Edited by H. Ryssel, P. Pichler - September 1995

A New Statistical Enhancement Technique in Parallelized Monte Carlo Device Simulation

K. Shigeta, K. Tanaka, T. Iizuka, H. Kato[a] and H. Matsumoto

Microelectronics Res. Labs., NEC Corp., [a]ULSI Device Dev. Labs., NEC Corp.
1120, Shimokuzawa, Sagamihara, Kanagawa 229, Japan

Abstract

A new statistical enhancement technique (split-and-remove technique) in Monte Carlo device simulation, which is suitable for parallel processing, has been developed. By using this technique, an accurate energy distribution function near the drain edge can be obtained within a reasonable CPU time.

1. Introduction

On the advent of parallel computing, Monte Carlo(MC) device simulation is about to recover its dream of being a practical and accurate tool. Even in such a rosy view, a statistical enhancement technique is mandatory for sampling rare events, such as hot carriers, often jammed with stochastic noise due to a limited number of samplings. Split-and-gather (SG)[1](Fig. 1(a)) and multiple refresh (MR)[2] are such techniques that manipulate MC particles belonging to a region #i, defined in phase space $(\vec{r}, \vec{k})$, through adjusting particle's statistical weight w. This manipulation allows more MC particles with "light" weight to be loaded in a region of interest than in others, which leads to detailed sampling in the region.

The SG technique has two drawbacks in its gathering process; (i) momentum and energy cannot be conserved simultaneously, which causes a distortion of carrier distribution in phase space, (ii) searching a partner is CPU timewise expensive. The MR technique avoids such difficulties. However, it would be unsuitable for parallel computing. The MR technique refers to the distribution of particle weights in each region to readjust the number of MC particles to a desired population N_i in the region. This requires fetching particle data handled by other processors, which should be avoided because interprocessor communication is usually a bottleneck in parallel processing. In this paper, the authors propose a "split-and-remove" (SR) technique (Fig. 1(b)), with particular attention to parallel computing.

2. Split-and-Remove Technique

The SR technique comprises two processes; splitting and removing MC particles. In the splitting process, a much "heavier" particle than the desired weight W_i given by $(\sum_{\text{in}\,\#i} w)/N_i$ is split into $int(w/W_i)$ particles. In the removing process, a much "lighter" particle than W_i, is randomly removed with probability $r(0 < r < 1)$, since it comes into a region of little interest and hence it should be labeled as surplus particle to be removed for saving the CPU time. The weight of the unremoved surplus particle is scaled up by a factor of $1/(1-r)$ (Fig. 2). Therefore, the obtained carrier distribution in phase space is preserved in a statistical manner.

In the removing process, the total charge is not exactly conserved because of random process while its expected value is conserved. Although the total charge conservation could be realized by adding a further adjustment process that refers to all the particles in each region, the SR technique omits this additional process on purpose in parallel processing to reduce the interprocessor communication. In this way, complete parallelization with respect to particles is achieved.

The previous MR technique randomly chooses as many particles as there should be within each region among all existent particles according to their weights, which allows any particle to be removed or be split by chance. In contrast, the random removal in the SR acts only on the "light" particles labeled as surplus. Therefore, the SR is more robust against stochastic noise than the MR.

The number of MC particles can be kept close to N_i in each region without *a priori* knowledge of the distribution, since W_i is automatically updated along with the transit in the carrier distribution. The characteristics of the three SR, SG and MR techniques are summarized in Table 1.

3. Simulation Results and Discussion

Fig. 3 shows influences of the gathering process in the SG and the removing process in the SR on energy distribution. In the gathering process, two MC particles within the same energy region are joined satisfying momentum conservation. The removing process is not accompanied by any distortions in the energy distribution that is observed for the gathering process. In addition, device simulation results reveal that our method is 5 times faster than the SG technique because finding partners to be gathered is unnecessary.

