Efficient and Accurate Full Band Semi-Classical Monte-Carlo Transport Simulation Using Smearing Method and Marching Tetrahedra Algorithm

Donghyeok Lee
Deptartment of Materials Science and Engineering
Yonsei University
Seoul 03722, South Korea
dju94@yonsei.ac.kr

Seong Woo Kang
Deptartment of System Semiconductor Engineering
Yonsei University
Seoul 03722, South Korea
dju94@yonsei.ac.kr

Leonard F. Register
Microelectronics Research Center
The University of Texas at Austin
Austin, TX 78758, U.S.A
register@austin.utexas.edu

Sanjay K. Banerjee
Microelectronics Research Center
The University of Texas at Austin
Austin, TX 78758, U.S.A
banerjee@ece.utexas.edu

Jiwon Chang
Deptartment of System Semiconductor Engineering
Yonsei University
Seoul 03722, South Korea
jiwonchang@yonsei.ac.kr

***Abstract*— We propose an enhanced Cellular Monte-Carlo (CMC) algorithm that merges the computational efficiency of traditional CMC with the high accuracy of conventional full-band Semi-Classical Monte-Carlo simulations. The algorithm incorporates the smearing method and Marching Tetrahedra algorithm to significantly reduce discretization errors in transition rates and ensure rigorous energy and momentum conservation. Electron drift velocity simulations in silicon demonstrate that our method closely aligns with both conventional Monte-Carlo results and experimental data, highlighting its suitability for efficient large-scale semiconductor transport simulations.**

Keywords*— *Cellular Monte-Carlo, Marching Tetrahedra, Smearing method, Carrier transport

I. Introduction

The Semi-Classical Monte-Carlo (SCMC) method remains a powerful computational approach for simulating carrier transport in semiconductors, particularly under non-equilibrium and high-field conditions where traditional drift-diffusion models are inadequate [1, 2]. The conventional SCMC (conventional MC) explicitly calculates scattering events dynamically, offering high physical accuracy but requiring substantial computational resources, thus limiting the practical applicability. The Cellular MC (CMC) method [3] improves computational efficiency by precomputing transition rates between discretized cells in the Brillouin zone but compromises accuracy due to discretization-induced errors in energy and momentum conservation. To address this issue, we present an enhanced CMC algorithm incorporating Marching Tetrahedra (MT) algorithm [4] and the smearing method [5], maintaining computational efficiency while significantly improving accuracy.

II. Proposed Algorithm

In the SCMC framework, carrier transport is simulated by following the motion of individual particles under external and internal forces. The particles drift under the influence of the fields, and upon a scattering event, their states are updated according to the transition rates. The transition rate $\Gamma^{\eta}_{nn'}(\boldsymbol{k},\boldsymbol{k}')$ is calculated via Fermi's golden rule as follows:

$$\Gamma^{\eta}_{nn'}(\boldsymbol{k},\boldsymbol{k}') = \frac{2\pi}{\hbar}\left|g^{\eta}_{nn'}(\boldsymbol{k},\boldsymbol{k}')\right|^2 \left(N_{\boldsymbol{q}} + \frac{1}{2} \mp \frac{1}{2}\right) \times \delta\left[E_n(\boldsymbol{k}) \pm \hbar\omega^{\eta}_{\boldsymbol{q}} - E_{n'}(\boldsymbol{k}')\right] \quad (1)$$

where $g^{\eta}_{nn'}(\boldsymbol{k},\boldsymbol{k}')$ denotes the matrix elements, n/n' and $\boldsymbol{k}/\boldsymbol{k}'$ represent the band index and wavevector of the initial/final state, respectively, η is the phonon mode index, $N_{\boldsymbol{q}}$ is the phonon occupation number, and E and ω correspond to electron and phonon energies, respectively. The upper and lower signs indicate phonon absorption and emission, respectively.

