Real-Space Treatment of the Electron-Electron and Electron-Ion Interactions in Particle-Based Device Simulators

Dragica Vasileska
Professor
Arizona State University

July 21st, 2011

In particle-based device simulation schemes one couples the Monte Carlo Transport Kernel with a Poisson equation solver as shown diagrammatically in Figure 1. Briefly, after the free-flight scatter sequence, particle-mesh coupling takes place that is followed by a Poisson equation solution for the electrostatic potential and the electric field needed in the subsequent free-flight scatter sequence.

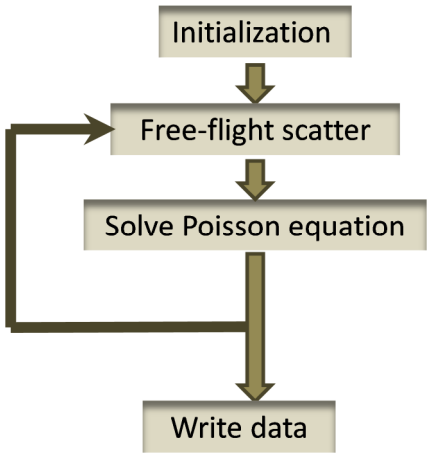

Figure 1. Typical flow-chart of a particle-based device simulator.

The Poisson equation is solved on a mesh that is determined by the Debye criterion. Namely, in critical device regions the mesh has to be smaller than the extrinsic Debye length [i]. If the mesh is infinitely small then the Coulomb potential is completely resolved. However, that would typically require a large number of node points. As in silicon devices, to get accurate results one has to solve the 2D/3D Poisson equation every 0.1 fs, and the total simulation time is on the order of 5-10 ps, that means that the Poisson equation solution, which is the bottleneck for 3D simulations, has to be solved many times which, in turn, requires very efficient Poisson solvers. The time to solve the Poisson equation limits the number of node points that has to be used in the Poisson mesh. As the mesh has to be coarser that, in turn, reduces the amount of the short-range Coulomb interaction that is accounted for via the solution of the 3D Poisson equation. The short-range portion of the Coulomb interaction is typically accounted for by considering Coulomb scattering as additional scattering mechanism in the k-space portion of the Monte Carlo transport kernel. The proper calculation of electron-electron scattering and electron ion scattering requires a proper screening model. Screening requires evaluation of the distribution function, which is typically noisy and time consuming task [ii]. Moreover, how much of the short-range Coulomb interaction and how much of the long-range Coulomb interaction is taken into account with the k-space approach is not really known and some overestimation or underestimation of the interaction usually occurs. Also, multiple scattering processes and dynamical screening are typically almost impossible to be accounted for.

To avoid the problem with the k-space treatment of the Coulomb interaction, a real space approach has been proposed by Lugli and Ferry [iii] in which the electron-electron and the electron-ion interactions are accounted for via real-space molecular dynamics routine. It is important to note that direct application of the real-space molecular dynamics can be used for bulk systems only where it is not required to solve the Poisson equation. This aspect has already been elaborated in the beginning of this section. Hence, an approach is needed that correctly accounts for the full Coulomb interaction in particle-based device simulators. The group from ASU has been in a sense a pioneer in this field and in our simulation modules we currently have implemented three approaches:

1. The Corrected Coulomb approach – an approach that we have introduced [iv],
2. The particle-particle-particle-mesh coupling method due to Hockney and Eastwood [v], and the
3. Fast Multipole Method [vi].

It is important to note that the Corrected Coulomb approach and the particle-particle-particle-mesh coupling methods are similar in philosophy. Namely, a correction force is calculated given the mesh and it is that correction force that is used in the molecular dynamics routine. The fast multipole method is completely different in philosophy in a sense that the Laplace equation is solved to account for charges at the ohmic contacts and afterwards only fast multipole method is used to account for the full Coulomb interactions between electrons and electrons and ions. The difference between these two ideologies is graphically shown in Figure 2.

