invited

Hierarchy of simulation approaches for hot carrier transport in deep sub–micron devices

Umberto Ravaioli

Beckman Institute and
Department of Electrical and Computer Engineering
Unversity of Illinois at Urbana-Champaign
Urbana, IL 61801, USA

1. Introduction

Semiconductor device integration has progressed rapidly in recent years, often surpassing predictions and roadmap objectives. Device scaling is now pushing the gate length of commercial MOSFET integrated transistors closer and closer to 0.1 μm, with experimental devices already well beyond that limit. In the conduction channel of such short silicon devices, very large fields are established even if every effort is made to reduce bias voltages, and the energy distribution of carriers is very appreciably out of equilibrium. In particular, the high energy tail of the carrier distribution becomes of great importance to understand the details of transport and device behavior. The presence of these so called *hot carriers* is typically linked to reliability problems.

Device simulation has evolved over the last thirty years, developing new models and simulation approaches of increasing complexity in the attempt of capturing new hot carriers and the quantum effects manifesting themselves as scaling is pushed down to its technological limits and new device structures are introduced. Such effects have included velocity overshoot, interface scattering, injection across heterointerfaces, ballistic transport, quantum tunneling, and interface damage. The purpose of this brief review is to survey the hierarchy of physical approaches for semiconductor transport and device simulation, giving an indication of the limits of applicability and approximations underlying the various approaches. The main focus is on the relevance of the approaches for the simulation of hot carrier effects in deeply scaled devices.

2. Simulation Hierarchy

A schematic diagram for the complete hierarchy of approaches that can be applied to device simulation is shown in Fig. 1. At the top of the hierarchy we have approaches based on a quantum description of transport. This area it is still far from maturity, being the subject of much research efforts. The levels below are based on the Boltzmann Transport equation or its simplifications, with a semi–classical (or classical) description of transport. All the simulation levels, except for the compact approaches at the bottom of the hierarchy, involve the solution of a set of coupled partial differential equations, where the transport equation can actually be an integro-differential equation as is the case for the Boltzmann equation and for quantum formulations including scattering. Analytical approaches are based on integral solutions of such semiconductor equations, suitable as building blocks for the solution of complete circuits. In the case of particle Monte Carlo approaches, the transport equation is emulated by a computer experiment. With few exceptions, self-consistency is obtained by coupling the transport equations with the Poisson equation of electrostatics. The various approache will be discussed below, starting from the bottom level of the hierarchy.

Compact Approaches - With this name we refer to a broad category of analytical or semi-analytical methods that are formulated to provide a simple solution for device behavior, whenever possible in closed analytical form. The starting point is often a spatial integration of a drift–diffusion formulation, that yields terminal current values. The integration can be performed if simplifications are made in the geometry and doping distribution of the model structure. The limitations of the physical model are the same of the drift–diffusion formulation adopted. Therefore, the computational advantage in the transition from a numerical drift–diffusion simulation to an analytical formulation, comes at the cost of loss in resolution of the structure features. In order to capture the details of terminal current characteristics as verified in experiments, compact models are modified through the addition of adjustable terms that must be calibrated appropriately. The limitation of this approach is in the lack of scalability. Since the empirical fitting has to be performed in a statistical sense, but ignoring the underlying non–linear physical transport, new sets of calibration parameters are necessary when dimensions and other variables (dopings, bias voltages) are changed.

Analytical or integrated formulations are still of enormous value, because fast approaches are essential for efficient circuit design. The main problem, when devices are scaled into the deep–submicron regime, is in the rapidly

0-7803-4369-7/98 $10.00

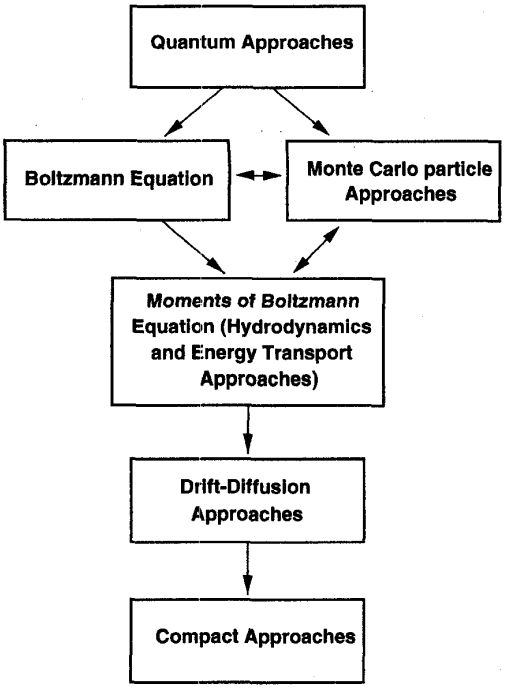

Figure 1: Schematic diagram illustrating the simulation hierarchy

soaring costs of experimental verification for model calibration. This stresses the importance of detailed physical approaches that are scalable and can provide sufficiently quantitative answers, to supplement or replace experimental verification for the purpose of calibration of simple models.

