Particle-based Full-band Approach for Fast Simulation of Charge Transport in Si, GaAs, and InP

MARCO SARANITI*, YIBING HUB and STEPHEN M. GOODNICKc

a Department of Electrical and Computer Engineering, Illinois Institute of Technology, 3301 South Dearborn, Chicago, IL 60616-3793, USA; b Agere Systems, Allentown, PA, USA; c Department of Electrical Engineering of Arizona State University, Tempe, AZ, USA

(Received 1 May 2001; Revised 1 April 2002)

We discuss the application of the fullband cellular automaton (CA) method for the simulation of charge transport in several semiconductors. Basing the selection of the state after scattering on simple look-up tables, the approach is physically equivalent to the full band Monte Carlo (MC) approach but is much faster. Furthermore, the structure of the pre-tabulated transition probabilities naturally allows for an extension of the model to fully anisotropic scattering without additional computational burden. Simulation results of transport of electrons and holes in several materials are discussed, with particular emphasis on the transient response of photo-generated carriers in InP and GaAs. Finally, a discussion on parallel algorithms is presented, for the implementation of the code on workstation clusters.

Keywords: Charge transport modeling; Monte Carlo; Cellular automaton; Semiconductors; Si; GaAs

INTRODUCTION

Particle-based methods [1] are well established for the simulation of charge transport in semiconductors. Their relative popularity is based on both physical and numeric considerations including: (i) the possibility of implementing sophisticated models without affecting algorithmic stability, (ii) the capability of microscopic modeling, and (iii) the numerical robustness. However, the price paid for accuracy and straightforward implementation is an inherent slowness of particle-based algorithms.

The computational burden continues to limit its use in commercial device modeling, which has not been alleviated by the evolution of faster computers, particularly when sophisticated (and computationally demanding) physical models are implemented. A typical example of the model evolution has been the increased use of fullband representation of the electronic structure [2–4], where the simple analytic representation of the band structure [5] is replaced by a much more realistic (but computationally expensive) model based on, for example, the empirical pseudopotential method (EPM) [6].

New algorithmic approaches are clearly needed both to increase the modeling capabilities of these methods, and to reduce the simulation time and so making them suitable as design tools. The so-called fullband cellular automaton (CA) was recently introduced [7] to alleviate the computational burden of fullband particle-based simulation without imposing severe physical approximations. The CA model (which is perhaps more appropriately referred to as a discrete k-space Monte Carlo [MC] algorithm) is based on a nonuniform discretization of the first Brillouin Zone (BZ) [8] of the crystal lattice. The CA approach allows the tabulation of the transition probabilities between all initial and final states on the mesh, greatly simplifying the final state selection of the conventional MC algorithm [4,5]. The tremendous speedup obtained by the CA approach allows to simulate slower phenomena, and/or to implement more sophisticated physical models.

The aim of this paper is to discuss and validate the results obtained with the CA approach, and to apply it to the simulation of ultrafast phenomena occurring in GaAs and InP after photo-excitation. In the following, we briefly introduce the basic theory of particle-based approaches in second section. Simulation results are subsequently
discussed of charge transport in bulk Si, GaAs and InP. Comparison with published results are shown, and an analysis is done in third section of the accuracy of the CA approach compared to the MC. In fourth section, simulation of the transient response of carriers to an externally applied electric field is compared with recently published experiments. Fifth section is devoted to some critical aspects of particle-based computation, while the final sixth section discusses future development of the algorithm.

CARRIER DYNAMICS

A detailed discussion of semi-classical particle-based treatment of carrier dynamics in phase-space has been extensively done elsewhere [4]. The fullband CA approach used in the present work and its comparison to fullband MC has been discussed in Ref. [7], which is briefly described here.

Basically, within the fullband CA (or discretized k-space) method, the electronic states are calculated on a uniform grid within the irreducible wedge of the BZ using an appropriate band structure calculation. Here we use the EPM [6] as parameters for common semiconductor materials that are well-known. Typically 916 points are used in the irreducible wedge. A much finer nonuniform tensor-product grid is then defined onto which the electronic structure is interpolated for use in generating the transition table used in the CA. The use of a nonuniform grid allows for refinement of the mesh in critical regions of the BZ such as band minima, while minimizing the number of tabulated points. The density of states is obtained by the Gilat-Raubenheimer algorithm [9]. The calculated band structure and corresponding DOS are shown for InP in Fig. 1 using nonlocal EPM parameters reported by Pötz and Vogl [10].

