Multi-dimensional Quantum Effect Simulation Using a Density-Gradient Model and Script-Level Programming Techniques

C.S. Rafferty(a), B. Biegel(b), Z. Yu(c), M.G. Ancona(d), J. Bude(a), R.W. Dutton(c)

May 22, 1998

(a) Bell Laboratories, Lucent Technologies
(b) NASA Ames Research Center
(c) Stanford University
(d) Naval Research Laboratory

ABSTRACT

A density-gradient (DG) model is used to calculate quantum-mechanical corrections to classical carrier transport in MOS inversion/accumulation layers. The model is compared to measured data and to a fully self-consistent coupled Schrödinger and Poisson equation (SCSP) solver. Good agreement is demonstrated for MOS capacitors with gate oxide as thin as 21Å. It is then applied to study carrier distribution in ultra short MOSFETs with surface roughness. This work represents the first implementation of the DG formulation on multi-dimensional unstructured meshes. It was enabled by a powerful scripting approach which provides an easy-to-use and flexible framework for solving the fourth-order PDEs of the DG model.

Introduction

The industry trend towards ever-thinner gate dielectrics makes quantum effects in both bulk and poly silicon an important limiter of device drive. Such effects have been modeled using either a rigourous coupled solution of the Schrodinger and Poisson equations, or by building analytical models based on such solutions. The density gradient method is a new approach to modeling quantum effects. At the cost of adding high order terms to the drift-diffusion model, it allows one to capture some of the physics of quantum confinement and exclusion in a non-empirical way.

A major challenge faced in the development of new simulation methods is the translation of advanced physics into useable tools for technology design. The DG method modifies the governing PDEs in a fundamental way and, as a result, its inclusion in a standard, highly-specialized device simulation package would inevitably require major revisions.

This work decsribes the implementation of a DG model in the script-level PDE solver PROPHET. The discussion shows how the model can be readily written in terms of available operators and applied to unstructured, multi-dimensional meshes. This allows the examination of quantum effects in nonplanar structures. Comparison with a rigorous 1D SCSP calculation shows excellent agreement for MOS capacitors with gate oxide as thin as 21Å. The DG model is then applied to calculate the 2D transport of electrons in a MOSFET inversion layer, and to examine surface roughness effects on the inversion layer charge density.

The approach demonstrates the power of the scripting paradigm for rapidly prototyping new models, testing choice of variables for system solution, and scaling up from simple model problems to full-scale multi-dimensional simulations where efficiency and stability are paramount.

Equation System for Density Gradient Model

The density gradient model [1] is an approximate approach for introducing QM corrections into a macroscopic electron transport description like the drift-diffusion approximation. It has the advantage of being easily extended to multi-dimensional problems and to non-equilibrium situations. The QM corrections, which are introduced by making the electron gas equation of state density-gradient dependent, introduce an extra term into the carrier flux expression:

$$\mathbf{F}_n = n\mu_n \nabla\psi - D_n \nabla n + 2\mu_n \nabla \left(b_n \frac{\nabla^2 \sqrt{n}}{\sqrt{n}} \right) \tag{1}$$

where $b_n = \hbar^2/(12qm_n^*)$ and all other symbols have the conventional meanings. The new term makes the electron continuity equation a fourth order PDE. This fact, together with the singularly-perturbed nature of the equations, suggests that a conventional discretization would require special numerical treatment. We propose to re-write the equations in terms of the generalized quasi-Fermi potential, ϕ_n, as follows:

$$\mathbf{F}_n = n\mu_n \nabla\phi_n \tag{2}$$

Dirichlet and Neumann boundary conditions for potential ψ and ϕ_n can be specified as usual on the ohmic and dielectric interfaces of the device. Dirichlet boundary conditions are applied to the electrons and holes at ohmic and dielectric interfaces, using charge-neutral, and vanishingly small values, respectively. Hence the final system of equations to be solved for device simulation with QM corrections is:

$$-\nabla \cdot (\epsilon \nabla \psi) = q(p - n + N_D^+ - N_A^-) \tag{3}$$

$$\nabla \cdot (n\mu_n \nabla\phi_n) = g - r \tag{4}$$

$$\nabla \cdot (-p\mu_p \nabla\phi_p) = g - r \tag{5}$$

$$\phi_n - \psi + \frac{kT}{q} \ln \frac{n}{n_i} = 2b_n \left(\frac{\nabla^2 \sqrt{n}}{\sqrt{n}} \right) \tag{6}$$

$$\phi_p - \psi - \frac{kT}{q} \ln \frac{p}{n_i} = -2b_p \left(\frac{\nabla^2 \sqrt{p}}{\sqrt{p}} \right) \tag{7}$$

where ψ, ϕ_n, ϕ_p, $\sqrt{n}$, and $\sqrt{p}$ are fundamental variables, g and r are generation and recombination rates, respectively, and only the steady state is considered.

Script-level Programming

Script-level simulation exploits the fact that most PDEs used in TCAD simulation are generalized conservation equations. Because of this and because of the requirements of coordinate-invariance, such equations contain only a small set of mathematical operators (such as the divergence) and physical terms (such as carrier flux). These operators and terms can be provided in a discretization library to allow their combination into a problem-specific model. The PROPHET simulator has finite difference and finite element discretizations of the common operators in 1D, 2D and 3D. It applies Newton's method for the solution of the nonlinear discretized PDEs, using an iterative/direct sparse matrix package to solve the resulting linear system.

To give a flavor of the scripting language, equations (3), (4) and (6) are listed below. Equation (6) has been multiplied through by the square root to make better use of existing operators.

```
system name=dg2_qf
+ sysvars=psi,phi_n,phi_p,sqrt_n,sqrt_p
+   nterm=14
+   term0=box_div.lapflux(psi|psi)@{silicon,oxide,poly}
+   term1=q*nodal.add(holes,netdope|psi)@{silicon,poly}
+   term2=-q*nodal.copy(electrons|psi)@{silicon,poly}
+   term3=-1*box_div.drift_qf(electrons,phi_n|phi_n)
+   @{silicon,poly}
+   term4=nodal.prod(sqrt_n,psi_qn|sqrt_n)@{silicon,poly}
+   term5=-1*box_div.lapflux(sqrt_n|sqrt_n)@{silicon,poly}
...
+ term10=dirichlet.device_dirichlet_dg(netdope|sqrt_n,sqrt_p)
+   @{silicon/back,poly/gate}
+ term11=dirichlet.default_dirichlet(0|sqrt_n)
+   @{silicon/oxide,poly/oxide}
+ term12=dirichlet.device_dirichlet_qf(netdope|psi,phi_n,phi_p)
+   @{silicon/back,silicon/source,silicon/drain,poly/gate}
+ term13=constraint.continuity(psi|psi)
+   @{silicon/oxide,poly/oxide}
+   nfunc=10
+ tmpvars=electrons,t11,t12,t13,psi_qn,holes,t21,t22,t23,psi_qp
+   func0=prod(sqrt_n,sqrt_n|electrons)@{silicon,poly}
+   func1=rni*copy(electrons|t11)@{silicon,poly}
+   func2=-kTq*log(t11|t12)@{silicon,poly}
+   func3=sub(t12,phi_n|t13)@{silicon,poly}
+   func4=add(t13,psi|psi_qn)@{silicon,poly}
...
```

Each operator is composed of a "geometrical" and a "physical" part, reads a set of input fields, and writes on a set of output equations. Bulk and surface terms are handled alike, with the regions or interfaces to which the term is being applied listed at the end of each term. Fields can be generated recursively by taking algebraic combinations and functions of other fields; the derivatives required for Newton's method are automatically generated.

Results - 1D capacitance

To test model validity, the DG model has been applied to a MOS capacitor from a state-of-the-art 0.18μm CMOS technology, with 31Å gate oxide thickness and non-uniform channel doping approaching $10^{18} cm^{-3}$. The simulation results using the DG model are compared to those from a SCSP solver described elsewhere [2] and to the measured data [3].

