Simulation of Ultra-Small Electronic Devices: The Classical-Quantum Transition Region

Bryan A. Biegel
MRJ Technology Solutions Inc.
IT Modeling and Simulation Group, Numerical Aerospace Simulation (NAS) Division
NASA Ames Research Center, MailStop T27A-1, Moffett Field, CA 94305-1000

Abstract

Concern is increasing about how quantum effects will impact electronic device operation as down-scaling continues along the SIA Roadmap through 2010. This document describes part of a new semiconductor device modeling (SDM) program at NAS to investigate these concerns by utilizing advanced NAS and third-party numerical computation software to rapidly implement and investigate electronic device models including quantum effects. This SDM project will investigate quantum effects in devices in the classical-quantum transition region, especially sub-0.1 μm MOSFETs. Specific tasks planned for this project include the use of quantum corrections to the classical drift-diffusion and hydrodynamic models of electron transport, and the use of nominally quantum models including significant scattering.

1: Introduction

Electronic devices have decreased in size and switching time by many orders of magnitude over the past three decades. In spite of this, the drift-diffusion (DD) model of electronic device operation is still used in nearly all line-of-business device simulations [1]. The reason is that the DD model has adequately explained or predicted the behavior of commercially important electronic devices through this rapid technology advancement. Because the DD model has maintained reasonable accuracy, the development of new (smaller) device generations using simple scaling laws and a few experimental iterations to optimize performance and yield has also worked very well through this advancement.

However, concern is increasing about how quantum effects will impact electronic device operation as progress continues along the SIA roadmap through 2010, which predicts that MOSFET gate lengths will then be only 70 nm [2]. The increasing significance of quantum effects in these ultra-small devices (see Figure 1), such as tunneling through gate oxides, inversion layer energy quantization, barrier proximity exclusion, and wave-like transport of electrons over short distances, has called into question the adequacy of the classical DD model (and other classical models) as down-scaling continues. Technology leaders

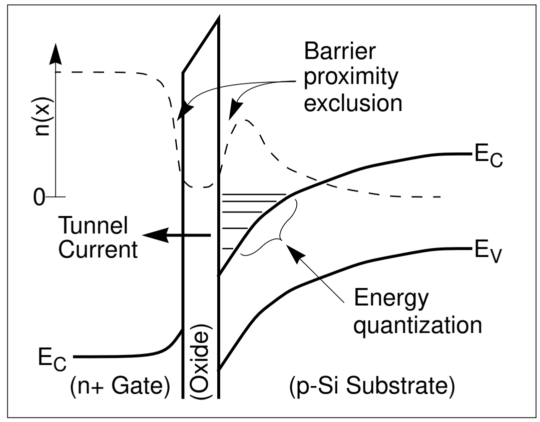

Figure 1: Quantum Effects in an n-MOSFET

now very much want to know how significantly parasitic quantum effects will degrade electronic device operation with each future device generation, how long these effects can be suppressed and by what means, and whether quantum effects might be used to actually improve device operation.

Experiment is not a suitable first line of attack in the investigation of these questions, since it can not view internal device operation or isolate particular physical effects, it has a very high (and increasing) cost, experimental structures and conditions are not precisely controllable, and turn-around time is very slow. Numerical simulation is a very viable alternative to experiment, since it does not suffer from these weaknesses. From the electronic device modeling community, two approaches are being followed in the attempt to answer questions about quantum effects in electronic devices: the addition of quantum corrections to conventional device models such as DD, and the development of fully quantum mechanical models for electronic devices. However, existing simulation tools currently can not provide the needed information for two reasons: 1) converting a new device model (including quantum effects) into functioning simulation software is very time-consuming, and 2) the required computational resources are immense. Both of these difficulties are directly addressed by this project, the goal of which is the rapid and accurate investigation of quantum effects in near-future electronic devices.

This project addresses the first issue by advancing the trend in software development away from writing huge single-purpose software packages, and towards the use of extensible software packages and generic modules, in order to rapidly implement and investigate new electronic device models including quantum effects. In particular, this project will draw upon the wide array of highly functional numerical simulation software and expert personnel that NAS has accumulated in its pursuit of advanced aerospace simulation and parallel numerical code development. Relevant NAS software resources include parallel equation solver routines for linear and non-linear systems, a 3-D Poisson equation solver, advanced dynamic gridding codes, computational fluid dynamics (CFD) codes, and data visualization codes. Use will also be made of appropriate third-party numerical computation tools and code modules.

With the formation of the SDM group, an additional resource is the combined knowledge within the group of many electronic device simulation approaches, including various classical models (drift-diffusion, hydrodynamic, and Monte-Carlo) and quantum models (Wigner function, Green’s function, transfer matrix, and density matrix). This knowledge and the associated codes will allow collaboration and code-sharing that will accelerate progress by each approach.

The objective of device modeling is to produce, as efficiently as possible, accurate predictions of device operation. Thus, the productive tasks of the device modeling physicist are developing accurate and computationally feasible models of the physics of interest, and analysis of simulation results. The goal of collaboration and code reuse in the NAS SDM Program is illustrated in Figure 2: to maximize the fraction of time spent on these “high-level” tasks, while minimizing the “low level” work of writing and debugging code. The traditional approach to electronic device modeling of spending years writing monolithic, “vertical” simulation codes (which only implement a single physical model) line-by-line from the ground up usually results in the opposite distribution of effort, and correspondingly slow progress.

The modular, collaborative approach is being increasingly used in software development, and will be extended as much as possible to device simulation tool development in this project. This goes well beyond the use of Netlib routines [3] to implement numerical functionality, since this project also seeks to leverage all of the NAS software resources listed previously.