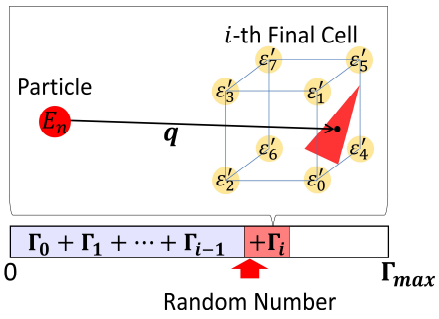

Fig. 1. Illustration of final-state selection in conventional MC. Transition rates are cumulatively summed over cells that satisfy energy and momentum conservation. The final cell is where the cumulative rate first exceeds the random number, and the final state is selected from the iso-energy surface within that cell.

In the conventional MC approach, the delta function in Fermi's golden rule is evaluated using the Gilat–Raubenheimer (GR) method [6]:

$$\delta\left[E_n(\boldsymbol{k}) \pm \hbar\omega^{\eta}_{\boldsymbol{q}} - E_{n'}(\boldsymbol{k}')\right] = \frac{1}{|\nabla\varepsilon'|} dS\Big|_{\varepsilon'=E_n(\boldsymbol{k})\pm\hbar\omega^{\eta}_{\boldsymbol{q}}} \quad (2)$$

where dS represents the iso-energy surface at the final state energy. As illustrated in Fig. 1, conventional MC computes transition rates one by one for all final cells that satisfy energy and momentum conservation, until the cumulative rate exceeds a random number drawn between 0 and Γ_{max}. Once

This research was supported by the MOTIE (Ministry of Trade, Industry & Energy) (1415185352) and KSRC (Korea Semiconductor Research Consortium) (20019450) support program for the development of the future semiconductor device. This work was also supported by the Technology Innovation Program ("RS-2023-00234828") funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea) (1415187391).

the final cell is selected, the final state is randomly chosen from the iso-energy surface dS within that cell. Since the initial state of the particle changes after each scattering event, the transition rates must be recalculated every time. Performing this procedure for every scattering event of every particle results in significant computational cost, which constitutes a major source of overhead in conventional MC simulations.

CMC achieves substantial speedup over conventional MC by precomputing transition rates on a fixed grid and organizing them into a hierarchical lookup table. This structure enables fast sampling: as illustrated in Fig. 2, once the initial cell is known, the algorithm selects final states by traversing the hierarchy based on transition rate weight. However, this efficiency comes at the cost of accuracy. Discretizing both the initial and final states introduces errors in energy and momentum conservation, and assigning the final state to the center of the selected cell exacerbates this issue.

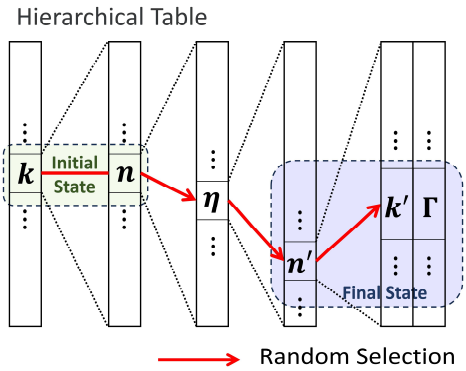

Fig. 2. Illustration of final-state selection in CMC. Random selection is performed through a hierarchical lookup table based on transition rate weights.

To correct the inaccuracies arising from the discretization of states, our enhanced CMC algorithm introduces two additional processes after selecting the final cell through the standard CMC procedure: a filtering step to enforce strict energy and momentum conservation, and the application of the Marching Tetrahedra (MT) algorithm to accurately determine the final state. As the first step, once the final cell is selected through the basic CMC procedure, we examine whether the chosen final cell contains the energy $E_n(\boldsymbol{k}) \pm \hbar\omega_{\boldsymbol{q}}^{\eta}$. If this condition is not satisfied, the scattering event is treated as self-scattering. This approach effectively yields the same result as using a pre-filtered transition rate table that includes only energy- and momentum-conserving transitions. If a final cell satisfying energy and momentum conservation is selected, the MT algorithm is applied to that cell. Fig. 3 illustrates the MT algorithm used to determine the final state. The final cell is divided into six tetrahedra, and each tetrahedron is examined to determine which case it corresponds to, as shown in Fig. 3(a). Although the MT algorithm defines 15 possible cases, symmetry reduces these to 8 unique configurations. Iso-surfaces are then identified within each tetrahedron. Fig. 3(b) illustrates the method for randomly selecting an iso-surface based on the area-weighted probabilities calculated from the iso-surface areas within each tetrahedron. Each of the six tetrahedra may contain an iso-surface, and after calculating the respective areas, an iso-surface is randomly chosen with a probability proportional to the area. Finally, Fig. 3(c) shows how the final state within the selected triangular surface is determined using random numbers R and S, which are multiplied by vectors $\boldsymbol{w}$ and $\boldsymbol{v}$, respectively, to specify the position within the triangle. If $R + S > 1$, the resulting state falls outside the triangle, so the values are adjusted to $1-R$ and $1-S$, respectively, ensuring that the final $\boldsymbol{k}$-point remains within the triangle. Furthermore, to enhance energy conservation and improve the accuracy of final-state momentum and energy calculations, each cell in the original 30×30×30 mesh was further subdivided into 3×3×3 sub-cells, resulting in an effective resolution of 90×90×90 sub-cells. The procedure described in Fig. 3 is then performed individually within each sub-cell, thus significantly increasing simulation precision.

