations. By using the Grid environment provided by the EGEE project middleware, we were able to reduce the computing time of Monte Carlo simulations of ultra-fast carrier transport in semiconductors.

The numerical results discussed in the Figures 1-3 are obtained for zero temperature and GaAs material parameters: the electron effective mass is 0.063, the optimal phonon energy is $36 meV$, the static and optical dielectric constants are $\epsilon_s = 10.92$ and $\epsilon_\infty = 12.9$. The initial condition at $t = 0$ is given by a function which is Gaussian in energy, scaled in a way to ensure, that the peak value is equal to unity [22].

Successful tests of the application were performed at the Bulgarian EGEE-GRID sites using the South Eastern Europe Resource Broker at the IPP-BAS.

Table 1. CPU times for SALUTE execution on MPI-enabled Grid site

Number of CPU	2	4	6
Time	9790	4896	3262

The MPI implementation was MPICH 1.2.6, and the execution is controlled from the Computing Element via the Torque batch system. The timing results for evolution time $t = 100$ femtoseconds are shown in Table 1. The parallel efficiency can clearly be seen from the results - with 4 CPUs, the running time is half less than the running time with 2 CPUs, with 6 CPUs the running time is 3 times less than with 2 CPUs.

4   Conclusion and Future Work

A Grid application that solves important problems in the field of nanotechnologies has been developed, studied and tested on MPI-enabled Grid sites. The test results show excellent parallel efficiency. Obtaining results for larger evolution times requires more computational power, which means that the application should run on larger sites or on several sites in parallel. The application can provide results for other types of semiconductors like Si or for composite materials.

The SALUTE application is able to fully utilize the computational power of Grids.

References

1. KALOS, M.A. AND P.A. WHITLOCK: *Monte Carlo Methods*. Volume 1: Basics, Wiley, New York, 1986.
2. KARAIVANOVA, A.: *Monte Carlo Methods for Parallel Processing*. PhD dissertation, IPP-BAS, 1996.
3. L. SMITH AND M. BULL: Development of mixed mode MPI/OpenMP applications. *Scientific Programming*, series 9 (**2/3**), 2001, 83–98.
4. B. R. BROOKS, R. E. BRUCCOLERI, B. D. OLAFSON, D. J. STATES, S. SWAMINATHAN, AND M. KARPLUS: CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamic Calculations. *J. Comp. Chem.*, **4**, 1893, 187–217.
5. J. A. RATHKOPF: *PMC: A shared short-cut to portable parallel power*. Lawrence Livermore National Laboratory, Livermore, CA, UCRL-112311, 1992.
6. J. FOSTER AND C. KESSELMANN: *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann, 1999.
7. J. FOSTER, C. KESSELMANN AND S. TUECKE: The Anatomy of the Grid. *International Journal of Supercomputer Applications*, **15(3)**, 2001.
8. Globus website, http://www.globus.org .
9. Entropia: *PC Grid Computing*. http://www.entropia.com, 2002.
10. F.SOLMS AND W.H.STEEB: Distributed Monte Carlo integration using CORBA and Java. *International Journal of Modern Physics*, **9(7)**, 1998: 903-915.
11. M. H. ZHOU: *A Scientific Computing Tool for Parallel Monte Carlo in a Distributed Environment*. PhD dissertation, School of Mathematics, University of Southern Mississippi, 2000.
12. M. MASCAGNI, D. SEPERLEY AND A. SRINIVASAN: SPRNG: A Scalable Library for Pseudorandom Numbar Generation. *ACM Transactions on Mathematical Software*, 2000.
13. SPRNG website, http://sprng.cs.fsu.edu.

14. R. Buyya, S. Chapin, and D. DiNucci: Architectural Models for Resource Management in the Grid. In: R. Buyya, M. Baker (Eds.): *Grid Computing*, GRID 2000, LNCS, vol. **1971**, Springer Verlag, 2000: 18–35.

15. R. Rossi, C. Jacoboni, M. Nedjalkov: A Monte Carlo Solution of the Wigner transport Equation. *Semiconductor Science Technology*, **9**, 1994: 934-940.

16. M. Nedjalkov, I. Dimov, F. Rossi, C. Jacoboni. Convergent of the Monte Carlo Algorithms for the Solution of the Wigner Quantum-Transp. Equation. *J. Math. Comp. Mod.*, Vol. **23** (8/9), 1996: 159–166.

17. J. Rammer: Quantum transport theory of electrons in solids: A single-particle approach. *Reviews of Modern Physics*, series **63(4)**, 1991, 781–817.

18. M. Nedjalkov, R. Kosik, H. Kosina, and S. Selberherr: A Wigner Equation for Nanometer and Femtosecond Transport Regime. In: *Proceedings of the 2001 First IEEE Conference on Nanotechnology*, IEEE, 2001.

19. J. Barker and D. Ferry: Self-scattering path-variable formulation of high field time-dependent quantum kinetic equations for semiconductor transport in the finite-collision-duration regime. *Physical Review Letters*, series **42 (26)**, 1979: 1779–1781.

20. T.V. Gurov, M. Nedjalkov, P.A. Whitlock, H. Kosina and S. Selberherr: Femtosecond relaxation of hot electrons by phonon emission in presence of electric field. *Physica B*, **314**, 2002: 301–304.

21. T.V. Gurov and I.T. Dimov: A Parallel Monte Carlo Method for Electron Quantum Kinetic Equation. *LNCS*, **2907**, Springer-Verlang, 2004: 153–160.

22. T.V. Gurov, P.A. Whitlock: An efficient backward Monte Carlo estimator for solving of a quantum kinetic equation with memory kernel. *Mathematics and Computers in Simulation*, **60**, 2000: 85–105.