Simulation of ULSI Silicon MOSFETs

M. R. Pinto, J. Bude and C. S. Rafferty
AT&T Bell Laboratories, Murray Hill, NJ 07974

INTRODUCTION

Process and device simulation have played a large role in the scaling of MOS technology. In order to remain predictive into the ULSI regime (L <0.25μm), a number of challenges must be met, both in terms of physical models and computational algorithms. This paper focuses on the status of simulation applied to silicon MOSFETs, highlighting the effects that must be captured and providing brief descriptions of optimal approaches as viewed at present. Areas requiring increased attention include process simulation, quantization and mobility in inversion layers, hot carrier reliability and 3D grid generation.

PROCESS SIMULATION

The accuracy of simulated electrical characteristics depends on both process and device models. Although the main focus of this work is on device issues, some observations are made regarding process simulations of ULSI MOSFETs. Of principal importance is the ability to accurately model dopant profiles (junction, channel and isolation), which require accurate treatment of transient enhanced diffusion (TED) for the low thermal budgets associated with ULSI processing. The effect of implant damage on junction profiles is modeled accurately at intermediate (LDD) doses [1], however the nonlinear dose-damage dependence must be taken into account at both lower and higher doses. In addition to the vertical junctions, source/drain implant damage can affect the channel profile of small devices. Both reverse short-channel effects in threshold voltage V_t [2] and reduced body coefficient [3] have been observed in sub 0.5μm NMOS devices. As shown in figure 1, the effect can be substantial on strongly nonuniform profiles. Given the emphasis on substrate doping engineering [4] and low voltage operation for ULSI, it is critical to model these effects accurately.

In addition to improving physical models, geometrical effects are particularly important for device isolation. Latchup sensitivity has a strong layout dependence, while gate-induced source-to-drain leakage depends critically on the isolation sidewall shape. Another geometrical leakage effect recently noted is shown in Figure 2. The upturned boron contours in the vicinity of the LOCOS perimeter creates a source/bulk junction at a higher boron level than in the bulk. This leads to exponentially more tunneling current at the perimeter than in the bulk[24]. 3D simulation shows that the effect is similar at a corner and an edge, thus edge leakage should dominate.

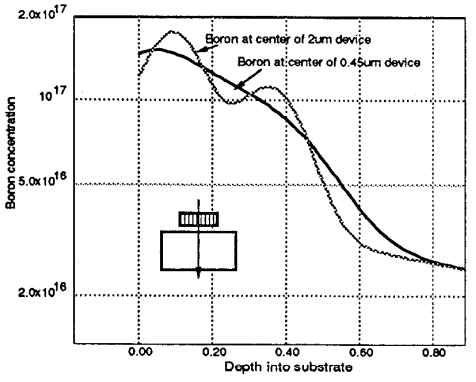

Figure 1 – Simulated channel boron profile for short- and long-channel NMOS devices illustrating redistribution from implant damage. The difference in profiles leads to a reduction in substrate bias sensitivity for the short device which is accurately reproduced by simulation [3].

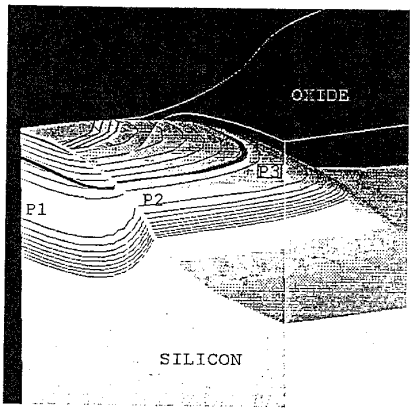

Figure 2 – Boron contours and arsenic junction in vicinity of 3D bird's beak. Tunneling is higher along the perimeter (P2) than in the bulk (P1), but is no worse at the corner (P3) than at the perimeter.