The second reason for the inability of device simulation tools to answer questions about quantum effects in electronic devices is that accurate simulation of quantum

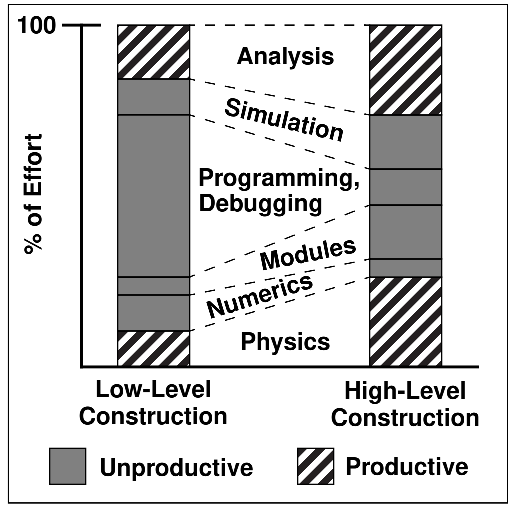

Figure 2: High-Level vs. Low-Level TCAD
The most productive TCAD research takes the shortest path between defining the physics of the system and analyzing simulation results. High-level TCAD tool development greatly reduces unproductive time, including derivation of the numerical model, programming, and debugging. Some unproductive tasks may increase somewhat, including researching and incorporating third-party code modules and computation time using more generic code.

effects in commercially important devices requires huge computational resources, both in terms of memory and CPU cycles [4]. Such large computations are most feasibly handled by vector or parallel supercomputers, since many numerical and graphical libraries which perform the vast majority of the computation can be (and have been) efficiently vectorized and parallelized.

This project addresses the second device modeling challenge by utilizing available computation resources at NAS, including Cray C90s and J90s vector machines, and IBM SP2 and SGI Origin 2000 parallel machines [5]. Many of the NAS software resources mentioned above are designed to take advantage of large parallel computation systems such as those at NAS. Note that the availability of very powerful computation hardware makes it reasonable to focus on rapid model implementation, even if resource usage is increased somewhat. This combination of rapidly-developed software and powerful hardware will bring previously infeasible computations such as 2-D and 3-D quantum simulations within reach. More importantly, it will finally allow the questions about quantum effects in current and future electronic devices to be answered.

The principal test device for this project will be the MOSFET, in which quantum effects are the highest concern, due to its dominance in electronics in the near future (the focal time-frame in this project) and to the wide range of quantum effects which are increasing in significance in this device (see Figure 1). Other projects in the NAS Semiconductor Device Modeling (SDM) Program [6] focus on longer term quantum simulation approaches and devices. One goal of the NAS SDM Program is cover the entire range from classical devices and physics to purely quantum computing. This spectrum is described in terms of particular (proposed or demonstrated) electronic devices in Table 1. Note from Table 1 that as quantum (wave transport) effects become more fundamental to device operation, classical effects (due to inelastic scattering) become more detrimental to proper device operation, and vice-versa.

The preceding paragraphs describe the motivation for this study of quantum effects in near-future electronic devices and the general approach that will be taken. The remainder of this document develops the specific tasks and plans for this project in more detail. The two specific tasks each pursue one of the approaches being taken by the electronic device modeling community to answer the industry's questions about quantum effects in these electronic devices: adding quantum corrections to classical electronic device models (Section 2), and using a fully quantum model (Section 3). Finally, Section 4 contains a discussion of issues which continue to shape our approach in this effort at NAS to develop a widely useful electronic device simulation capability in general, and a quantum effect and quantum device simulation capability in particular.

Table 1: Classical to Quantum Electronic Devices
Classical (inelastic scattering) effects become more detrimental, and quantum (wave transport) effects more essential, as devices transition from classical to quantum. Note: HET = hot electron transistor, QWLD = quantum well laser diode, QUIT = quantum interference transistor, SQUID = superconducting quantum interference device, SET = single electron transistor.

Electronic Device	Classical Effects	Quantum Effects
MOSFET, BJT	Dominant	Parasitic
MODFET, HET, QWLD	Dominant	Useful
RTD, RTT	Significant	Significant
QUIT, SQUID	Parasitic	Dominant
Quantum Dot, SET	Parasitic, Negligible	Dominant
Quantum Computer	Computa-tion-killer	Exclusive

2: Quantum Corrected Transport Models

As stated above, quantum effects such as those depicted in Figure 1 will increasingly affect electronic device operation as devices are scaled to smaller and smaller dimensions. As a result, electronic device models based on classical mechanics, such as drift-diffusion (DD), hydrodynamic (HD), and Boltzmann transport equation (BTE), are becoming progressively less accurate. At the same time, the increasing cost of experimental R&D with small devices makes it imperative to use device modeling to a greater extent in the advancement of electronics into the future. One way to reconcile these incompatible trends (maintain accuracy of device models in the face of increasing quantum effects) is to add some form of quantum correction to the classical models. The main strength of this approach is that it retains all of the accumulated experience and refinement that has made classical device models efficient, robust, and acceptably accurate for past and current electronic devices such as the MOSFET. The main weakness is that an independent approach is required to determine when the quantum correction is accurate, and under what conditions it too breaks down. This weakness is addressed by the task discussed in Section 3.

Many forms of quantum corrections to classical electronic device models have been proposed or implemented. These include MOSFET-specific quantum corrections [7, 8, 9, 10, 11, 12] and generic quantum corrections to the drift-diffusion [13], hydrodynamic [14, 15, 16], and Boltzmann transport equation [17] models.

It is impossible for a researcher to single-handedly implement and adequately investigate more than one or two of these quantum correction approaches. [This fact is further discussed in Section 4.] Unfortunately, choosing one or a few approaches from the wide array is not straightforward. The main trade-off between models is computational efficiency versus accuracy and generality. Different quantum correction approaches are undoubtedly preferable for various device sizes, device types, or simulation tasks. Given that the strength of NAS in large numerical computations is to be exploited in this project, the choice in this case will favor accuracy and generality, at the probable expense of higher computational cost. The two methods that appear to best fit this description are a 3-D density-gradient quantum correction to the DD model, and a 3-D quantum-corrected HD model. These will be the first quantum correction models implemented in this task.

Even having chosen just two quantum correction models to investigate, these models should be implemented as expeditiously as possible. This task will therefore use a general PDE solver called PROPHET [18] (and possibly similar tools) to quickly implement and investigate these models. Rapid model implementation is further discussed in Section 4.

