Nanoworld Semiconductor Industry - State and Future Challenges of Technology Computer Aided Design

Dr. Helmut Puchner

LSI Logic Corporation, 3115 Alfred Street
Santa Clara CA-95054, USA, helmut@lsil.com

ABSTRACT

Several application examples of nanoscale techniques used to influence or enhance the understanding of material properties as well as processing behavior are presented. We will also present a review of first principle density functional theory calculations used to investigate the scaling trends of high-k gate dielectrics. Since conventional as well as high-k dielectrics may still require a strong diffusion barrier against boron penetration we will present ab-initio quantum chemical calculations for the diffusion of boron within a thin gate oxide layer. Plasma-nitridation can be applied to alter the diffusion behavior of boron inside the gate oxide dielectric and to completely block boron penetration. We will present a detailed view on the nitridation mechanism and a combination of reactor-feature scale simulations in combination with Monte Carlo implantation simulations to describe the nitridation process. Finally, kinetic Monte Carlo applications are presented for diffusion processes.

Keywords: High-k, plasma nitridation, reactor-scale / feature-scale, kinetic Monte Carlo

1 Introduction

Technology Computer Aided Design (TCAD) is widely used in the semiconductor industry to design and optimize submicron transistors. With the appearance of the first reliable and dependable device simulators in the early 1980s [1][2] simulation tools then could be employed to investigate small geometry effects and the device phenomena associated with the downscaling of device technologies. Increasingly powerful computers allowed simulation of multidimensional realistic device geometries and sophisticated device effects. Early device simulation codes generated the doping profiles within their program flow before the device simulator was loaded with the simulation structure. Primitive geometry manipulations as well as analytical doping profile descriptions [3] were implemented to generate the input doping concentrations and geometries for the device simulator. With the increasing complexity of device structures and the onset of sophisticated diffusion mechanisms separate process simulation codes were developed independently [4][5] to model the complete process flow from the silicon substrate to the passivation of the surface. It is the ultimate goal to replace the experimental trial-error process with TCAD simulations to limit capital expenditures as well as to shorten the development cycle. Classical TCAD simulation tools solve PDEs within a specific macroscopic simulation domain. Common semiconductor manufacturing steps are simulated and modeled by continuum equations and numerical algorithms, e.g in order to simulate the ion implantation process one can either employ analytical distribution functions or Monte Carlo algorithms to describe the final doping profile in the semiconductor material. Anneal processes are commonly described by Fick's first and second law and several additional recombination/generation terms to account for the complicated defect evolution mechanisms during the annealing process. The application of these TCAD simulation tools for reactor scale simulation domains is limited by the number of required discretization points (mesh points) and, hence, by the speed as well as memory requirements of nowadays computer systems. Successful wafer scale simulations could only be demonstrated for fluid dynamic simulations within the processing reactor environment including the semiconductor surface. In order to link the reactor scale simulations with the detailed feature scale simulations of the semiconductor surface certain approximations have to be taken into account. Since the wafer size is increasing and the feature size of the minimum gate length transistors are decreasing, the gap between reactor scale and feature scale simulations is growing with future generations. This leads to the conclusion that reactor scale simulations will plays only a minor role for feature scale TCAD simulations. On the other hand, atomistic scale simulation methods have been introduced for semiconductor materials quite successfully for some time. Ab-initio calculations could provide insights into the migration of single atoms within a super-cell with periodic boundary conditions or within a cluster approximation. The ultimate goal of these atomic scale simulations is to extract energy barriers for continuum models, which as a consequence can be applied more efficiently for future feature scale problems.

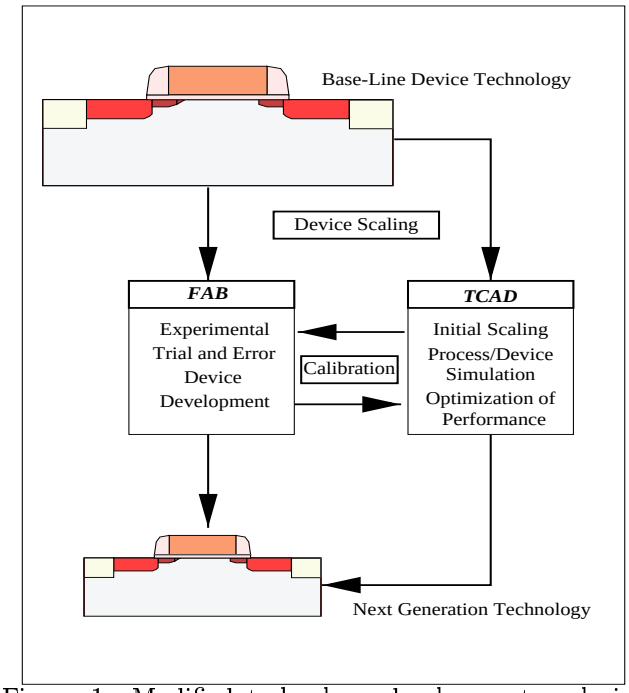

Figure 1: Modified technology development cycle including the role of TCAD. The use of calibrated simulation tools results in better performing devices.