VELOCITY OVERSHOOT EFFECTS

Experimental and theoretical evidence suggests velocity overshoot is observable in NMOS transconductance (g_m) at deep-submicron channel lengths. Energy balance (EB) models, based on moment expansions of the Boltzmann transport equation and a carrier energy-temperature relation [5,6], represent an attractive means of modeling overshoot at an

incremental cost over conventional drift-diffusion. However accurate momentum and energy relaxation rates are required, obtainable from Monte Carlo (MC) simulations of 1D homogeneous slabs. Using both the field E_H and unheated mobility μ_0 as parameters, velocity-field $v(\mu_0,E_H)=\mu(\mu_0,E_H)\cdot E_H$ and energy-field $w(\mu_0,E_H)$ relations can be extracted. Figure 3 shows $w(\mu_0,E_H)$ obtained from full-band MC simulations; the results are relatively insensitive to $\mu_0 > 150\,\mathrm{cm}^2/\mathrm{V\text{-}s}$. By inverting the energy-field relation to obtain $E_H(\mu_0,w)$, the EB mobility is thus evaluated locally using $\mu[\mu_0,E_H(\mu_0,w)]$. Since the true driving force F can be much larger than E_H, the drift velocity $v=\mu(\mu_0,E_H)\cdot F$ can locally exceed v_{sat}.

MC fits can also be obtained for other energy dependent parameters, e.g. nonparabolicity coefficients and ionization rates. For conventional EB formulations, it is critical to fit the heat conduction proportionality factor κ which can strongly affect the local solution in regions with abrupt energy/field transitions. For some devices, the optimal κ can have structural and bias dependencies, arising from non-Maxwellian carrier distributions which would be better treated by a two temperature EB model [8]; improved results have are also observed by deriving the energy transport system from moment integration after making the relaxation approximation [9]. For conventional EB, we find that MOSFETs are relatively insensitive to κ, and excellent agreement can be achieved with MC for a number of different types of impurity profiles (including LDD) using $\kappa \approx 0.7$. Figure 4 shows a comparison of PADRE EB [10] and MC simulations with self-consistency [11] of an NMOSFET, including proper surface scattering models, demonstrating excellent agreement. Comparisons with terminal characteristics are equally good [10].

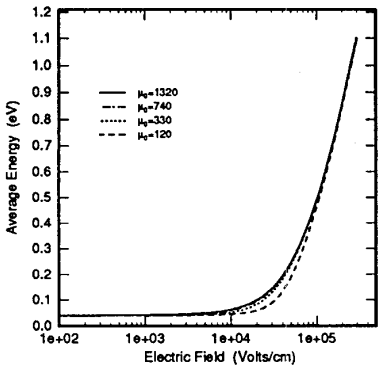

Figure 3 – Electron energy-field characteristics obtained from Monte Carlo simulations of 1D homogeneous slabs using low field mobility μ_0 as a parameter.

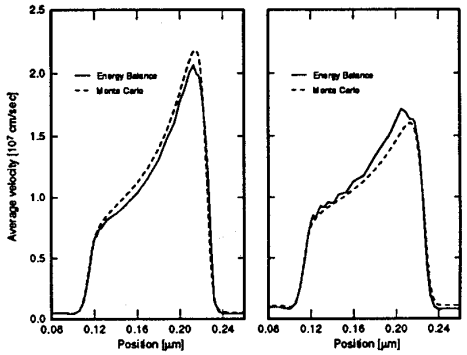

Figure 4 – Comparison of average electron velocity, integrated normal to the channel, for an NMOSFET with $L_{eff}=0.1\mu m$, $t_{ox}=4$nm, $x_j=30$nm and a nonuniform channel profile (see figure 5). Bias voltages are V_{GS},V_{DS}= (a) 1.5V,1.5V; (b) 2.5V,1.5V.