The CA transition table here is computed using the scattering rates detailed by Fischetti and Laux [4], adapted for the discrete k-space method [7]. In particular, the nonpolar transition rate from a region of band \( \nu \) centered in the point \( k \) to a region \( \Omega_{k'} \), in band \( \nu' \) centered around the point \( k' \) is approximated by

\[
P_{\nu,\nu'}(k, k') = \frac{\pi}{\rho \omega_{0q}} |\Delta_{\nu,\nu'}(q)|^2 \omega_{0q} \times |I(\nu, \nu'; k, k')|^2 D_{\nu}(E', \Omega_{k'}) \times \left(n_{0q} + \frac{1}{2} \pm \frac{1}{2}\right),
\]  

where \( \rho \) is the semiconductor density, \( q \) the scattered wave vector, \( \Delta_{\nu,\nu'}(q) \) the nonpolar matrix element as defined (and approximated) in Ref. [4], \( I \) is the overlap integral between Bloch states, \( D_{\nu}(E', \Omega_{k'}) \) is the density of states (DOS) in \( \Omega_{k'} \) at energy \( E' = E(k) \pm (\hbar \omega_{0q}) \) in band \( \nu' \), and, finally, \( n_{0q} \) is the phonon occupation number at the lattice temperature.

In polar semiconductors, the rate for the polar interaction of electrons with phonons is given by the Fröhlich expression

\[
P_{\text{pol},\nu,\nu'}(k, \Omega_{k'}) = \frac{\pi \epsilon_0^2 \omega_{0q}}{2q^2} \left[ \frac{1}{\epsilon_{\infty} - \epsilon_0} \right] \times |I(\nu, \nu'; k, k')|^2 D_{\nu}(E', \Omega_{k'}) \times \left(n_{0q} + \frac{1}{2} \right). \tag{2}
\]

where \( \epsilon_{\infty} \) and \( \epsilon_0 \) are the optical and static dielectric functions, respectively.

The phonon dispersion, \( \omega_{0q} \), is calculated using an empirical shell model [12], where \( \eta \) designates the possible phonon modes (i.e. longitudinal acoustic, etc.), and \( \omega \) is the frequency of vibration. The method has been implemented to allow the input parameters to be formatted for the so-called “14-parameter shell model” [13] as well as for the “valence shell model” [11]. The phonon dispersion of InP is shown in Fig. 2 using parameters reported by Borchers and Kunc [14].

The phonon transition rates as computed from the Eqs. (1) and (2), can be integrated [7] over the final states \( \Omega_{k'} \) and, successively, over constant energy surfaces, to obtain the familiar representation of the scattering rates.
as a function of the carrier energy shown in Fig. 3, where both the electron and hole scattering rates for InP are shown.

In an analogous way, the impact ionization rate is computed for each portion of the discretized BZ. The model used here to implement impact ionization is a simple isotropic, double-threshold [15] model

\[
\frac{1}{\eta_{II}(E)} = \sum_{i=1}^{2} a_i (E - E_i)^{r_i}
\]

where \(E\) represents the carriers kinetic energy, \(E_i\) are threshold energies, and the exponents \(r_i\) and the prefactors \(a_i\) are fitting parameters indicated for both electrons and holes in Ref. [15]. In the present work, the implemented impact-ionization model does not allow for generation of secondary carriers, and thus affects the simulation results at high fields (see “Comparison of Fullband EMC and CA Simulations” Section).

The CA algorithm tabulates the total transition rate for every initial state to every final state in the entire BZ, hence full anisotropic scattering rates (as calculated for example using the rigid ion model [16]) may be included with no additional computational burden once they are tabulated. At every time step in the simulation, a random number is used to determine if scattering occurs or not based on the total scattering rate for a particle in a particular momentum state. The selection of the final state in the CA algorithm is reduced to the generation of a single random number, which allows to select the final state from among the pre-tabulated final states on the CA mesh. In contrast, for a traditional MC approach, a second random number is needed to chose the mechanism for scattering, and then the final state after scattering for this mechanism must be chosen stochastically from the probability of scattering from the initial to all possible final states, which generally requires the numeric inversion of the energy dispersion relation in a fullband scheme.

The CA algorithm discussed above basically provides greatly improved performance in terms of execution speed, at the cost of significantly increased memory requirement (to store the transition table), as well as discretization error due to the finite mesh in terms of the final state energy (which is reduced by reducing the coarseness of the mesh). The discretization error in the CA may lead to artificial heating, which typically manifests itself at low fields as an inaccuracy in the average energy of carriers compared to thermal equilibrium. This inaccuracy can be controlled by refinement of the nonuniform mesh, but within the constraints of the available memory. For unipolar simulation, acceptable accuracy is obtained using 2 GB of available RAM per process, which is typical for a 32-bit processor, such as the Pentium 3 and 4.

