Full-scale Whole Device EMC/MD Simulation of Si Nanowire Transistor Including Source and Drain Regions by Utilizing Graphic Processing Units

Akito Suzuki, Takanobu Watanabe
Faculty of Science and Engineering
Waseda University
Tokyo, Japan
suzuki@watanabe.nano.waseda.ac.jp

Takefumi Kamioka
Semiconductor Laboratory
Toyota Technological Institute
Nagoya, Japan

Yoshinari Kamakura
Graduate School of Engineering
Osaka University
Osaka, Japan

***Abstract*—We have realized the full-scale whole device EMC/MD simulation including source and drain regions by utilizing graphic processing unit. The transfer characteristic of a gate-all-around nanowire Si MOSFET is simulated by reproducing the field effect of the surrounding gate electrode with spreading charged particles on the gate insulator layer. We have found an appreciable impact of the random dopant distribution (RDF) in source and drain regions on the drain current variability. Furthermore, the dynamic fluctuation of the drain current is found to be increase as the channel length decreases. The EMC/MD simulation powered by GPU is a useful method to investigate the dynamic fluctuation as well as the statistical device-to-device variability of nano-scale FETs.**

Keywords—Si nanowire Transistor; ensemble Monte Carlo / molecular dynamics (EMC/MD); Graphic Processing Unit (GPU); random telegraph noise (RTN); random dopant fluctuation (RDF)

I. INTRODUCTION

The limit of Si CMOS scaling and integration is considered to be determined by the device-to-device variability, e.g. random dopant fluctuation (RDF), line-edge roughness, and the fluctuation in the metal gate work function [1-5]. Dynamic current fluctuation induced by the random telegraph noise (RTN) and thermal/shot noise also emerges as the concern to the integration, because the total transported charges in one clock cycle varies from time to time even in the identical devices [6]. To analyze these static and dynamic variability issues, a particle-based carrier transport simulation technique, e.g. molecular dynamics (MD), is useful because the discreteness of carriers and impurity ions can be described intuitively. However, the MD simulation requires huge computation resources to calculate the point-to-point Coulomb interactions for each pair of particles, that has prevented a full-scale simulation of the whole device structure with source and drain regions which include large number of carries and impurities.

In recent years, graphic processing unit has attracted much attention due to its large scale parallelism ability. Many studies have been reported that MD and Monte Carlo (MC) algorithms are accelerated by utilizing GPU [7, 8]. In our previous work [9], an accelerated ensemble Monte Carlo / molecular dynamics (EMC/MD) [10, 11] simulation code utilizing parallel computing on GPU was developed, and the execution speed of the EMC/MD calculation was enhanced by 10 times and more.

In this work, we carried out a full-scale whole device EMC/MD simulation of a Si nanowire (NW) transistor including source and drain regions utilizing by GPU. The field effect caused by the surrounding gate electrode is reproduced by spreading particle with fractional charges over the gate insulator layer. Thus troublesome problems on the mesh cutting are avoided completely. The I_D-V_G characteristics of the Si NW MOSFET are successfully reproduced. The impact of the device variability due to RDF in source and drain regions and the transported charge variation due to dynamic current fluctuation is investigated.

II. SIMULATION METHODS

Figure 1 shows the simulation model of gate-all-around (GAA) n-i-n Si NW model. Table.1 shows the device parameters determined based on ITRS 2012 [12]. The periodic boundary condition is adopted in the longitudinal direction of the NW. A quantum confinement potential is adopted in the cross sectional plane [6, 13], which is given by (1), (2), and (3):

$$V(r,N_{ele}) = -\frac{2k_bT}{q}\left[\ln\left[\sin\left(\frac{\pi}{\phi}\left(r+\frac{\phi}{2}\right)\right)\right]+a+b\right] \quad (1)$$

$$a(N_{ele}) = \ln\left(\frac{A(N_{ele})N_{ele}}{\pi L_G n_i \phi\sqrt{2\phi}}\right) \quad (2)$$

$$b(r, N_{ele}) = \ln\left[\exp\left(-\frac{B(N_{ele})}{\phi}\left(r+\frac{\phi}{2}\right)\right)+\exp\left(\frac{B(N_{ele})}{\phi}\left(r-\frac{\phi}{2}\right)\right)\right] \quad (3)$$

where ϕ is the channel diameter, r is a position along channel diameter. N_{ele} is the number of electrons in the channel. L_G is the gate length. n_i is the intrinsic carrier density of Si. $a(N_{ele})$ and $b(r, N_{ele})$ are functions depends on N_{ele}, and these parameter are obtained by (4), (5).

