3-1 (Invited)

A Strategy for Enabling Predictive TCAD in Development of sub-100nm CMOS Technologies

C. F. Machala, S. Chakravarthi, D. Li*, S.-H. Yang and C. Bowen

Texas Instruments, Inc., PO Box 650311, Dallas, TX 75265

*University of Texas at Austin

Ph: 972-995-7973. Fax: 972-995-2037. E-mail: c-machala@ti.com

Abstract - CMOS technology development is an expensive undertaking. Predictive TCAD can be a key tool in reducing the time and cost of CMOS development by acting as a virtual wafer fab where experimental lots are first simulated and the results analyzed before the first process step is ever run. In this way errors may be uncovered, poor splits eliminated and inadequate experimental designs improved. Each lot yields more information through better design and less waste. However, in order to achieve these improvements TCAD simulators must be quickly calibrated with new models that simulate the new physics for sub-100nm technology nodes. Because of the rapid pace of CMOS development, the available commercial simulators do not usually contain all the models needed in order to be predictive. This paper describes our efforts in upgrading the commercially available simulators TSUPREM4[1] and Dessis[2] for the purpose of aiding in the development of sub-100nm high performance and low stand-by power CMOS technologies.

I. INTRODUCTION

Technology computer aided design (TCAD) can be an important tool in CMOS development. Simulations have the potential to reduce cost, save silicon and decrease development time. Simulations can be used to determine key processes early in design so that suitable manpower may be focused on process developments that have the highest impact. TCAD can help find problems in a proposed lot design so that silicon is not wasted on splits that will not yield any information. And TCAD can be used to quickly explore a design space and find the optimal design.

The challenge is that the pace of development is so fast that simulators rarely have all the necessary models to accurately simulate a future technology. Lower energy implants, amorphizing implants and spike anneals are but a few processes now being widely used in new technologies. For these and other processes, no accurate models exist. On the device side, transistors now under development show significant quantum and high dose effects that up to now have not been encountered. Given a limited amount of resources, companies using TCAD must focus their development efforts on those processes that have the biggest impact on device performance and at the same time leverage vendors, universities and research consortia so that models come on line when needed.

In order to simulate technologies currently under development we found we needed to create and calibrate models in several areas. First, in implant we needed to extend existing table based implant models to lower energy. At the same time, we needed to include the effects of a preamorphization implant on both the implant channeling and the damage number (i.e. +n). In diffusion we needed to improve existing models so they could accurately simulate spike anneals. We also needed to include implant, damage and diffusion effects of other species, such as fluorine and antimony as well as clustering, activation and deactivation effects.

To improve device simulation we found it necessary to include solving the energy balance equations for transistors with gate lengths less than 100nm. A recalibration of mobility for the ultra thin nitrided gate oxides was also required in order to accurately simulate drive current and the effects of pocket dose on mobility. For simulating low stand-by power CMOS technologies, we needed to calibrate the band to band tunneling models in order to accurately simulate the gate edge diode leakage. Also, the effects of line edge roughness needed to be taken into account when device performance variability was simulated.

II. IMPLANT MODELING

Modeling shallow junctions requires that the initial implanted profile be modeled accurately. Pre amorphous implants are commonly used to achieve shallow junctions by reducing the channeling of the implanted dopants and by reducing transient enhanced diffusion. Amorphous silicon regrows

epitaxially at low temperatures, leaving a defect free silicon lattice. Transient enhanced diffusion is reduced in this layer since the only point defects are those that diffuse in from outside the regrown layer.

In order to model shallow junctions formed in this manner, both the reduced channeling and transient enhanced diffusion needs to be taken into account. This requires not only accurate modeling the dopant profiles but also modeling of the amorphous layer thickness and remaining damage. In a joint project with the University of Texas we were able to accurately model these effects using UT-MARLOWE [3]. We ran 48 samples using a range of pre-amorphous implants using antimony as the amorphizing species. The implants were done through bare silicon or silicon capped with either 20Å or 50Å of thermally grown oxide. Boron was then implanted into some of the samples and the dopant/antimony profiles and damage profiles were measured using secondary ion mass spectroscopy (SIMS) and Rutherford back-scattering (RBS).

Figure 1 shows antimony profiles implanted into crystalline silicon through varying oxide layer thicknesses from zero to 50Å. In order to get this good agreement we had to use the pair specific interatomic potential for antimony and silicon[4]. The universal interatomic potential did not give satisfactory results. Figure 2 shows measured and simulated RBS damage profiles for an antimony implant with a dose of $3x10^{13}$ cm^{-3} at an energy of 15keV. The simulated damage profile is obtained from a calibrated version of UT-MARLOWE. We calibrated the damage profile using the Kinchin-Pease damage model [5] coefficient f_{rec}.

