SCALED ENSEMBLE MONTE CARLO STUDIES OF IMPACT IONIZATION

A. M. Kriman[a] and R. P. Joshi[b]
[a] *Elec. &Comp. Eng'ng & CEEM., Univ. at Buffalo, NY, 14260 (kriman@acsu.buffalo.edu)*
[b] *Elec. &Comp. Eng'ng, Old Dominion Univ., Norfolk, VA, 23529 (rpj@ee.odu.edu)*

Abstract

We introduce a scaled ensemble Monte Carlo (SEMC) technique to simulate high field electron transport. This technique is designed to improve the accuracy of the phase-space statistics of the non-equilibrium carrier distribution. As recognized in weighted-ensemble Monte Carlo, increasing the number of simulated particles is inefficient, since it substantially increases the computation times without ensuring adequate representation of the sparsely populated regions of phase space. A scaled scheme is especially important for these transient simulations, since one cannot average the electronic trajectories over sufficiently long times.

The SEMC technique we propose overcomes this problem by redistributing the computational effort to weight the low-density regions of k-space more heavily. This is done through a formulation which uses an energy-dependent factor to scale the distribution functions and the scattered rates. The technique has the advantage of flexibility and simplicity in coding, is very similar to the traditional non-weighted approach, and doesn't involve particle-splitting. In the present work, the SEMC procedure is applied to simulate impact ionization for high-field transport, using a soft threshold. Since the impact ionization coefficient is strongly affected by the high-energy tail of the distribution function, we apply SEMC to evaluate this. In general, results depend sensitively on band structure and on coupling to other energy dissipation modes.

I. INTRODUCTION

The simulation of impact ionization is, in principle, not very different from the simulation of other scattering processes in semiconductors: an instantaneous, energy-dependent microscopic scattering rate $R(E)$ is determined used directly in Monte Carlo simulation; or a parameter α determined from $R(E)$ and the non-equilibrium distribution function is used in drift-diffusion simulation. The main practical difficulty is that the total scattering rate, and α, depend sensitively on the distribution function at very high energy. This is true even for the low fields where α is approximately field independent. Since only a small fraction of particles are found at the relevant high energies, ordinary Monte Carlo techniques produce results with large statistical uncertainties, and special modifications must be used which enhance the accuracy of the simulation in the low particle-density regimes. Here we describe the use of a new method, *scaled* ensemble Monte Carlo (SEMC), to perform this function.

Like most Monte Carlo treatments, the present one is based on an essentially classical description. In the absence of significant interparticle correlations, the system is described formally by a time-dependent single-particle distribution function $f(\mathbf{r},\mathbf{p};t) = f(\mathbf{x};t)$ [$\mathbf{x} = (\mathbf{r},\mathbf{p})$ is a phase space coördinate]. The distribution function obeys the Liouville equation:

$$\frac{\partial f}{\partial t} = \{H,f\} + \left(\frac{\partial f}{\partial t}\right)_{\text{coll}}, \tag{1}$$

where $H = H(\mathbf{x})$ is the time-independent single-particle Hamiltonian, and $\{\bullet,\bullet\}$ is the Poisson bracket. In our simulations, we used a multi-valley Hamiltonian for electrons in GaAs bulk [2], with central Γ and satellite L and X valleys. In our band model, the different satellite minima communicate only via deformation scattering—the wave vector is unrestricted in principle, although the region of crystal momenta between minima is essentially unoccupied for energetic reasons. This standard model is augmented by a uniform electric field.

We make the usual assumptions that collisions or scattering events take place on time scales much shorter than the time between collisions, so effects such as collision broadening can be neglected. Further, the scattering events are in fact approximated as instantaneous, so that intracollisional field effects can also be ignored. In this approximation we treat acoustic and optical phonon scattering by deformation potential, and polar optical scattering. All important sources of potential and phonon scattering can be written in the form

$$\left(\frac{\partial f(\mathbf{x};t)}{\partial t}\right)_{\text{coll}} = -\Gamma(\mathbf{x};t)\, f(\mathbf{x};t) + \int \Gamma(\mathbf{x},\mathbf{x}';t) f(\mathbf{x}';t)\, d\mathbf{x}' \ . \tag{2}$$

II. SCALING FORMALISM

Our approach, which has been developed previously for the zero-field case, depends primarily on the observation that an energy- and time-dependent scaling factor commutes with noncollisional term on the right-hand side of (1):

$$s(H,t)\ \{H,f\} = \{H,s(H,t)f\} \ . \tag{3}$$

Thus, we define a *scaled* distribution function defined by

$$\hat{f}(\mathbf{x};t) \equiv s(H(\mathbf{x}),t)\, f(\mathbf{x};t) \ . \tag{4}$$

The principal advantage of multiplying by an energy-dependent factor arises from counting statistics. If we are interested in the distribution function in some region about the phase space point x, we consider as a function of time the number of simulation particles N in the vicinity of that point. The fractional error in $\hat{f}$ is then $1/\sqrt{N}$, which is the fractional error in $f \equiv s^{-1}\times\hat{f}$ as well. For regions of low phase-space density, N is proportional to $f(\mathbf{x};t)$. Thus, for example, to examine regions where the phase space density is down by seven orders of magnitude from the maximum, one needs many times of ten million simulation particles to keep the error from exceeding the estimate.