For the purposes of comparison, electron and hole masses of 0.916 m_e and 0.49 m_e were assumed in both the DG model and the SCSP model. Zero conduction band (CB) offset was applied between degenerate polysilicon and bulk silicon. The agreement between the DG model and the more physically rigourous SCSP approach is rather good using these identical masses, while the fit to data is only moderately good.

To improve the fit to data, the SCSP model used transverse/longitudinal electron masses of 0.19/0.916 m_e, while for the holes the effective masses were tuned to optimize parabolic band fit to a full band-mixing calculation between heavy and light holes. The DG model used an electron mass of $0.19 m_e$ while retaining the hole mass of 0.49 m_e in calculating b_n and b_p. As is shown in Fig. 2, the agreement between the two models remains good in accumulation and inversion. The agreement between both simulations to data is now more satisfactory, with the DG actually being a little closer to data in accumulation and inversion. Around threshold, the DG curve is shifted -20 mV relative to SCSP, which is in turn shifted -83mV relative to the experimental data. The origin of these shifts are not yet fully understood; they be due to genuine model disagreement, to remaining small differences in the SCSP and DG simulations, or to ignoring workfunction shifts due to degenerate band-tailing and grain size.

A number of capacitors with oxide thicknesses from 21Å to 79.5Å were compared with the DG model in Figure 3. As a measure of the fit at each thickness, the accumulation capacitance at -2.5V was expressed in terms of the "effective thickness", that is, the thickness t_{eff} which would reproduce the measured or simulated capacitance according to $C/A = \epsilon_r \epsilon_0 / t_{eff}$. The effective thickness is always greater than the measured thickness due to the vanishing electron wavefunctions at the oxide interface and the displacement of the electron peak to the bulk. As can be seen, the effective thicknesses (or equivalently, the accumulation capacitance) is in excellent agreement between experiment and density gradient model over the entire range of oxide thicknesses, using a single value of b over all thicknesses.

Results - Current Flow in 2D

Figures 4 and 5 show subthreshold and linear current in a MOSFET with $L_{gate} = 300$Å . For this calculation, the mobility is held constant in order to examine effects arising solely from the different electron distributions. The DG model shows a smaller subthreshold slope than the classical simulation. The origin of the difference can be seen in the electron contours in Figure 6 and 7; the channel thickness is larger due to quantum effects, leading to more charge sharing. At high gate bias the DG current crosses over the classical current. This effect, which is important only at very short gate length, is due to strong screening in the classical model. The gate enhances the conductance of the source/drain extensions by its fringing field, but strong screening due to channel electrons against the gate surface reduce this effect in the classical model as compared to the DG model.

Results - Surface Roughness

Gate oxides are imperfect. The flat surface shown in Figure 6 is an idealization of the true interface. The density gradient model was used to explore the effects of surface roughness on the electron distribution below the gate oxide. A less ideal structure is shown in Figure 8, where a 40Å oxide is grown on a surface with 4Å peak-to-peak roughness and 20Å wavelength. The thickness of the oxide is assumed for this calculation to be constant; other assumptions are possible. The calculation was done on a single period of the structure using an unstructured triangular mesh to represent the curved shapes.

Figures 9 and 10 show the electron density curves underneath the rough oxide. The classical model shows large local enhancements in the electron concentration becuase its electrons are able to take advantage of the hollows to get electrons to a higher potential. The density gradient model, whose electrons are unable to enter the small hollows, shows a smaller influence of the roughness. The final figure compares the peak electron concentrations for the two models, along the interface for classical electrons, and along a line about 10Å below the interface for quantum electrons. The results suggest that scattering from short wavelength roughness should be quite limited due to the exclusion of carriers from the surface.

Results - Numerical

Simulations with the DG model took about three times as many Newton loops as with the classical drift-diffusion model for 0.1V bias steps. Each linear solve took about 20 per cent longer. The absolute times were about a minute for two-carrier solutions on

a 1598-node grid, using a Sun Ultrasparc-30/300. An equivalent SCSP solution including carrier transport in two dimensions would be an intimidating proposition. It was noted that convergence with the DG model was always linear, and difficult at high drain bias. Since the model uses only existing operators which have been tested and which give quadratic convergence on other problems, it appears likely that in this formulation the Jacobian is either close to singular or poorly conditioned, leading to the poor convergence. Further work is needed in this area.