Figure 2: Comparison of measured and simulated damage for an antimony implant.

Figure 3 shows boron profiles implanted into pre-amorphized silicon. We used different antimony implants to amorphize the silicon and then used the calibrated UT-MARLOWE to test the damage calibration. UT-MARLOWE was able to accurately predict the dechanneling of boron caused by the damage from the antimony PAI.

III. ARSENIC CLUSTERING

To achieve ultra shallow junctions not only changes in implant but changes in anneals are also being employed. The use of short, high temperature Rapid Thermal Processing (RTP) anneals has greatly
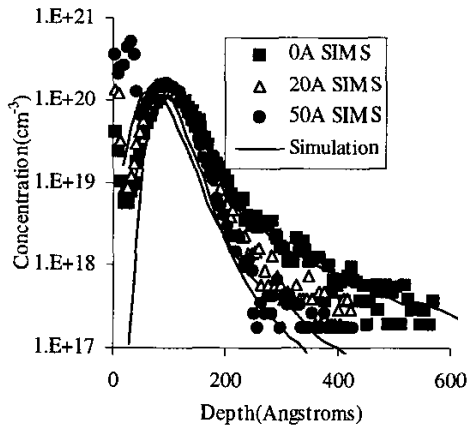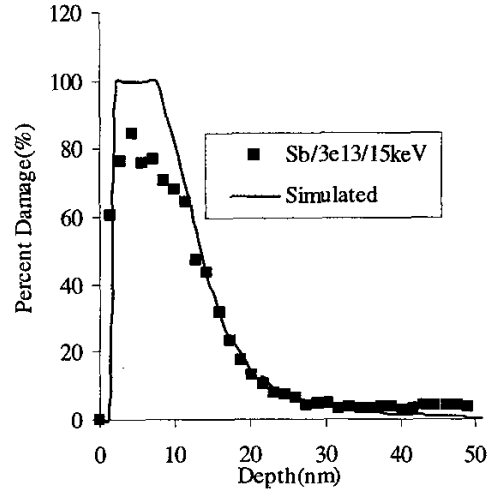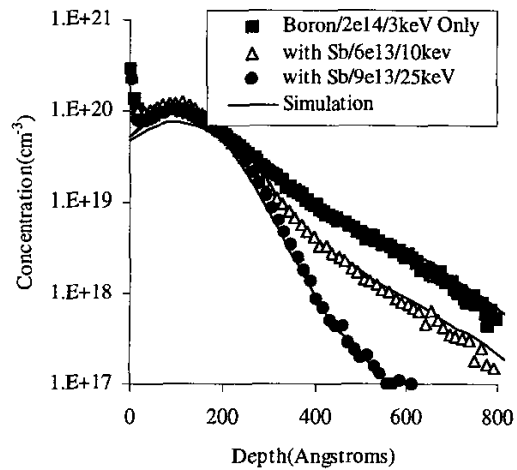

Figure 1: Antimony profiles implant through varying screen oxide layer thicknesses. Thinner oxide allows more channeling and a slightly deeper amorphous layer.

Figure 3: Measured an simulated boron profiles implanted through varying amorphous layers.

34

diminished the importance of Transient Enhanced Diffusion (TED) in ultra-shallow arsenic junctions. However, annealing of ultra-shallow arsenic implants suggests other effects controlling arsenic diffusivity and activation [6]. Similarly we have observed anomalous activation and diffusion of arsenic at low energies [7]. Solmi *et.al.* have suggested formation of arsenic clusters based on experimental data [8]. Recent *ab-initio* calculations have shown that the critical process in arsenic deactivation is a rearrangement of arsenic atoms via diffusion leading to interstitial ejection [9]. We have modeled this process by a single reaction in TSUPREM4. Further, we consider the formation of a precipitate phase at very high arsenic concentrations. The precipitate is modeled as a large cluster with no point defect interaction[8].