$$A\left(B(N_{ele})\right) = 2\left[2\exp(-B)+\frac{\pi^2\left(1-\exp(-2B)\right)}{B\left(B^2+\pi^2\right)}\right]^{-\frac{1}{2}} \quad (4)$$

$$B(N_{inv}) = \phi\left[\frac{qm^*\pi^2}{4\varepsilon_{Si}\hbar^2}\left(\frac{5q}{6\pi\phi L_G}N_{ele}\right)\right]^{\frac{1}{3}} \quad (5)$$

N_{ele} is determined self-consistently in the simulation. Only conducting electrons are considered as carriers. Figure 2 is an electron potential landscape in the GAA NW transistor at on-state. All electrons and impurity ions are randomly placed inside the source and drain region, and no impurity ion is included in the channel region. The number density of electrons and impurity ions is set to 10^{20} cm^{-3}. A single EMC/MD simulation is lasted for 50 ps or 1 ns with the time step of 10^{-16} s.

TABLE I. DEVEICE PARAMETER FROM ITRS 2012 [12]

Channel Length [nm]	Device Parameter			
	Drain Voltage [V]	*EOT [nm]*	*Diameter [nm]*	*Number of Electron*
30	1.00	1.00	20.0	1280
20	0.85	0.88	13.3	566
10	0.65	0.60	6.7	144
7	0.60	0.48	4.7	72
5	0.55	0.42	3.3	36

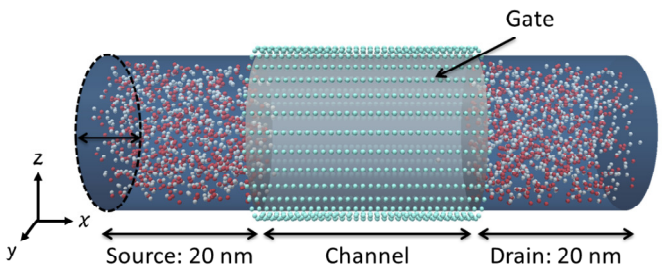

Fig. 1. Example of a figure captionSimulation model for the GAA n-i-n Si NW model.

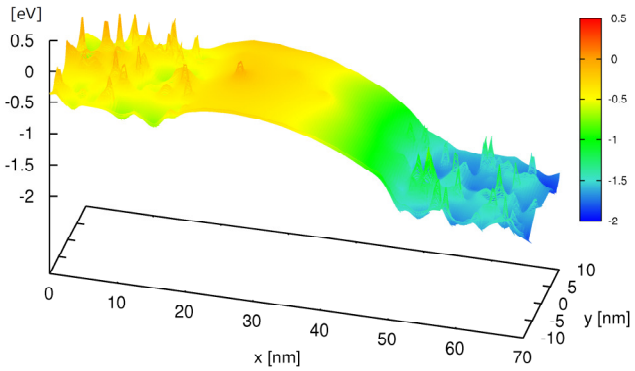

Fig. 2. The electron potential landscape in the GAA NW at on-state.

In the EMC/MD method (Fig. 3), 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 [11]. Carriers and impurity ions are treated as point charges in the EMC/MD method, and thus a singular point of the Coulomb potential appears at zero distance and the Coulomb force diverges to infinity. To solve the problem, we employed softened Coulomb potential between electron and positively charged ions [9]:

$$V(r) = -\frac{e^2}{4\pi\varepsilon}\frac{1}{\sqrt{r^2+\alpha^2}} \quad (6)$$

Equation (6) is Coulomb potential with softening factor. In this EMC/MD simulation of Si NW transistor, softening factor is 7.0 x 10^{-10}, which is determined by low field mobility calculation of n-type Si bulk.

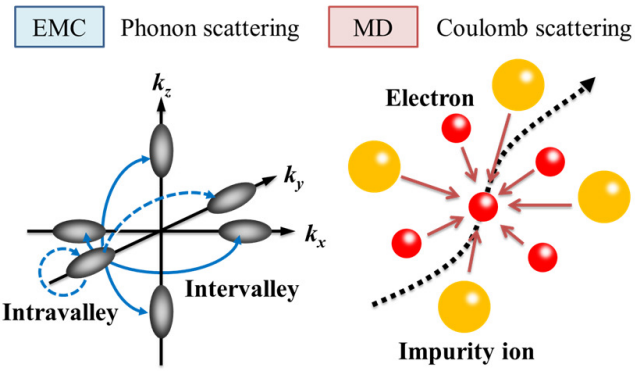

Fig. 3. Scattering processes described in the EMC/MD method. (a) Phonon scattering based on EMC method. (b) Coulomb scattering based on MD method.

