Accelerated parallel computing of carrier transport simulation utilizing graphic processing units

A. Suzuki [1], T. Kamioka [1, 3], H. Imai [1], Y. Kamakura [2, 3], and T. Watanabe [1, 3]
[1] Faculty of Science and Engineering, Waseda University, Japan,
[2] Graduate School of Engineering, Osaka University, Japan,
[3] Japan Science and Technology Agency (JST), CREST, Japan.
e-mail: suzuki@watanabe.nano.waseda.ac.jp

INTRODUCTION

As the scaling down of Si MOSFET proceeds, the discreteness of the impurity distribution and the number of transported carriers emerges, so that the variability of device characteristics and intrinsic current noise arise as serious concerns [1, 2]. To address these issues, particle-based carrier transport simulation technique, e.g. molecular dynamics (MD), is a powerful method which allows us intuitive insight into the effect of the discreteness of impurity ions and carriers. However, the MD simulation requires huge computation resources to calculate the interaction forces acting on each pair of particles. This is a great obstacle to simulate the entire device structure including source and drain regions with high carrier and impurity density.

In recent years, graphics processing unit (GPU) has attracted much attention due to its large-scale parallelism ability. Many successful implementations on GPU have been reported for MD and Monte Carlo (MC) algorithms [3, 4]. However, to our knowledge, no application to the carrier transport simulation has been reported so far.

In this work, we demonstrate that the Ensemble Monte-Carlo/Molecular Dynamics (EMC/MD) method [5, 6] is successfully accelerated by GPU. The present result shows that EMC/MD simulation can be easily applied to the practical nano-scale device simulation including source and drain regions.

SIMULATION METHOD

In the EMC/MD method, carriers are treated as classical particles, and their real-space trajectories under the Coulomb point-to-point potentials are calculated by the MD algorithm. The acoustic and optical phonon scattering are described as stochastic changes in the momentum of carriers according to the standard energy-dependent formulations [6].

In this work, the EMC/MD algorithm is parallelized by utilizing GPU. Single GPU thread is assigned to the calculations associated with one electron, so that the number of thread is equal to the number of electrons. We employed NVIDIA GeForce GTX560Ti and GeForce GTX690, and compared the execution time with the single core calculation with Intel core i7 3930k CPU.

Figure 1 shows the simulated bulk n-type Si model. All electrons and ions are randomly placed inside the unit cell, on which the 3D periodic boundary condition is adopted. An external electric field of 1kV/cm is applied along the <100> axis. The time step is 10^{-17} s. We evaluate the execution time changing the number of electrons with keeping the total electron density. The execution time is defined as the duration to complete the 10^5 steps calculation.

Since carriers and impurity ions are treated as point charges in EMC/MD, a singular point of the Coulomb potential appears at zero distance. This is problematic in deal with the majority carriers in the source and the drain regions. To solve this problem, softened Coulomb potential is employed between electron and positively charged ions:

$$\varphi = -\frac{e^2}{4\pi\varepsilon}\frac{1}{\sqrt{r^2+\alpha^2}} \qquad (1)$$

where e is the elementary charge, ε is the permittivity of semiconductor, and α is the softening factor. Eqn. 1 is commonly used formula in the field of gravity calculation [7].

RESUITS AND DISCUSSION

Figure 2 shows the execution time of the EMC/MD simulation using CPU and GPU. In the case of CPU calculation, the execution time is $O(N^2)$. By parallelizing with GPU computing, the execution time is successfully reduced to $O(N)$. A cross point of the execution times appears at 200

978-3-901578-26-7

electrons. Above 200 electrons, the computation speed with GPU exceeds that of CPU, and the benefit of the parallel computation increases as the number of elections increases. For small number of electrons, GPU computation gives no speed up, due to the overhead of parallelism and the poor performance of single GPU thread compared with that of the single CPU core.

Figure 3 shows the speedup rate, which is the ratio of the execution time with GPU to that with CPU. The GPU parallel computation speed is enhanced by more than 10 times in the case of several thousand particles. Thus the GPU computation is suitable for the practical nano-scale device simulation including source and drain regions. GeForce GTX690 shows better performance and scalability than GeForce GTX560Ti, mainly due to the difference in the number of Stream processors of these GPUs.