The arsenic clustering model has been calibrated over a range of experimental conditions. For e.g., Figures 4 and 5 show comparison to different thermal budgets. Using this model we can satisfactorily explain the anomalous arsenic diffusion observed[7]. During diffusion, there is a continuous formation of new As_4V clusters due to dissolution of the precipitate phase. The interstitials ejected by the deactivating arsenic lead to further enhancement of the diffusing arsenic. This enhancement increases with a decrease in energy. This is because more As_4V clusters form at lower energies driven by the higher arsenic concentration. Also note the active concentration (open symbol) is much lower than the total chemical concentration. In figure 6 we calculate
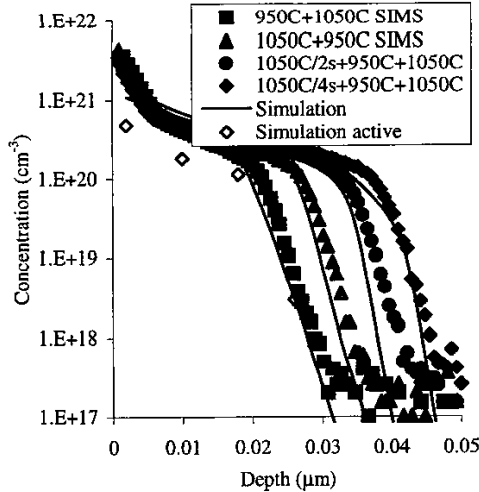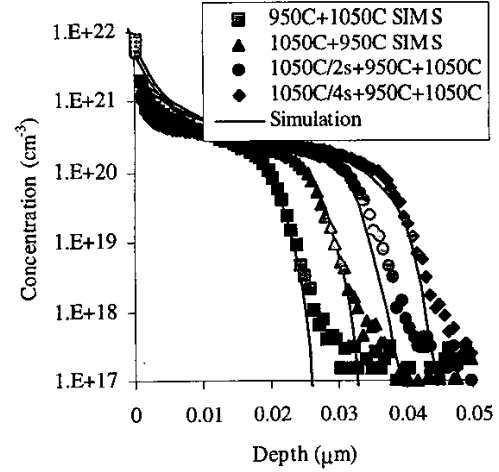

Figure 5: Experimental and simulated arsenic profiles from a 1.6e15, 2keV arsenic implant.

the sheet resistance values predicted from our model. As can be seen it compares very favorably with experimental data.

IV. MODELING RETARDING SPECIES

Silicon technology development forces new dopant/foreign atoms to be explored for potential advantages. TCAD offers a unique platform to understand and thus better leverage the learning from experimental lots. However, one of the major challenges to process TCAD is the complexity associated with the introduction of new impurities. This is because it requires (a) an understanding of the
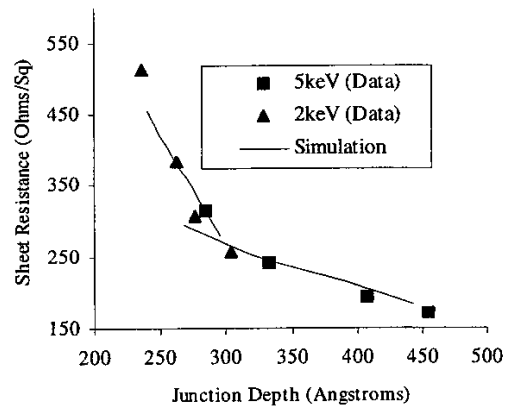

Figure 4: Arsenic diffusion experiments and simulated profiles with new clustering model for different anneals.

Figure 6: Measured and simulated Xj/Rs for arsenic drain extensions. The model agrees well with experiment showing good agreement to chemical and active dopant profiles.

physical interactions between the impurity and dopant atoms, and (b) a flexible simulation framework that supports addition of these new models/interactions. For example, carbon is a relatively new impurity that has been studied by different groups [10,11]. It is well known that carbon is incorporated as a substitutional atom but migrates rapidly as a carbon interstitial. The primary reactions associated with carbon diffusion are the kick-out [12]:

Cs + I = Ci
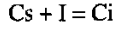
and the dissociative mechanism:

Ci + V = Cs,
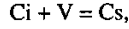
Where, Cs is the substitutional form of carbon and Ci is interstitial carbon. This requires addition of two more solution variables (fields) Cs and Ci. Interstitial carbon has been found to have a low migration energy [13,14]. Hence similar to other dopants, we can assume Ci is in equilibrium with substitutional carbon. This allows us to reduce the system of equations to the familiar full-coupled model. Annealing of high concentrations of carbon suggest the presence of carbon interstitial complexes [10]. Recent experimental and simulation results indicate the formation of dicarbon interstitial (C_2I) complexes [13,14,15]. Hence, in our modeling effort we include a C_2I complex formed by a simple clustering reaction:

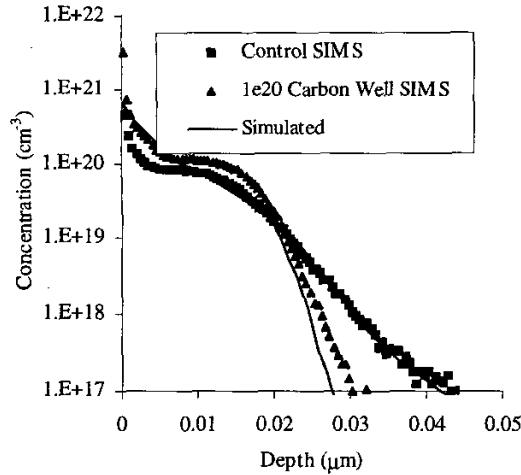$$2C_s + I = C_2I.$$
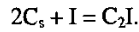
TSUPREM4 offers two approaches to adding

Figure 7: Measured and simulated boron profiles with and without carbon. The model is able to capture the reduction in junction depth and abruptness.

new impurities/dopants in the simulator. It is possible to add new species/reactions using either the User Interface (USEIT) or with built-in functions. We have incorporated the above reactions for carbon diffusion in TSUPREM4. The generic dopant-interstitial cluster model was used to save computational time for modeling the C_2I cluster.

Diffusion of dopants that are interstitial diffusers (e.g. B) can be retarded by the presence of carbon. Carbon primarily decreases the net interstitial concentration and hence retards these dopants. It can be expected that the net effective diffusivity will be a function of the concentration of carbon in silicon. For example, figure 7 shows the effect of including a high concentration carbon layer on final boron junctions. In these experiments boron was implanted into two wafers. Wafer 1 has no carbon whereas wafer 2 has a carbon well with a concentration of $1e20 cm^{-3}$. It is apparent that the boron diffusion is severely retarded in the carbon well. Furthermore, the abruptness of the boron (decay of the boron concentration) increases with the introduction of carbon. Our model is able to capture both the changes in junction depth and abruptness of the boron profile very well.

V. DEVICE SIMULATION FOR SUB-100NM CMOS

In order to accurately simulate a technology both process and device simulation improvements are needed. Sub-100nm CMOS requires high halo concentrations in order to meet off-state current leakage specifications. At high channel concentrations, mobility can degrade significantly due to Coulomb scattering. Mobility can also be degraded by charge in gate dielectrics. For these reasons it is important to calibrate electron and hole mobility for the particular oxide being used in the device at channel concentrations seen by the sub-50nm devices.

We ran split wafers where uniform channels with concentrations varying from $1x10^{17}$ to $4x10^{18}$ cm^{-3} were created. Mobility curves were extracted from experimental data measured on long channel devices using gate oxides reflected in the technology under development. Figure 8 shows the measured mobility. The figure shows significant mobility reduction as the channel concentration is increased. The channel concentration of sub-100nm CMOS devices with high halo concentrations can exceed $2x10^{18}$ cm^{-3}. If the electric field in the channel is approximately 1MV/cm we see the mobility is less than half of what the universal mobility would be. The Coulomb scattering region of the mobility curve can also shift depending on other charges, such as fixed charge in
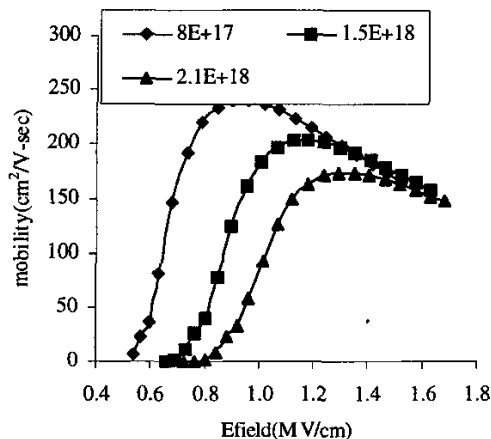
Figure 8: Electron mobility for higher channel concentrations with nitrided gate oxides. The Coulomb scattering regime is moving into higher electric fields with higher channel doping.
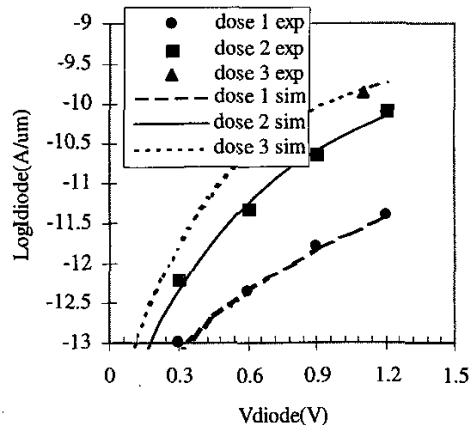