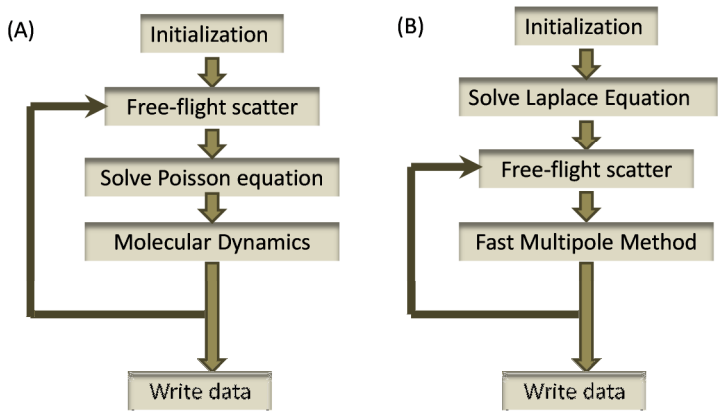

Figure 2. Philosophy behind the (A) corrected Coulomb approach, where correction force is used in the molecular dynamics routine, and (B) the fast multipole method where the full Coulomb interaction is being considered to get the force on the electrons in the free-flight portion of the Monte Carlo transport kernel.

In what follows, each of these methods is explained in more details. We first discuss the corrected Coulomb approach. Next the particle-particle-particle-mesh coupling method is discussed. Finally, the ideology behind fast multipole method is explained.

1. Corrected Coulomb Approach

Within the Corrected Coulomb approach the separation of the short-range and the long-range Coulomb interaction is accomplished in the following manner: a target and a fixed electron are placed in a 3D box and the separation between the target and the fixed electron is varied. For each separation of the target and the fixed electron the 3D Poisson equation is solved which gives the Hartree potential. The Hartree potential is used to calculate the Hartree force on the electron. Simultaneously, given the separation between the electrons, the Coulomb force is calculated and the Hartree force is subtracted from the Coulomb force. This gives a correction force, which in general diverges when the separation between the target and the fixed electron is zero. Modification to the correction force has to be made to account for this divergence. The way that is accomplished is the following one. For distances smaller than the Bohr radius, linear interpolation of the force to zero is assumed. Since the correction force is significant for few mesh spacing, an outer radius is defined and all the electrons and/or ions that fall within the outer radius of the fixed electron are being considered using the electron-electron and electron-ion interaction to get the short range force on the target electron. That target force is added to the Hartree force and used in the subsequent free-flight portion of the Monte Carlo routine. Using this methodology, excellent agreement is achieved for the doping dependence of the low-field electron mobility between the simulation and the available experimental data. Results of these simulations can be found in Ref. [vii]. Also given in Ref. [iv] are the implementation details of the corrected Coulomb approach.

2. Particle-Particle-Particle-Mesh Method

The particle-particle-particle-mesh (P^3M) algorithms are a class of hybrid algorithms developed by Hockney and Eastwood [v]. These algorithms enable correlated systems with long-range forces to be simulated for a large ensemble of particles. The essence of P^3M algorithms is to express the interparticle force as a sum of a short-range part calculated by a direct particle-particle force summation and a long-range part approximated by the particle-mesh (PM) force calculation. Using the notation of Hockney, the total force on a particle *i* may be written as

$$\boldsymbol{F}_i = \sum_{j \neq i} \boldsymbol{F}_{ij}^{coul} + \boldsymbol{F}_i^{ext}. \qquad (1)$$

$\boldsymbol{F}_i^{ext}$ represents the external field or boundary effects of the global Poisson solution. $\boldsymbol{F}_{ij}^{coul}$ is the force of particle *j* on particle *i* given by Coulomb's law as

$$\boldsymbol{F}_{ij}^{coul} = \frac{\boldsymbol{q}_i \boldsymbol{q}_j}{4\pi\varepsilon} \frac{\left(\boldsymbol{r}_i - \boldsymbol{r}_j\right)}{\left|\boldsymbol{r}_i - \boldsymbol{r}_j\right|^3}, \tag{2}$$

where $\boldsymbol{q}_i$ and $\boldsymbol{q}_j$ are particle charges and $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$ are particle positions. In a P^3M algorithm, the total force on particle *i* is split into two sums

$$\boldsymbol{F}_i = \sum_{\substack{j \neq i \\ SRD}} \boldsymbol{F}_{ij}^{sr} + \sum_{\substack{j \neq i \\ GD}} \boldsymbol{F}_{ij}^{m}. \tag{3}$$