A standard solution to this problem — weighted EMC — defines distinct regions of phase space with different ratios of simulation points to actual particles (the ratio is constant within each region). When particles cross between regions, particle trajectories must be "split" (for multiple simulation) or decimated. If the density falls smoothly, there is no single efficient place to draw the boundary between regions, and many regions and multiple particle-splitting interfaces must be defined [3].

Our approach is based upon scaling the distribution function that is simulated by EMC, rather than upon a weighting the EMC simulation of an unscaled distribution function. The choice of s determines the trajectory density. In principle, these two approaches may be equivalent in particular cases. However, a weighted EMC eliminates the usual identification between individual initial particles and individual trajectories sampled (in what is the Monte Carlo integration of the Boltzmann equation). As a result, one loses the intuitive simplicity of regarding sampled trajectories as individual particles of a large ensemble. In a scaled EMC, on the other hand, one preserves a one-to-one correspondence of initial condition to trajectory, and it remains possible to regard the trajectories sampled as the actual paths of individual particles. As we describe below, however, in order to redistribute the statistical sampling weight, one pays the price that the trajectories do not follow the paths of ordinary particles.

III. TIME-DEPENDENT SCALING

Particle-number conservation imposes an important constraint on how $s(H(\mathbf{x}),t)$ is allowed to be chosen. By appropriate normalization, the total number N of simulation particles in the simulation of $\hat{f}$ is the integral of the scaled distribution function:

$$\hat{N}(t) = \int \hat{f}(\mathbf{x})\, d\mathbf{x} = \int s(\mathbf{x}) f(\mathbf{x})\, d\mathbf{x} \ . \tag{5}$$

If s is chosen to emphasize high-energy regions which have low density, then during a relaxation, thermalization will cause a transfer of (real) electrons to lower-energy regions where s is smaller. If s is not allowed a time-dependence to compensate, the total number of simulation particles of $\hat{f}$ must decrease — degrading the statistics in $\hat{f}$. Conversely, a heating mechanism would increase the number of simulation particles, improving statistical precision but possibly requiring undesirable computational expense. By allowing s to have a time-dependence, we accomplish in the time domain what the energy dependence of s accomplishes for the energy domain: reduce variations in particle number so that fractional errors can be kept at an acceptable level throughout the region of interest, with the least computational effort.

We chose a simple form for the joint energy-and-time dependence of s: we let them be independent factors. This can be written

$$s(H,t) = \exp\big(\alpha(t)+\gamma(H)\big) \ . \tag{6}$$

Furthermore, we let the function $\gamma(H) = -H \,/\, k_B T^{eff}$. This is appropriate for distributions f which are approximately characterized by effective temperatures below T^{eff} [Higher temperatures lead to normalization problems with $\hat{f}$.] This choice is also convenient computationally: a linear function γ implies that inelastic scattering rates (described below) are space-position-independent, and can be stored efficiently in look-up tables.

The time evolution of α is then defined specifically so that the total number of simulation points is constant.

$$\dot{\alpha} = \frac{-1}{\hat{N}} \int dx\ dx'\ G(x,x')\ \hat{f}(x';t) \ , \tag{7}$$

where

$$G(x,x') \equiv s(x)\ \Gamma^I(x,x')\ [s(x')]^{-1} \ = \ \exp\big[\gamma(H(x))-\gamma(H(x'))\big]\ \Gamma^I(x,x') \tag{8}$$

is essentially the unscaled total in-scattering rate, adjusted for the change in scaling factor experienced by a particle that undergoes a change in energy. (The I-subscript is explained below.) Equation (7) is a necessary condition for $\hat{f}$ to be simulable by scattering events alone, without independent generation and recombination (or other gain/loss) mechanisms. That it is sufficient is demonstrated constructively by finding the appropriate scattering rates.