Drift-Diffusion Approaches - The drift-diffusion formulation is the simplest transport model obtained from moments of Boltzmann equation. The model adds to the Poisson equation continuity equations that account for charge flow of electrons and holes [1]. The current formulation in the continuity equations contains all the main transport information through the mobility and the diffusivity. The underlying assumption is equilibrium of the carrier gas. One can go beyond this equilibrium regime, by introducing a field dependent mobility and diffusivity, which *empirically* extends the validity of the model into the hot carrier regime. In this way one introduces only a local steady-state (bulk-like) equilibrium of the carrier population with the given field [2]. In order to introduce in physical way the effect of overshoot, additional terms dependent on the field derivative of the field must be added (augmented drift-diffusion). This approach introduces severe non-linearities which make a multi-dimensional numerical solution difficult [3]. However, the augmented drift-diffusion model is suitable for inclusion in compact models, where it has been found to be useful [4]. Despite its limitations, the basic drift-diffusion approach remains the workhorse of industry because of the well established solution methodologies and the availability of robust vendor supported implementation.

Moments of Boltzmann Equation - The basic drift-diffusion approach is improved with additional flow equations which are obtained through higher order moments of the Boltzmann equation. The so-called hydrodynamic model adds a flow equation for average momentum and one for average energyi [5]. Despite considerable increase of information, the truncation of the moments sequence must be appropriately closed with the addition of an appropriate closed form equation for the heat flow. It has been recognized that the classical Franz-Wiedemann law, as introduced in th original model, incorrectly represents the heat flow in typical device channels, particularly at the drain junction where most hot carriers are [6]. A number of empirical [6] or physical approaches [7] have been proposed to improve closure of the hydrodynamic model, which for reason of space cannot be fully reviewed here. In alternative, it has also been proposed to introduce an additional moment equation of higher order. There are also numerical challenges, due to the fact that the hydrodynamic model may have hyperbolic (i.e. wave) character, with the formation of numerical shocks in the solution which need to be adequately resolved [8]. A very large mathematical literature has been devoted to the solution of this kind of problems. Yet, although the hydrodynamic model has been known in its original form since 1970, the wide variety of formulations and numerical approaches have not yielded a standardized approach, which has somewhat hampered its widespread adoption for industrial investigations.

An alternative approach, usually referred to as Energy Transport model, only adds an energy flow equation and differs from the hydrodynamic model in the way the moment averaging is carried out [9, 10]. A microscopic relaxation time is defined for the distribution function, rather than average momentum and energy relaxation time obtained after the moment operation in the hydrodynamics formulation. The main advantage of the energy transport model is in the fact that from a numerical point of view it retains the same structure of drift-diffusion [10]. Therefore, it is quite easy to implement it in the well established and robust numerical platforms for drift-diffusion simulation.

The models at this level of the hierarchy provide a physical way to account for average hot carrier effects, which leads to improved potential solution and terminal currents estimation. However, these models do not provide any information regarding the energy dependence of the distribution function.

Solution of Boltzmann Equation - The Boltz-

mann transport equation (BTE) is a formidable integro-differential problem. for a 3-D time-dependent simulation, one has to keep track of 7 coordinates because both momentum and real space are involved. Only recently it has been possible to attempt the direct solution of the BTE, mainly within the spherical harmonic expansion approximation for the distribution function [11]. Other methods of solution use discretized numerical strategies, most notably the Scattering Matrix approach [12] and the Cellular Automata approach [13]. Since the BTE provides information on the distribution function, one needs to worry about the correct inclusion of band structure effects. In all the lower levels of the hierarchy the band structure is heavily simplified. Usually, a single parabolic or non-parabolic band band is considered, although multiple valleys can be included at the hydrodynamic level, at the cost of including an extra set of transport equations per valley. A similar problem exists when the Boltzmann equation is solved as a partial differential equation. Inclusion of a complete bandstructure is easier to implement in the case of numerical approximations, since large tables can be precalculated which include the detailed bandstructure effects.