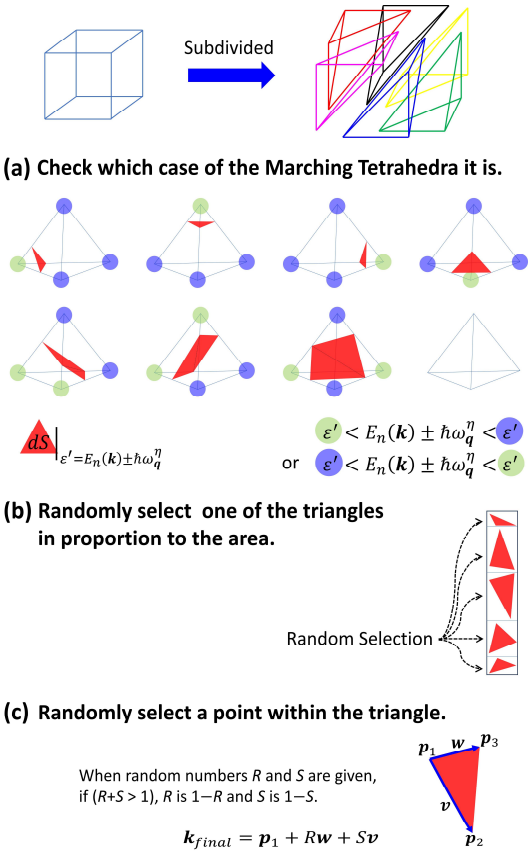

Fig. 3. Illustration of the MT method used to determine the final state. (a) Representative configurations of the MT cases used to identify iso-surfaces within each of the six tetrahedra in a final cell. (b) Area-weighted random selection of an iso-surface among candidate tetrahedra. (c) Selection of the final particle position within the selected triangle.

While the MT approach significantly improves the accuracy of energy and momentum conservation in the final state, it does not correct the erroneous transition rates caused by the discretized transition rate table. Fig. 4 shows the electron distributions in silicon under an electric field of 0.1 kV/cm along [100] directions. Fig. 4(a) is the electron distribution obtained from a conventional MC simulation performed on a 30×30×30 mesh in the Brillouin zone of silicon. Since conventional MC calculates transition rates explicitly based on the energy and momentum of each particle, the resulting electron distribution is naturally smooth and physically accurate. In contrast, Fig. 4(b) exhibits a markedly different behavior. It shows the electron distribution obtained by applying the MT method after the CMC procedure, using transition rates precomputed on a 30×30×30 mesh via the GR

method. As seen in the Fig. 4(b), the distribution is artificially sharp. This is because the transition rates calculated strictly at the cell center by the GR method can deviate significantly from the true transition rates corresponding to the actual particle states.