2 TCAD Development Cycle

With the appearance of TCAD tools the classical development cycle could be modified as depicted in Figure 1. TCAD tools can predict the downscaled new device technology before the production equipment is in place to actually manufacture these new devices. The first silicon lot of a new device generation in development can be already verified and optimized with TCAD tools, which represents a tremendous savings in material and processing costs. Once the first initial devices are fabricated, TCAD tools are used to further optimize the devices for the succeeding experimental wafer lots. This task is achieved by calibration of the TCAD tools with the experimental data from the first silicon devices. Since several revisions of the process technology are explored before the process development for a technology node is completed, several recalibrations also may be necessary between revisions.

In order to predict and replace some of the very cost intensive experimental development tasks with computer simulations the TCAD tools that are used have to be able to handle a wide variety of problems in an accurate fashion. A tremendous range of process and device physics must be captured. The speed of technology changes in the manufacturing process in recent years has dramatically increased. These changes included the addition of copper interconnects as major backend metal, silicon germanium devices to improve signal to noise ratios, the introduction of silicon-on-insulator (SOI) technology, low permittivity backend dielectrics, and new type of photoresists for deep UV steppers. Accurately simulating each significantly new process requires new simulation models, otherwise the simulation tools lose their predictability. Depending on the available resources some companies program their own models into proprietary as well as commercial available tools based on experimental measurements. Some of these model get implemented into commercial simulation tools and need to be further adapted to the local production environment. Not having an internal model development team limits the flexibility of simulation models to the available commercial codes, but avoids delays because of software development issues on the other hand. There is a need for a coordinated and ongoing model and code development effort within the TCAD community.

3 Transistor Engineering

It is important for the TCAD tools to maintain their predictability and accuracy as device dimensions shrink. State-of-the-art devices show increasing small device geometry effects. The most important ones from the viewpoint of TCAD are the reverse short channel effect, reverse narrow width effect, drain induced barrier lowering, gate induced drain leakage, and junction leakage currents. All these effects play an important role when simulating the electrical characteristics of deep sub-micron devices. Devices are also starting to exhibit significant quantum mechanical effects, such as tunneling effects and inversion layer quantization. The gate tunneling current causes a significant contribution to the overall leakage current for gate oxides below 20Å. Therefore, it is extremely important to accurately simulate leakage currents. Modern submicron devices also require an additional large angle pocket implant to prevent punchthrough at short channel lengths. The pocket implant therefore adds to the total channel dose and can be considered a major dopant source for the "reverse short channel effect" (RSCE)[6], [7]. As the pocket implant dose increases, we observe a drastically increased RSCE. TCAD tools must be able to capture this dramatic rise in threshold voltage by adjusting the interstitial recombination rate at the Si/SiO_2-interface to achieve the right amount of channel dopant redistribution on one hand, but not to overdiffuse the source/drain junctions in the lateral as well as vertical direction on the other hand. Especially, the lateral diffusion is critical in modeling the roll-off behavior of the threshold voltage vs. gate length characteristic. In order to estimate the device behavior at deep submicron device dimensions ($< 100nm$) one has to be able to calibrate the simulation tools to real silicon experimental data beforehand. Only in this case it is possible to sustain predictability of the process and device simulation tools.