INVERSION LAYER QUANTIZATION

The higher channel doping concentration N_{ch} and reduced oxide thickness t_{ox} implied by conventional scaling of ULSI MOSFETs leads to higher fields and hence decreased inversion layer mobility [12], accounted for in device simulators by empirical relations involving the local field and doping such as [13]. Recently however it has been shown the quantum mechanical (QM) splitting of energy levels which occurs at $N_{ch} \approx 5\times10^{17}$ produces V_t shifts of 0.1–0.2V [14,15]. These results imply that self-consistent QM calculations must be performed within the device simulator, especially for nonuniform N_{ch}.

Figure 5 shows integrated electron density as a function of gate bias for three 1D MOS structures simulated using classical Poisson and fully self-consistent Poisson-Schrödinger QM procedures. For thin t_{ox}, we observe *both* a V_t shift and a change in strong-inversion slope. The change in slope, which is strongest for the nonuniform N_{ch}, results from the QM capacitance induced by the offset of charge density from the interface as shown in figure 6. Using self-consistent QM models in device simulators, one must therefore take great care in separating this purely capacitive effect on g_m from traditional scattering mechanisms in field-dependent mobility models.

HOT CARRIER RELIABILITY

To achieve the optimal tradeoff between device lifetime and performance, accurate modeling of hot carrier degradation is most desirable. Voltage scaling in the ULSI regime provides an additional degree of freedom which needs to be carefully considered along with the applicability of extrapolated aging. Unfortunately physically-based degradation models remain a challenging problem, and most MOS design uses substrate current I_{subs} as a monitor. Even MC-calibrated EB simulations of I_{subs} begin to fail for $L_{eff} \approx 0.3$–$0.5\mu m$ [16], and modeling higher energy processes such as gate current I_g and oxide damage strictly through an average T seems hopeless.

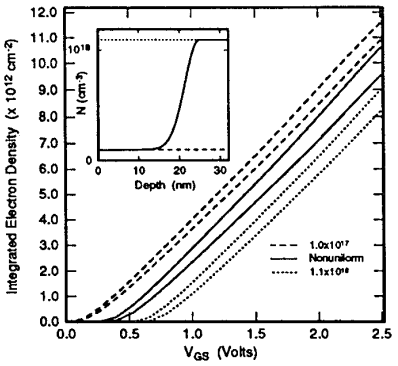

Figure 5 – Electron density versus gate voltage for 1D MOS capacitors, computed using classical Poisson (solid) and quantum mechanical Schrödinger-Poisson (dashes) models. Each device had t_{ox}=4nm, and the doping profiles are given in the inset.

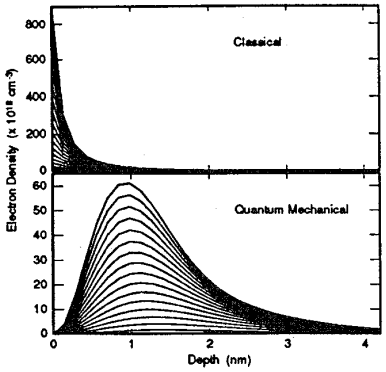

Figure 6 – Electron density versus depth for the nonuniform profile for varying gate voltage. Note the offset of the profiles from the surface which for small t_{ox} leads to a noticeable quantum capacitive effect. Gate voltages correspond to the data of figure 5 ranging from 0V to 2.5V with 0.13V steps.

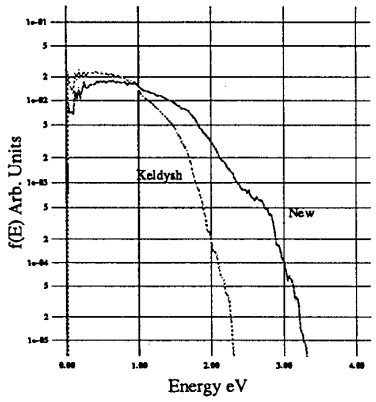

Figure 7 – Energy distribution functions obtained using full-band Monte Carlo simulation of a 1D bulk sample with 400kV/cm electric field. The new ionization rates [19], derived over the full band structure, result in more high energy electrons.