To mitigate these distribution errors, we applied the smearing method, which approximates the delta function in Fermi's golden rule with a normalized Gaussian:

$$\delta\left[E_n(\boldsymbol{k}) \pm \hbar\omega_{\boldsymbol{q}}^{\eta} - E_{n'}(\boldsymbol{k}')\right]$$

$$= \frac{1}{\sigma\sqrt{\pi}} \exp\left[-\left(\frac{E_n(\boldsymbol{k}) \pm \hbar\omega_{\boldsymbol{q}}^{\eta} - E_{n'}(\boldsymbol{k}')}{\sigma}\right)^2\right] \quad (2)$$

where σ is the smearing width. This approach smooths the final-state distribution and reduces errors caused by the discretized transition rates. Fig. 4(c), which uses smearing-based transition rates, shows a distribution that closely matches the result from conventional MC. This similarity arises because the smearing method produces smoothly varying transition rates, ensuring that the rate evaluated at the cell center better approximates the rate at arbitrary points within the cell.

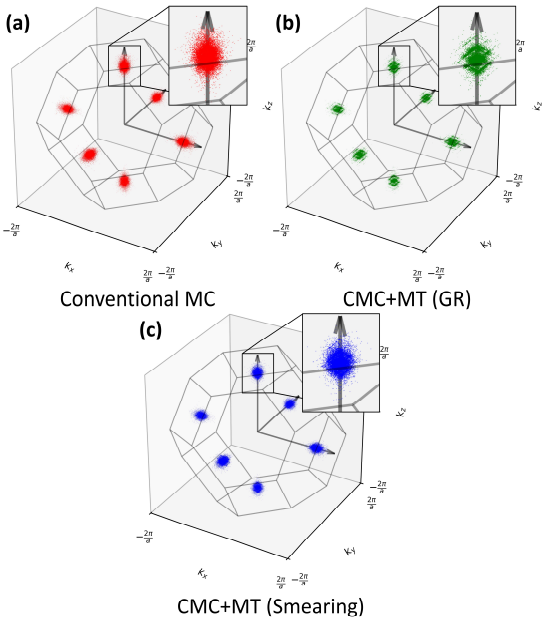

Fig. 4. Comparison of three-dimensional electron distributions in the Brillouin zone of silicon under a 0.1 kV/cm electric field along the [100] direction: (a) conventional MC, (b) CMC with MT using GR-based transition rate table, and (c) CMC with MT using smearing-based transition rate table.

III. SIMULATION AND RESULTS

A. First-principles Based Preprocessing

The parameters required to calculate the scattering rates—matrix elements g, phonon energy $\omega_{\boldsymbol{q}}$, and electron energy E—were obtained using Quantum ESPRESSO [7, 8] and the EPW package [9, 10]. Density Functional Theory (DFT) calculations using Quantum ESPRESSO were performed with the pseudopotentials from the "SSSP PBE Precision v1.3.0" library [11] and the input parameters summarized in Table 1. Subsequently, matrix elements were interpolated from coarse $\boldsymbol{k}$- and $\boldsymbol{q}$-point grids (16×16×16 and 8×8×8, respectively) to finer 30×30×30 mesh using Wannier interpolation in EPW. Both the electronic band structure and phonon dispersion were similarly interpolated using EPW. These quantities were further interpolated on even finer grid, such as 90×90×90 $\boldsymbol{k}$-point mesh. All tables were generated on a Γ-centered grid in the crystal coordinate system defined by reciprocal lattice vectors. To accelerate the calculation of group velocity, we also generated an additional band structure table on a 100×100×100 $\boldsymbol{k}$-point grid in Cartesian coordinates (k_x, k_y, k_z).

TABLE I. INPUT PARAMETERS USED IN QUANTUM ESPRESSO

Input parameter	Value
Wave function cutoff	30 Ry
Charge density cutoff	240 Ry
Smearing method	Marzari-Vanderbilt
Gaussian spreading	0.05 Ry
$\boldsymbol{k}$-point grid	12×12×12
$\boldsymbol{q}$-point grid	6×6×6

Due to the large size of the matrix element data, it is essential to remove unused entries during the table generation stage. We prefiltered matrix elements that do not conserve energy to reduce the memory footprint of the simulation [12]. Additionally, the tables were stored in HDF5 format [13], which improves loading speed and reduces storage requirements.