Summary

The density gradient (DG) model is a useful method of exploring quantum effects in carrier distribution and transport in multiple dimensions. Without adjusting parameters over a range of oxide thicknesses, it gives a good account of carrier displacement from the oxide interface in MOS capacitors. In our experience to date, the flatband voltage is not as accurately predicted as by a self-consistent Schrodinger-Poisson (SCSP) solution. However it offers new possibilities for predicting quantum effects in structures which cannot be readily simulated in any other way.

Script-level programming allowed rapid prototyping of the model, including a number of experiments in choosing variables until the system was successfully solved. It offered a substantial acceleration over traditional hard-coded discretization of the equations.

Acknowledgment

The authors would like to thank Dr. P. V. Voorde of HP Labs for discussions and data. MGA thanks ONR for support.

References

[1] M.G. Ancona, Phys. Rev. B42(2) p.1222 (1990)
[2] K.S. Krisch et al., Electron Dev. Lett. p.521 (1996)
[3] P. Vande Voorde et al., IEDM p.811 (1996)

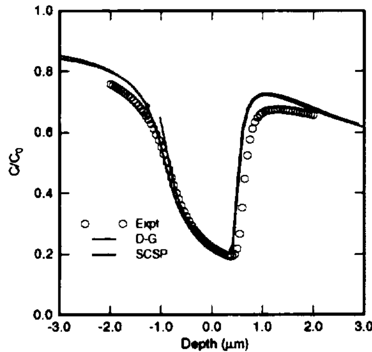

Figure 1: Comparison of DG and SCSP models using identical effective masses, for a MOS capacitor with $t_{ox} = 31\text{Å}$.

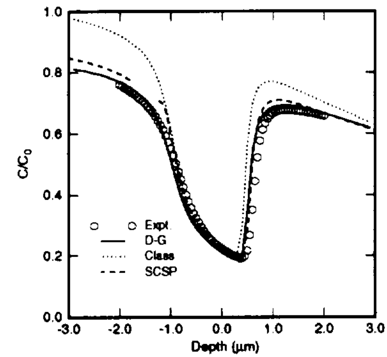

Figure 2: Comparison of simulation and measurement for MOS capacitor with $t_{ox} = 31\text{Å}$, using best parameters in SCSP and in DG models.

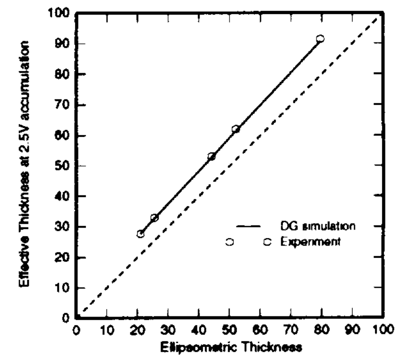

Figure 3: The exclusion of carriers from the interface region decreases the capacitance and increases the apparent thickness of oxide. The apparent thickness with the DG model agrees well with experiment over a range of oxide thicknesses with no parameter fitting.

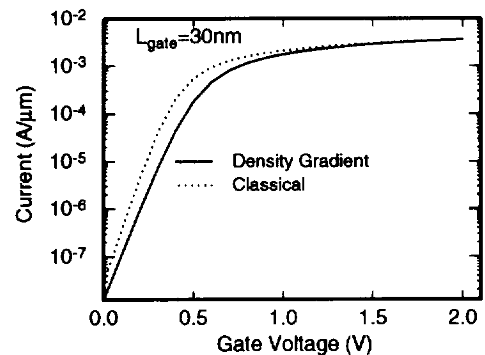

Figure 4: Subthreshold current of 30nm MOSFET (Vd=0.1V)