The density-gradient quantum correction to the DD model will be investigated first in this task, since it should require the less time to implement and fewer computational resources. Indeed, implementation of the 3-D DG model in PROPHET is already underway. The DG model formalizes the quantum mechanical requirement that wavefunctions, and thus carrier densities, can not change abruptly versus position. For example, classically, carriers can reach very high densities directly against the MOS gate oxide in the inversion layer, and drop to zero just inside the oxide. Quantum mechanically, carrier wavefunctions are near zero in the oxide, and they must decrease smoothly towards zero in the neighboring inversion layer. The result is that the inversion charge is smoothed out and forced away from the gate oxide by some (classically unknown) distance, decreasing gate capacitance and thus MOSFET transconductance. The DG model can also model quantum tunneling.

The mathematical description of the DG model is a simple extension of the classical DD model. The classical DD model can be written:

$$\begin{aligned} \nabla \cdot (\varepsilon \nabla \psi) &= -\rho = -q(p - n + N_D^+ - N_A^-) \\ \frac{\partial n}{\partial t} &= \frac{\nabla \cdot \boldsymbol{J}_n}{q} = \nabla \cdot (-n\mu_n \nabla \psi + D_n \nabla n) \\ \frac{\partial p}{\partial t} &= -\frac{\nabla \cdot \boldsymbol{J}_p}{q} = \nabla \cdot (p\mu_p \nabla \psi + D_p \nabla p) \end{aligned} \quad , \qquad (1)$$

which equations are solved for electrostatic potential ψ, electron density n, and hole density p. Mathematically, the DG model [13] modifies the two continuity equations by adding a "quantum potential" (the Bohm potential [19]) to ψ:

$$\begin{aligned} \frac{\partial n}{\partial t} &= \nabla \cdot (-n\mu_n \nabla \psi_n^* + D_n \nabla n) \\ \frac{\partial p}{\partial t} &= \nabla \cdot (p\mu_p \nabla \psi_p^* + D_p \nabla p) \end{aligned} \quad , \qquad (2)$$

where

$$\begin{aligned} \psi_n^* &\equiv \psi + \frac{\hbar^2}{6m_n^* q}\left(\frac{\nabla^2 \sqrt{n}}{\sqrt{n}}\right) \\ \psi_p^* &\equiv \psi - \frac{\hbar^2}{6m_p^* q}\left(\frac{\nabla^2 \sqrt{p}}{\sqrt{p}}\right) \end{aligned} \quad . \qquad (3)$$

The DG model has only been implemented in 1-D, usually assuming an infinite oxide band gap. This task will study the DG model in 3-D, with a physically correct (finite) oxide band gap. In this way, gate oxide tunnel current can be investigated, along with other quantum effects in MOSFETs (see Figure 1) and other "classical" electronic devices.

After implementing and investigating the density-gradient model, the quantum hydrodynamic (QHD) model will be implemented in 3-D using PROPHET. The idea of the HD transport model (classical or quantum) is that, rather than resolve the momentum distribution of carriers exactly, the momentum distribution is assumed to be a mathematically simple modification of the equilibrium. Under this assumption, the distribution can be described by a few characteristic values. The equations for these values are derived by taking one or more moments of the Boltzmann (classical) or Wigner function (quantum) transport equation. The standard HD electronic device model uses three moments, with the resulting characteristic values being density, average velocity, and average energy.

We now describe the QHD model mathematically. Several forms of the HD and QHD transport equations have been proposed. For illustration purposes, we present one which has been written in both classical and quantum corrected form. In classical form, for a spatially-independent effective mass, and in 1-D, this HD model is [17]:

$$\begin{aligned} &\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}\left(\frac{\rho p}{m}\right) = 0 \\ &\frac{\partial(\rho p)}{\partial t} + \frac{\partial}{\partial x}\left(\frac{\rho p^2}{m}\right) + \rho\frac{\partial U}{\partial x} + \frac{\partial}{\partial x}(\rho k T) = \left(\frac{\partial(\rho p)}{\partial t}\right)_c \\ &\frac{\partial W}{\partial t} + \frac{\partial}{\partial x}\left(\frac{pW}{m}\right) + \frac{\partial}{\partial x}\left(\frac{\rho p k T}{m}\right) + \frac{\rho p}{m}\frac{\partial U}{\partial x} = \left(\frac{\partial W}{\partial t}\right)_c \\ &W = \frac{3\rho k T}{2} + \frac{\rho p^2}{2m} \end{aligned} \quad , \qquad (4)$$

which are solved for carrier density ρ, average momentum p, and average energy W. A full mathematical HD description including three moments each for electrons and holes requires seven equations (including the Poisson equation).

The quantum-corrected system of HD transport equations corresponding to (4), as derived from the Wigner function-corrected BTE [17], are:

$$\begin{aligned} &\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}\left(\frac{\rho p}{m}\right) = 0 \\ &\frac{\partial(\rho p)}{\partial t} + \frac{\partial}{\partial x}\left(\frac{\rho p^2}{m}\right) + \rho\frac{\partial}{\partial x}\left(U + \frac{Q}{3}\right) + \frac{\partial}{\partial x}(\rho k T) \\ &\qquad = \left(\frac{\partial(\rho p)}{\partial t}\right)_c \\ &\frac{\partial W}{\partial t} + \frac{\partial}{\partial x}\left(\frac{pW}{m}\right) + \frac{\partial}{\partial x}\left(\frac{\rho p k T}{m}\right) + \frac{\rho p}{m}\frac{\partial}{\partial x}\left(U + \frac{Q}{3}\right) \\ &\qquad - \left(\frac{\rho \hbar^2}{12m}\right)\frac{\partial}{\partial x}\left(\frac{1}{\rho}\frac{\partial \rho}{\partial x}\right)\frac{\partial}{\partial x}\left(\frac{p}{m}\right) = \left(\frac{\partial W}{\partial t}\right)_c \end{aligned} \quad , \qquad (5)$$

$$W = \frac{3\rho kT}{2} + \frac{\rho p^2}{2m} - \left(\frac{\rho\hbar^2}{24m}\right)\frac{\partial}{\partial x}\left(\frac{1}{\rho}\frac{\partial\rho}{\partial x}\right), \tag{6}$$

where Q is the Bohm quantum potential. This (or a similar) QHD model should give more accurate electronic device simulations than purely classical models, but the expense of the computation may be very high, and numerical robustness may suffer. Analysis of such expectations will be an important aspect of the investigation of quantum-corrected classical models in this task.