Figure 9: Measured and simulated diode leakage current as a function of diode voltage and halo implant dose. With calibration the increase in diode leakage due to higher dose and higher electric field is simulated well.

the oxide or interface charges. If the device simulator mobility model is not calibrated this mobility reduction due to high channel concentrations and charge in the gate stack will not be predicted. The device simulator will mislead and predict high halo concentrations improve device performance when in fact the high halos will degrade device performance.

VI. MODELING LOW STAND-BY POWER TECHNOLGIES

Technologies offering low stand-by power for hand-held and battery operated applications require accurate modeling of the various leakage mechanisms found in CMOS devices. Several areas must be addressed. The primary concerns are gate tunneling and subthreshold leakage. For sub-100nm devices that require ultra-shallow junctions and high halo concentrations, the electric fields and therefore the junction leakage are of major concern. Also, for very small gate lengths line edge roughness can increase the magnitude and variability in the transistor off-state leakage current[16].

Gate-induced diode leakage (GIDL) is very important for low power technologies and TCAD can help optimize devices to reduce this component if the dopant profiles are correct, the electric fields are modeled well and the band-to-band tunneling model is calibrated. Figure 9 shows part of this calibration. Here the GIDL current is calibrated as a function of halo implant dose and gate voltage using the band to

band tunneling model in the Dessis device simulator. Further calibration and device optimization using the simulator is shown in [17].

VII. CONCLUSIONS

In order to simulate new CMOS technologies featuring gate lengths on the order of 50nm and low off-state leakage, several models must be improved and calibrated. The higher channel concentrations seen in these technologies increase the electric field and leakage currents and also decrease mobility. If new models and calibration are not done the simulator will not predict these effects and can direct development efforts in the wrong direction.

VIII. ACKNOWLEDGEMENTS

IX. References

[1] TSUPREM4, Avant! Corporation.
[2] Dessis, ISE Integrated Systems Engineering.
[3] D. Li, S.-H. Yang, C. Machala, L. Lin and S.K. Bannerjee, SISPAD 2002.
[4] J. F. Ziegler, Private Communication.
[5] G. H. Kinchin and R. S. Pease, Rep. Progress. Phy. 18, pp. 1-51, 1955.

[6] R.Kasnavi, P.B.Griffin, and J.D.Plummer, in Sim of Semicond. Proc. and Dev. Tech. Dig. (SISPAD '98),48, 1998.

[7] S.Chakravarthi, P.R.Chidambaram, C.Machala, A.Jain and X.Zhang, in Proc. of Mater. Res. Soc. (MRS '02), C3.7, 2002.

[8] D. Nobili and S. Solmi in J. Appl. Phys. 83 (5) 2481 (1998); D. Nobili, S. Solmi, A. Parisini, M. Derdour, A. Armigliato and L. Moro, in Phys. Rev. B 49 (4) 2477.

[9] P. A. Stolk, D. J. Eaglesham, H. J. Gossmann, and J. M. Poate, *Appl. Phys. Lett.* 66, 1370 (1995).

[10]]H. Rucker, B. Heinemann, W. Ropke, R. Kurps, D. Kruger, G. Lippert, and H. J. Osten, *Appl. Phys. Lett.* 73, 1682, (1998).

[11] U. Gosele, A. Plobl and T.Y. Tan in Defects in Semiconductors, *Electrochem. Soc. Proc.*, 309 (1996).

[12] G.D Watkins and K.L Brower, *Phys Rev. Lett.* 36, 1329 (1976).

[13] C.L. Liu, W. Windl, L. Borucki, S. Lu, X.Y. Liu, *Appl. Phys. Lett.* 80, 52 (2002).

[14] S. Mirabella, A. Coati, D. De Salvador, E. Napolitani, A. Mattoni, G. Bisognin, M. Berti, A. Carnera, and A. V. Drigo, S. Scalese, S. Pulvirenti, A. Terrasi, and F. Priolo, *Phys. Rev. B* 65, 045209 (2002).

[15] J.L. Ngau P.B. Griffin and J.D. Plummer, *J. of Appl. Phys.*, 90, 1768 (2001).

[16] J. Wu, J. Chen and K. Liu, SISPAD 2002.

[17] S. Zhao et. al., SISPAD 2002.