Advances in Numerical Methods for Convective Hydrodynamic Model of Semiconductor Devices

N. R. Aluru, K. H. Law and R. W. Dutton

Integrated Circuits Laboratory, Stanford University,
Stanford, California 94305 USA

Abstract

The convective hydrodynamic model of semiconductor devices is analyzed employing parallel and stabilized finite element methods. The stabilized finite element method for the two-carrier hydrodynamic equations and the parallel computational model are briefly described. Numerical results are shown for a bipolar transistor. A comparison of drift-diffusion, energy-transport and the hydrodynamic models is presented for a 0.1μm channel n^+-n-n^+ silicon diode.

1. Introduction

Comprehensive semiconductor device simulations employing the hydrodynamic model involve the solution of the Poisson equation, electron and hole hydrodynamic conservation laws and the lattice thermal diffusion equation [1]. The electron and hole hydrodynamic equations are derived by considering the zeroth, first and second moments of the Boltzmann transport equations and are summarized as follows:

$$\frac{\partial c_\alpha}{\partial t} + \nabla \cdot (c_\alpha \mathbf{u}_\alpha) = [\frac{\partial c_\alpha}{\partial t}]_{col} \tag{1}$$

$$\frac{\partial \mathbf{p}_\alpha}{\partial t} + \mathbf{u}_\alpha(\nabla \cdot \mathbf{p}_\alpha) + (\mathbf{p}_\alpha \cdot \nabla)\mathbf{u}_\alpha = (-1)^\alpha \epsilon c_\alpha \mathbf{E} - \nabla(c_\alpha k_b T_\alpha) + [\frac{\partial \mathbf{p}_\alpha}{\partial t}]_{col} \tag{2}$$

$$\frac{\partial w_\alpha}{\partial t} + \nabla \cdot (\mathbf{u}_\alpha w_\alpha) = (-1)^\alpha \epsilon c_\alpha(\mathbf{u}_\alpha \cdot \mathbf{E}) - \nabla \cdot (\mathbf{u}_\alpha c_\alpha k_b T_\alpha) - \nabla \cdot \mathbf{q}_\alpha + [\frac{\partial w_\alpha}{\partial t}]_{col} \tag{3}$$

where $\mathbf{u}_\alpha$, $\mathbf{p}_\alpha$, T_α, w_α and $\mathbf{q}_\alpha$ are the velocity vector, momentum density vector, temperature, energy density and heat flux vector of the carrier α. (For electrons, α = n or $\alpha = 1$; for holes, α = p or $\alpha = 2$). The terms $[]_{col}$ represent the rate of change in the particle concentration, momentum and energy due to collision of the carriers; the collision terms can be approximated by their respective relaxation times and the expressions can be found in [1].

Numerical studies employing the convective hydrodynamic model cannot be performed trivially since conventional numerical methods often fail when convective terms are included. The classical Scharfetter-Gummel (SG) method for discretization, that can be extended for the simplified hydrodynamic (neglecting convective terms) and the energy-transport models, does not work well for the hydrodynamic model. We have developed new stabilized finite element methods and they are summarized in this paper. This paper also presents a parallel computational model for the finite element method on distributed memory parallel computers and provides a comparison of drift-diffusion, energy-transport and hydrodynamic models for a 0.1μm channel n^+-n-n^+ silicon diode.

2. Stabilized Finite Element Methods

The Poisson and the lattice thermal diffusion equations are elliptic in nature and standard Galerkin finite element discretization can be shown to be stable. Galerkin finite element formulation is, however, unstable for electron and hole hydrodynamic equations as the solutions to these equations contain steep layers. Hence, advanced Galerkin/least-squares finite element formulations are developed. The temporal behavior of the equations is accounted by employing a discontinuous Galerkin method in time. With a discontinuous Galerkin in time and a Galerkin/least-squares in space, the discretization technique is referred to as a space-time Galerkin/least-squares finite element method. The space-time Galerkin/least-squares finite element method can be shown to be stable for the electron and hole hydrodynamic equations. Hence this method is also referred to as a stabilized finite element method. The important steps in the space-time Galerkin/least-squares formulation are summarized as follows:

1. A least-squares term of a residual type is introduced to the weak form of the hydrodynamic equations so that the numerical stability of the system is enhanced. Furthermore, a discontinuity-capturing term is added to overcome the undershoot and overshoot phenomena near steep gradients. The least-squares and discontinuity capturing terms vanish when the exact solution is substituted, thus making the method consistent.