3: Quantum Models with Scattering

Even with the near-term device modeling focus of this project, more accurate models of quantum effects in electronic systems are needed to complement quantum correction models such as those described above. These more accurate models will be used to determine the accuracy and limitations of the quantum correction models, and possibly to derive more accurate and computationally efficient quantum correction models. The second specific task in this project, described in this section, will use a fully quantum model to accomplish these goals.

Figure 3 shows many of the quantum formulations that have been used for electronic device modeling. As with the quantum correction models, it is only possible for a single researcher to implement and adequately investigate one or two of these models, so a choice must be made among these formulations. Because classical devices inherently exhibit significant inelastic scattering, a quantum model which can efficiently include scattering must be chosen for this task. The Wigner function formulation (WFF) of quantum mechanics meets this requirement. Other crucial characteristics of the WFF which are important for accurate electronic device simulation include the ability to easily treat open boundaries, transient operation, and Poisson self-consistency. Therefore, the WFF will be used in this task. Because the WFF is a fully quantum model, macroscopic quantum devices (i.e., quantum devices which typically operate with significant scattering) such as the resonant tunneling diode (RTD) and resonant tunneling transistor (RTT) will be investigated along with conventional electronic devices.

The WFF was originally described [20] as a quantum correction to the BTE. Thus, the WFF also serves as a potential means of deriving more accurate and computationally efficient quantum correction models. In fact, both the DG and QHD models described in the previous section can be derived as simplifications of the Wigner function transport equation (WFTE) model [13, 15], in an analogous manner to derivations of the DD and HD models from the BTE [21, 22].

The mathematical description of the WFTE will be pre-

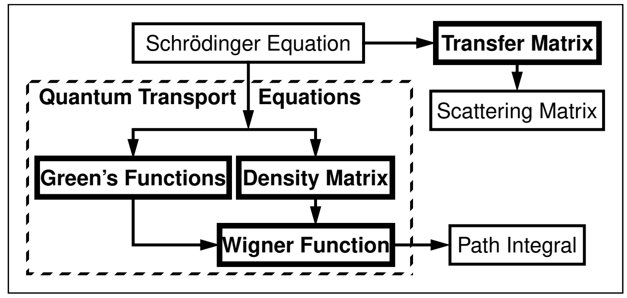

Figure 3: Partial Quantum Mechanics Family Tree
Relationships between quantum mechanics formulations relevant to the simulation of electronic devices are shown. Models implemented in existing NAS 1-D simulation software tools are shown in bold.

sented in an analogous manner to that of the DG and QHD models in the previous section. Specifically, the BTE will be shown first, and then the WFTE as the quantum analog. In 1-D, the BTE is:

$$\frac{\partial f_c}{\partial t} = -\underbrace{v\frac{\partial f_c}{\partial x}}_{\text{diffusion}} - \underbrace{\frac{F}{\hbar}\frac{\partial f_c}{\partial k}}_{\text{drift}} + \underbrace{\left(\frac{\partial f_c}{\partial t}\right)_{\text{coll}}}_{\text{scattering}}, \tag{7}$$

where $f_c(x, k, t)$ is the classical carrier distribution function and F is the force on the carriers. By comparison, the 1-D WFTE is:

$$\frac{\partial f_w}{\partial t} = -\underbrace{\frac{\hbar k}{m}\frac{\partial f_w}{\partial x}}_{\text{diffusion}} - \underbrace{\frac{1}{\hbar}\int\frac{dk'}{2\pi}V(x, k-k')f_w}_{\text{drift}} + \underbrace{\left(\frac{\partial f_w}{\partial t}\right)_{\text{coll}}}_{\text{scattering}}, \tag{8}$$

where $f_w(x, k, t)$ is the Wigner distribution function, $\hbar k/m = v$ is the carrier velocity, and $V(x, k)$ is called the non-local potential. Since the only difference between the BTE and WFTE is in the drift term, all of the quantum mechanics is contained in this term. Note that although this equation is 1-D in position, the domain is 2-D (position and momentum), with is the reason for the high computational cost of WFF simulations. As an example, Figure 4 shows the 1-D Wigner function of an RTD at high bias, showing a beam of electrons traveling into the right contact.

If computational cost and time were immaterial, one might always solve the BTE for classical device simulations, and the WFTE for quantum device simulations. However, computational costs *are* significant (in fact, overwhelming) in these computations. For example, it is just becoming feasible with huge computational resources to

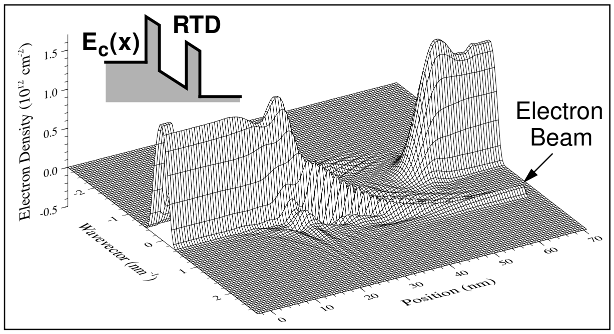

Figure 4: RTD Wigner Function at High Bias
The Wigner function shows the particle density versus position and momentum (wavenumber). Electrons are seen tunneling through the RTD and into the right contact. The inset shows the conduction band profile of the simulated RTD.

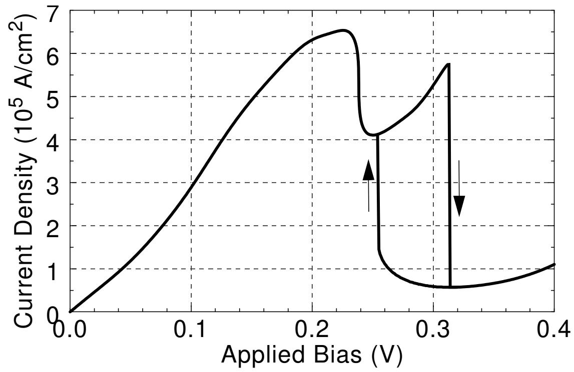

Figure 5: Intrinsic Hysteresis and Bistability
Self-consistent WF simulations of an RTD show hysteresis and bistability in the I-V characteristic.