4 Plasma Nitridation

One of the most critical building block of a very deep submicron technology is the gate dielectric layer. The continuous down-scaling trend in device dimensions drives the further decrease of the gate oxide thickness. Deep sub-micron CMOS technologies require gate oxide thicknesses well below 20Å [8]. One of the most severe problems caused by the employment of thin gate oxide layers is the boron penetration from the heavily doped p+ polysilicon gate trough the underlying gate oxide and into the channel region of the PMOS transistor. This leads to the decrease in threshold voltage down to a surface source-drain punchthrough [9]. Boron penetration also degrades the device reliability by generation of defects in the gate oxide. One possible solution to suppress boron penetration is gate oxide hardening by incorporation of nitrogen into the gate oxide [18], [11]. Different approaches to nitrogen incorporation into the gate oxide layer have been developed. Among them there are formation of a nitrogen atom monolayer at the Si/SiO_2 interface between gate oxide and substrate[11], incorporation of nitrogen atoms in gate oxide bulk, and formation of a thin heavily nitrogen doped layer at the top of the gate oxide [12]. Each of these approaches addresses a specific problem with respect to device reliability and performance. For example, a monolayer nitrogen incorporation at the Si/SiO_2 interface reduces the interface roughness and therefore improves the high field mobility and reduces defect generation under hot carrier stressing. Bulk nitridation reduces the electrical oxide thickness by maintaining the same optical thickness resulting in the employment of thicker gate oxides and the advantages associated with it, e.g. lower direct tunneling current. Top surface nitridation is the most effective method of avoiding boron penetration. It offers several advantages compared to the creation of a nitrogen diffusion barrier at the bottom of the gate oxide. Depending on the method used to create the barrier the nitrogen at the Si/SiO_2 interface will increase the interface trap density causing possible mobility degradation. Additionally, the boron trapped in the bulk of the gate oxide increases the electron trap density [13] throughout the gate oxide. The latter will increase the defect generation under Fowler-Nordheim and hot carrier stressing[14]. Top surface nitridation of the gate oxide is almost free from these problems. Conventional high temperature oxidation in N_2O or NO ambient results in a nitrogen incorporation at the Si/SiO_2 interface of less than 5% [15]. Two different techniques allow much higher level of nitrogen incorporation at the gate oxide top surface: deposition of an ultra-thin silicon nitride layer on the top of the gate oxide [14] and the ultra-low energy ion bombardment from a nitrogen plasma source [12]. The strength of the diffusion barrier will depend on the nitrogen concentration as well as on the distribution of the incorporated nitrogen within gate oxide.

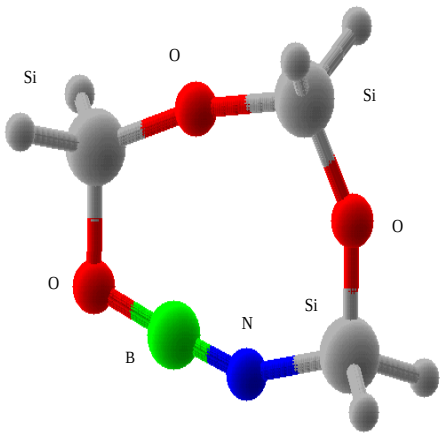

Figure 2: Model cluster for boron insertion into the nitrogen containing siloxane ring

Ab-initio quantum chemical calculations on model systems containing several siloxane bonds have been employed to get insight into the mechanisms of boron diffusion in silicon oxide [16] and suppression of boron penetration into the gate oxide by plasma-induced nitridation (see Fig. 2). Calculated energies of insertion of various dopants into the siloxane bond show a certain correlation with experimental diffusion activation energies through silicon oxide. Plasma induced nitridation leads to incorporation of nitrogen atoms into siloxane bond. Energy gain for Boron-insertion into a regular siloxane bond (approx. $3eV$) dramatically increases for insertion into a nitridized siloxane bond ($10eV$). This might be a plausible explanation of the Boron trapping in the gate oxide after plasma nitridation. Semi-empirical quantum chemical methods showed a qualitative agreement with ab-initio calculations and have been applied to larger model systems[17]. Since a nitrogen plasma contains a higher amount of neutral nitrogen than ionized nitrogen, model calculations confirmed the absence of nitridation for the neutral nitrogen interaction with silicon oxide surface bonds. In order to achieve sufficient surface nitridation the ionized nitrogen needs to be injected into the surface region to create lattice damage. This allows the neutral nitrogen to be incorporated into the surface regions. Increasing the ionized nitrogen flux will lead to sputtering of the surface and needs to be carefully optimized.

5 High-k Gate Dielectrics

Another major challenge will be to meet the equivalent gate oxide thickness requirement for different technology nodes. During the last decade the only competitive gate dielectric material was silicon dioxide because of its manufacturability and adhesion properties. With a required 10Å electrical oxide thickness, which

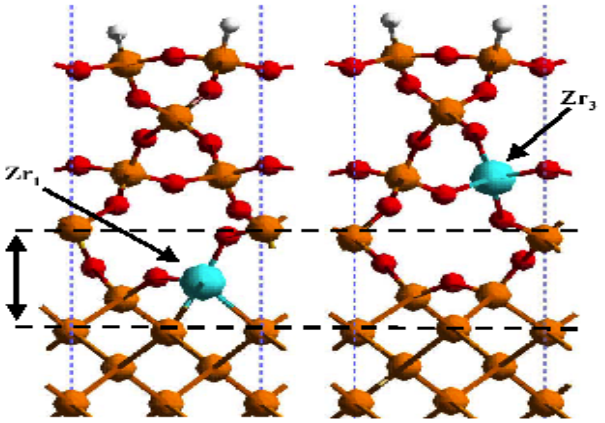

Figure 3: Interface bonding simulation structure for different Zr positions. The Zr3 position has a 0.7eV lower energy compared to the Zr1 position.