For bipolar simulation where more memory is required to store the transition table of both the conduction and valence bands, a memory-saving, mixed “hybrid” MC/CA algorithm discussed in Ref. [7] has been reported, in which the CA algorithm is only used for portions of the band whereas close to band extremum and at high energies, a conventional fullband MC algorithm is employed. In the present work, only unipolar simulation is considered, so all results discussed are for the CA code and not the modified hybrid approach.

### SIMULATION RESULTS

In this section, we compare simulation results using the CA algorithm discussed in “Carrier Dynamics” Section for an ensemble of carriers with published fullband ensemble Monte Carlo (EMC) results reported using the code DAMOCLES [4]. Here, we attempt to use the same physical models in terms of EPM band structure, valence shell phonon dispersion, and scattering mechanisms as used in DAMOCLES to compare the accuracy of
the present code. Then in “Transient Response” Section, we compare recent experimental ultrafast data to the predictions of the present CA code.

Comparison of Fullband EMC and CA Simulations

Comparisons between the CA and EMC approaches in modeling electron transport in Si are shown in Fig. 4, where the average energy and velocity versus field are shown. While the agreement with published data is excellent, the saturation velocity computed with the CA approach is somewhat lower than the one obtained via EMC simulations. This phenomenon is believed to be possibly due to the lack of secondary electron creation in the present CA implementation.

Figure 5 illustrates the calculated steady state electron distribution function obtained from the CA method. On the same plot, the energy averaged scattering rates calculated over constant energy surfaces in $k$-space are shown. Both are in good agreement with results found from fullband EMC simulation.

Within the same CA algorithm, hole transport may be simulated as well using the EPM band structure for Si (Fig. 6). Results are shown for the average energy and velocity field characteristic compared with results reported by Fischetti and Laux [4]. Here evidence of inadequate mesh refinement are illustrated, as the average energy of carriers at low energy is somewhat above the thermal equilibrium value at 300 K. The agreement of the velocity-field characteristic on the other hand is quite good.

For ionic materials such as III–V compounds, the CA algorithm is equally accurate. The main difference is the inclusion of polar optical phonon scattering through the Fröhlich interaction given by Eq. (2). Figure 7 shows the comparison of the average energy and drift velocity for InP with fullband EMC results [4]. The agreement is excellent over most of the range of electric field strengths.

TRANSIENT RESPONSE

Recently, ultrafast measurement of terahertz radiation emission due to short pulse excitation of electron-hole pairs in GaAs and InP pin structures have been reported [17].
The time variation of the emitted radiation is directly related to the time varying Hertzian dipole arising from the acceleration of electrons and holes in opposite directions generated in the intrinsic region of the pin structures. This measurement gives a direct time-resolved measurement of the transient overshoot velocity in these materials, which may be compared to calculation. In the interpretation of these experiments, it is assumed that the transient velocity of electrons is primarily responsible for the observed optical pulse, and therefore direct comparison is made to the transient velocity associated with carriers accelerated in the constant electric field of the intrinsic region. Figure 8 shows the calculated electron velocity in the direction of the electric field for GaAs simulated using the fullband CA algorithm (here shown as negative). The maximum velocity corresponding to the peak magnitude of the data in Fig. 8 may be compared to the same quantity measured in the ultrafast experiment, as shown in Fig. 9 for both GaAs and InP. Within the error bars of the experiment indicated in this figure, the results are in excellent agreement with this experimental measurement of the transient overshoot velocity.

**COMPUTATIONAL ISSUES**

In this section, the main algorithmic aspects of the CA approach are discussed and compared with the strategies adopted within the EMC method.

**Discretization Schemes**

As suggested by their denomination, particle-based approaches require the tracking of carriers both in momentum and position space. For any fixed field, the computational load clearly scales linearly with the carrier population. Both domains in momentum space
(i.e. the BZ) and in real space (i.e. the device layout) are complicated, and a tetrahedral (or triangular, in the case of 2D devices) grid would allow to accurately reproduce their geometric features. However, tracking particles in tetrahedral grids requires a large computational effort, and makes the simulation programs impractical. A commonly adopted compromise to track particles in position space is to use tensor-product inhomogeneous grids with cell boundaries matching some finer and homogeneous discretization schemes. This well known technique, described for example in Ref. [7], allows for computations between integer numbers, and reduces the CPU time spent for particle tracking.