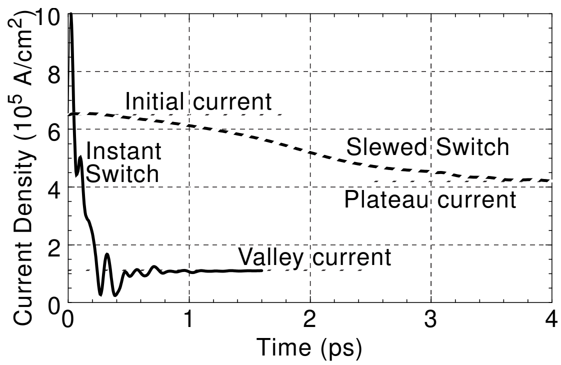

Figure 6: Slew-Rate Dependent Operation of RTD
Self-consistent WF simulations of an RTD switched into a bistable region show that the final operating current depends on the applied bias slew rate.

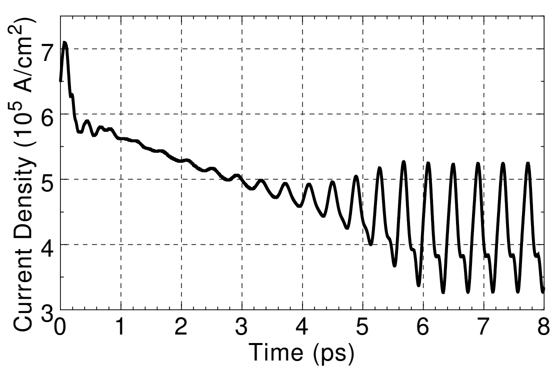

Figure 7: 2.5 THz Self-Oscillation in RTD
Self-consistent WF simulations of an RTD switched into the negative differential region of the plateau show high frequency self-oscillations.

simulate electronic devices with reasonable accuracy in 2-D with the WFTE (although this has never been implemented or attempted). To make the best use of available computational resources, the less costly (and usually less accurate) quantum correction methods should be used for exploratory simulations, while the WFTE should be used only in cases where greater accuracy is required (e.g., to test the quantum correction models), and where reduced dimensionality is acceptable.

To study the full WFTE for electronic device simulation, an existing quantum device simulation tool called SQUADS [23] was adopted. This tool includes 1-D simulation capability with both the Wigner function and transfer-matrix formulations. SQUADS has already produced new results in self-consistent [23] and transient [24] quantum device simulation. For example, the self-consistent RTD simulations in Figures 5-7 show hysteresis and bistability, slew-rate dependent operation, and 2.5 THz self-oscillations. Future investigations with the WFTE in SQUADS are planned into the use of very fine momentum gridding, new transport equation discretizations, and more detailed scattering models.

Based on knowledge acquired from 1-D electronic device simulations, a 2-D Wigner function code will eventually be developed. As stated above, 2-D quantum device simulation of this accuracy (including open boundaries, scattering, self-consistency, and transient operation) has never been attempted due to the high computational cost. The computational requirements for such 2-D simulations will be several orders of magnitude higher than for the 1-D case, requiring that more sophisticated solution methods be employed. Where appropriate, NAS expertise and numerical code (which has successfully solved this scale of CFD simulation) will be utilized. 2-D simulation is the minimum necessary for modeling the operation of the MOSFET, which device has been chosen as the focus of this project. However, macroscopic quantum devices such the RTD and RTT will continue to be investigated for the longer-term.

4: Discussion

4.1: The TCAD Challenge

As electronic device operation becomes increasingly complicated by quantum and small-geometry effects, and experiment-based device research becomes more expensive and slower, device simulation should offer some help to maintain the high rate of semiconductor technology advancement into the future. However, even after over 30 years of effort, device modeling tools are currently far too rudimentary to provide much help. For TCAD tools to meet industry's requirements, and provide measurable assistance to electronics advancement in the future, they must:

- accurately handle much more complicated physics, device structures, material systems, and simulation modes,
- accommodate much larger computations with more robustness and faster execution,
- exploit more advanced hardware architectures,
- permit plug-and-play user selection of the optimal physical models, numerical methods, and solvers at run-time, and eventually automate this selection,
- have an intuitive graphical interface, to decrease training cost and broaden the user base,
- have high-quality graphical output (1-D, 2-D, 3-D, steady-state or video) of any desired quantity, and
- greatly simplify interaction between simulation tools (process, device, circuit, etc.).

Obviously, producing TCAD tools with enough sophistication to have a significant impact on the rate of progress will require a tremendous investment of time and effort. Historically the necessary convergence of resources has been lacking, device modeling tool capabilities have lagged production by at least a device generation, and practical application of device modeling has been limited. Allowing this trend to continue is very dangerous. An inefficient device simulation effort will ensure that this endeavor remains largely irrelevant to industry's needs, resulting in a drastic decrease in the pace of technology advancement, with resulting deleterious effects on economies and standards of living [25]. To address this emerging crisis, the third and final task in this project will be to help define and initiate a much more productive device modeling development effort.

The first step in this task is to understand in sufficient detail why device simulator advancement has been so slow. The fundamental challenges were described in Section 1: the difficulty of converting a device model into a useful simulation package, and the expense of acquiring sufficient computation hardware to accurately simulate devices of relevance. In the following subsections, these issues are considered in more detail, and long-term solutions are described for each. These build on the partial solutions that are being taken by the NAS SDM group, as discussed in Section 1.

4.2: The Global TCAD Framework

The first challenge with the current electronic device simulator development process is to more effectively bring together the immense software development resources necessary to convert a physical device model into a widely useful simulation package. The creation of such software tools requires expertise in device physics, numerical methods, general and computational programming, linear and non-linear system solution algorithms, graphical interfaces, and graphical visualization. In the past, collaboration between device modeling researchers has been rare, so that every device modeling researcher had to either find appropriate software to provide these functions, or to develop the expertise in each field to create the necessary software. Unfortunately, the available software in any of these areas rarely meet all of the requirements for use in a device simulator. Software with desired functionality is often difficult to locate, poorly documented and supported, too inflexible, too expensive to justify its use, or has no API (application programming interface) at all. As a result, most simulation researchers have developed a mindset where essentially all code for their TCAD tool must be developed by them and from scratch.

