A full–band cellular automaton for charge transport simulation in semiconductors.

M. Saraniti, and S. M. Goodnick[1]

Department of Electrical and Computer Engineering
Illinois Institute of Technology, 3301 S. Dearborn, Chicago, IL 60616-3793
Phone: (312) 567-8813, Fax: (312) 567-8976
E-mail: saraniti@ece.iit.edu

[1]Department of Electrical Engineering
Arizona State University, Tempe (AZ) 85287-5706
Phone: (602) 965-6410, Fax: (602) 965 3837
E-mail: stephen.goodnick@asu.edu

1. Introduction

The Ensemble Monte Carlo (EMC) method [1] is presently considered a mature technique for semiconductor device simulation based on the paradigm of semi-classical charge transport. However, the relatively heavy computational burden of the EMC approach has limited its use to mainly academic environments, particularly when non analytical, realistic band structures are used [2]. In order to reduce the computational demands of EMC simulation, the cellular automaton (CA) approach was developed in the context of semiconductor device simulation [3]. In the CA approach, both k-space and real space are discretized, which simplifies the description of scattering and the particle motion in real and momentum space. This technique was successfully demonstrated using an analytical, non-parabolic band model [3], where significant speed-up was observed compared to more traditional EMC methods.

The aim of the present work is to demonstrate a new simulation approach, based on the CA method, that includes a full-band representation of the electronic structure and of the phonon spectra, while maintaining the short simulation times typical of the previous non-parabolic CA. Simulation results are shown for charge transport in bulk Si to demonstrate the equivalence of this new approach with full-band EMC simulation results, as well as its efficiency.

2. Physical model

The band structure used in the full-band CA is computed using the empirical pseudopotential (EPM) method [4], while phonon spectra are obtained via the valence shell method [5]. The non-polar transition rate from a region centered in the point $\mathbf{k}$ in band ν to a region $\Omega_{\mathbf{k}'}$ centered around the point $\mathbf{k}'$ in band ν' is then approximated by

$$P_{\nu,\nu'}(\mathbf{k},\Omega_{\mathbf{k}'}) = \frac{1}{\tau(\mathbf{k},\nu;\Omega_{\mathbf{k}'},\nu')} = \frac{\pi}{\rho\omega_{\eta\mathbf{q}}}|\Delta_{\eta,\nu'}(\mathbf{q})|^2 \times$$

$$\times|\mathcal{I}(\nu,\nu';\mathbf{k},\mathbf{k}')|^2 D_{\nu'}(E',\Omega_{\mathbf{k}'})(n_{\eta\mathbf{q}} + \frac{1}{2} \pm \frac{1}{2}) \quad (1)$$

where ρ is the semiconductor density, $\omega_{\eta\mathbf{q}}$ is the frequency of a phonon of type and polarization η and wave vector $\mathbf{q} = \mathbf{k}' - \mathbf{k}$, $\Delta_{\eta,\nu'}(\mathbf{q})$ is the non-polar matrix element as defined (and approximated) in [2], $\mathcal{I}$ is the overlap integral, $D_{\nu'}(E',\Omega_{\mathbf{k}'})$ is the density of states in $\Omega_{\mathbf{k}'}$ at energy $E' = E(\mathbf{k}) \pm (\hbar\omega_{\eta\mathbf{q}})$ in band ν', and, finally, $n_{\eta\mathbf{q}}$ is the phonon occupation number at the lattice temperature. The density of states is evaluated using the well known linear analytical method of integration of the Brillouin zone introduced by Gilat and Raubenheimer [6] in its version for orthorhombic cells [7].

The transition rate in Eq. 1 is computed and tabulated for each cell in the grid of each energy band included in the model. In an analogous way, the impact ionization rate is computed and tabulated for each portion of the discretized first Brillouin Zone (BZ). The model used to implement impact ionization in Si is a simple isotropic, multi-threshold [8] model:

$$\frac{1}{\tau_{II}(E)} = \sum_{i=1}^{3} \theta(E - E_{th}^{(i)})P^{(i)}\left(\frac{E - E_{th}^{(i)}}{E_{th}^{(i)}}\right)^2 \quad (2)$$

were E is the electron kinetic energy, $E_{th}^{(i)} = 1.2, 1.8$, and 3.45 eV, $P^{(i)} = 6.25 \times 10^{10}, 3.0 \times 10^{12}$, and $6.8 \times 10^{14} \, s^{-1}$, for $i = 1, 2$, and 3, respectively, and θ is the step function.