Figure 4 shows the calculated low field mobility in n-type Si plotted versus the impurity concentration. The mobility is determined by the mean travel distance along <100> under the external field of 1kV/cm. To reproduce the experimental mobility [8], the softening parameter should be changed between lower and higher electron densities.

CONCLUTION

We demonstrate that the GPU parallel computing is effective to accelerate the EMC/MD simulation involving large number of electrons and impurity ions. The nano-scale devices including source and drain regions can be fully simulated in a reasonable execution time by utilizing GPU.

ACKNOWLEDGEMENT

This work was supported by Grant-in-Aid for Scientific Research (B) (24310082) from Japan Society for the Promotion of Science and CREST from Japan Science and Technology Corporation (JST).

REFERENCES

[1] T. Kamioka et al.: IEDM Extd. Abs. 17.2.4 (2012).
[2] N. Sano et al.: IEDM Tech. Dig., 275 (2000).
[3] E. Alerstam et al.: J. Biomed Opt., 13:060504 (2008).
[4] J.A. Anderson et al.: J. Comp. Phys., 227, 5342 (2008).
[5] C. Jacoboni et al.: Rev. Modern Phys., **55**, 645 (1983).
[6] Y. Kamakura et al.: IEICE Trans. Elec. E-86-C, 357 (2003).
[7] A.E. Evrard, Mon. Not. R. astr. Soc., 235, 911 (1988).
[8] D.M. Caughey et al.: Proceeding of the IEEE, 55, 12, 2192 (1967).

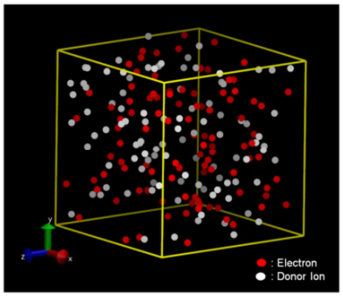

Fig. 1. Schematic view of the simulation system for electron transport in bulk-Si. Electrons and impurity ions are randomly placed inside a box.

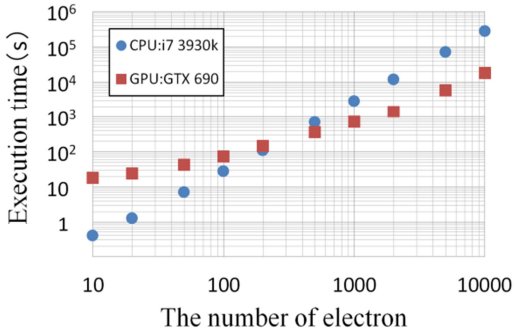

Fig. 2. The results compering the execution time of the CPU used only single core and GPU. The time step is 10^{-17}, and 10^5steps simulated. The electron density D is 1.0×10^{18} cm^{-3}, and the number of electron is from 10 to 10000.

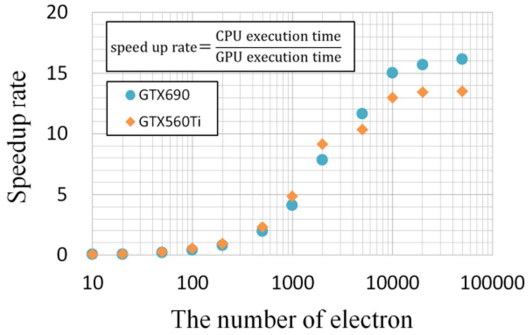

Fig. 3. The speed up rate of which is ratio between the execution time of CPU and that of GPU. The speedup rate reaches more than 10 times in several thousand particles.

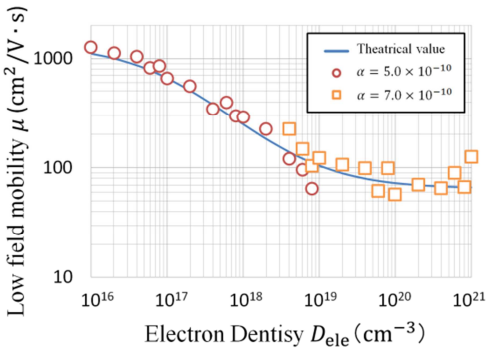

Fig. 4. The impurity concentration dependence of the low field mobility. The simulation is performed for a long time enough to reach the convergence. The low field mobility is simulated 100 ps with 100 elections and impurity ion.