Particle Monte Carlo Approaches - The most successful technique for the solution of the BTE has been to-date the particle Monte Carlo approach [14-17]. A detailed computer experiment follows individual carriers inside the simulation domain, with free flights interrupted by scattering events. When done carefully, this approach can be shown to be equivalent to the Boltzmann equation, and actually one can push the validity of Monte Carlo beyond the limits of definition of the BTE. Since numerical tables are used to provide all the physical information about the semiconductor, a complete band structure can be readily accounted for simply at the cost of sufficiently large tables. The numerical methods mentioned above (scattering matrix, cellular automata) are similar conceptually to the standard Monte Carlo approach, but differ in the fact that they operate directly on a distribution function rather than on the actual ensemble of partciles. Another method, the Mutation Operator Monte Carlo (MOMC) [18] which was developed as part of an evolutionary algorithm optimization, also operates directly on the distribution function and will be discussed elsewhere in these proceedings.

One of the advantages of Monte Carlo is the relative maturity of the approach, since it has been applied to practical device simulation since the mid-70's. The obvious disadvantage of the technique is the somputational cost, both in terms of CPU and memory. However, the rapid advanced of computer platforms are making Monte Carlo applications more affordable and practical today.

Quantum Approaches - Quantum device simulation is the more immature area due to the intrinsic difficulties in even formulating a complete device model that could be realistically solved. While the solution of a self-consistent transport problem is relatively feasible in the case of ballistic motion, inclusion of scattering phenomena at normal device operating temperature remains a very difficult task. A variety of formulations have been used to solve quantum problem: Schrödinger equation, Wigner function and density matrix formalism, non-equilibrium Green's functions, quantum hydrodynamic model [19-23]. All of these techniques have advantages and disadvantages, and tend to be more suitable to study specific aspects or features of transport rather than complete devices. Most of the attention has been paid to heterojunction devices, because of the more pronounced quantum effects, in part due to the lower effective mass. While a complete review of quantum approaches is beyond the scope of this work, it is worth noticing the emerging importance of atomic level calculations which go beyond the typical effective mass approximation used in the formulation of quantum transport models. As attention is focusing on the nanometer regime, system granularity and details of interfaces must be dealt with the approaches typical of material science and physical chemistry, since the approximations of smooth dopings and carrier densities tend to brak down. For traditional integrated devices in the extreme scaling limit, quantum effects are useful when combined with the semi-classical techniques introduced above. For instance, quantum features can be used to dress a Monte Carlo particle (e.g. tunneling when the particles reach an interface) while the simulation is carried on in many respect still classically.

3. Simulation Needs for Deep-Scaling

During the first decade of the next century, industry is likely to begin large scale production of integrated circuits with MOS devices having channel length at or below $0.1\mu m$ (100 nm). In order to continue the scaling trends followed until now, the range from 0.1 to 0.01 μm needs to be better understood if predictive physical simulation tools must be made available. For silicon, the mean free path of carriers at room temperature should remain below 0.01 μm (100 $\AA$) and semi-classical approaches at the level of the BTE can still be valid. However, in this size range, profound changes take place in the underlying material models. Consider for instance a doping of 10^{18} cm^{-3}. On the average, only one dopant ion is present in a cube with 100 $\AA$ side. Even at 10^{20} cm^{-3} the size of this cube would be about 22 $\AA$. It is clear that dopings cannot no longer be treated as a smooth jellium, rather, granular effects of the material need to be incorporated in a model. Similar issues are associated to detailed interaction of carriers with interfaces.

Contacts will also need to be revisited, either because the Schottky barrier at the metal/semiconductor interface will play a role in the device behavior or because hot carrier effects will become more pronounced in contact doping regions. Another aspect of scaling will involve the lateral device direction, which eventually will become small enough to generate true 3-D effects in MOS devices. Until now, little or no use has been made of multidimen-

sionality of devices for the purpose of creating new logic functions, while memory cell have exploited in some measure 3-D effect for some time. It is conceivable that novel device concepts could be developed as the current flow could be diverted from the typical quasi-1-D flow of standard MOS structures. Basic changes of device topology would have to go hand-in-hand with changes in circuit architecture, while facing increasing challenges from the point of view of power consumption and dissipation as well as manufacturability.

Quantum effects are expected to creep in gradually as devices are further scaled. While the standard compact and drift-diffusion approaches can (and will) be calibrated to duplicate the device characteristics of deeply scaled devices, by artful tuning of empirical parameters, physical simulation approaches will need to carefully consider quantum tunneling and size quantization. Within a semi-classical framework, the main question is how to couple classical trajectories with quantum dominated regions. An interesting question is related to the lateral limits of scalability of conduction channels. It is found that it is possible to obtain well confined and *quasi-monomode* quantum wire channels at room temperature in silicon, when the gate contact is shape as a T, through an appropriate etching of the oxide. While an undoped substrate is unable to isolate a deep quantum wire channel and a high substrate doping may be undesirable, an acceptable confinement can be obtained if an undoped region containing the conduction channel is sandwiched between the oxide interface and a highly doped ground plane.