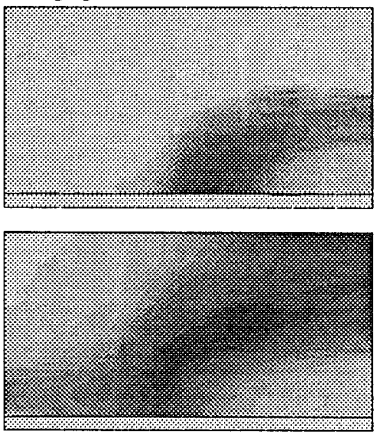

Figure 8 – Average electron (top) and hole (bottom) energies in the drain region of a 0.4μm nMOSFET from full band MC (white=0eV, black=1.5eV). Holes generated by impact ionization are heated along the drain depletion region but away from the oxide (bottom vertical bar), while electrons are heated along the oxide surface.

Full-band MC calculations are an attractive alternative, however they have been hampered by uncertainties in physical mechanisms and their exorbitant computational burden. Physically for instance, it is difficult to explain observed I_g at V_{DS} below the Si–SiO_2 barrier and I_{subs} at V_{DS} below the bandgap [17,18]. Some progress is made by reconsidering the impact ionization rate, starting from first-principles [19], which results in a softer energy dependence than the traditional Keldysh formulation [20]. Because both models yield equal ionization in bulk samples, internal energy distributions must differ, pushing more carriers into higher energy states even for homogeneous fields (figure 7); the difference for nonhomogeneous fields should be even more substantial, although it remains uncertain whether this difference is enough to explain all low V_{DS} anomalies.

Significant progress has also been made recently in the computational throughput and numerical accuracy of full-band MC calculations using refinable simplex discretization in both real and momentum space [21]. Speed-ups of over 2 orders of magnitude over previous approaches yield full-band MC simulations which are actually faster than isotropic analytic band MC codes, making physically rigorous hot carrier simulations practical. Figure 8 shows electron and hole average energies simulated in a 0.4μm nMOSFET using the full-band simplex algorithm, highlighting non-local ionization and carrier heating. These effects are prominent in low voltage, high field situations, and can only be properly treated by solution of the Boltzmann Equation (eg., by full-band MC). Because of the computational efficiency of the simplex algorithm, ionization generation and aging can be studied over wide bias and parameter ranges in practical time periods, reducing the uncertainty in the forms of scattering mechanisms and so improving predictive power.

NOVEL MOS DEVICES

Finally, consideration should also be given to alternative MOS structures for ULSI applications. Most important of these would likely be silicon-on-insulator MOS (SOI) and thin-film transistors (TFT). Nontraditional physical models that may be required for these devices include polysilicon grain effects and self heating (also required for bulk MOSFETs in ESD I/O circuits). TFT devices in high-density SRAM cells can exhibit significant 3D effects due to underlying topography or nonplanar gate overlap.

Figure 9 shows electron density contours in Delta SOI-MOSFET [22], obtained using integrated process and device simulation [23]. This device has several attractive features for ULSI applications, including lithographic control of film thickness, nearly complete gate wrap-around and high drive to surface area performance. In addition to accurate physical models, such simulations require sophisticated numerical algorithms, including both linear solvers and grid generation. Similar to gate overlaps in recessed oxide MOSFETs [8], stress effects during oxide growth can lead to non-Manhattan corners which require fine mesh to resolve the inversion layer. Figure 10 shows a portion of a nonrectangular mesh which is ideally suited to these corners. Given realistic constraints on cpu time, accurate simulation of structures such as these still require manual intervention in the grid generation process.