It should be stressed that while the simplified scattering rates used here are essentially isotropic (apart from the overlap intergral in (1)), the full-band approach presented here is completely general in that fully anisotropic

rates in momentum space may be used without any increase in the computational burden.

3. Implementation
Choice of the final states

The algorithm used to build the transition (or scattering) table is directly derived from that proposed by Kunikiyo *et al.* [9], and is organized in two parts: *i*) find all the final states for a given initial state, and *ii*) compute the probability for each of the final states.

The first part of the task is accomplished by finding all cells which span an energy in the interval $[\epsilon_0 \pm \epsilon_{\eta,min}, \epsilon_0 \pm \epsilon_{\eta,max}]$ in the case of absorption (+ sign) or emission (− sign) of a phonon of type and polarization η; $\epsilon_{\eta,max}$, where $\epsilon_{\eta,min}$ are the maximum and minimum values of the energy of the phonon in BZ, and, ϵ_0 is the energy of the carrier. The resulting set of cells represents all the final states that satisfy energy conservation.

Once all possible final state candidates are known, the transition probability is computed for each of them from Eq. 1, taking the wave vector of the phonon as *the vector connecting the centers of the initial and the final cell.*

This process is repeated for all the phonon modes considered, for both absorption and the emission. The final states are sorted and recurrent states are grouped together, adding their probabilities. In other words, if two different scattering mechanisms result in the same final cell for a given initial state, the two final state scattering probabilities are added, and only the sum stored in the table. This grouping procedure allows the size of the transition table to be dramatically reduced. However, as a consequence, information on the type of mechanism (i.e. the type of phonon mode and, consequently, its energy) which generated a transition is *lost*. This procedure is different than the traditional Monte Carlo technique where this information is retained, and precludes any "post-scattering" correction of the carrier wave vector, which can result in an unacceptable error in energy conservation, as clearly pointed out by Fischetti and Laux in [2]. The way in which we address this potential problem is to use an inhomogeneously spaced grid, as discussed in the next section.

The computational advantage of this method is that once the transition table is built, no information concerning the carrier energy is required to simulate scattering events. In other words, scattering processes are simulated entirely in a discrete (cellular) momentum space, without the requirement of inverting the energy-momentum dispersion relation.

Grid in momenta space

The basic idea of the full band CA is to tabulate the total probability of scattering from each initial state $\mathbf{k}$ to any possible final $\mathbf{k'}$. It is clear that the grid used to discretize the first BZ plays a crucial role in keeping the table dimensions manageable while at the same time accurately representing the energy dispersion in momen-

tum space.

The use of a homogeneous rectangular grid with typical spacing $l = 0.05(2\pi/a)$ (where a represents the lattice constant of the material) requires a computationally costly correction of the carrier wave vector after each scattering event in Si using a traditional Monte Carlo approach. This correction is even more difficult for the conduction band minima of GaAs [2] where the effective mass is much lighter.

These problems have been addressed by using an inhomogeneous rectangular grid to represent momentum space in the transition table used in the present CA algorithm. The algorithms for computing the density of states and to track the particle motion in momentum space have been accordingly modified, and a simple but powerful user interface has been developed to allow easy input of the grid data (as well as other material and simulation parameters). Of course, different grids have been defined for the different bands of each material. For example, the grid spacing in the regions around the minima of the conduction bands of Si is $l_{Si} = 0.02(2\pi/a)$ with a cell density one order of magnitude higher than the grid density used in [2].

Preliminary results clearly show that the use of an even finer grid (with spacing $l_{GaAs} = 0.012(2\pi/a)$, and a grid seventy times more dense than the one in [2]) around the Γ-point allows the low field behavior of GaAs to be modeled within a full-band representation of the dispersion relation.

A problem that we consider still open is to find an *automatic* method to generate an adequate grid. Currently, the grids are generated heuristically, accounting for the features of the 3D dispersion relation. This produces two negative consequences: 1) the energy discretization error is not constant over the BZ and 2) the process of generating the grid is long and computationally costly.

Tests have been done with a recursive *branch-on-need* bisection algorithm, driven by the energy spanned by a region of momentum space, and by its first derivative (to find the optimal direction of bisection). The algorithm, which recursively bisects a region of BZ if the interval of spanned energies is bigger than a given threshold, produced excellent grids form the point of view of energy conservation and symmetry. However, the local nature of the branch process produces grids which are completely unsuitable for tracking particle along their Newtonian trajectories. Several attempts were performed, but no algorithm was found to *efficiently* follow a particle from one cell to another, when the number of neighbors at each side of each cell is arbitrary.