Due to granularity of the system, it is also important to resolve carefully coulombian effects. An interesting way to go beyond the standard numerical approaches for Poisson equation is to adopt a *meshless* formalism, where points, rather than meshes, are sprinkled in the device domain to generate the appropriate approximation to the differential operator by a finite element or collocation procedure. This is a natural way to treat Monte Carlo problems in 3-D since the collocation points can be associated to the particles themselves, generating a procedure which is automatically adaptive since the particle locations are tracked exactly at each simulation step. This approach will also be introduced elsewhere in these proceddings.

4. Conclusions

The standard simulation hierarchy has served well the device community and industry for practical devices that are approaching the 0.1 μm channel length limit. A gap exists between quantum models and the more advanced semi-classical approaches based on BTE. The main challenge for device simulation in the 0.1 μm - 0.01 μm range will be to bridge this gap since we expect that silicon devices in this space scale range will present a mixture of quantum and semi-classical quantum features. Since granularity of the material system will become very important, the development of atomistic level models will also be an essential component to well resolve issues of device reliability and carrier-interface interaction.

Acknowledgments - This work was partially supported by the Semiconductor Research Corporation, contract 98-SJ-406, and by the National Center for Computational Electronics (NCCE), National Science Foundation grant ECS 95-26127.

References

1. S. Selberherr, *Analysis and Simulation of Semiconductor Devices* Springer–Verlag (1994).
2. K.K. Thornber, *IEEE Electron Dev. Lett.* **3**, p. 69 (1982).
3. E.C. Kan, U. Ravaioli and T. Kerkhoven, *Solid–State Electron.* **34**, p. 995 (1991).
4. K. Sonoda, K. Taniguchi, and C. Hamaguchi, *IEEE Trans. Electron Dev.* **38**, p. 2662 (1991).
5. K. Blotekjaer, *IEEE Trans. Electron Dev.* **17**, p. 38, (1970).
6. M.A. Stettler, M.A. Alam, and M.S. Lundstrom, *IEEE Trans. Electron Dev.* **40**, p. 733 (1993).
7. P. Golinelli, L. Varani, and L. Reggiani, *Phys. Rev. Lett.* **77**, p. 1115 (1996).
8. J.J. Jerome and C.-W. Shu, *VLSI Design* **3**, p. 131 (1995).
9. R. Stratton, *IEEE Trans. Electron Dev.* **19**, p. 1288 (1972).
10. D. Chen, E.C. Kan, U. Ravaioli, C.-W. Shu, and R.W. Dutton, *IEEE Electron Dev. Lett.* **13**, p. 26 (1992).
11. H. Lin, N. Goldsman and I. Mayergoyz, *Solid-State Electron* **35**, p. 769, 1992.
12. M.A. Alam, M.A. Stettler, and M.S. Lundstrom *Solid–State Electron.* **36**, p. 263 (1993).
13. G. Zandler, A. Di Carlo, K. Kometer, P. Lugli, P. Vogl, and E. Gornik, *IEEE Electron Dev. Lett.* **14**, p. 77 (1993).
14. W. Fawcett, A.D. Boardman, and S. Swain, *J. Phys. Chem. Solids* **31**, p. 1963 (1970).
15. C. Jacoboni and P. Lugli, *The Monte Carlo method for Semiconductor Device Simulation*, Springer–Verlag (1989).
16. K. Hess, ed., *Monte Carlo Device Simulation: Full Band and Beyond*, Kluwer (1991)
17. M.V. Fischetti and S.E. Laux, *IEEE Trans. Electron Dev.* **38**, p. 634 (1991).
18. J. Jakumeit, U. Ravaioli, and K. Hess, *J. Appl. Phys.* **80**, p. 5061 (1996).
19. H.L. Grubin, T.R. Govindan, J.P. Kreskovsky, and M.A. Stroscio, *Solid–State Electron.* **36**, p. 1697 (1993).
20. U. Ravaioli, M.A. Osman, W. Pötz, N. Kluksdahl, and D.K. Ferry, *Physica B* **134**, p. 36 (1995).
21. W.R. Frensley, *Phys. Rev. B* **36**, p. 1570 (1987).
22. K.L. Jensen and F.A. Buot, *J. Appl. Phys.* **65**, p. 5248 (1989).
23. C.L. Gardner and C. Ringhofer, *Phys. Rev. E* **53**, p. 157 (1996).