The SR approach is implemented in our 2-carrier self-consistent MC device simulator. Fig. 4 shows MC particle distribution in MOSFET obtained by the simulation using parameters shown in Table 2. This parameter set is intended for allocating MC particles mainly to the drain edge region (#3). In fact, MC particles placed in #3 is about 40 times as many as the case without the technique. Statistical enhancement in energy space is also designed by dividing region #3 into 81 subregions as indicated under the x-axis in Fig. 5 and by allocating the same number of particles in each subregion.

As shown in Fig. 5, the accurate energy distribution function is obtained by using the SR technique (b), while it is jammed with stochastic noise without the SR technique (a), although almost the same number of MC particles are used. Moreover, the higher energy tail over 1.7eV with the SR is more accurate than the reference (dashed line) which consists of almost the same number of particles sampled during a much longer sampling period (4.0ps = 4000 time steps). This shows that the SR technique not only in real space but also in energy space is important for statistical enhancement of higher energy tail near the drain.

In our numerical experiments, the SR process in every time step takes up only 10% of the total CPU time. On a parallel machine Cenju-3[3](16PEs) which has a VR4400SC (75MHz) processor and 64MB local memory in each PE, the MC simulation for 1.0 ps (1,000 time steps) takes about 50 minutes.

4. Conclusion

The split-and-remove (SR) technique was proposed and implemented in our parallelized Monte Carlo device simulator. Its efficiency in statistical enhancement was demonstrated in the energy distribution function in a MOSFET. This technique drives MC device simulator toward a daily-use tool using a parallel machine.

References

[1] F. Venturi, R. K. Smith, E. C. Sangiorgi, M. R. Pinto and B. Ricco, "A General Purpuse Device Simulatior Coupling Poisson and Monte Carlo Transport with Applications to Deep Submicron MOSFET's," *IEEE Trans.Computer-Aided Design*, Vol. 8, pp. 360–369, 1989.

[2] R. Thoma, J. Peifer, W. L. Engl, W. Quade, R. Brunetti and C. Jacoboni, "An Improved Impact-ionization Model for High-energy Electron Transport in Si with Monte Carlo Simulation," *J. Appl. Phys.*, Vol. 69, pp. 2300–2311, 1991.

[3] K. Muramatsu, S. Doi, T. Washio and T. Nakata, "Cenju-3 Parallel Computer and its Application to CFD," in *Proc. 1994 Int. Symp. on Parallel Architectures, Algorithms and Networks (ISPAN)*, Kanazawa, pp.318–325, 1994.

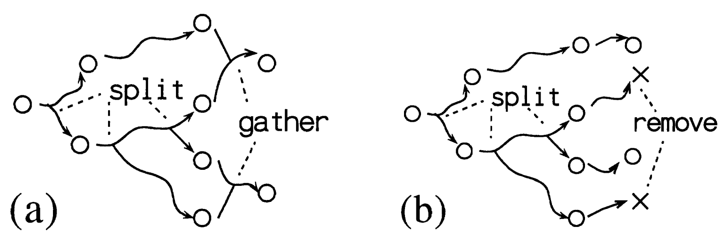


Figure 1: (a) Split-and-gather technique. The particles which are much "heavier" than the desired weight W_i are split, while much "lighter" particles are gathered into fewer particles. (b) Split-and-remove technique. Instead of gathering, "lighter" particles are randomly removed as illustrated in Fig. 2.

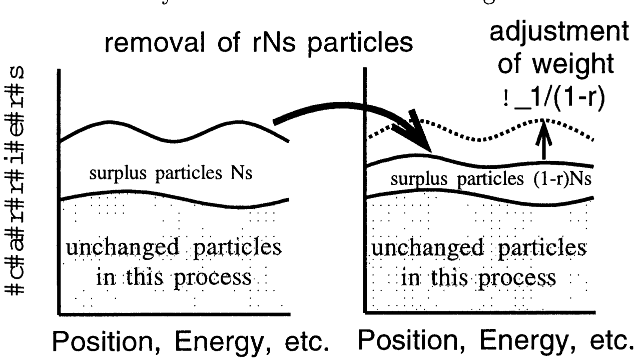


Figure 2: The removing process. This process is applied only to the surplus MC particles. They are randomly removed with probability r, and then the weights of the remaining surplus particles are scaled by a factor of $1/(1-r)$. Since rN_S particles are removed without changing the distribution of the surplus particles, the distribution of the total particles is obviously unchanged.