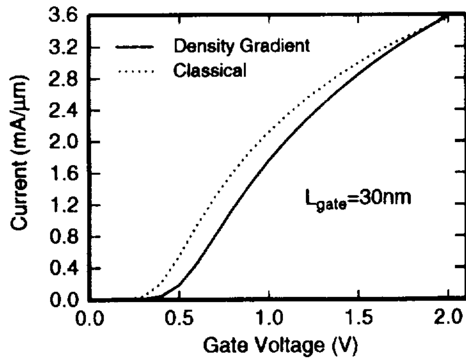

Figure 5: Turn-on characteristic of 30nm MOSFET (Vd=0.1V)

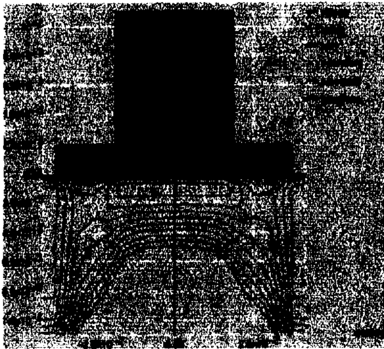

Figure 6: 20nm MOSFET Electron density in density gradient model

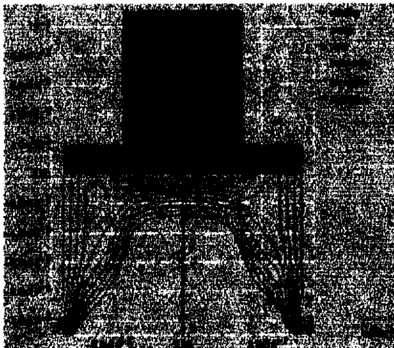

Figure 7: 20nm MOSFET Electron density in classical model

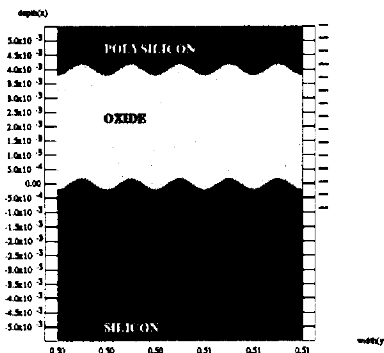

Figure 8: Rough MOS interface. The simulation included one period of the roughness which has a 2Å amplitude and 20Å wavelength

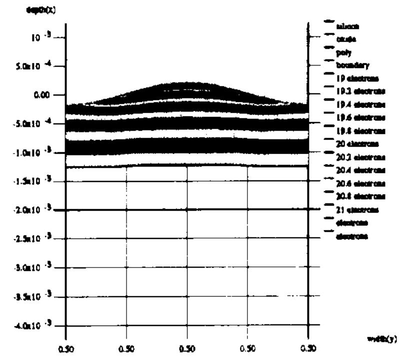

Figure 9: Classical Electron Distribution. Electrons concentrate in the hollows where the potential is highest

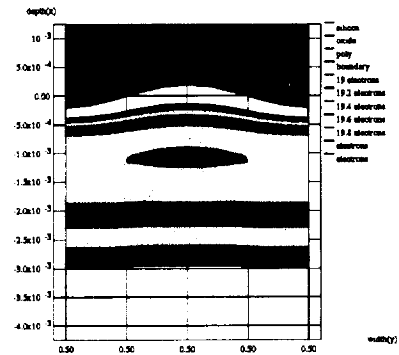

Figure 10: Quantum Electron Distribution. Electrons excluded from the hollows show little influence of the surface roughness.

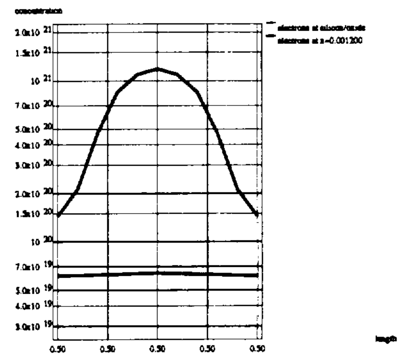

Figure 11: Peak Electron Distributions. The upper curve is the classical distribution along the interface. The lower curve is the quantum distribution along a line 10Å below the interface where the peak lies.