B. Simulation Setup

We assessed the accuracy and computational efficiency of our method by calculating the electron drift velocity in silicon under electric fields applied along the [100] direction, using conventional full-band SCMC and CMC simulations with either the GR or smearing method. Each simulation employed 100,000 particles with a time step of 10 femtoseconds, and unless otherwise specified, a 30×30×30 $\boldsymbol{k}$-point mesh is used. The drift velocity at each electric field was obtained as the 250-step moving average, once the relative error between the 500-step and 250-step averages fell below 0.001.

C. Results

Fig. 5 presents the simulated electron drift velocities in silicon under electric fields ranging from 0.1 kV/cm to 100 kV/cm along the [100] crystallographic direction, compared with experimental data [14]. Fig. 5(a) shows results from simulations performed on a 30×30×30 $\boldsymbol{k}$-point mesh using conventional MC, GR-based CMC, and smearing-based CMC. Significant differences between conventional MC and both CMC methods are observed. Compared to experimental data, conventional MC also exhibits deviations. To improve the accuracy in conventional MC, simulations were performed again with a denser 90×90×90 $\boldsymbol{k}$-point mesh, resulting in much closer agreement with the experimental data, as shown in Fig. 5(b). As discussed previously, in the case of CMC+MT simulations, the final cell was first determined using the transition rate tables on 30×30×30 $\boldsymbol{k}$-point mesh. Each final cell was then subdivided into 3×3×3 sub-cells, resulting in an effective resolution equivalent to a 90×90×90 $\boldsymbol{k}$-point mesh. The smearing-based CMC+MT approach demonstrates close agreement with the conventional MC results. In contrast, the GR-based CMC+MT still shows noticeable error, primarily due to the transition rates strictly calculated at the cell centers. The smearing method, by generating smoothly varying

transition rates across cells, allows the resulting drift velocities to better match those of conventional MC.

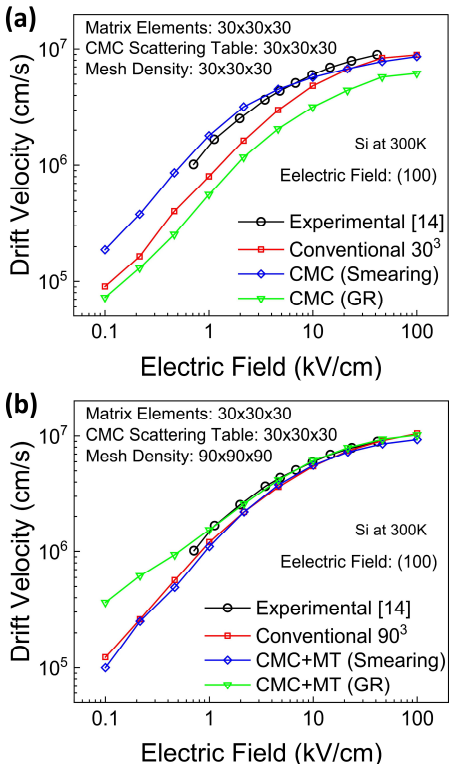

Fig. 5. Comparison of simulated electron drift velocities in silicon under electric fields along the [100] crystallographic direction. (a) Results from simulations on a 30×30×30 mesh using conventional MC, GR-based CMC, and smearing-based CMC, shown with experimental data [14]. (b) Results from simulations on a 90×90×90 mesh using conventional MC, and CMC combined with the MT method, also shown with experimental data [14].

Table 2 summarizes the average computational time per step for each simulation method. For conventional MC, increasing the mesh density from 30×30×30 to 90×90×90 dramatically increases computational time whereas applying MT to the CMC method introduces negligible additional computational cost. Notably, under low electric field conditions, the computational time for conventional MC increases by more than an order of magnitude. This efficiency arises from the straightforward algorithmic structure of the MT method. The difference in computation time between the GR and smearing methods primarily arises from the time required to access the scattering table. In the smearing method, transition rates are computed for all possible final cells, even if the values are negligible or close to zero. The size of the scattering table depends on how aggressively such insignificant transition rates are pruned. In our implementation, transition rates smaller than 1/100 of the maximum rate in each initial cell were excluded when generating the scattering table. As a result, the scattering table for the smearing method was smaller than that of the GR method, leading to shorter lookup times.