2. Within each space-time slab, the trial and test functions are approximated by linear in space and constant in time basis functions.

3. The nonlinear system is solved using a Newton iterative scheme by linearizing the nonlinear equations with respect to the unknown trial solution.

Comprehensive semiconductor device simulations are performed employing the stabilized methods discussed above. The boundary conditions required for the electron and hole hydrodynamic equations are more complicated compared to the drift-diffusion or the energy-transport model. They are discussed in greater detail in [1]. Figure 1 and Figure 2 show, respectively, the electron concentration and the lattice temperature for a bipolar transistor. The numerical results indicate that stabilized finite element methods produce extremely robust and accurate solutions.

3. Parallel Computational Model

The convective hydrodynamic model demands enormous computations due to both the advanced nature of the model and the numerical method. A single-program-multiple-data (SPMD) programming model is designed and implemented on distributed memory parallel computers such as iPSC/860, Touchstone Delta and IBM SP-1. The power of the programming model lies in the quick adaptation of the serial code to the parallel code. The parallel algorithms are shown to be scalable and excellent efficiencies are reported. Figure 3 is a description of the parallel programming model and the various features available in the element library. A pre-processor reads/generates a mesh, reads boundary conditions and several other input parameters for the program. In a parallel program the pre-processor includes a domain decomposition algorithm, which partitions the domain into several subdomains. The pre-processor also prepares the input data for each subdomain and sends it to the corresponding processor. The same serial program is then executed on each processor with changes made only to accomodate inter-processor communication. Inter-processor communication is needed when iterative solvers perform vector-vector and matrix-vector products. The calculation of currents is also done in parallel and the final step is visualization which is done on a host. Figure 4 presents a comparison of the CPU times on serial and parallel computers. The CPU time on serial machines increases exponentially as the mesh size increases. The CPU time is shown to be significantly lower employing 32 processors of iPSC/860. The hydrodynamic simulations are shown to be very efficient employing just 8 processors of SP-1. These results indicate that serial computers are bottlenecks for grand challenge device applications and parallel computers can enable solution of the hydrodynamic model in a reasonable time.

4. Matrix-Free Techniques

The non-symmetric system of equations obtained from discretizing the convective hydrodynamic equations are solved using a matrix-free generalized minimal residual (GMRES) algorithm. The GMRES algorithm primarily involves vector-vector and matrix-vector products. The matrix-vector products can be evaluated by matrix-free algorithms to reduce storage costs and improve convergence of the fractional-step solution algorithm. In a matrix-free algorithm, the Jacobian is never formed and a matrix-vector product involving the Jacobian, $\mathbf{J}(\mathbf{v})$, and a vector, $\mathbf{u}$, can be obtained directly using the residual $\mathbf{R}$ as shown below

$$\mathbf{J}(\mathbf{v})\mathbf{u} = \lim_{\varepsilon \to 0} \frac{\mathbf{R}(\mathbf{v} + \varepsilon \mathbf{u}) - \mathbf{R}(\mathbf{v})}{\varepsilon} \tag{4}$$

where ε is taken as a small but finite value. The choice of ε is important for accurate determination of the matrix-vector product $\mathbf{J}(\mathbf{v})\mathbf{u}$ and is discussed in greater detail in [1].

5. Comparison of Transport Models

The significance of the convective terms in the transport equations can be revealed by a comparison of currents (see Figure 5) obtained from the drift-diffusion, energy-transport and the hydrodynamic models for a 0.1μm channel n^+-n-n^+ silicon diode. The results indicate that the convective hydrodynamic model produces non-negligible effects on submicron terminal characteristics. For large applied voltages the currents obtained from the hydrodynamic model are almost twice greater than the currents obtained from the energy-transport model. While the accuracy of the transport models could not be established because of the lack of experimental results for the test structure, an analysis of the energy resulted in the hydrodynamic model shows that the kinetic energy is approximately 50% of the total energy in the channel region (see Figure 6). The non-negligible amount of kinetic energy in the channel region could indicate that the results obtained from the hydrodynamic model are more accurate since the simplified hydrodynamic and the energy-transport models neglect convective effects and assume that the kinetic energy is negligible compared to the thermal energy.

References

[1] N. R. Aluru, *Parallel and Stabilized Finite Element Methods for the Hydrodynamic Transport Model of Semiconductor Devices*, Ph.D Thesis, Stanford University, June 1995.