Symmetry and structure of the transition table

It is clear that the critical parameter of this full-band CA approach is the dimension of the transition table. The irregular grid, as well as the grouping of the final states which appear more than one time for a given initial state, are due to the need to keep the transition table as small as possible. Another crucial reduction is

obtained by tabulating the final states (wherever they occur in the BZ) related to only those initial states which are contained in the irreducible wedge (IW) of the BZ. Initial states outside of the irreducible wedge are "rotated" inside the IW, processed, and the resulting final states are finally back-rotated. The transformation matrix from a cell outside the IW to the corresponding cell inside the wedge is obviously pre-computed, and requires a negligible amount of memory to be stored.

Finally, it has to be observed that even if the grid is inhomogeneously spaced, it is built in such a way that symmetry is conserved for its nodes. In other words, each grid-point outside the IW is rotated to a point which is still a grid-point. This fact also ensures isotropy of the energy discretization error.

Memory requirements

The grid used to obtain the results shown in this paper spans the momentum space from the bottom of the conduction band in Si to an energy of 5.1 eV. The first two bands are discretized on ca. 79000 and 49000 inhomogeneously spaced cells, respectively. Only cells in the first BZ are considered and stored. This energy range is spanned by the first 5 conduction bands, and, again, only regions of momentum space with energy within the mentioned range are considered. The resulting table is stored in less than 800 Mbyte of RAM.

Each final state is represented in memory by 8 bytes containing a floating point representation of the probability (4 Bytes), and the indices of the final state cell and band, as well as one one-bit flag used to discriminate states produced by impact ionization. Impact ionization processes are treated separately from all other mechnisms due to the multiplication effect which requires knowledge of the type of scattering occuring. Therefore, the final states probabilities due to this process are not combined with the other scattering mechanisms so that the multi-body final state can be properly accounted for.

4. Results

Figure 1 shows a comparison between results of published full-band EMC simulation [2] (lines), and the present full band CA (dots) for the energy and drift velocity as a function of the electric field in the (100) direction in Si at $300K$. Good agreement is found both in the energy and the drift velocity at all fields, particularly at low field where discretization effects are expected to be most pronounced.

Figure 2 illustrates the distribution function versus energy at various electric fields, again showing good agreement with published full band EMC results. The energy dependent scattering rate averaged over all momentum is shown as well.

5. Benchmarks

Measures of the execution time on a typical workstation (here a 500 MHz DEC Alpha processor) are shown in Table 1. The number of simulated carriers (electrons)
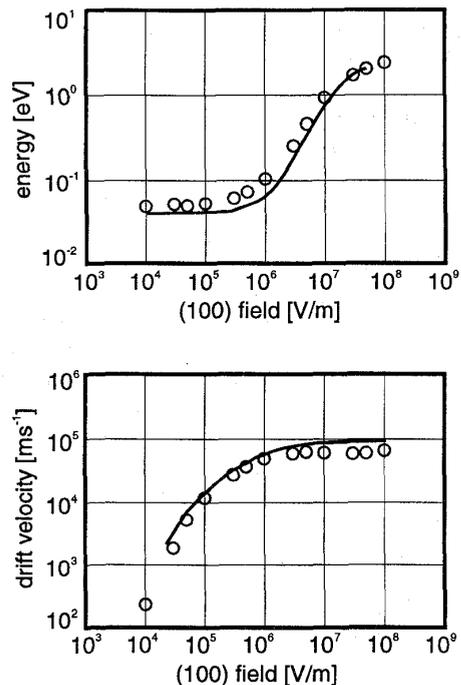
Figure 1: Energy-field (left) and Velocity-field (right) curves in Si at $300K$ as computed with full-band EMC [2] (lines), and with CA (circles).

is 10^4, and the electric field applied on the (100) direction is 10^7 V/m. Two sets of times are shown, for two different ways of computing the drift velocity. Since the present k-space results do not require the carrier velocity to simulate the dynamics, the drift velocity can be pre-tabulated and taken from the nearest grid point rather than being interpolated. While this is an acceptable procedure for k-space simulation, a more accurate determination of the velocity has to be computed for device simulation when the CA is self-consistently coupled with a Poisson solver. Table 1 shows the CPU time per time step (1.2 fs) when the velocity is simply copied from the closest grid point (*from grid* row in Table 1) compared with a more exact computation by a linear interpolator (*interpolated* row). The results – shown in the *effective* row, given by the