The first sum represents the direct forces of particles *j* on particle *i* within the short-range domain (SRD), while the second sum represents the mesh forces of particles *j* on particle *i* over the global problem domain (GD) as well as the effect of material boundaries and boundary conditions on particle *i*. $\boldsymbol{F}_{ij}^{sr}$ is the short-range particle force of particle *j* on particle *i*, and $\boldsymbol{F}_{ij}^{m}$ is the long-range mesh force of particle *j* on particle *i*. The short-range Coulomb force can be further defined as

$$\boldsymbol{F}_{ij}^{sr} = \boldsymbol{F}_{ij}^{coul} - \boldsymbol{R}_{ij} \tag{4}$$

where $\boldsymbol{F}_{ij}^{coul}$ is given by Eq. (2) and $\boldsymbol{R}_{ij}$ is called the reference force. The reference force in Eq. (4) is needed to avoid double counting of the short-range force due to the overlapping domains in Eq. (3). The reference force should correspond to the mesh force inside the short-range domain (SRD) and equal to the Coulomb force outside the short-range domain. In other words, a suitable form of reference force for a Coulombic long-range force is one which follows the point particle force law beyond the cutoff radius $\boldsymbol{r}_{sr}$, and goes smoothly to zero within that radius. Such smoothing procedure is equivalent to ascribing a finite size to the charged particle. As a result, a straightforward method of including smoothing is to ascribe some simple density profile $\boldsymbol{S}(\boldsymbol{r})$ to the reference inter-particle force. Examples of shapes, which are used in practice, and give comparable total force accuracy are the uniformly charged sphere, the sphere with uniformly decreasing density, of the form given in Eq. (5) and the Gaussian distribution of density. The second scheme gives marginally better accuracies in 3D simulations. For this case the reference force can be obtained [[viii]] as in Eq. (5). Hockney advocates pre-calculating the short-range force, $\boldsymbol{F}_{ij}^{sr}(\boldsymbol{r})$ defined in Eq. (4) including the reference force above for a fixed mesh. The reference, short-range and Coulomb force are each represented in Figure 3. It is important to extend the P^3M

algorithm to nonuniform meshes for the purpose of semiconductor device simulation since practical device applications involve rapidly varying doping profiles and narrow conducting channels which need to be adequately resolved. A method similar to that used in Ref. [viii] is depicted in Figure 3. Since the mesh force from the solution to the Poisson equation is a good approximation within about two mesh spaces, r_{sr} is locally chosen as the shortest distance which spans two mesh cells in each direction of every dimension of the mesh at charge *i*.

$$S(\mathrm{r}) = \begin{cases} \dfrac{48}{\pi r_{sr}^{\,4}}\left(\dfrac{r_{sr}}{2} - r\right), & r \le r_{sr}/2 \\ 0, & otherwise \end{cases},$$

$$\begin{cases} R_{ij}(r) = \dfrac{q_i q_j}{4\pi\varepsilon} \times \dfrac{1}{35 r_{sr}^2}(224\xi - 224\xi^3 + 70\xi^4 + 48\xi^5 - 21\xi^6) & \xi = \dfrac{2r}{r_{sr}} \quad \text{and} \quad 0 \le r \le r_{sr}/2 \\ R_{ij}(r) = \dfrac{q_i q_j}{4\pi\varepsilon} \times \dfrac{1}{35 r_{sr}^2}\left(\dfrac{12}{\xi^2} - 224 + 896\xi - 840\xi^2 + 224\xi^3 + 70\xi^4 - 48\xi^5 + 7\xi^6\right) & r_{sr}/2 \le r \le r_{sr} \\ R_{ij}(r) = \dfrac{q_i q_j}{4\pi\varepsilon} \times \dfrac{1}{r^2} & r > r_{sr} \end{cases} \tag{5}$$