With regard to mesh in momentum space, full-band EMC programs have been implemented both with tetrahedral [18] and homogeneous tensor-product grids [4]. In the present implementation, the CA code also uses a tensor-product grid, but with irregular spacing, to reduce to acceptable values the otherwise prohibitive errors in energy conservation (see next section).

**Choice of the Final State**

As mentioned in “Carrier Dynamics” section, the main difference between the CA and the EMC method is in the algorithm for the choice of the new electron state after a scattering has occurred. In fact, while the EMC approach minimizes the use of fast RAM, the CA tends to store many precomputed quantities, to reduce the runtime computing time. This difference is mainly expressed by the structure of the scattering table: an EMC code stores the probability of carrier to be scattered somewhere in the momentum space, while the CA stores that probability for each small region of the discretized BZ. Clearly, the CA table requires an enormous amount of fast memory, while the EMC approach must calculate the final state (i.e. the carrier wave vector) after each scattering. This computation makes the EMC approach prohibitively slow (i.e. the carrier wave vector) after each scattering. This computation makes the EMC approach prohibitive slow (i.e. the carrier wave vector). The final wave-vector must then be chosen somewhere within the discrete cell. Besides reducing the grid spacing, we studied the effect of different placement strategies on the electron distribution function. Result of the analysis can be seen in Fig. 10, which has been computed using a high resolution in energy, to identify small features in the carrier distribution. Full-band EMC results have been included for comparison.

During the simulation, the final wave-vector was randomly placed in regions of different size around the cell’s center. As the volume of the central region was increased, the computed electron distribution function was smoother, but an anomalous diffusion is noticeable at the high energy tails of the curves. The chosen placement strategy randomly maps the wave-vector in a region with side 0.25 times the size of the side of the final cell.

**FUTURE WORK**

This section shortly describes the features which are currently being implemented in the CA code, and are particularly suited for the particle-based simulation methods.

**Anisotropic Scattering Rates**

Since the transition tables used in the CA are precomputed, the execution time of the transport kernel is independent of the model used, whether it is approximate parabolic band rates, or completely anistropic scattering rates derived from first principles electronic structure calculations. There are various approaches to the problem of the calculation of the electron–phonon interaction directly from the electronic structure and phonon dispersion, but for semiconductor materials, the main results reported to date for transport
calculations are based on the so-called rigid ion model [16]. Empirical models may be employed in the rigid ion model for the electron and phonon states, such as the EPM and valence shell models discussed earlier. Such calculations have been reported for Si [19] as well as ZnS [20], which are calculated for every initial and final state in the first BZ. Such anisotropic rates will provide the input for the present CA algorithm in future calculations.

Parallel Computing

The local nature of the carrier interactions (excluding carrier–carrier interactions) naturally leads to parallel computing as a strategy to reduce execution time in particle-based simulation [21]. Recent developments in distributed computing have led to inexpensive alternatives to the use high-end parallel supercomputers. In particular, the availability of powerful workstations and relatively fast computer network has stimulated the development of efficient and reliable interprocess communication software [22] to be used for parallel applications.

Using a population decomposition approach, we modified the scalar version of the particle-based algorithm in order to improve the efficiency of our particle-based simulation engine. However, preliminary results show that the performance of this parallel algorithm is limited by the bandwidth of the communication network, if a standard “fast Ethernet” (100 Mb/s) network is used. The recent introduction of hardware capable of bidirectional communication at speed up to 2 Gb/s should address the problem, and allow to increase the performance nearly linearly with a relatively large number of concurrent processes.

CONCLUSIONS

Herein we have reviewed fullband particle-based approaches for the simulation of charge transport in semiconductors. Both the EMC and the so-called CA methods have been discussed, and their main algorithmic differences have been compared. In particular, the maturity of the CA method has been demonstrated, and its role as a faster substitute to the traditional EMC approach has been outlined. The speed-up of the CA method for calculating frequency distribution functions in solids, “Hybrid full-band Cellular Automaton/Monte Carlo approach for fast simulation of charge transport in semiconductors”, IEEE Transactions on Electron Devices 47(10), 1909–1915.

Brillouin, L. (1946) Wave Propagation in Periodic Structures, International Series in Pure and Applied Physics, 1st Ed. (McGraw-Hill, New York).

Gilat, G. and Rauhenheimer, L.J. (1966) “Accurate numerical method for calculating frequency distribution functions in solids”, Physical Review 144(2), 390–395.

Pötz, W. and Vogl, P. (1981) “Theory of optical-phonon deformation potentials in tetrahedral semiconductors”, Physical Review B 24(4), 2025–2037.