corresponds to approximately two atomic silicon dioxide layers, the limits of scaling are reached. The integration of alternative higher dielectric gate oxide materials (high-κ) allows physically thicker dielectric layers which reduce the gate current significantly. Unfortunately, these high-κ materials exhibit unexpected interfacial growth mechanisms [18] and are thermally unstable. Junction annealing temperatures have to be significantly lowered or a *gate damascene* process has to be used. Both measures increase the manufacturing complexity and costs dramatically. Significant advances in first principle methods, together with the increase in computational power have allowed the characterization of some promising high-κ candidates using first principles method based on density functional theory (DFT). Kawamoto et al. recently presented results for the silicate-silicon interface band offset of Zr and Hf silicates. They have found that the conduction band edge is determined by the metal d-state and the valance band edge of the silicate by the oxygen p-state. DFT calculations also confirmed that Hf atoms are significantly less likely to diffuse into the silicon substrate from the silicate film. Figure 3 gives the model cluster for different positions within the silicate/silicon interface. The Zr3 location results in a $0.7eV$ lower energy than the Zr1 location [19], which confirms the tendency of Zr to stay inside the silicate material layer. Further calculation enhancement, such as the calculation of the dielectric constant purely from simulations, will be used to provide input parameters for PDE process and device simulators.

6 Kinetic Monte Carlo Diffusion

Kinetic Monte Carlo codes have been used since quite some time to gain theoretical understanding of defect-dopant interaction in silicon and to develop a theoretical understanding of diffusion mechanisms and extended defects during annealing processes [20]. First principle calculations were used to obtain the energetics for boron-interstitial complexes and incorporated into kinetic Monte Carlo algorithms to estimate the dopant profile evolution as well as the dopant activation. The link between first principle calculations and kinetic Monte Carlo algorithms represents a promising avenue for further investigations of other dopant-defect complexes in silicon.

7 Conclusions

Atomistic first principle and kinetic Monte Carlo simulation are extremely powerful tools to understand and improve the material properties needed for the next generation of MOS transistors. The understanding of defect evolution and shallow junction creation directly relates to the transistor performance. TCAD will still play an important role in the definition as well as optimization of future device technologies. The combination of atomistic modeling and continuum modeling will allow the incorporation of new materials into the mainstream process technology on much faster scale.

REFERENCES

[1] S. Selberherr et al, IEEE Trans. on Electr. Dev.,(ED-27), pp. 1540–1550, 1980
[2] M. Pinto, PISCES IIB, Stanford University, 1985
[3] J. Gibbons et al., Appl. Phys. Letters, 22(11), pp. 568–569, 1973
[4] G. Hobler et al., PROMIS 1.5, Technische Universität Wien, 1991
[5] D. Antoniadis et al., SUPREM II, Stanford University, 1978
[6] H. Jacobs et al., IEDM Techn. Digest, pp. 307–310, 1993.
[7] C. Rafferty et al., IEDM Techn. Digest, pp.311–314, 1993.
[8] G. Lucovsky, IBM J.Res.Develop., 43 (3), 301 (1999)
[9] J. Pfiester et al., IEEE Elec. Dev. Lett. 11, 247 (1990).
[10] D. Lee et al., J. Vac. Sci. Technol. A 13, 607 (1995).
[11] G. Lucovsky et al., J. Vac. Sci. Technol. A 14, 2832 (1996).
[12] S. Hattangady et al., Appl. Phys. Lett. 66, 3495 (1995).
[13] J. Pfiester et al., IEEE Trans. Elec. Dev. 37, 1842 (1990).
[14] Y. Wu et al., IEEE 98CH36173, Proceedings of 36th Annual Intern. Reliability Physics Symp. 70 (1998).
[15] Z. Ma et al., IEEE Elec. Dev. Lett., 15, 109 (1994).
[16] V. Zubkov et al., Mat. Science Semicond. Processing 3, 41 (2000).
[17] V. Sukharev et al., J.Vac.Sci.Technol. 17 (1999).
[18] C. Lee et al., IEDM Techn. Digest, 27 (2000).
[19] A. Kawamoto et al., IEEE Electron Dev. Lett. 22, 14 (2001).
[20] M. Caturla et al., IEEE Vol 1276, 1022 (1998).