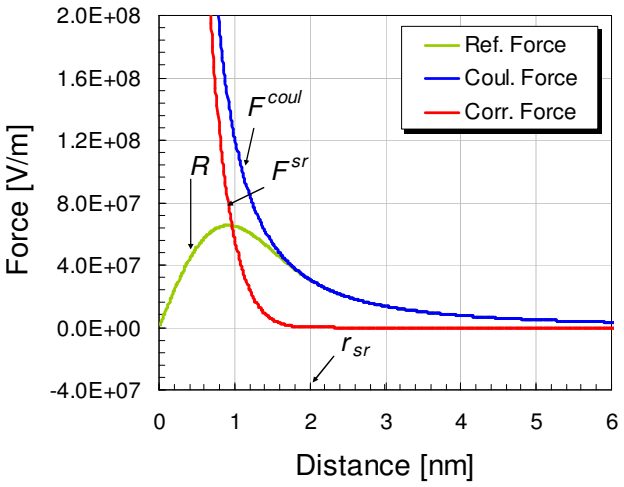

Figure 3. Illustration of the P^{3}M approach.

3. Fast Multipole Method

FMM was initially introduced by Rokhlin [ix] as a fast solution method for integral equations for two-dimensional Laplace's equation. In Rokhlin's paper the term *FMM* did not appear but the main framework of FMM was constructed. After Rokhlin's work, Greengard [x] refined the algorithm, applied FMM to two and three-dimensional *N*-body problems whose interactions are Coulombic or gravitational in nature and showed the applicability of FMM to various fields. Greengard's 1987 Yale dissertation "*The Rapid*

Evaluation of Potential Fields in Particle Systems" won an ACM Distinguished Dissertation Award. In a system of N particles, the decay of the Coulombic or gravitational potential is sufficiently slow that all interactions must be accounted for, resulting in CPU time requirements of the order $O\left(N^2\right)$. Whereas, the FMM algorithm requires an amount of work proportional to N to evaluate all interactions to within roundoff error, making it practical for large-scale problems encountered in the fields of plasma physics, fluid dynamics, molecular dynamics, and celestial mechanics.

There have been a number of previous efforts aimed at reducing the computational complexity of the *N*-body problem. As mentioned in the previous sections particle-in-cell methods have received careful study and are used with much success, most notably in plasma physics. Assuming the potential satisfies Poisson's equation, a regular mesh is laid out over the computational domain and the method proceeds by:

1. interpolating the source density at mesh points,
2. using a fast Poisson solver to obtain potential values on the mesh, and
3. computing the force from the potential and interpolating to the particle positions.

The complexity of these methods is of the order $O\left(N + M \log M\right)$, where M is the number of mesh points. The number of mesh points is usually chosen to be proportional to the number of particles, but with a small constant of proportionality so that $M \ll N$. Therefore, although the asymptotic complexity for the method is $O\left(N \log N\right)$ the computational cost in practical calculations is usually observed to be proportional to *N*. Unfortunately, the mesh provides limited resolution, and highly non-uniform source distributions cause a significant degradation of performance. Further errors are introduced in step (3) by the necessity for numerical differentiation to obtain the force. To improve the accuracy of particle-in-cell calculations, short-range interactions can be handled by direct computation, while far-field interactions are obtained from the mesh, giving rise to so-called particle–particle/particle–mesh (P^3M) methods described in section 2 above. While these algorithms still depend for their efficient performance on a reasonably uniform distribution of particles, in theory they do permit arbitrarily high accuracy to be obtained. As a rule, when the required precision is relatively low, and the particles are distributed more or less uniformly in a rectangular region, P^3M methods perform satisfactorily. However, when the required precision is high (as, for example, in the modeling of highly correlated systems), the CPU time requirements of such algorithms tend to become excessive.

In FMM Rokhlin uses *multipole moments* to represent distant particle groups and introduces a *local expansion* to evaluate the contribution from distant particles in the form of a series. The multipole moment associated with a distant group can be *translated* into the coefficient of the local expansion

associated with a local group (See Figures 4 and 5). Interactions with particles which are nearby are handled directly.

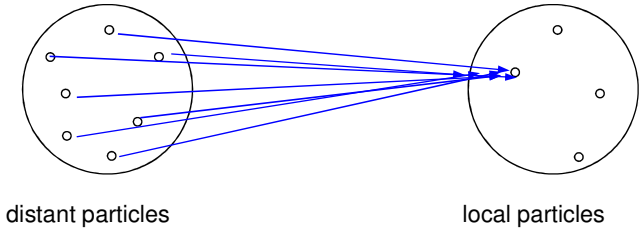

Figure 4. Conventional evaluation of contribution from distant particles: $O(N^2)$ algorithm.