Table 1: CPU time, in seconds, per iteration, for a workstation equipped with a 500 MHz DEC Alpha processor.

	v_d interpolated	v_d from grid
free flight	0.533	0.191
scattering	0.051	0.050
averages	0.142	0.141
total	0.726	0.382
effective	**0.584**	**0.241**
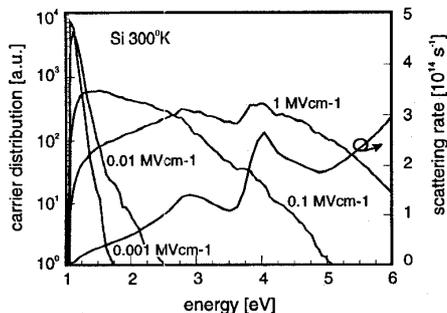
Figure 2: Electron distribution function in Si at $300K$ for various electric field strengths. The momentum averaged scattering rate is also shown as a function of energy on the right scale.

total time minus the time spent with averages – clearly show that the scattering part of the carrier dynamics requires a small fraction of the total CPU time, while this part is typically the computationally more demanding component of a traditional full-band EMC. The main improvement in execution time in the CA is due to the fact that the final state after scattering is selected in a single operation due to the pre-tabulated scattering table, with no costly inversion of the dispersion relation required after the scattering.

6. Final remarks and future work

In the present paper we have shown a new approach for particle-based simulation of charge transport in semiconductor materials. Results were shown which are in excellent agreement with previous full band Monte Carl results from the literature. Benchmarks are also shown that demonstrate the high computational efficiency of the new method.

Improvements of the present algorithm are aimed to two main goals: to extend the materials modeled by the simulator, and to include the real space representation of the carrier dynamics. The latter project involves the inclusion of a fast Poisson solver [10] and an efficient carrier tracking algorithm in real space, such that complex boundary conditions in real devices can be accounted for.

Finally, the authors would like to comment on the appropriateness of the name used for this method. In computational electronics, the denomination "cellular automaton" was initially chosen by P. Vogl at the Technical University of Munich (in [3] and previous papers) to describe their method to simulate charge transport in semiconductor devices. That choice was justified by the many similarities of Vogl's approach with the definition of CA offered by Wolfram [11][1] and with the method of Frish et al. [12] for solution of hydrodynamics problems.

---

[1]curiously, the fundamental papers written by Carlo Jacoboni et. al. on Monte Carlo [1], and the one written by Stephen Wolfram on cellular automata [11] were adjacent in the same issue of the journal where they appeared.

The method presented in this paper is an evolution of the work initiated by the Munich group. However, many of the "cellular automaton" features of the algorithm have been modified (for example the free particle dynamics) such that the resulting code is more properly regarded as a hybrid CA-Monte Carlo algorithm. However, as frequently happens the denomination of the method survived during its evolution, and, for the moment, we have chosen not to change it.

7. Acknowledgments

References

[1] C. Jacoboni and L. Reggiani, Reviews of Modern Physics 55, 645 (1983).

[2] M. V. Fischetti and S. E. Laux, Physical Review B 38, 9721 (1988).

[3] K. Kometer, G. Zandler, and P. Vogl, Physical Review B 46, 1382 (1992).

[4] J. R. Chelikowsky and M. L. Cohen, Physical Review B 14, 556 (1976).

[5] K. Kunc and O. H. Nielsen, Computer Physics Communications 17, 413 (1979).

[6] G. Gilat and L. Raubenheimer, Physical Review 144, 390 (1966).

[7] L. Raubenheimer and G.Gilat, Physical Review 157, 586 (1967).

[8] E. Cartier, M. Fischetti, E. Eklund, and F. McFeely, Applied Physics Letters 62, 3339 (1993).

[9] T. Kunikiyo, M. Takenaka, Y.Kamakura, M.Yamaji, H. Mizuno, M. Morifuji, K.Taniguchi, and C. Hamaguchi, Journal of Applied Physics 75, 297 (1994).

[10] M. Saraniti, A. Rein, G. Zandler, P. Vogl, and P. Lugli, IEEE Transaction on Computer–aided Design of Integrated Circuits and Systems 15, 141 (1996).

[11] S. Wolfram, Reviews of Modern Physics 55, 601 (1983).

[12] U. Frish, B. Hasslacher, and Y. Pomeau, Physical Review Letters 56, 1505 (1986).