To reproduce the filed effect caused by the gate electrode, many fixed charges are placed over the gate insulator layer, as shown in Fig.1. These fixed charges represent accumulated charges at metal/oxide interface, and they bear a fractional

charge determined by the gate voltage and oxide film capacitance. To satisfy the charge neutrality condition of the whole device, the number of electron in the source and drain regions are adjusted. The work function of the gate electrode is assumed to be identical to that of the Si channel region. The drain current is calculated by counting the number of electrons passed through the center of channel region at every MD step.

In this work, calculation of the Coulomb force and calculation of the MD and MC is parallelized on GPU. One thread of GPU is assigned to the calculation with respect to one electron. It is possible to parallel computing the number of threads in the GPU.

Figure 4 compares the for-loop structures of the simulation code for a single core CPU and GPU calculation. In the case of CPU calculation, the execution time is $O(N^2)$. By parallelizing with GPU computing, the execution time is reduced to $O(N)$. The benefit of the parallel computation increases as the number of electrons increases. For example, in the case of a NW transistor model with a diameter of 10 nm and the source and drain regions of 20 nm long, the device model includes 3,840 particles. To perform the EMC/MD simulation of 10^7 steps (1 ns), it takes about 4 days by using CPU. On the other hand, GPU can reduce the calculation time to about 18 hours. Therefore, we are allowed to perfume the statistical analysis by repeating the EMC/MD simulation by utilizing GPU. In this work, we used Intel core i7 3930k CPU and NVIDIA GeForce GTX690 and GTX560Ti GPUs.

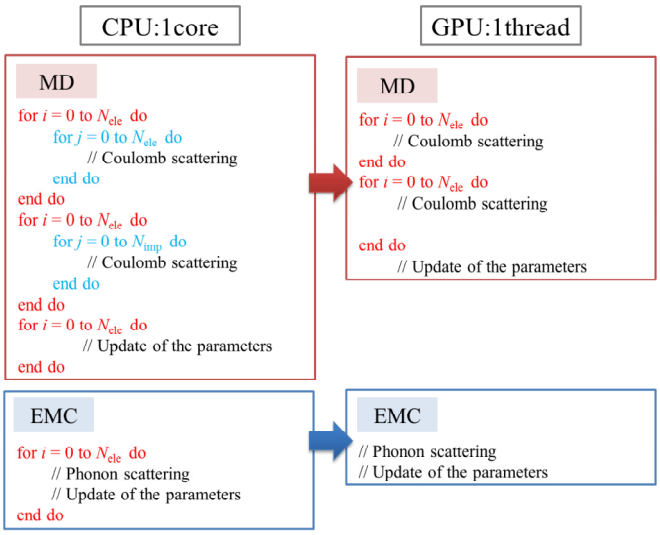

Fig. 4. The change of computational complexity due to parallelization.

III. Results and Discuttion

Figure 5 shows the I_D-V_{GS} characteristics of the GAA n-i-n NW transistor estimated by the EMC/MD simulation. Here the drain current is normalized by the channel diameter. At a drain and gate bias condition, we repeated the simulation 30 times by changing the random number sequence to generate statistical samples. The obtained characteristic shows that the off-leakage current between the source and drain increases as the channel length shortens. Especially, appreciable deterioration is observed when the channel length is decreased from 10 nm to 7 nm. It is indicate that the NW diameter should be more scaled down than the ITRS 2012 expectation.

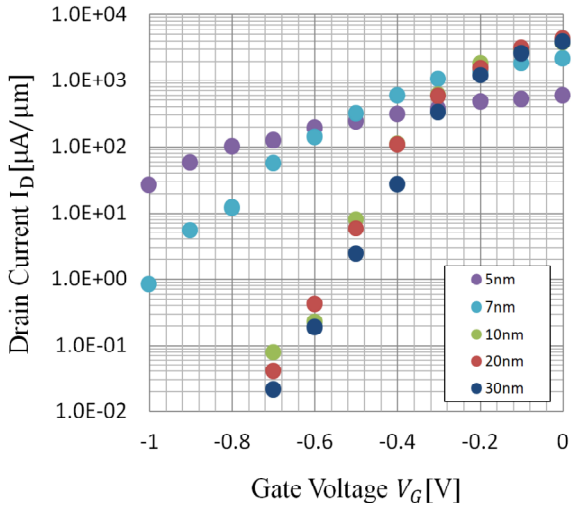

Fig. 5. The electron potential landscape on the cross-sectional plane in the GAA NW at on-state.