	SR	SG	MR
accuracy	○	×	○
calculation time	○	×	○
parallel processing	○	×	×
total charge conservation	△*	○	○

Table 1: Comparison of the three techniques. * The expected value of total charge is conserved.

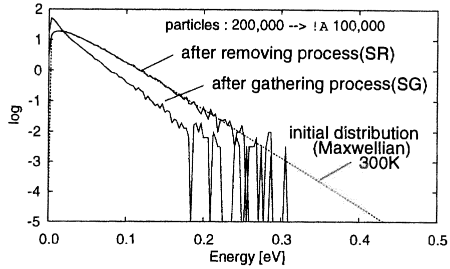


Figure 3: Influence of particle reduction on energy distribution. Only energy space is taken into account.

number of particles	
electron:	~ 220,000 (adjusted weight)
hole:	~ 110,000 (constant weight)
number of regions	
real space:	4 (denoted as #1, #2, etc.)
energy space:	6 (in #1), 11(in #2), 81(in #3), 11(in #4)
distribution of particles	
real space: $N'_{\#1} : N'_{\#2} : N'_{\#3} : N'_{\#4}$ =1:1:10:1 ($N'_{\#i}$ =number of particles in $\#i$)	
energy space: $N'_{\#i}$ divided by number of energy space division for $\#i$	
removing probability	$r = 0.05$
time step width	$\Delta t = 1 fsec$

Table 2: Simulation parameters. The SR is applied only to electrons.

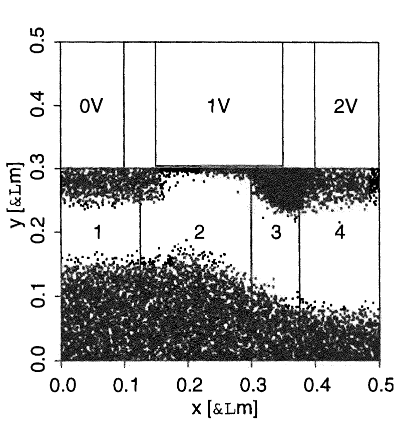
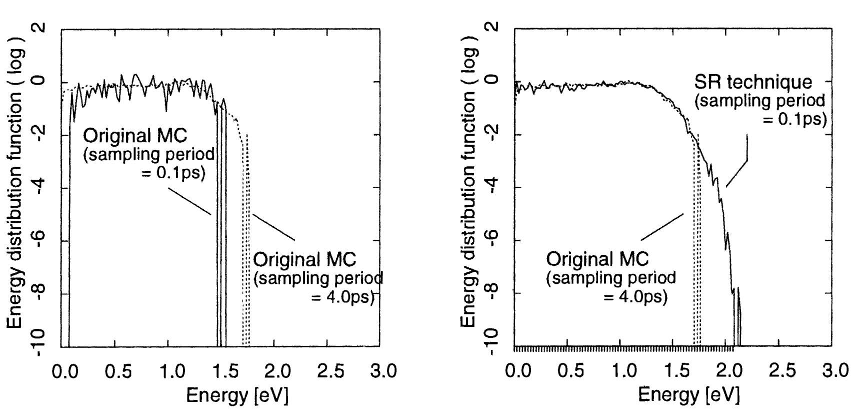


Figure 4: MC particles' distribution.



(a) Original MC (b) SR technique

Figure 5: Energy distribution function (a) without and (b) with the SR technique, sampled for a period of 0.1ps. The dashed line is reference data without the SR, sampled for 4.0ps. Before sampling, both required about 1.0ps to reach steady state.