IV. CONCLUSION

We developed an enhanced CMC algorithm employing smearing and MT methods, effectively addressing CMC limitations in accuracy and energy-momentum conservation. Simulation results confirm that our method closely matches conventional full-band MC simulations and experimental data, while providing significant computational advantages. This enhanced CMC approach offers a promising computational framework for efficient and accurate large-scale semiconductor transport simulations.

TABLE II. AVERAGE ELAPSED TIME (MS) PER STEP (10 FS)

δ -Function Integration	MC Method	k-Point Mesh Density	Electric field (kV/cm)		
			1	10	100
GR	Conventional	30^3	1192.4	1466.2	10281.7
		90^3	15787.1	16704.6	20278.0
	CMC	30^3	263.0	261.0	254.0
	CMC+MT	90^3	259.3	258.0	264.0
Smearing	CMC	30^3	19.5	19.4	25.8
	CMC+MT	90^3	20.0	20.0	33.0

REFERENCES

[1] X. Wang, V. Chandramouli, C. M. Maziar, and A. F. Tasch Jr., "Simulation program suitable for hot carrier studies: An efficient multiband Monte-Carlo model using both full and analytic band structure description for silicon," J Appl Phys, vol. 73, no. 7, pp. 3339–3347, Apr. 1993.

[2] M. V Fischetti and S. E. Laux, "Monte-Carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects," Phys Rev B, vol. 38, no. 14, pp. 9721–9745, Nov. 1988.

[3] M. Saraniti and S. M. Goodnick, "Hybrid fullband cellular automaton/Monte-Carlo approach for fast simulation of charge transport in semiconductors," IEEE Trans Electron Devices, vol. 47, no. 10, pp. 1909–1916, 2000.

[4] A. K. Akio DOI, "An Efficient Method of Triangulating Equi-Valued Surfaces by Using Tetrahedral Cells," IEICE TRANSACTIONS on Information, vol. E74-D, no. 1, pp. 214–224, Jan. 1991.

[5] M. Methfessel and A. T. Paxton, "High-precision sampling for Brillouin-zone integration in metals," Phys Rev B, vol. 40, no. 6, pp. 3616–3621, Jun. 1989.

[6] G. Gilat and L. J. Raubenheimer, "Accurate Numerical Method for Calculating Frequency-Distribution Functions in Solids," Physical Review, vol. 144, no. 2, pp. 390–395, Jun. 1966, doi: 10.1103/PhysRev.144.390.

[7] P. Giannozzi et al., "QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials," Journal of Physics: Condensed Matter, vol. 21, no. 39, p. 395502 (19pp), 2009.

[8] P. Giannozzi et al., "Advanced capabilities for materials modelling with QUANTUM ESPRESSO," Journal of Physics: Condensed Matter, vol. 29, no. 46, p. 465901, 2017.

[9] S. Poncé, E. R. Margine, C. Verdi, and F. Giustino, "EPW: Electron–phonon coupling, transport and superconducting properties using maximally localized Wannier functions," Comput Phys Commun, vol. 209, pp. 116–133, 2016.

[10] H. Lee et al., "Electron–phonon physics from first principles using the EPW code," NPJ Comput Mater, vol. 9, p. 156, 2023.

[11] G. Prandini, A. Marrazzo, I. E. Castelli, N. Mounet, and N. Marzari, "Precision and efficiency in solid-state pseudopotential calculations," NPJ Comput Mater, vol. 4, no. 1, p. 72, 2018.

[12] A. Kumar and U. Singisetti, "Full-band Monte-Carlo simulation of two-dimensional electron gas in (AlxGa1−x)2O3/Ga2O3 heterostructures," J Appl Phys, vol. 132, no. 20, p. 205701, Jun. 2022.

[13] The HDF Group, "Hierarchical Data Format, version 5." [Online]. Available: https://github.com/HDFGroup/hdf5

[14] C. Canali, C. Jacoboni, F. Nava, G. Ottaviani, and A. Alberigi-Quaranta, "Electron drift velocity in silicon," Phys. Rev. B, vol. 12, no. 6, pp. 2265–2284, Sep. 1975.