As discussed in Section 1, the partial solution of the NAS SDM group to this self-limiting mindset is to utilize appropriate NAS software and collaborate with NAS simulation experts to the greatest extent possible. Based on this, the general solution for the entire TCAD community is to institute a TCAD development environment which enables and encourages collaboration and code-sharing worldwide. This environment would allow a TCAD researcher to contribute to the TCAD development effort in a very focused way, such as physical model development or numerical methods, and to easily use the complementary contributions of others. This environment will be called the Global TCAD Framework (GTF).

The main function of the GTF from a TCAD developer's point of view would be to provide a standard, dependable interface into which code functionality (numerical computation modules, graphical output modules, etc.) could be plugged. The GTF would provide services such as database access (for retrieval of physical models, numerical algorithms, material data, etc.), memory management, file I/O, facilities for tool interaction, and a basic GUI. It would also provide templates for describing physical quantities, physical structures, materials, activities performed (e.g., process steps or device tests), and so on.

The foregoing description of the GTF shows how it will conquer the programing complexity challenge by dividing the implementation of a single physical model into many parts. However, many physical models need to be implemented in the GTF. These include drift-diffusion, energy-balance, hydrodynamic, Boltzmann transport equation, Monte-Carlo, all quantum correction models (see Section 2), all quantum models (see Section 3), all optoelectronic models, and all compatible combinations of these. Each of these models has some range of usefulness, depending on its computational cost and accuracy for various device types and sizes.

Because so many models need to be implemented in the GTF, the physical model developer has one further requirement of the GTF: the physical model definition must be separate and independent (to the extent possible) from all other code functionality. New physical models can then be specified at run-time (usually as a set of PDEs), and solved using "generic" discretization, simulation, and visualization code. From the physical model developer's perspective, the ideal device simulation tool requires only the specification of the physical model and a device structure to test the model, as depicted in Figure 8. In this way, the model developer need not be an expert in programming, numerical methods, solving systems of equations, or any other "generic" functionality of simulation software.

A few emerging TCAD tools, particularly PROPHET [18] and ALAMODE [26], have demonstrated that the physical model can be specified at run-time, and that computational cost need not increase unacceptably with generic computation modules. As pointed out in Figure 2, computation is usually a small part of the TCAD tool development process, and this approach dramatically decreases several other unproductive tasks. For the implementation of many physical models, the separation of model from code not only increases the modularity (and thus collaboration potential) of TCAD tool development, it also reduces the redundancy of the effort, since functionality is shared by many different models.

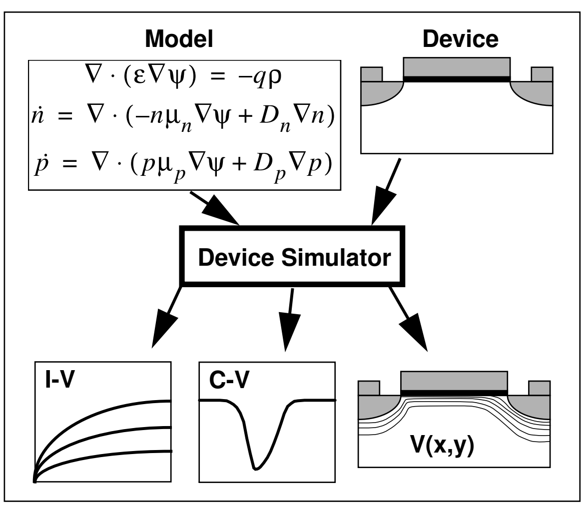

Figure 8: Model Developer's Ideal Simulator
Ideally, a model developer should only have to specify the physical model and a device in order to investigate a new model. Traditionally, each model developer must program all other simulator functionality as well.

To get a better idea of what the GTF should be like, Table 2 lists several existing software packages that provide some of the features of the GTF. For example, Mathematica and Maple V allow a model (mathematical equations) to be specified at run-time, and the appropriate solution methods are selected automatically. However, these packages can not handle the complex geometries of an electronic device, nor the large problem sizes of interest. Commercial TCAD tools such as those from TMA [27] and Silvaco [28] have the opposite strengths and weaknesses. New TCAD tools like PROPHET [18] and ALAMODE [26] allow new models to be specified independent of the code, and can handle large computations and complex device structures, but they have inadequate graphical capabilities and interactions between tools (e.g., process and device simulations).

To conclude this description of the GTF, a few more points should be made. As stated above, the current mindset of TCAD tool developers is one of developing (or re-inventing all over the world) huge, incompatible, "vertical" (i.e., which only solve one device model in a fixed way) simulation codes. Perhaps the biggest challenge of a GTF will be to demonstrate that there is at least as much work to be done and recognition to be gained in contributing highly tuned yet widely applicable software modules to a global TCAD development effort. Of course, by coordinating and combining the efforts of many researchers, the GTF will certainly have a greater impact on the future of electronic technology than current TCAD research. Realization of the GTF would have the following specific benefits:

- reduce duplication of development effort, since code of a given functionality would need to be created only once for all developers and users,
- produce faster TCAD tool progress, with developer and user access to best-of-breed code in each functionality category,
- make collaboration the rule, not the exception,
- make simulation *results* (not code) the goal,
- make the interaction of simulation tools, the use of a hierarchy of models in a single computation, the comparison of accuracy and efficiency of different physical models, and the implementation of other high level functionality more feasible,
- dramatically reduce the initial time investment in TCAD research to begin making contributions, and
- make TCAD tools more visible, accessible, stan-

Table 2: Features of GTF vs. Existing Software

The features of various existing software are compared to the proposed GTF. Software packages are: M = Mathematica/Maple V, P = PROPHET, A = ALAMODE, C = commercial packages (TMA, Silvaco, etc.), G = GTF. Functionality features are: GUI interface, modular upgradeability, high-quality graphical output, physical model independent from numerics, auto selection of numerical solution methods, complex simulation domains (devices), large computation capability, interaction between TCAD tools, and a framework for collaboration. "--" = poor.