Using (2), we find that $\hat{f}$ obeys a modified Liouville equation

$$\frac{\partial \hat{f}}{\partial t} \ = \ \{H,\hat{f}\} + \left(\frac{\partial \hat{f}}{\partial t}\right)_{coll} , \tag{9}$$

in which the "collision" or scattering term is defined by

$$\left(\frac{\partial \hat{f}}{\partial t}\right)_{coll} - \frac{1}{s}\left(\frac{\partial s}{\partial t}\right)\hat{f} = s\left(\frac{\partial f}{\partial t}\right)_{coll} . \tag{10}$$

Equations (2) and (10) specify completely the modified scattering term $(\partial \hat{f}/\partial t)_{coll}$. However, in order to implement a Monte Carlo time-evolution, one must determine scattering rates for the scaled problem which are analogous to the out-scattering rates $\Gamma(x;t)$ and the in-scattering rates $\Gamma(x,x';t)$ of the unscaled problem. There is some freedom in way this is done. One well-known degree of freedom is associated with self-scattering:

$$\left\{ \begin{array}{ll} \Gamma(x;t) & \rightarrow \Gamma(x;t) + \Delta\Gamma(x;t) \\ \Gamma(x,x';t) & \rightarrow \Gamma(x,x';t) + \Delta\Gamma(x;t)\ \delta(x-x') \end{array} \right\} , \tag{11}$$

where ordinarily $\Delta\Gamma$ is chosen to make the total out-scattering a positive constant. This is a kind of gauge transformation, in which the physically-significant total scattering rate is fixed, while unobservable components of the in- and out-scattering rates make the numerical implementation tractable. Another choice of gauge (the in-scattering gauge used in (8) and indicated by a superscript I) uses $\Delta\Gamma(x) = -\Gamma(x)$, to set the out-scattering to zero. In any case, the off-diagonal part of the out-scattering is a gauge invariant. Making some further transformations, we arrive at

$$\hat{\Gamma}(x,x') = G(x,x') + K(x)\,\frac{1}{\hat{N}}\hat{f}(x) \quad \text{for } x \neq x', \tag{12}$$

where

$$K(x) \equiv \int G(x',x)\ dx' \ . \tag{13}$$

The second term on the right-hand side of (12) leads to a kind of attractive interparticle scattering. This performs a rôle similar to that of trajectory iteration in weighted-EMC approaches: simulation particles entering critical regions are given greater weight, and are effectively caused to perform multiple traversals. However, in SEMC this weighting is implemented smoothly, rather than abruptly at the boundary of a region of interest, and it is accomplished with a fixed number of particles undergoing essentially ordinary scattering.

IV. SIMULATIONS

We have applied the SEMC technique to bulk GaAs semiconductor at 300 K. We used parameters (deformation potentials, phonon energies, band structure, etc.) that have been confirmed empirically in previous simulations [2]. The SEMC simulation used an effective temperature of 400 K. The system was allowed to reach a steady state under a uniform field. The figure below illustrates a typical result. The total particle density (Γ, L and X valleys) is plotted as a function of local energy (that is, for each particle, the energy is measured from the local position of the conduction band minimum). Results are shown for 70 kV/cm. The striking feature is the range of densities estimated—sixteen (16) orders of magnitude.

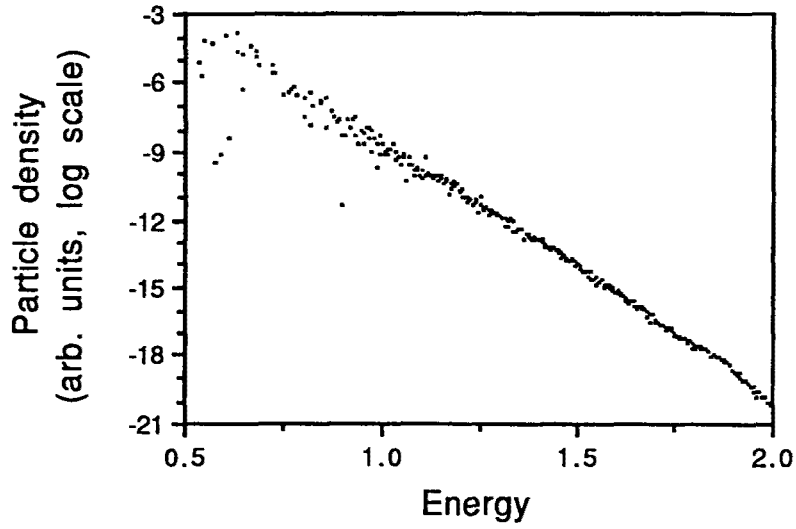

Fig. 1: Particle distribution function $f(E)$, including all valleys, as a function of local energy above the conduction band, in a 70 kV/cm field.

REFERENCES

[1] D. K. Ferry, *Semiconductors*, (Macmillan, 1991).

[2] M.-J. Kann, A. M. Kriman and D. K. Ferry, "Effect of Electron-Electron Scattering on Intervalley Transition Rates of Photoexcited Carriers in GaAs," Phys. Rev. B Vol. 41, No. 18, pp. 12659-12665, 15 June 1990.

[3] C. Jacoboni and P. Lugli, *The Monte Carlo Method for Semiconductor Device Simulation*, (Springer-Verlag, 1989).