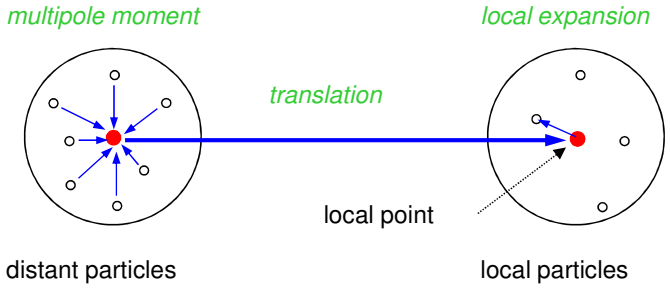

Figure 5. Evaluation with the multipole moment and the local expansion: $O(N)$ algorithm.

In addition to Rokhlin's work, Greengard introduces a hierarchical decomposition of a data-space with a quad-tree in two dimensions and an oct-tree in three dimensions to carry out efficient and systematic grouping of particles with tree structures. The hierarchical decomposition is used to cluster particles at various spatial lengths and compute interactions with other clusters that are sufficiently far away by means of the series expansions.

For a given input configuration of particles, the sequential FMM first decomposes the data-space in a hierarchy of blocks and computes local neighborhoods and *interaction-lists* involved in subsequent computations. Then, it performs two passes on the decomposition tree. The first pass starts at the leaves of the tree, computing *multipole expansion coefficients* for the Columbic field. It proceeds towards the root accumulating the multipole coefficients at intermediate tree-nodes. When the root is reached, the second pass starts. It moves towards the leaves of the tree, *exchanging* data between blocks belonging to the neighborhoods and interaction-lists calculated at tree-construction. At the end of the downward pass all long-range interactions have been computed. Subsequently, nearest-neighbor computations are performed directly to take into consideration interactions from nearby bodies. Finally, short- and long-range

interactions are accumulated and the total forces exerted upon particles are computed. The algorithm repeats the above steps and simulates the evolution of the particle system for each successive time-step.

References

[i] D. Vasileska, S. M. Goodnick and G. Klimeck, *Computational Electronics: Semiclassical and Quantum Transport Modeling*, Taylor and Francis, June 2010.

[ii] M. V. Fischetti and S. E. Laux, "Long-range Coulomb interactions in small Si devices. Part I: Performance and reliability," *Journal of Appl. Phys*., vol. 89, pp. 1205–1231, 2001.

[iii] P. Lugli and D. K. Ferry, *IEEE Trans. Electron Dev*. **32**, 2431 (1986).

[iv] W. J. Gross, D. Vasileska and D. K. Ferry, "A Novel Approach for Introducing the Electron-Electron and Electron-Impurity Interactions in Particle-Based Simulations," *IEEE Electron Device Lett.* 20, No. 9, pp.463-465 (1999)

[v] R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles* (New York, McGraw-Hill, 1981).

[vi] L. Greengard and V. Rokhlin, "A fast algorithm for particle simulations," *J. Comput. Phys*., vol. 135, no. 2, pp. 280–292, 1997.

[vii] W. J. Gross, D. Vasileska and D. K. Ferry, "3D Simulations of Ultra-Small MOSFETs with Real-Space Treatment of the Electron-Electron and Electron-Ion Interactions", *VLSI Design*, Vol. 10, pp. 437-452, 2000.

[viii] C. J. Wordelman, U. Ravaioli, "Integration of a particle-particle-particle-mesh algorithm with the ensemble Monte Carlo method for the simulation of ultra-small semiconductor devices", *IEEE Trans. Electron Devices*, vol. 47, pp. 410-416, 2000.

[ix] V. Rokhlin, "Rapid solution of integral equations of classical potential theory", *J. Comp. Phys*., vol. 60, pp. 187-207, 1985.

[x] R. Beatson and L. Greengard, "A short course on fast multipole methods," in Wavelets, Multi-level Methods and Elliptic PDEs (Leicester, 1996), ser. *Numer. Math. Sci. Comput*. New York: Oxford Univ. Press, 1997, pp. 1–37.