Feature	Software Package				
	M	P	A	C	G
GUI	Y	N	N	Y	Y
Modularity	N	Y	Y	N	Y
Graphic Out	Y	--	--	Y	Y
Model Indep.	Y	Y	Y	N	Y
AutoNumeric	Y	N	N	N	Y
Devices	N	Y	Y	Y	Y
Large Comp.	N	Y	Y	Y	Y
Interaction	N	--	N	Y	Y
Framework	N	--	N	N	Y

dardized, and highly functional for users, thereby helping to maintain the rate of technology progress.

4.3: The Information Power Grid

The second challenge with the current electronic device simulator development process is to provide inexpensive and transparent access to sufficient computation resources for useful TCAD work. The current approach of providing expensive and difficult access for privileged researchers at a few supercomputing centers is not compatible with the collaborative development and use of the GTF. Scarce and expensive compute resources necessitate compromises in the physical model, numerical implementation, and simulation execution. Ideally, any TCAD researcher should be able to access, at a reasonable cost, the computational resources needed. The productivity of the best and brightest researchers should not be shackled by their physical location or the stature of their organization.

The NAS SDM group's partial solution (Section 1) to the scarce hardware challenge is to utilize NAS and NAS-connected computational hardware to aggregate the necessary compute power. The general solution for the global TCAD community is to make the GTF an application in the Information Power Grid (IPG), a NASA initiative [29] to achieve global compute resource pooling. Currently, research organizations are largely isolated, and so must supply all of their computational needs. The result is high cost and low utilization. The IPG, like the analogous electric power grid, allows users to transparently, inexpensively, and universally purchase needed power from a computation grid, or sell excess capacity to the grid. By increasing utilization and providing universal access, the IPG will make extremely compute-intensive applications like the GTF accessible to virtually all researchers.

The IPG is essentially an intelligent, highly-scalable network operating system. Its main function is to dynamically manage tasks to maximize utilization of the available computational resources (regardless of machine type, number, and location). To meet its goals of inexpensive, universal, and transparent computational resource access, several characteristics of the IPG are apparent. It must be open (platform-independent), so that it can be implemented on any existing or future hardware and operating system. Since applications are run on hardware which is dynamically selected by the IPG, applications must be compiled just before execution, similar to the Java model. In fact, the Web browser interface is a perfect model for the IPG, making its use transparent and universal. Each researcher would not be required to purchase and install very complex applications such as the GTF in order to use them. They would simply interact with the graphical interface to the application, while the IPG managed the myriad computational resources in the background.

Many projects are underway around the world to create IPG-like functionality. Some of the more high-profile efforts include those at Sun [30], JavaSoft, [31], Lucent Technologies [32], IBM [33], and Microsoft [34]. One project, the Purdue University Network Computing Hub [35], provides some of the proposed IPG functionality and benefits, and includes a TCAD-specific "laboratory" called the Semiconductor Simulation Hub [36]. This project does not solve most of the challenges that developers face in creating new and more sophisticated TCAD tools, as discussed in Sections 1 and 4.1.

The IPG is much more advanced in its development, and more certain in its eventual success, than the GTF. Indeed, we already use remote servers and supercomputers through the Internet to access Java applications, run search engines, play interactive games, and so on. The obvious utility of network computing will, without question, provide the necessary driving force to make the IPG happen. The main question to the TCAD community is whether something like the GTF will be created as a first-class IPG application, or will TCAD tool development continue to be disorganized, redundant, and inefficient to the point of irrelevance?

4.4: NASA Interests in Device Modeling

NASA interests in helping to advance the state-of-the-art in semiconductor device modeling (and thus the devices themselves) are obvious. Current and future NASA mission rely critically on advanced information technology, as well as advanced analog electronics and optoelectronics (including infrared detectors, millimeter and submillimeter wave sensors, UV and x-ray CCDs, photonic devices, optoelectronic integrated circuits, micromagnetic devices, and electronic neural networks [37]). The development of all of these applications would benefit greatly from an accurate and efficient device simulation capability.

NASA Ames Research Center (ARC) specifically has a compelling interest in supporting the effort to create a successful TCAD framework. ARC has been designated the Center of Excellence for Information Technology. As such, ARC will facilitate NASA's effort to create more powerful computation hardware, as exemplified by the petaflops and Information Power Grid initiatives. The GTF is an ideal application for the IPG, since it depends on both distributed collaboration and powerful computation. Device modeling and the GTF are also important to NASA because they will help to continue the current high rate of technology advancement into the 21st century, making the NASA petaflops computing initiative a realistic goal.

The positioning of SDM Program within NAS is quite appropriate, since NAS has a number of unique resources to bring to bear. NAS supercomputing and parallel computation hardware, advanced numerical computation software, and numerical and parallel computation experts will allow meaningful in-house prototyping of both the IPG and GTF, as well as making direct contributions to the device modeling projects. Finally, the significance of the GTF project to the productivity of the NAS SDM Group is difficult to overstate. A highly functional GTF prototyped at NAS will translate directly into higher productivity of the entire NAS SDM group.

5: Summary

This document has described plans for the three tasks in one of the projects in the recently initiated Semiconductor Device Modeling Program at NAS. The first task is the investigation of quantum corrections to the classical drift-diffusion and hydrodynamic transport models, using the PDE solver PROPHET for rapid implementation. The second task is the investigation of the Wigner function and transfer matrix quantum transport models, building on an existing quantum device simulation tool called SQUADS. The third task is the exploration of a TCAD framework as an application for the Information Power Grid. Success in this final task would greatly accelerate progress in semiconductor device modeling, and thereby help to maintain the rapid advance of information technology.

Acknowledgements

The author would like to thank M.P. Anantram and Cun-Zheng Ning for reviewing this manuscript. This work was supported by MRJ, Inc. under CALSPAN contract WTD12197 and NASA contract NAS2-14303.

[1] E. Wang. "Advanced transport model development requirements in industrial environment for deep submicron technologies." In *Second NASA Device Modeling Workshop*, NASA Ames Research Center, Moffett Field, CA 94035-1000, Aug. 7-8, 1997.

[2] The SIA Roadmap is available at http://www.sematech.org/public/roadmap/index.htm.

[3] The NetLib home page is at http://www.netlib.org.