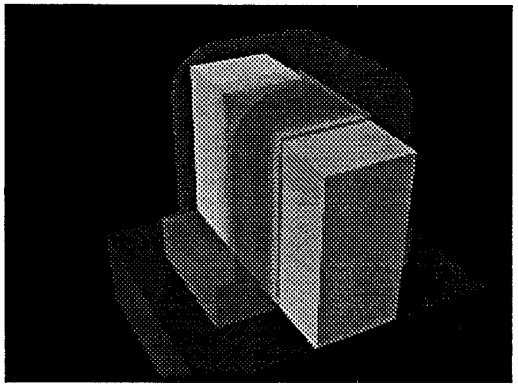

Figure 9 – Electron density contours in a Delta SOI-MOSFET simulated using integrated 3D process and device simulation. Light is high, dark is low and the semitransparent material is gate poly. The device is biased in accumulation to better illustrate corner effects.

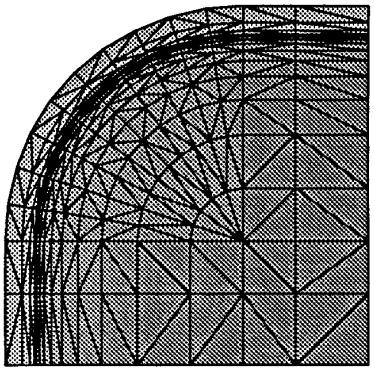

Figure 10 – Example of a portion of a 2D cross-section of a full 3D nonrectangular grid in a corner covered by gate oxide. In order to properly model the inversion layer, a very fine grid is required ($<0.5nm$) normal to the surface.

ACKNOWLEDGEMENTS

The contributions of W. M. Coughran, Jr., E. Sangiorgi and R. K. Smith are gratefully acknowledged.

REFERENCES

[1] M. Giles, *J. Electrochem. Soc.*, **138**, p.1160 (1991).
[2] T. Kunikiyo et.al., *Proc. NUPAD IV*, p.51 (1992).
[3] C. Rafferty et.al., *VPAD 1993*, (submitted).
[4] R.-H. Yan et.al., *Appl. Phys. Lett.*, **59**, p.3315 (1991).
[5] K. Bløtekjaer, *IEEE Trans. Elect. Dev.*, **17**, p.38 (1970).
[6] R. Stratton, *Phys. Rev.*, **126**, p.2002 (1962).
[7] A. Gnudi et.al., *EET*, p.307 (1990).
[8] M. Pinto et.al., in *Computational Electronics*, Kluwer, p.3 (1990).
[9] D. Chen et.al., *Proc. NUPAD IV*, p.109 (1992).
[10] M. Pinto, *Electrochem. Soc. Proc.*, **91-11**, p.43 (1991).
[11] F. Venturi et.al., *IEEE Trans. CAD ICs*, **8**, p.360 (1989).
[12] A. Sabnis and J. Clemens, *IEDM Tech. Dig.*, p.18 (1979).
[13] C. Lombardi et.al., *IEEE Trans. CAD ICs*, **7**, p.1164 (1988).
[14] Y. Ohkura, *Solid-St. Elect.*, **33**, p.1581 (1990).
[15] M. van Dort et.al., *IEDM Tech. Dig.*, p.495 (1991).
[16] H. Peifer et.al., *IEDM Tech. Dig.*, p.131 (1991).
[17] J. Chung et.al., *IEEE Trans. Elect. Dev.*, **34**, p.1651 (1990).
[18] L. Manchanda et.al., *IEDM Tech. Dig.*, p.994 (1992).
[19] J. Bude et.al., *Phys. Rev. B*, **45**, p.10958 (1992).
[20] L. Keldysh, *Sov. Phys. JTEP*, **21**, p. 1135 (1965).
[21] J. Bude et.al., to appear.
[22] D. Hisamoto et.al., *IEEE Trans. Elect. Dev.*, **34**, p.1419 (1991).
[23] M. Pinto et.al., *IEDM Tech. Dig.*, p.923 (1992).
[24] B. Park et.al., to appear.