Figure 6 shows the effect of the RDF in S/D regions on the drain currents of on-state for various channel lengths. In this study, the on-state is defined as the gate voltage of -0.2 V so as to satisfy the recommended on-state current of 1000 – 2000 μA/μm in ITRS2012 [12]. Each data point shows the time-averaged drain current in the course of a simulation for 1 ns. Current variability caused by the S/D RDF also increases as the channel length shortens.

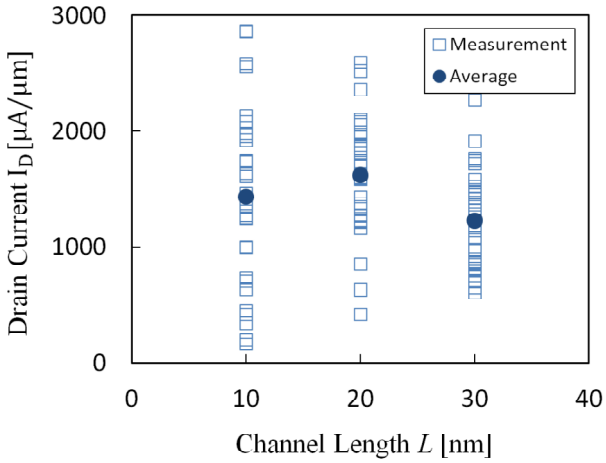

Fig. 6. The summary of the drain currents of on-state for various channel lengths.

Figure 7 shows how the time average of the drain current converges as the total simulation time increases, with respect to various channel lengths. The time to convergence decreases as the channel length L_G shortens. In the case of L_G = 10 nm, the mean drain current almost converges within about 1 ns. In other words, one pulse time of 1 ns is required to suppress the

total transported charge variation, so that the crock speed is limited to the order of a few GHz. The result indicates that the dynamic on-current fluctuation is largely affected by the separation between S/D regions, and it may become a serious obstacle to limit the operation speeds.

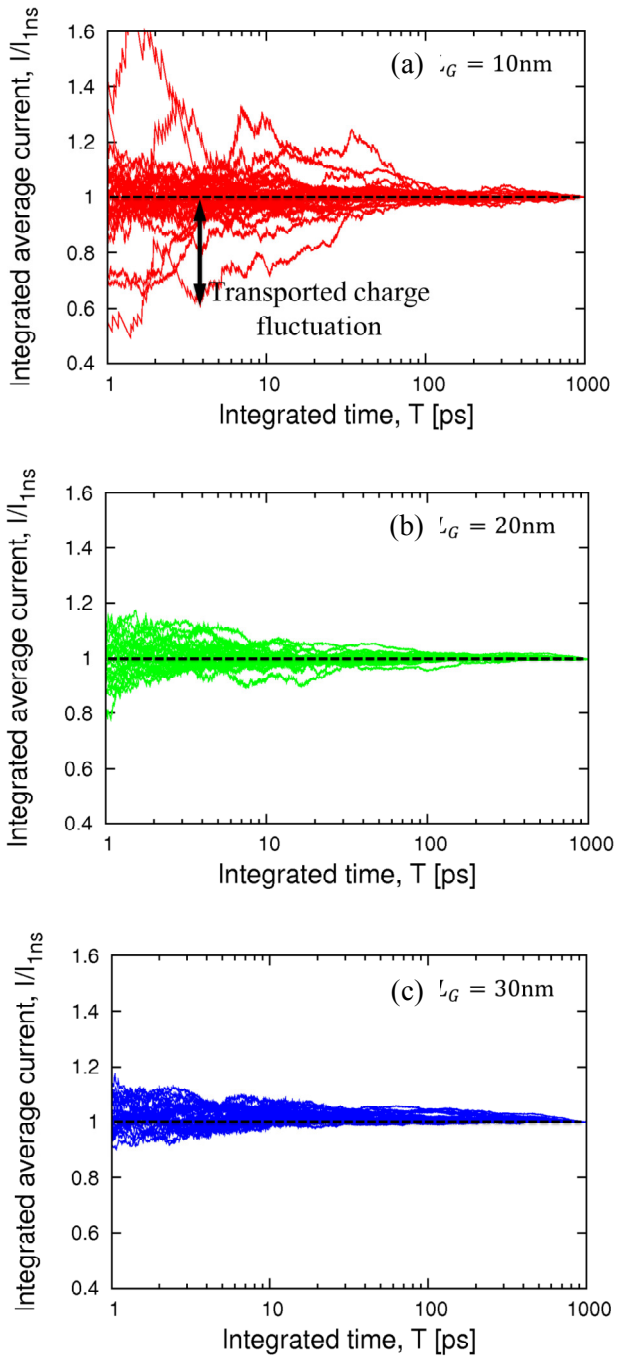

Fig. 7. The time averaged on-current vs the total simulatin time. (a) L_G = 10 nm, (b) L_G = 20 nm, (c) L_G = 30 nm.