[4] For example, useful modeling of MOSFET operation requires 2-D simulation at a minimum. Using the Wigner function model (see Section 3) in 2-D, where 100 points in each of the 4 dimensions (x, y, kx, ky) is minimally sufficient. Even storing the 2-D Wigner function would require 800 MB of space. The most memory-efficient solution approach (e.g. Lanczos) would require at least 3.2 TB of working space, 1 GFLOP of computation per iteration, and perhaps 1000 iterations per time step or bias point.

[5] Descriptions of some of the NAS supercomputing resources are at http://www.nas.nasa.gov/Technology/Supercomputing/.

[6] The NAS Semiconductor Device Modeling Program is described at http://www.nas.nasa.gov/ biegel/sdm_group.html.

[7] W. Hansch, T. Vogelsang, R. Kircher, and M. Orlowski. "Carrier transport near the Si/SiO2 interface of a MOSFET." *Solid State Electronics*, 32(10):839–849, 1989.

[8] Y. Ohkura. "Quantum effects in Si n-MOS inversion layer at high substrate concentration." *Solid State Electronics*, 33(12):1581–1585, 1989.

[9] M. J. van Dort, P. H. Woerlee, A. J. Walker, C. A. H. Juffermans, and H. Lifka. "Influence of high substrate doping levels on the threshold voltage and the mobility of deep-submicrometer MOSFET's." *IEEE Transactions on Electron Devices*, 39(4):932–938,

1992.

[10] N. D. Arora, R. Rios, and D. A. Antoniadis. "Capacitance modeling for deep submicron thin gate oxide MOSFETs." In *European Solid State Devices Research Conference (ESSDERC)*, pages 569–572, 1995.

[11] R. Rios, N. D. Arora, C.-L. Huang, N. Khalil, J. Faricelli, and L. Gruber. "A physical compact MOSFET model, including quantum mechanical effects, for statistical circuit design applications." In *International Electron Devices Meeting (IEDM) Technical Digest*, pages 937–940. IEEE, 1995.

[12] P. Vande Voorde, P. B. Griffin, Z. Yu, S.-Y. Oh, and R. W. Dutton. "Accurate doping profile determination using TED/QM models extensible to sub-quarter micron nMOSFETs." In *International Electron Devices Meeting (IEDM) Technical Digest*, pages 811–814. IEEE, 1996.

[13] M. G. Ancona. "Quantum correction to the equation of state of an electron gas in a semiconductor." *Physical Review B*, 39(13):9536–9540, 1989.

[14] D. L. Woolard, M. A. Stroscio, M. A. Littlejohn, R. J. Trew, and H. L. Grubin. "A new nonparabolic hydrodynamic model with quantum corrections." In K. Hess, J. Leburton, and U. Ravaioli, editors, *Computational Electronics: Semiconductor Transport and Device Simulation*, pages 59–62. Kluwer Academic Publishers, Boston, 1991.

[15] J.-R. Zhou and D. K. Ferry. "Simulation of ultra-small GaAs MESFET using quantum moment equations." *IEEE Transactions on Electron Devices*, 39(3):473–478, 1992.

[16] J. P. Kreskovsky and H. L. Grubin. "Electron transport using the quantum corrected hydrodynamic equations." *VLSI Design*, 3(2):179–200, 1995.

[17] H. L. Grubin and J. P. Kreskovsky. "Quantum moment balance equations and resonant tunnelling structures." *Solid State Electronics*, 32(12):1071–1075, 1989.

[18] The PROPHET home page is currently at http://gloworm.stanford.edu/ conor/prophet/guide.html.

[19] H. L. Grubin, J. P. Kreskovsky, T. R. Govindan, and D. K. Ferry. "Uses of the quantum potential in modelling hot-carrier semiconductor devices." *Semiconductor Science and Technology*, 9:855–858, 1994.

[20] E. Wigner. "On the quantum corrections for thermodynamic equilibrium." *Physical Review*, 40:749–759, June 1, 1932.

[21] M. R. Pinto. *Comprehensive semiconductor device simulation for silicon ULSI*. PhD thesis, Stanford University, Aug. 1990. p. 23.

[22] D. L. Woolard, R. J. Trew, and M. A. Littlejohn. "Hydrodynamic hot-electron transport model with Monte Carlo-generated transport parameters." *Solid State Electronics*, 31(3/4):571–574, 1988.

[23] B. A. Biegel and J. D. Plummer. "Comparison of self-consistency iteration options for the Wigner function method of quantum device simulation." *Physical Review B*, 54(11):8070–8082, 1996.

[24] B. A. Biegel and J. D. Plummer. "Applied bias slewing in transient Wigner function simulation of resonant tunneling diodes." *IEEE Transactions on Electron Devices*, 44(5):733–737, 1997.

[25] An analysis of the importance of TCAD to the future of information technology can be found at http://www.nas.nasa.gov/ biegel/pubs/nas_sdm_why.ps.

[26] The ALAMODE home page is at http://gloworm.stanford.edu/tcad/programs/alamode.html.

[27] The TMA home page is at http://www.tmai.com.

[28] The Silvaco home page is at http://www.silvaco.com.

[29] A list of links about Information Power Grid-related work is at http://www.nas.nasa.gov/Groups/Tools/IPG/.

[30] The home page for Sun's Spring distributed OS is at http://www.sun.com/tech/projects/spring/index.html.

[31] The home page for JavaSpaces, JavaSoft's effort to create a distributed Java-based OS, is at http://chatsubo.javasoft.com/javaspaces/.

[32] The home page for Inferno, Lucent Technologies's apparent successor to Plan 9, is at http://plan9.bell-labs.com/inferno/.

[33] IBM's Open Blueprint Network Computing Framework initiative is described at http://www.software.ibm.com/openblue/papers/obncfr2a.htm.

[34] Microsoft's position paper on their Millennium distributed OS initiative is at http://131.107.1.182:80/research/os/Millennium/mgoals.html.

[35] The Purdue University Network Computing Hub (PUNCH) home page is at http://punch.ecn.purdue.edu:8000/.

[36] The Semiconductor Simulation Hub home page is at http://punch.ecn.purdue.edu:8000/SimHub/.

[37] From the JPL Microdevices Laboratory home page at http://mishkin.jpl.nasa.gov/MDL.html.