Feldkamp, L.A., Steinman, D.K., Vagelatos, N., King, J.S. and Venkataraman, G. (1971) “Lattice dynamics of cubic Zinc Sulfide by neutron scattering”, Journal of Physical and Chemical Solids 32(7), 1573–1583.

Kunc, K. and Nielsen, O.H. (1979) “Lattice dynamics of Zincblende structure compounds. II. Shell model”, Computer Physics Communications 17(4), 413–422.

Price, D.L., Rowe, J.M. and Nicklow, R.M. (1971) “Lattice dynamics of Grey Tin and Indium Antimonide”, Physical Review B 3(4), 1268–1279.

Borchers, P.H. and Kunc, K. (1978) “The lattice dynamics of indium nitride”, Journal of Physics C: Solid State Physics 11, 4145–4155.

Fischetti, M.V., Sano, N., Laux, S.E. and Natori, K. (1996) “Full-band-structure theory of high-field transport and impact ionization of electrons and holes in Ge, Si, and GaAs”, Proceedings of the International Conference of Semiconductor Processes and Devices (SISPAD’96), (Tokyo, Japan).

Fischetti, M.V. and Higman, J.M. (1991) “Theory and calculation of the deformation potential electronic–phonon scattering rates in semiconductors”, Chapter 5, In: Hess, K., eds, Monte Carlo Device Simulation: Full Band and Beyond The Kluwer International Series in Engineering and Computer Science, (Kluwer Academic Publishers, Dordrecht, The Netherlands), pp. 123–160.

Leitenstorfer, A., Hunsche, S., Shah, J., Nuss, M.C. and Knox, W.H. (1999) “Ultrafast high-field transport in semiconductors”, Physica B 272, 348–352.

Acknowledgements

The authors would like to thank Manfred Dür and Massimo Fischetti for their comments and assistance related to this research. This work has been partially supported by the National Science Foundation Grant ECS-9976484.
[18] Bude, J. and Smith, R.K. (1994) “Phase-space simplex Monte Carlo for semiconductor transport”, Semiconductor Science and Technology 9(5S), 840–843.
[19] Yoder, P.D., Natoli, V.D. and Martin, R.M. (1993) “Ab initio analysis of the electron–phonon interaction in silicon”, Journal of Applied Physics 73(9), 4378–4383.
[20] Dür, M., Goodnick, S.M., Redmer, R., Reigrotzki, M., Fitzner, N. and Stadler, M. (1999) “First principles modeling of high field transport in wide-band-gap materials”, Physica B 272(1–4), 295–298.
[21] Ranawake, U., Huster, C., Lenders, P.M. and Goodnick, S. (1994) “Pmc-3d: a parallel monte carlo 3d device simulator”, IEEE Transactions on Computer Aided Design 13, 712–724.
[22] Walker, D.W. (1994) “The design of a standard message passing interface for distributed memory concurrent computers”, Parallel Computing 20(4), 657–673.

Marco Saraniti graduated in Physics at the University of Modena (Italy) on April 1991. In 1992 he joined the Walter Schottky Institut (WSI) of the Technical University of München (Germany) where he received his Ph.D. in 1996. From 1996 to 1998 he worked as a Research Associate at Arizona State University. Since August 1998 he is an Assistant Professor at the Electrical and Computer Engineering Department of the Illinois Institute of Technology. His current research activity focuses mainly on computational electronics applied to semiconductor device simulation. He is author/co-author of more than 30 publications and several technical reports.

Yibing Hu graduated in Electrical Engineering at Tsinghua University in Beijing (China) on July 1999. He then joined the Department of Electrical and Computer Engineering at the Illinois Institute of Technology where he received his M.S. degree in July 2001. He is currently with Agere Systems.

Stephen M. Goodnick received his B.S. degree in Engineering Science from Trinity University in 1977, and his M.S. and Ph.D. in Electrical Engineering from Colorado State University in 1979 and 1983, respectively. He was an Alexander von Humboldt Fellow at the Technical University of Munich, Germany and the University of Modena, Italy in 1985 and 1986. He was the Melchor visiting chair at the University of Notre Dame in 1991. He was faculty member from 1986 to 1997 in the Department of Electrical and Computer Engineering at Oregon State University, Corvallis, Oregon, where he held the rank of Professor. He is presently Chair and Professor of Electrical Engineering at Arizona State University. Dr Goodnick is a senior member of IEEE, a member of the American Physical Society, Sigma Xi, and Eta Kappa Nu. He has co-authored over 100 journal articles and book chapters related to transport in semiconductor devices and microstructures.
Submit your manuscripts at http://www.hindawi.com