The current variability due to the RDF in the source and drain regions is about ±1000 μA/μm, which is much larger than that of the transport charge variation during the one pulse time of 1 ns. The result suggests that RDF in the source and drain regions will emerges as a major factor to limit the Si CMOS scaling.

IV. Conclutions

The full-scale whole device EMC/MD simulation including source and drain regions are realized by utilizing GPU. The GPU computing drastically accelerates the EMC/MD simulation, and it is useful for the statistical analysis on the nano-scale device characteristics. We have investigated the static and dynamic variabilities induced by the RDF in the source and drain, and found that both of the mean current variability and dynamic fluctuation increases as the separation between source and drain shortens. These results implies that RDF in the source and drain regions will emerges as a major concern of the statistical variability, and it determine the limit of Si CMOS scaling and its operation speed.

References

[1] Y. Ye, T. Liu, M. Chen, S. Nassif, and Y. Cao, "Statistical modeling and simulation of threshold variation under random dopant fluctuations and line-edge roughness," VLSI Systems, IEEE Transactions on, 19.6, pp. 987-996, 2011.

[2] C. Riddet, C.Alexander, A. R. Brown, S. Roy, and A. Asenov, "Simulation of "Ab Initio" Quantum Confinement Scattering in UTB MOSFETs Using Three-Dimensional Ensemble Monte Carlo," Electron Devices, IEEE Transactions on, 58.3, pp. 600-608, 2011.

[3] X. Wang, A. R. Brown, B. Cheng, and A. Asenov, "Statistical variability and reliability in nanoscale FinFETs," IEDM Tech. Dig., pp. 5.4.1, 2011.

[4] S. Markov, B. Cheng, and A. Asenov, "Statistical variability in fully depleted SOI MOSFETs due to random dopant fluctuations in the source and drain extensions," Electron Device Letters, IEEE, 33.3, pp. 315-317, 2012.

[5] S. Markov, A. S. M. Zain, B. Cheng, and A. Asenov, "Statistical variability in scaled generations of n-channel UTB-FD-SOI MOSFETs under the influence of RDF, LER, OTF and MGG," SOI Conference, 2012.

[6] T. Kamioka, H. Imai, Y. Kamakura, K. Ohmori, K. Shiraishi, M. Niwa, K. Yamada, and T. Watanabe, "Current fluctuation in sub-nano second regime in gate-all-around nanowire channels studied with ensemble Monte Carlo/molecular dynamics simulation," IEDM Tech. Dig., pp. 17.2.1, 2012.

[7] E.Alerstam, T. Svensson, and S. Andersson-Engels, "Parallel computing with graphics processing units for high-speed Monte Carlo simulation of photon migration," Journal of biomedical optics, 13.6, pp. 060504-060504, 2008.

[8] J. A. Anderson, C. D. Lorenz, and A. Travesset, "General purpose molecular dynamics simulations fully implemented on graphics processing units," Journal of Computational Physics, 227.10, pp. 5342-5359, 2008.

[9] A. Suzuki, T. Kamioka, H. Imai, Y. Kamakura, and T. Watanabe, "Accelerated parallel computing of carrier transport simulation utilizing graphic processing units," IWCE Extended Abstract, pp. 166, 2013.

[10] C. Jacoboni, and L. Reggiani, "The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials," Reviews of Modern Physics, 55.3, pp. 645, 1983.

[11] Y. Kamakura, H. Ryouke, and K. Taniguchi, "Ensemble Monte Carlo/Molecular Dynamics Simulation of Inversion Layer Mobility in Si MOSFETs--Effects of Substrate Impurity," IEICE transactions on electronics, 86.3, pp. 357-362, 2003.

[12] ITRS2012, http://www.itrs.net/, (2013).

[13] L.Ge, and J. G. Fossum, "Analytical modeling of quantization and volume inversion in thin Si-film DG MOSFETs," Electron Devices, IEEE Transactions on, 49.2, pp. 287-294, 2002.